Genotyping of *Anaplasma phagocytophilum* in natural endemic cycles

von Julia Fröhlich

Inaugural-Dissertation zur Erlangung der Doktorwürde der Tierärztlichen Fakultät der Ludwig-Maximilians-Universität München

Genotyping of *Anaplasma phagocytophilum* in natural endemic cycles

von Julia Fröhlich

aus Memmingen

München 2016

Aus dem Veterinärwissenschaftlichen Department der Tierärztlichen Fakultät der Ludwig-Maximilians-Universität München

Lehrstuhl für Vergleichende Tropenmedizin und Parasitologie

Arbeit angefertigt unter der Leitung von: Priv.-Doz. Dr. med. vet. Cornelia Silaghi

Gedruckt mit Genehmigung der Tierärztlichen Fakultät
der Ludwig-Maximilians Universität München

Dekan:	Univ.-Prof. Dr. Joachim Braun
Berichterstatter:	Priv.-Doz. Dr. med. vet. Cornelia Silaghi
Korreferent/en:	Univ.-Prof. Dr. Eckhard Wolf

Tag der Promotion: 11. Februar 2017

Für meine Familie und Freunde

“There is nothing permanent except change.”

Heraclitus of Ephesus

INDEX

ABBREVIATIONS

A. alces	*Alces alces*
A. agrarius	*Apodemus agrarius*
acc. no., *pl.*: acc. nos.	accession number, *pl.*: accession numbers
A. phagocytophilum	*Anaplasma phagocytophilum*
A. marginale	*Anaplasma marginale*
A. flavicollis	*Apodemus flavicollis*
app.	approximately
BLAST	Basic Local Alignment Search Tool
B. bonasus	*Bison bonasus*
B. primigenius taurus	*Bos primigenius taurus*
bp	base pairs
C. aegagrus hircus	*Capra aegagrus hircus*
C. capreolus	*Capreolus capreolus*
CDC	Centers for Disease Control and Prevention
C. elaphus	*Cervus elaphus*
CGA	Canine Granulocytic Anaplasmosis
C. lupus familiaris	*Canis lupus familiaris*
C. nippon	*Cervus nippon*
CSF	Cerebrospinal Fluid
D. albipictus	*Dermacentor albipictus*
D. dama	*Dama dama*
D. occidentalis	*Dermacentor occidentalis*
D. silvarum	*Dermacentor silvarum*
D. variabilis	*Dermacentor variabilis*
DNA	Desoxyribonucleic acid
E. chaffeensis	*Ehrlichia chaffeensis*
E. europaeus	*Erinaceus europaeus*
E. ferus caballus	*Equus ferus caballus*
e.g.	*exempli gratia*, *engl.*: for example
et al.	*et alia*, *engl.:* and others
EGA	Equine Granulocytic Anaplasmosis
ExPASy	Expert Protein Analysis System
Fig.	Figure
FRET	Fluorescent Resonance Energy Transfer
F. silvestris catus	*Felis silvestris catus*
g	gram
HE-staining	Hematoxylin-Eosin-staining
HGA	Human Granulocytic Anaplasmosis
H. inermis argyropus	*Hydropotes inermis argyropus*
H. megaspinosa	*Haemaphysalis megaspinosa*
H. sapiens	*Homo sapiens*
H. inermis argyropus	*Hydropotes inermis argyropus*
ICU	Intensive Care Unit
IFA	Immunofluorescence Antibody
i.e.	*id est*, *engl.*: that means
I. dendatus	*Ixodes dentatus*
I. persulcatus	*Ixodes persulcatus*
I. nipponensis	*Ixodes nipponensis*
I. ricinus	*Ixodes ricinus*
I. scapularis	*Ixodes scapularis*
I. spini-palpis	*Ixodes spinipalpis*
I. ventalloi	*Ixodes ventalloi*
kDa	kilodalton
lat.	latin
Mbp	mega basepair
mg	milligramm
min	minute
msp	*major surface protein*
MODS	Multiple Organ Dysfunction Syndrome
NCBI	National Center for Biotechnology Information
neg.	negative

O. gmelini aries	*Ovis gmelini aries*
nm	nanometer
O. virginianus	*Odocoileus virginianus*
P. leucopus	*Peromyscus leucopus*
pos.	positive
rpm	revolutions per minute
sec	second
SIRS	Systemic Inflammatory Response Syndrome
PCR	Polymerase Chain Reaction
Tab.	Table
TBF	Tick-borne fever
U. arctos	*Ursus arctos*
USA	United States of America
US	United States
°C	degree celsius
µl	microliter

I. INTRODUCTION

The alpha-proteobacterium *Anaplasma phagocytophilum* is the causative agent of an emerging disease affecting both animals and humans, namely granulocytic anaplasmosis. Although known in humans since 1994, this bacterium has recently gained importance, as more and more people in the United States of America (USA) and recently also in Asia are symptomatically infected (Chen et al., 1994; Kim et al., 2014). The Centers for Disease Control and Prevention (CDC) in the USA confirm a steady rise of annual human granulocytic anaplasmosis cases (HGA) since 1994 (CDC, 2013; Chen et al., 1994). From 2000 – 2007, the hospitalization rate of affected patients in the USA was as high as 36.0% and the case-fatality rate reached 0.6% (Dahlgren et al., 2011). In contrast, only scattered cases of HGA are known in humans in Europe, counting 66 cases in total until 2005 (Dumler et al., 2005). Nevertheless, recent cases in Poland (Welc-Faleciak et al., 2015), Germany ex Scotland (Hagedorn et al., 2014), France (Edouard et al., 2012), Austria (Vogl et al., 2010) and Slovakia (Novakova et al., 2010) confirm the continuing occurrence of HGA in European countries. The first case in Belgium was detected in 2000 and amounted 60-90 cases in the years 2005-2009, which led to the description of Belgium as a hot spot for HGA (Cochez et al., 2011). Underdiagnosing of HGA cases in Europe could be due to a lack of awareness by physicians, whereas granulocytic anaplasmosis is notifiable in the USA. However, variability in the genome of *A. phagocytophilum* could also be the reason for a differing pathogenicity in Europe compared to the USA. Besides humans, susceptible hosts developing clinical symptoms during *A. phagocytophilum* infections are ruminants, like cattle (*Bos primigenius taurus*) (Nieder et al., 2012), goats (*Capra aegagrus hircus*) (Silaghi et al., 2011e) and sheep (*Ovis gmelini aries*) (Stuen et al., 2013b), dogs (*Canis lupus familiaris*) (Kohn et al., 2008), horses (*Equus ferus caballus*) (Silaghi et al., 2011d) and cats (*Felis silvestris catus*) (Shaw et al., 2005). In order to maintain the endemic cycle of *A. phagocytophilum* in nature, the role of wild ruminants (Kang et al., 2011; Massung et al., 2005; Silaghi et al., 2011b), small mammals (Cao et al., 2006; Foley et al., 2008b; Silaghi et al., 2012b) or birds (Daniels et al., 2002; De La Fuente et al., 2005b) have frequently been discussed as reservoir hosts. *A. phagocytophilum* is transmitted to vertebrate hosts by ticks of the *Ixodes*

persulcatus complex (Woldehiwet, 2010). Host species potentially developing clinical symptoms, reservoir hosts and ticks as vectors form the basis for several putative enzootic cycles of *A. phagocytophilum*. In order to uncover natural life cycles of *A. phagocytophilum*, genetic analyses comparing strains from different potential hosts have previously been performed. For example, Jahfari et al. (2014) suggested four distinct ecotypes in Europe comparing *groEL* sequences. Another study claimed two separate cycles for *A. phagocytophilum*, ticks and either rodents or ruminants in Europe on the basis of the *16S rRNA*, the *msp4* and the *DOV1*(a noncoding region) gene (Bown et al., 2009), whereas a study from Asia proposed one common cycle for livestock and small rodents based on the *16S* r*RNA* and the partial *p44ESup1* gene (Zhan et al., 2010b). In the USA, two main variants based on the *16S rRNA* gene of *A. phagocytophilum* strains are discussed, differing in their pathogenicity for humans (Massung et al., 2005; 2006; 2002). A general distinction between ruminant and non-ruminant *A. phagocytophilum* strains was proposed and confirmed (Chastagner et al., 2014; Massung et al., 2005; 2006; Scharf et al., 2011a).

By analyzing *16S rRNA*, *groEL*, *msp4* and/or *msp2* nucleotide sequences of *A. phagocytophilum*, Silaghi et al. previously provided information of diverse *A. phagocytophilum* strains occurring in multiple animal species. *A. phagocytophilum* from goat flocks and cattle varied both on geographic and interspecies level (Silaghi et al., 2011e). Besides, few *A. phagocytophilum* variants were responsible for causing Equine Granulocytic Anaplasmosis (EGA) in clinically apparent horses (Silaghi et al., 2011d). Wild ungulates showed remarkably high variation in infectious *A. phagocytophilum* strains (Overzier et al., 2013a; Silaghi et al., 2011b). In contrast, *A. phagocytophilum* from hedgehogs (*Erinaceus europaeus*) showed high uniformity in the *16S rRNA*-gene sequences (Silaghi et al., 2012a).

Nevertheless, a broad analysis of the *A. phagocytophilum* strains infecting different animal species, which could contribute to a better understanding of possible natural life cycles of *A. phagocytophilum* in Europe, is lacking so far.

Therefore, the objectives of this study were to investigate potential host-pathogen-associations by:

I.) genotype characterization on the basis of four genes (*16S rRNA*, *groEL*, *msp4*, *msp2*) of *A. phagocytophilum* from different mammalian host animals with a

comparison of *A. phagocytophilum* strains from the GenBank.

II.) statistical and phylogenetic analysis of relationships of different *A. phagocytophilum* variants.

By providing information on the genetic variability of the pathogen, this study enhances the knowledge of the possible endemic cycles and potential reservoir hosts of *A. phagocytophilum*.

II. LITERATURE

1. Taxonomy

1.1. Taxonomy

Anaplasma phagocytophilum belongs to the class of the *α-Proteobacteria*. Part of this bacteria class is the order *Rickettsiales*, which includes the family *Anaplasmataceae*. Among others, the family *Anaplasmataceae* contains the genus *Anaplasma*. *Anaplasma phagocytophilum* is a species of the genus *Anaplasma* (Fig. 1).

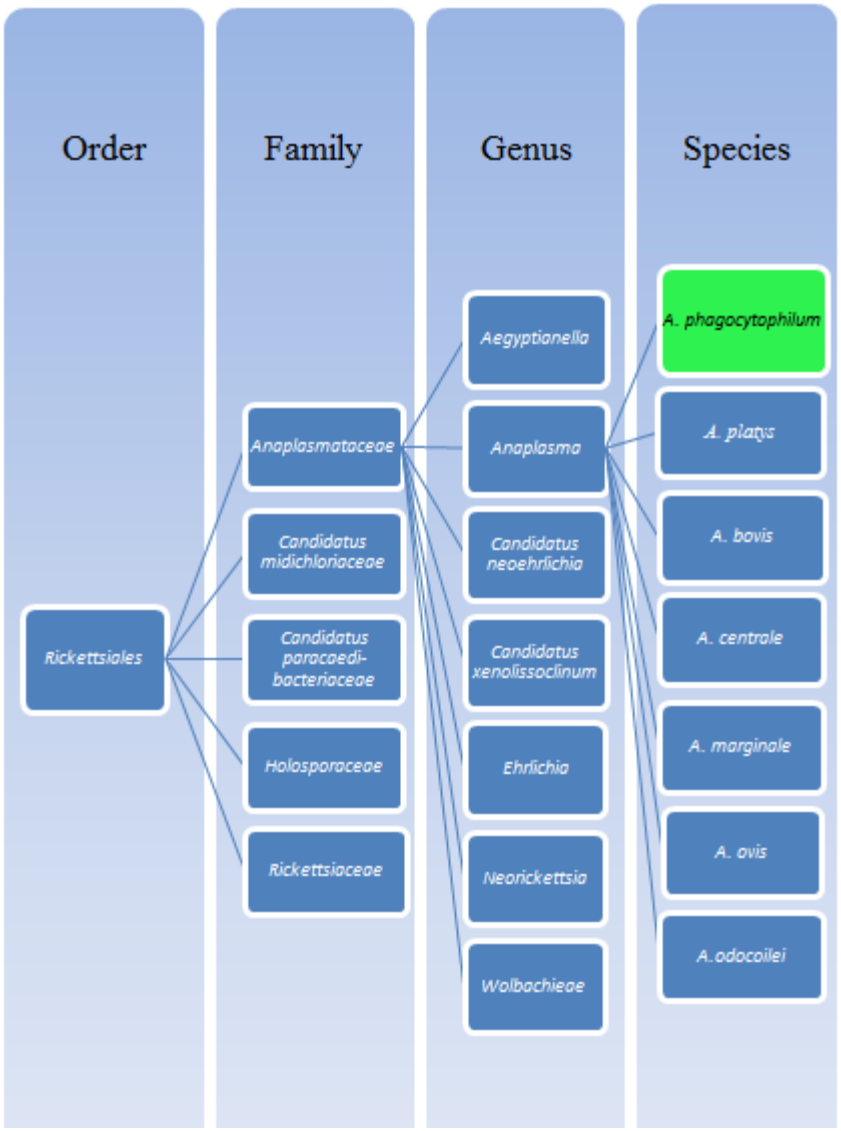

Figure 1: Taxonomy of the order *Rickettsiales* (based on NCBI GenBank information)

Before the taxonomic reorganization in 2001, *A. phagocytophilum* consisted of three different species: the Human Granulocytic Ehrlichiosis (HGE) agent, *Ehrlichia equi* and *Ehrlichia phagocytophila*. All three species infected neutrophilic granulocytes in their vertebrate hosts, but differed in their target animal species. The HGE agent caused disease in humans, *Ehrlichia equi* in horses and *Ehrlichia phagocytophilum* in ruminants like cattle, sheep and bison (Rikihisa, 1991). However, with the development of molecular tools, remarkable similarities in the *16S rRNA* gene and the *groESL* gene were detected in the order *Rickettsiales*. Therefore, Dumler et al. (2001) proposed a complete reorganization within the order *Rickettsiales* and the HGE agent. *Ehrlichia equi* and *Ehrlichia phagocytophila* were united to one single species *A. phagocytophilum.* Nevertheless, heterogeneity was discovered in several investigated genes of *A. phagocytophilum* infecting different animal species (Morissette et al., 2009; Rymaszewska, 2010). Thus, the validity of the latest reclassification of the *A. phagocytophilum* group might be questionable (Foley et al., 2008a).

1.2. Morphology and host cell modulation

A. phagocytophilum is a relatively small (0.5–1.5 µm), coccoid to ellipsoidal, often pleomorphic and gramnegative bacterium (Dumler et al., 2001). As an obligate intracellular pathogen, it parasitizes in cytoplasmatic vacuoles within cells deriving from mammalian bone marrow (Fig. 2). Thereby, *A. phagocytophilum* predominantly targets neutrophilic granulocytes, but also in rare cases eosinophilic granulocytes and monocytes and forms microcolonies, so-called *morulae* (lat.: *morus* – mulberry), in eukaryotic host cell vacuoles (Dumler et al., 2001).

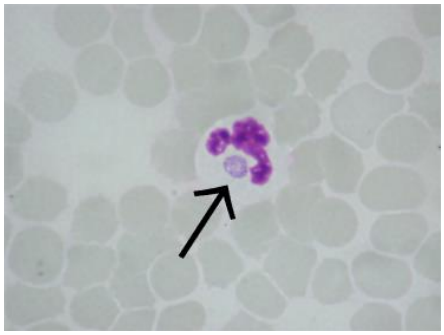

Figure 2: Morula in a neutrophilic granulocyte (Giemsa), HE-staining (Image provided by B. Kohn, Clinic for Small Animals, Faculty of Veterinary Medicine, Free University of Berlin, Berlin, Germany)

Intracellularly, two different ultrastructural forms of *A. phagocytophilum* are distinguished: dense-cored cells and reticulate cells. Dense-cored forms dominate during bacterial receptor-mediated endocytosis and the subsequent release from the host cell. In contrast, reticulate cells are mostly present during bacterial replication by binary fission within the target cells. Once infected by *A. phagocytophilum*, cell functions like apoptosis or programmed cell death are disrupted enabling intracellular survival of the bacterium (Yoshiie et al., 2000). Additionally, neutrophilic functions including degranulation, respiratory outburst or phagocytosis are influenced by *A. phagocytophilum* (Dumler, 2005).

2. Vectors of *A. phagocytophilum*

2.1. Ticks as vectors for *A. phagocytophilum* infection

The main vectors of *A. phagocytophilum* are ticks of the genus *Ixodes* (Fig. 3).

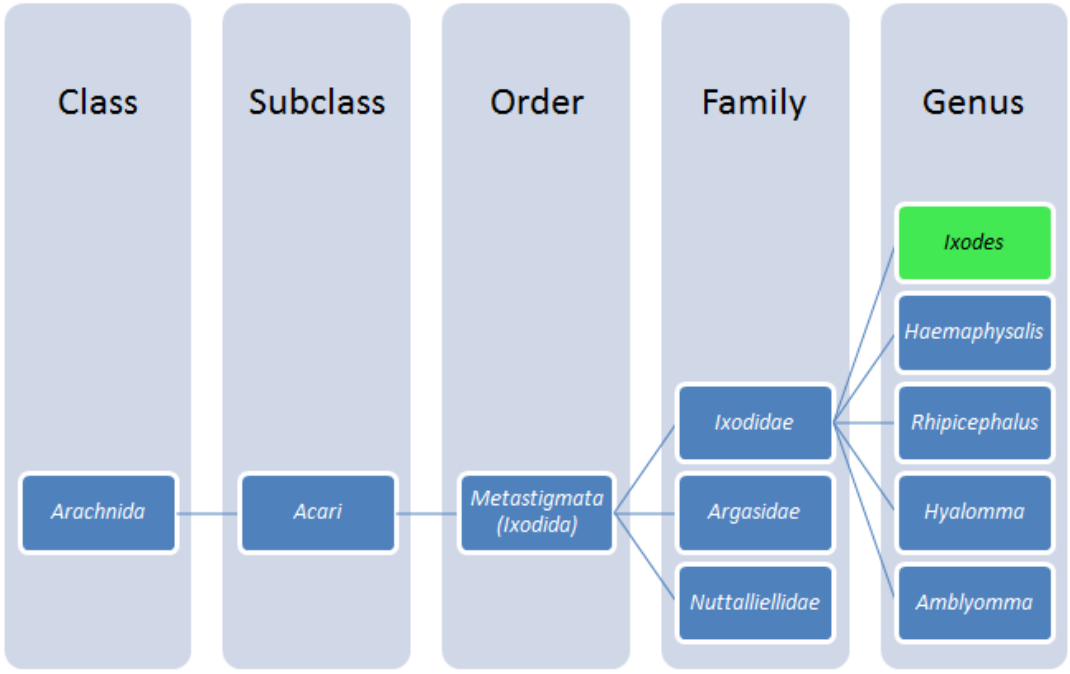

Figure 3: Taxonomy of the genus *Ixodes* [modified from Deplazes et al. (2013)]

According to the distribution of infected ticks, *A. phagocytophilum* is predominantly spread by the American deer tick (*I. scapularis*) in Northeastern and North Central USA, by the western black-legged tick (*I. pacificus*) on the West Coast of the USA, the European sheep tick (*I. ricinus*) in Europe and the taiga tick (*I. persulcatus*) in Asia/Russia (Swanson et al., 2006). Prevalences of

A. phagocytophilum in questing *I. ricinus* vary considerably. Several studies in Europe detected molecular prevalences ranging from 0.3% (Cotte et al., 2010) to 19.2% (Stanczak et al., 2002). Apart from *I. ricinus*, DNA of *A. phagocytophilum* was also occasionally detected in other tick species in Europe including *Dermacentor reticulatus*, *Haemaphysalis concinna* and *I. ventalloi* (Paulauskas et al., 2012; Santos et al., 2004; Tomanovic et al., 2013). Based on molecular detection of *A. phagocytophilum*, questing adult ticks of *I. scapularis* from New York state reached prevalences of 42.0% (Moreno et al., 2006). *I. pacificus* showed a molecular prevalence of 6.2% in a coastal region in California (Holden et al., 2003). Other questing tick species acting as possible vectors for *A. phagocytophilum* could be *Amblyomma americanum, D. variabilis* and *D. occidentalis* (Clark, 2012; Holden et al., 2003). Questing *I. persulcatus* ticks revealed a prevalence of 3.0% in Russia (Rar et al., 2011), 4.0% in China (Cao et al., 2006) and 7.1% in Japan (Murase et al., 2011). Additionally, DNA of *A. phagocytophilum* occurred in a wide range of different tick species in Asia, for example *D. silvarum, H. megaspinosa* or *I. nipponensis* (Cao et al., 2006; Chae et al., 2008; Yoshimoto et al., 2010). The relatively wide range of prevalences might indicate a differing availability of suitable reservoir hosts.

Although transovarial transmission of *A. phagocytophilum* was shown experimentally in *D. albopictus* (Baldridge et al., 2009), the main path of transmitting *A. phagocytophilum* within a tick population is assumed transstadially (Dumler et al., 2001). Consequently, ticks are infected by feeding on infected host animals and are subsequently capable of transferring *A. phagocytophilum* to other hosts successfully. Persistently infected hosts or "carriers" of the pathogen serve as reservoir hosts for *A. phagocytophilum* (De La Fuente et al., 2005b). Carriers are animals infected with *A. phagocytophilum* without developing symptoms of the disease. Since adult ticks take an additional blood meal, the occurrence of *A. phagocytophilum* in adult ticks is much higher compared to larval or nymphal stages. Determination of reservoir competence of diverse animal species helps assessing the potential risk of *A. phagocytophilum* infections in a certain region (Keesing et al., 2012). To date, reservoir competence was experimentally shown only for small ruminants in Europe and the USA (Granquist et al., 2008; Kocan et al., 2012; Massung et al., 2006) and dusky-footed woodrats (*Neotoma fuscipes*) in the Western USA (Rejmanek et al., 2012). Nevertheless, many other animal species including wild ruminants, hedgehogs,

red foxes, rodents and birds are frequently discussed as reservoir hosts of *A. phagocytophilum* (De La Fuente et al., 2005b; Foley et al., 2008b; Hartwig et al., 2014; Silaghi et al., 2012a).

2.2. The life cycle of ticks as vector of *A. phagocytophilum*

The endemic cycle of *A. phagocytophilum* is intensely connected to the three-host life cycle of ixodid ticks serving as vectors for the pathogen (Fig. 4). After hatching out of the embryonated eggs, the larva infests its first host in order to take a blood meal. The blood meal ends with the larva dropping off and developing into the stage of the nymph. Nymphs already carrying the pathogen from the first blood meal serve as vectors for *A. phagocytophilum* during their next blood meal. These immature stages of ixodid ticks mainly target small mammals and birds, but can also infest larger animals, such as horses or dogs. The third and last host of the tick life cycle mainly consists of larger mammals. After mating on its last host animal, female ticks lay thousands of eggs in a sheltered microenvironment and die. With the hatching of the embryonated eggs, the life cycle starts from the beginning (Deplazes et al., 2013; Sonenshine, 1991).

Main target species for immature stages of *I. scapularis* in the Eastern USA are rodents and small mammals. Especially the white-footed mouse (*Peromyscus leucopus*) was shown to be frequently infested. Therefore, these rodents are considered as important reservoir hosts for *A. phagocytophilum* in the Eastern USA (Johnson et al., 2011). On the West coast of the USA, adult stages of *I. pacificus* infect mostly mammals, including domestic and wild animals. In contrast, the immature stages also target birds and reptiles (Castro and Wright, 2007). In Europe, the larvae of *I. ricinus* mostly infest small mammals and rodents. Nevertheless, these animal species are considered as accidental hosts of *A. phagocytophilum* according to very low prevalence rates (De La Fuente et al., 2005b; Foley et al., 2008b; Obiegala et al., 2014; Silaghi et al., 2012b). Nymphal and adult stages target a wide range of host species including small and large mammals, birds and reptiles. Humans represent accidental hosts of nymphs and adult ticks, but are at risk of developing clinical disease (Demma et al., 2005).

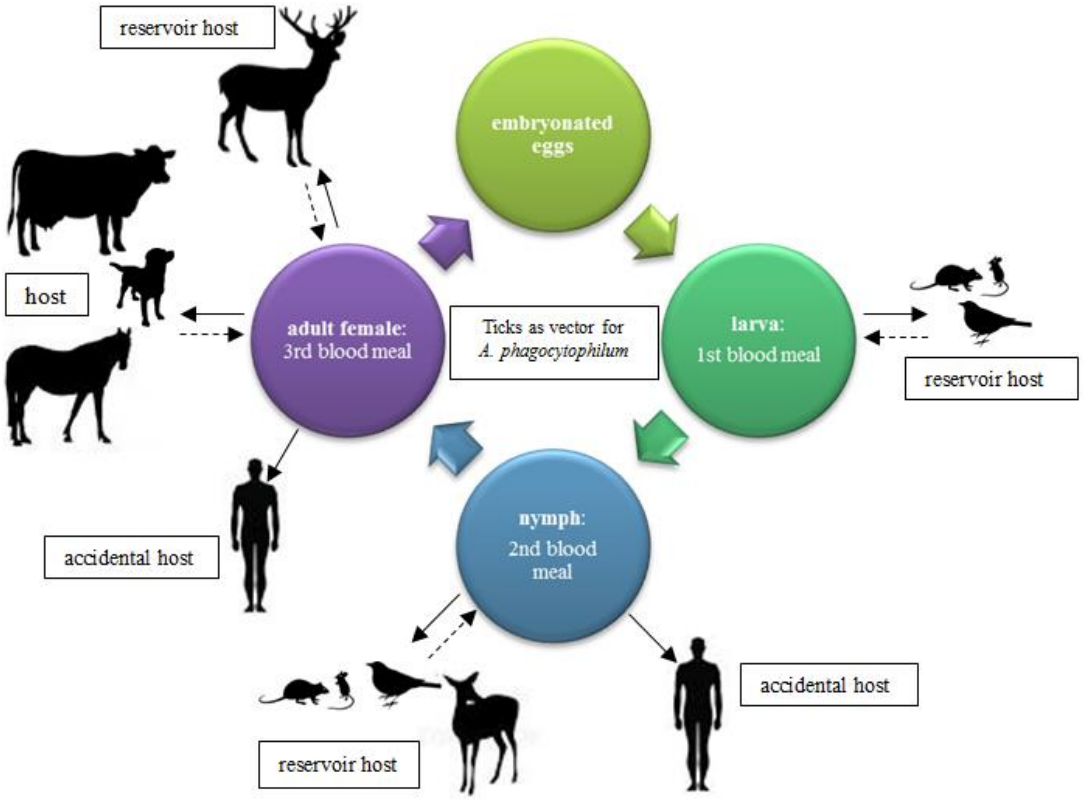

Figure 4: Epizootiology of *A. phagocytophilum* in context of the life cycle of 3-host-ticks [modified from Deplazes et al. (2013)]; dashed arrows show potential reservoir hosts (e.g. wildlife species, birds and rodents), solid arrows show natural hosts possibly developing symptoms of the disease (e.g. domestic animals)

The development of *Ixodes* ticks depends on temperature and relative humidity. Optimal temperature values range from 17°C – 20°C and a relative humidity of at least 80.0% (Medlock et al., 2013). This results in a peak of population activity in spring and fall, although variations can be observed due to climate changes or different geographic regions. The occurrence of human disease caused by *A. phagocytophilum* clearly reflects these activity peaks (Ismail et al., 2010).

3. Host animals of *A. phagocytophilum*

3.1. Reservoir hosts

In order to maintain the tick-transmitted pathogen *A. phagocytophilum* in nature, reservoir hosts and competent tick vectors are required (Dugat et al., 2015). Therefore, potential reservoir hosts are characterized by frequent tick infestation with at least two life stages and a persistence of *A. phagocytophilum* infection in order to transmit the bacteria to other following ticks (Stuen et al., 2013a).

3.1.1. Wild ruminants

Wild ruminants represent well-documented reservoir hosts of *A. phagocytophilum* in the USA and Europe (Dugan et al., 2006; Overzier et al., 2013a). The white-tailed deer (*Odocoileus virginianus*) in particular is considered an important reservoir host in the USA, with molecular prevalence rates reaching up to 47% (Johnson et al., 2011). Nevertheless, Massung et al. (2005) detected the apathogenic strain "Ap-variant 1 (Ap-V1)" in white-tailed deer samples from the USA, which differed from the strain causing HGA in humans from Minnesota ("Ap-ha"). Therefore, *A. phagocytophilum* in white-tailed deer from the USA might not contribute to the distribution of human infection. The available information on *A. phagocytophilum* infections in wild ruminants from the USA is strongly concentrated on white-tailed deer. With respect to *A. phagocytophilum* infections in other cervid species, only scarce information is available. For example, Foley et al. (1998) detected *A. phagocytophilum* in black-tailed deer (*Odocoileus hemonius columbianus*), mule deer (*Odocoileus hemonius hemonius*) and elks (*Cervus elaphus nannodes*) from California.

In Europe, several wild ruminant species are discussed to play a role as reservoir hosts of *A. phagocytophilum*. To date, only red deer (*Cervus elaphus*) have been experimentally shown to act as reservoir for *A. phagocytophilum* strains pathogenic for sheep (Stuen et al., 2010). Molecular prevalences of *A. phagocytophilum* in red deer range from 10.2% in Poland to 87.5% in Norway (Michalik et al., 2009; Stuen et al., 2013b). Although experimental confirmation is lacking to date, roe deer (*Capreolus capreolus*) also seem to play an important role in the distribution of *A. phagocytophilum* infections, reaching molecular prevalence rates as high as 98.9% in Germany (Overzier et al., 2013a). In contrast, Michalik et al. (2009) reported a molecular prevalence rate of only 9.6% concluding high variation in the occurrence of *A. phagocytophilum* infection in roe deer. Other deer species associated with *A. phagocytophilum* infections in Europe include fallow deer (*Dama dama*), sika deer (*Cervus nippon*) or Swedish moose (*Alces alces*). The molecular prevalences of *A. phagocytophilum* in fallow deer varied extensively from 1.5% in Poland to 72.4% in Italy (Ebani et al., 2007; Hapunik et al., 2011). Less data is available for infections in sika deer, but rather uniform rates of app. 40.0% were detected (Hapunik et al., 2011; Zeman and Pecha, 2008). Besides, Swedish moose showed a molecular prevalence of 26.3% (Malmsten et al., 2014).

In Asia, water deer (*Hydropotes inermis argyropus*) from Korea showed a high molecular infection rate of 63.6% (Jilintai et al., 2009; Kang et al., 2011; Kawahara et al., 2006) and sika deer from Japan a rate of 96.3% (Wu et al., 2015).

3.1.2. Small mammals

Tick larvae and nymphs especially feed on small mammals and are thus possibly exposed to *A. phagocytophilum* by their hosts. Therefore, rodents and insectivores are also frequently discussed as reservoir hosts for *A. phagocytophilum*. In the USA, the white-footed mouse (*P. leucopus*) is considered a main reservoir host for the *A. phagocytophilum* strain "Ap-ha" causing HGA in humans (Massung et al., 2003; Walls et al., 1997). In the Eastern USA, molecular prevalences of 20.0–46.8% were shown for white-footed mice (Johnson et al., 2011). Molecular prevalences in other discussed reservoir hosts for *A. phagocytophilum* were 88.4% for chipmunks (*Tamias spp*), 15.1% for red-backed voles (*Clethrionomys gapperi*), 14.3% for meadow voles (*Microtus pennsylvanicus*) and 17.2% for short-tailed shrews (*Blarina* spp.) (Johnson et al., 2011). In the Western USA, Foley et al. (2008b) detected DNA of *A. phagocytophilum* in several mouse (4.0 – 6.3%), chipmunk (6.9 – 50%), squirrel (15.8 – 18.8%), and woodrat species (4.3%).

In Europe, *A. phagocytophilum* infection rates in small mammels seem low compared to the USA. Molecular prevalences of 2.9% in yellow necked-mice (*Apodemus flavicollis*) and 4.2% in wood mice (*Apodemus sylvaticus*) were detected in Switzerland (Liz et al., 2000). Solely, bank voles (*Myodes glareolus*) showed relatively high molecular prevalences with 5-13.3% (Bown et al., 2003; Hulinska et al., 2004). Therefore, the role of rodents as reservoir hosts in Europe is discussed controversially (Bown et al., 2011; Silaghi et al., 2012b). Rodents were even described as probable accidental hosts (Obiegala et al., 2014). In contrast, a high infection rate of 61.8% for *A. phagocytophilum* in European hedgehog (*Erinaceus europaeus*) samples suggested a role as reservoir host in Germany (Silaghi et al., 2012a). Similar prevalence data (76.1%) was shown for the northern white-breasted hedgehog (*Erinaceus roumanicus*) in Hungary (Földvári et al., 2014).

Results from studies from Asia propose a reservoir function for small mammals. For example, Zhan et al. (2010b) showed a molecular prevalence of 14.5% in 159 small mammals of five different species in China. Among all five species, the

brown house rat (*Rattus norvegicus*) showed the highest infection rate of 55.5%. In Korea, 24% of the examined black-striped field mice (*A. agrarius*) and 64% in white-toothed shrews (*Crocidura lasiura*) showed *A. phagocytophilum* infections (Kim et al., 2006).

3.1.3. Birds

Birds constitute a great potential of distributing ticks and potentially also *A. phagocytophilum* over long distances. In the USA, Daniels et al. (2002) detected *A. phagocytophilum* DNA in ticks infesting an American robin (*Turdus migratorius*) and a veery (*Catharus fuscescens*). In Spain, the highest molecular prevalence was shown for blackbirds (*Turdus merula*) (De La Fuente et al., 2005b). Since *Turdus spp.* are ground-nesting birds, they get into contact with ticks quite easily. In contrast, Bjoersdorff et al. (2001) examined migrating birds in Sweden and were not able to detect *A. phagocytophilum* infections in birds. Consequently, birds might rather serve as carriers of ticks infected with *A. phagocytophilum* over far distances, than representing important reservoir hosts.

3.1.4. Other hosts

Several other wildlife species are discussed as hosts or reservoir hosts for *A. phagocytophilm*. In California (USA), the American black bear (*Ursus americanus perniger*) showed a molecular prevalence of 4.0% in 80 investigated animals (Drazenovich et al., 2006), whereas 24.3% out of 74 European brown bears (*U. arctos*) in Serbia were positive (Vichova et al., 2010).

Red foxes (*Vulpes vulpes*) are frequently discussed as reservoir host for *A. phagocytophilum* in Europe with infection rates ranging from 2.7% to 16.6% (Ebani et al., 2011; Hartwig et al., 2014; Karbowiak et al., 2009). Six of 70 gray foxes (*Urocyon cinereoargenteus*) (9.0%) from the USA were also positive (Gabriel et al., 2009).

Wild boars (*Sus scrofa)* tested positive in several studies with infection rates ranging from 3.6% in Japan and 12.5% in Europe (Masuzawa et al., 2011; Michalik et al., 2012; Silaghi et al., 2014). Nevertheless, other studies showed very low prevalence rates, suggesting a limited impact on the distribution of *A. phagocytophilum* as reservoir hosts for these animal species (De la Fuente and Gortazar, 2012).

3.2. Hosts

A. phagocytophilum causes clinical disease in different mammals including dogs, horses, livestock and humans. In contrast, competent reservoir hosts are infected with *A. phagocytophilum* but generally do not show clinical symptoms.

3.2.1. Livestock: Tick-borne fever

The causative agent of tick-borne fever (TBF), initially named *Rickettsia phagocytophila*, was first detected in Scottish sheep (*O. gmelini aries*) and goats (*C. aegagrus hircus*) in 1940 (Gordon et al., 1940). TBF had been mentioned eight years earlier as a distinct tick-transmitted disease in sheep (MacLeod, 1932). The same infective agent also caused disease in cattle (*B. primigenius taurus*) (Hudson, 1950). To date, TBF has been verified in several parts of Europe (Woldehiwet, 2006). For example in Central Europe, molecular prevalences reached 4.0% in German sheep and 5.6% in Swiss goats (Scharf et al., 2011a; Silaghi et al., 2011e; Stuen et al., 2013a). Torina et al. (2008b) examined ruminants in Sicily, Italy, and detected molecular prevalences of 17.0% in cattle and 3.0% in sheep. Lambs from Norway had a relatively high molecular prevalence of 37.5% (Stuen et al., 2013a; Stuen et al., 2013b). In the Czech Republik, 5.5% of the examined cattle were infected with the bacterium (Hulinska et al., 2004). European bisons from Poland showed a molecular prevalence rate of 58.0% (Scharf et al., 2011a). Although high prevalences were detected for *A. phagocytophilum* in Europe, clinical cases are likely underdiagnosed resulting in high economic losses because of reduced milk yield and lowered health condition of infected ruminants (Woldehiwet, 2006). A nationwide investigation of ruminants in China revealed a prevalence of 26.7% for goats and 23.4% for cattle (Zhang et al., 2012). Although no confirmed case of TBF is published in the USA, serological evidence of *A. phagocytophilum* in cattle kept on tick-infested pastures was shown (Magnarelli et al., 2002). Extensive farming of cattle in the USA could be a reason for an underdiagnosis of the disease. By nature, the supervision of single animals might be limited in beef cattle raised in huge pastures, compared to dairy cattle.

The course of TBF can vary from subclinical to severe disease symptoms (Silaghi et al., 2011e; Stuen et al., 2003). Typical clinical findings in infected ruminants are high fever (>41°C), depression, weakness and anorexia (Stuen et al., 2013a). In cattle, decreased milk production and lower limb edema with stiff walking was

observed (Nieder et al., 2012). Due to immunosuppressive effects of *A. phagocytophilum*, secondary infections may cause abortions, tick pyaemia or lameness in sheep (Littlejohn, 1950; McEwen, 1947; Stamp et al., 1950). Additionally, laboratory abnormalities like leucopenia and thrombocytopenia are observed (Gokce and Woldehiwet, 1999). Especially young animals or ruminants newly introduced to tick-infested pasture for the first time are likely to develop symptoms (Nieder et al., 2012). Persistence of *A. phagocytophilum* was detected in cattle (Foggie, 1951). According to Nieder et al. (2012), TBF in infected cows might be self-limiting, if affected ruminants are not treated with antibiotics.

3.2.2. Horses (*E. ferus caballus*): Equine Granulocytic Anaplasmosis

Equine Granulocytic Anaplasmosis (EGA) was first described in 1969 in Northern California (Gribble, 1969) and was subsequently diagnosed in other states of North America, as well as in Europe and Asia. Within Europe, confirmed cases of *A. phagocytophilum* infections occurred in several countries like Germany (Silaghi et al., 2011d), the Netherlands (Butler et al., 2008), Sweden (Bjoersdorff et al., 2002) and the Czech Republik (Jahn et al., 2010). Molecular prevalence of *A. phagocytophilum* in Italy investigating 300 horses reached 6.7% (Laus et al., 2013). Although information concerning prevalence data of *A. phagocytophilum* in Africa is limited, a study from Tunisia revealed a molecular prevalence of 13.0% for *A. phagocytophilum* in blood samples from 60 examined horses (M'Ghirbi et al., 2012).

The clinical course of EGA is either subclinical or acute. After approximately ten days of incubation period, acute symptoms like high fever, depression, anorexia, ataxia, icterus and lower limb oedema may occur (Silaghi et al., 2011d). Laboratory abnormalities include anaemia, leucopenia and especially thrombocytopenia possibly resulting in petechiae on mucous membranes like the inner surface of the lips or the gum. Franzen et al. (2009) described the persistence of *A. phagocytophilum* in blood of horses recovering from experimentally infected acute EGA. The course of infection largely depends on the age and condition of the infected animal (Dziegiel et al., 2013). Most of the infections are self-limiting with fatalities being uncommon for EGA (Franzen et al., 2007).

3.2.3. Dogs (*C. lupus familiaris*): Canine Granulocytic Anaplasmosis

In 1982, an infection agent resembling the causative agent of EGA was identified in two dogs from California (Madewell and Gribble, 1982). Today, Canine Granulocytic Anaplasmosis (CGA) is endemic in Europe, the USA and Asia. On the West Coast of the USA, the molecular prevalence for *A. phagocytophilum* reached 7.6% in 184 investigated dogs (Henn et al., 2007). Healthy dogs from Minnesota showed *A. phagocytophilum* infection in 3.0% of the samples, whereas dogs with clinical symptoms had a prevalence of 37.0% (Beall et al., 2008). In Europe, molecular prevalence rates of *A. phagocytophilum* in dogs vary from 5.7% in Germany (Kohn et al., 2011), 2.8 – 21.7% in Italy (Torina et al., 2008a) to 0.8% in the United Kingdom (Shaw et al., 2005). In Eastern European countries positive samples for *A. phagocytophilum* were detected in 3.4% of dogs originating from the Czech Republic (Kybicova et al., 2009) and 1.9% in stray and pet dogs from Hungary and Romania (Hamel et al., 2012). In a nationwide study in China, Zhang et al. (2012) showed that 10.9% of the 101 investigated dogs were infected with *A. phagocytophilum*. In Tunisia, two of the 228 examined dogs were infected with *A. phagocytophilum* (0.9%) (M'Ghirbi et al., 2009).

After an incubation time of app. 1 – 2 weeks, infected dogs show unspecific symptoms like acute fever, lethargy or inappetence. Additionally, lameness, coughing, polydipsia, intermittent vomiting and hemorrhages may occur (Carrade et al., 2009). Laboratory findings consist of thrombocytopenia, lymphopenia, anemia, hypoalbuminemia and an increased plasma alkaline phosphatase activity (Kohn et al., 2008). Nevertheless, most dogs show an asymptomatic course of disease proving subclinical infections caused by *A. phagocytophilum* (Carrade et al., 2009).

3.2.4. Humans (*Homo sapiens*): Human Granulocytic Anaplasmosis

In recent years, Human Granulocytic Anaplasmosis (HGA) has been regarded as an emerging zoonosis (Parola et al., 2005). The first confirmed clinical case was reported in 1994 in Minnesota, USA (Chen et al., 1994). Since then, the number of affected patients in the USA is rising steadily (Doudier et al., 2010), making it the third most common tick-borne infection after Lyme disease and Rocky mountain spotted fever (Dumler, 2005). From 2000 – 2007, the incidence of HGA reported to the CDC in the USA increased from 1.4 to 3 cases/million persons/year (Dahlgren et al., 2011). Since 1998, the active surveillance of HGA

by the CDC clearly increased the number of reported HGA cases (CDC, 2013). Especially the north eastern and upper mid-western regions of the USA are identified as endemic regions for HGA (Parola et al., 2005). Highest incidence rates were detected in Rhode Island (36.5 cases per million), Minnesota (12.3 cases per million), Connecticut (8.1 cases per million), New York (2.3 cases per million) and Maryland (1.6 cases per million) (Chapman et al., 2006). In 1997, the first clinical case was confirmed in Europe (Slovenia) (Petrovec et al., 1997). Several more cases followed in different European countries including the Netherlands (Van Dobbenburgh et al., 1999), Spain (Oteo et al., 2000), Sweden (Karlsson et al., 2001), Poland (Tylewska-Wierzbanowska et al., 2001; Welc-Faleciak et al., 2015), Austria (Walder et al., 2003), Italy (Ruscio and Cinco, 2003), Estonia (Prukk et al., 2003), Austria (Haschke-Becher et al., 2010), Belgium (Cochez et al., 2011), France (Koebel et al., 2012) and Germany ex Scotland (Hagedorn et al., 2014). Nevertheless, less clinical cases were reported in Europe compared to the USA. Data of *A. phagocytophilum* infections in Asia are very limited. Available information indicates a more severe course of disease in Asian HGA patients compared to European or patients from the USA (Wang et al., 2013). Li et al. (2011) described a mortality rate of 26.5% in a cohort of 83 Chinese HGA patients. Especially HGA patients developing complications like SIRS (Systemic inflammatory response syndrome) and/or MODS (multiple organ dysfunction syndrome) were at high risk of death (Li et al., 2011). The seroprevalence of *A. phagocytophilum* in healthy Chinese farm workers with a frequent exposure to vectors, hosts and potential reservoir hosts of the pathogen reached 8.8%, whereas in Korean patients with acute febrile disease a seroprevalence of 1.8% was detected (Heo et al., 2002; Zhang et al., 2008). In Japan, four serologically diagnosed cases of HGA occurred in clinically manifest patients, whereby three of the four patients were confirmed by a paired serological testing in the convalescent phase (Yoshikawa et al., 2014). In 2003, the first case of HGA was detected in Russia (Sidel'nikov Iu et al., 2003).

Main clinical symptoms of HGA are fever, headache, myalgia and malaise. Gastrointestinal and respiratory symptoms occur less frequently. Laboratory abnormalities include thrombocytopenia, leukopenia, anemia and increased serum transaminase activities (Dumler, 2005). The severity of clinical disease varies from mild self-limiting fever to fatality depending on risk factors like age or immunosuppression. The hospitalization rate of HGA was observed in 36.0% of

the cases, Intensive Care Unit (ICU) admission in 7.0% and death in 0.6% of the cases in the USA (Dumler, 2012). Nevertheless, high seroprevalences suggest the existence of either high subclinical infection rates or an underdiagnosing of HGA cases (Dumler et al., 2005). However, due to cross-reactivity with other *Anaplasma spp.* like *Ehrlichia chaffeensis*, high seroprevalences might also be responsible for an overestimation of HGA (Ismail et al., 2010).

3.2.5. Hosts with a potential to develope clinical symptoms

Although only few reports are published, cats (*Felis silvestris catus*) are considered possible hosts for *A. phagocytophilum*. The first clinical case of Feline Granulocytic Anaplasmosis was reported in Sweden (Bjoersdorff et al., 1999). In Poland, three clinical cases of cats were published (Adaszek et al., 2013). A German study investigating 265 cats showed a molecular prevalence of 0.5% (Morgenthal et al., 2012). Clinically morbid cats from Massachusetts, USA, were confirmed positive for *A. phagocytophilum* infection by molecular and serological analysis (Lappin et al., 2004). Another study from the Northeastern US showed 16 cats infected with *A. phagocytophilum* (Savidge et al., 2015). Cats with clinical cases of *A. phagocytophilum* infection showed fever, anorexia and lethargy. Less common clinical findings included conjunctivitis, swollen joints and ataxia. The treatment with antibiotics resolved the clinical signs in all affected cats rapidly (Adaszek et al., 2013; Lappin et al., 2004; Savidge et al., 2015).

Although wild ruminants are considered as reservoir hosts for *A. phagocytophilum,* two clinical cases of naturally infected roe deer calves were described by Stuen et al. (2001); (2006) in Norway. Both animals were heavily infested by *I. ricinus* ticks and died. Furthermore, a reindeer (*Rangifer tarandus tarandus*) was shown to develop severe illness after experimental infection with *A. phagocytophilum* (Stuen, 1996).

4. Diagnostics, Treatment and Prevention

Clinical findings and a history of tick-bite or tick exposure are the basis for suspicion of an *Anaplasma*-infection. The CDC provides guidelines for HGA diagnosis in Case Definitions for Infectious Conditions Under Public Health Surveillance (Centers for Disease Control and Prevention, 1997):

- In the acute and convalescent phase: Fourfold or greater change in antibody titer by immunofluorescence antibody (IFA) test, ideally taken

≥ 4 weeks apart.

- Positive PCR assay.

Probable cases include clinical compatible symptoms with either a single immunofluorescence antibody (IFA) serologic titer ≥ 64 or intracytoplasmic morulae identified in blood, bone marrow, or cerebrospinal fluid (CSF) leukocytes in the acute and convalescent phase. Further laboratory diagnostics in addition to clinical compatible symptoms are required for confirmation of a probable case (Centers for Disease Control and Prevention, 1997).

During the acute febrile phase, blood smears of infected animals show cytoplasmic inclusions in monocytes and leucocytes, especially in neutrophils (Ismail et al., 2010). Serological tests may support the diagnosis, but due to cross-reactivity among *Anaplasmataceae* species the reliability is limited (Ismail et al., 2010). The most specific and sensitive verification for an infection with *A. phagocytophilum* is the PCR.

For treatment, drug of choice is the antibiotic tetracyclines and within this antibiotic family especially doxycycline is frequently used for both pediatric and adult clinical cases (Ismail et al., 2010). This antibiotic group shows bacteriostatic effects on bacteria by inhibiting the protein biosynthesis. Most clinical cases improve within 24 – 48 hours after medication. The most successful prevention of an *Anaplasma* disease consists of avoidance of tick bites or in case of an infestation, a removal of the tick as soon as possible. An effective transmission of the pathogen requires less than 24 hours of tick infestation (Des Vignes et al., 2001). To date, no vaccines against *A. phagocytophilum* are available.

5. Genetic diversity of *A. phagocytophilum*

5.1. Genetics of *A. phagocytophilum*

5.1.1. Whole genome sequencing of *A. phagocytophilum*

A. phagocytophilum has a relatively small genome (1.47 Mbp) consisting of a single circular chromosome (Dunning Hotopp et al., 2006). The relatively small size is typical for obligate intracellular bacteria focusing on essential genes for basic genetic functions (Andersson and Kurland, 1998). Therefore, the genome of *A. phagocytophilum* is highly concentrated on genes encoding nucleotide biosynthesis, cofactor and vitamin biosynthesis and protein synthesis. Genes for transport or regulatory functions are less developed (Dunning Hotopp et al.,

2006). *A. phagocytophilum* shows numerous repeats in its genome. Especially repetitive sequences encoding immunodominant outer membrane proteins or cofactor and vitamin biosynthesis were shown by Dunning Hotopp et al. (2006) in the comparison of different representatives of the family *Anaplasmataceae*.

To date, twenty genomes of *A. phagocytophilum* have been published (www.ncbi.nlm.nih.gov/genome/genomes/1083?). The hosts of the investigated strains included eight humans (samples: HZ, HZ2, HGE1, HGE2, ApWI1, ApNYW, Webster, NCH-1), three dogs (sample: Dog2, ApMUC09, ApNP), two horses (sample: MRK, Annie), a sheep (sample: Norway variant2), a cow (sample: BOV_10-179), two rodents (sample: JM, CR1007) and three ticks (samples: CRT38, CRT35, CRT53-1) (Barbet et al., 2013; Dugat et al., 2014b; Dunning Hotopp et al., 2006). Nevertheless, only the genomes of three human strains have been fully sequenced without gaps. All samples originated from the USA except from a sheep sample from Norway, a cow from France, a dog from the Netherlands and a dog from Austria. While sequencing the whole genome of *A. phagocytophilum* originating from the French bovine sample, Dugat et al. (2014b) detected nine gene clusters unique to domestic ruminant strains from Europe and four additional clusters for the bovine sample. These clusters could be responsible for special host tropism or specific differences due to different origins of the samples (Dugat et al., 2014b).

5.1.2. Protein-coding genes

5.1.2.1. *Major surface protein* – genes

Major surface proteins play an important role for the adaptive immune response against *A. phagocytophilum* of susceptible hosts. As variable immunodominant antigens, they basically determine the success of an infection with the tick-borne pathogen and its persistence in the host (Brown, 2012).

5.1.2.1.1. *Major surface protein 2* – gene

A. phagocytophilum shares an immunodominant protein called *major surface protein* 2 (*msp2*) (app. 1.098 bp) with the obligate intraerythrocytic pathogen *A. marginale* (Rurangirwa et al., 1999). This major outer membrane protein is encoded by polymorphic multigene families and amounts the size of 44 kDa. The *msp2* gene is characterized by conserved terminal sequences flanking a hypervariable region (Murphy et al., 1998). There are several copies of the *msp2*

gene, so-called pseudogenes (app. 100 pseudogenes), distributed all over the genome of *A. phagocytophilum.* By recombination of these variable pseudogenes, new antigenic variants of the major surface protein are created enabling evasion of susceptible host immune systems (Brayton et al., 2001). The recombination of these pseudogenes follows no specific pattern except that some variants are expressed more often than others (Rejmanek et al., 2012). In the course of a chronic infection with *A. phagocytophilum* cyclic waves of bacteraemia, occurring approximately two to three weeks apart, were detected in experimentally infected sheep (Granquist et al., 2008). Thus, *A. phagocytophilum* temporarily circumvents the cellular and humoral defence mechanism of the (reservoir) host`s immune response enabling a reappearance at a later stage of infection.

Comparing the structure of different *msp2* strains, similarities were shown among strains from different animal hosts of US origin. In contrast, great differences were revealed when comparing the same US strains with European canine and ovine strains (Barbet et al., 2006). Therefore, worldwide diversity of the *msp2* gene is assumed. Several studies from Asia investigating the *msp2* gene in ticks and naturally infected deer also revealed great differences among each other, possibly due to different geographic origin or unselective expression of the gene (Gaowa et al., 2012). Further studies are required to detect variation patterns for the occurrence of specific *msp2* strains in order to understand the variability of *A. phagocytophilum.* Due to its high ability of recombination, a clustering of *msp2* sequences seems rather difficult. For example, the alignment of paralogue sequences from ruminants and ticks originating from the United Kingdom (UK) compared to US strains did neither cluster by country, host species nor by isolate (Casey et al., 2004).

5.1.2.1.2. *Major surface protein 4* – gene

The *major surface protein* 4 (*msp4*) (app. 849 bp) is one of the immunodominant major surface proteins expressed on the outer membrane of *A. phagocytophilum.* To date, its exact biological function is unknown, but the gene supposedly plays a role in host-pathogen interactions (De la Fuente et al., 2005a). Due to selective pressure of the immune system of the host, more intense evolution of the expression of the *msp4* gene in comparison to nuclear genes is expected (De La Fuente et al., 2003). The *msp4* gene is part of the *msp2* superfamily (Brayton et al., 2006) and shows great similarities to the well explored *msp4*-gene of

A. marginale (De la Fuente et al., 2005a). In contrast to the *msp2*-gene, the *msp4* is encoded by a single gene. Therefore, genetic diversity seems to be limited making it a useful and stable tool for phylogenetic and phylogeographic analysis (De La Fuente et al., 2003; De la Fuente et al., 2002). The investigation of the *msp4* gene from European and US strains resulted in a clustering of ruminant strains versus strains from other hosts including humans (De la Fuente et al., 2005a). Due to its limited potential of variation, the *msp4* also represents a possible target for the development of future vaccines against *A. phagocytophilum* (Brayton et al., 2006).

5.1.2.2. Heat shock protein *groEL* – gene

The protein of *groEL* is part of the so-called heat shock protein-group (*HSP*) and is equivalent to the eukaryotic *HSP60*, named after its molecular weight. Heat shock proteins are upregulated in physiological stress situations, for example during increased temperatures or toxic burden, and therefore function as a protection tool for cells (Dasch et al., 1990). The *groEL*-gene (app. 573 bp) is said to be highly conserved and is shared by both prokaryotic and eukaryotic organisms and organelles. The *groEL* is one of the larger products of two genes (*groEL* ["l" standing for large] and *groES* ["s" standing for small]) unified in the *groESL* gene. The term "gro" was chosen for the experimentally verified limit of growth of bacteriophages induced by mutation of the λ*E*-gene (Georgopoulos and Welch, 1993). The sequencing of the *groEL*-gene is considered a useful tool for phylogenetic analysis of *A. phagocytophilum.* Particularly in cases where the *16S rRNA* analysis is limited because of high conservation, the *groEL* gene may support and expand phylogenetic results (Dasch et al., 1990; Jahfari et al., 2014; Sumner et al., 1997; Vichova et al., 2014). For example, Jahfari et al. (2014) clustered European *A. phagocytophilum* strains on the basis of the *groEL* gene from diverse hosts into four different ecotypes with differential enzootic cycles. In Asian studies, the *groEL* gene was compared examining *A. phagocytophilum* from wild ruminants. Thereby, new *groEL* sequences were detected distinct from the sequences occurring in Europe and the USA (Kang et al., 2011; Kawahara et al., 2006). About 50 variants of the gene were detected and deposited in the GenBank. According to Rar and Golovljova (2011), most of these variants can be clustered in two groups. The first group includes *A. phagocytophilum* from various vertebrate hosts and humans and the second mainly originates from roe deer and

ticks.

5.1.2.3. Ankyrin repeat protein (*ankA*) – gene

The *ankA* gene of the human *A. phagocytophilum* HZ strain encodes eleven ankyrin repeats (Rikihisa and Lin, 2010). These ankyrin repeat proteins form scaffolds for protein-protein interactions like cell-cell signaling or cell-cycle regulation (Mosavi et al., 2004). *AnkA* of *A. phagocytophilum* in particular influences the transcription of gene expression in infected host cells, like the neutrophilic granulocytes. Thereby, the protein accumulates in the nucleus of the host cell and modifies the chromatin structure evoking transcriptional changes (Garcia-Garcia et al., 2009).

In order to gain information about genetic diversity of *A. phagocytophilum* infecting different animal species, the *ankA* gene (app. 3.690 bp) has been determined in several studies (Majazki et al., 2013; Michalik et al., 2012; Scharf et al., 2011a). For example, Majazki et al. (2013), analyzed the *ankA* gene of infected shrews and voles and detected a DNA strain cluster completely distinct from other host animals, possibly developed from gene recombination. Scharf et al. (2011a) detected four distinct clusters of *A. phagocytophilum* strains infecting diverse animal species on the basis of sequencing of the *ankA* gene. The majority of roe deer samples clustered in a group distinct from domestic animals suggesting great variation in the strains occurring in roe deer (Scharf et al., 2011a). According to *ankA* gene sequences, Massung et al. (2000) discriminated three clades of distinct HGA agents on the basis of their origin: Northeastern US, Upper Midwest US and Europe and concluded independent evolution of the diverse strains of *A. phagocytophilum* in these three regions.

5.1.3. Non-protein-coding

5.1.3.1. Partial *16S rRNA*-gene

The *16S ribosomal RNA* is part of the 30S subunit of prokaryotic ribosomes and takes a structural and functional role in the protein synthesis of bacteria. The *16S rRNA* gene (app. 1.4 kbp) is a highly conserved gene shared by all bacteria (Woese et al., 1975). PCR amplification of this gene is traditionally used for both epidemiological and diagnostic purposes. Because of its broad distribution, its functional constancy and its slow nucleotide sequence evolution, phylogenetic relationships of different bacteria species can be determined easily (Woese and

Fox, 1977). The determination of the *16S rRNA* gene seems to be a good method for an initial classification of *A. phagocytophilum* infective strains, but provides too little information for in-depth investigation (Massung et al., 2000; Silaghi et al., 2011d). Based on studies of the *16S rRNA* gene and supported by *groEL* analyses, Dumler et al. (2001) proposed a reorganization of the family *Anaplasmatacea* resulting in unifying *Ehrlichia phagocytophilum, Ehrlichia equi* and the agent of *human granulocytic ehrlichiosis* (HGA) in one species: *A. phagocytophilum.* Information about phylogenetic and evolutionary relationships among different *A. phagocytophilum* strains on the basis of the *16S rRNA* gene is limited due to high conservation and unspecificity of the *16S rRNA* gene (Rar and Golovljova, 2011; Roux et al., 2011).

Compared to other genes common for phylogenetic analysis of *A. phagocytophilum*, less *16S rRNA* variants are found in the GenBank database (Rar and Golovljova, 2011). Three of the variants were detected all over the world in several animal hosts and ticks (Chen et al., 1994; Massung et al., 2002; Rar et al., 2011). One of them is known as the human-pathogenic "prototype" of *A. phagocytophilum* (accession number: U02521). In 1994, the latter *A. phagocytophilum* variant was the first to be detected as causative agent for HGA (Ap-ha) (Chen et al., 1994). Since then, other human cases but also host animals and ticks infected by the prototype followed (Massung et al., 2002; Scharf et al., 2011a; Stuen et al., 2006; Von Loewenich et al., 2003a). In contrast, the Ap-variant 1 (acc. no.: AY193887) has only been detected in ruminants and ticks so far and is therefore considered to be apathogenic for humans (Courtney et al., 2003; Massung et al., 2002). The other wide spread partial *16S rRNA* sequence described in several host species and ticks is assigned to the accession number AF093789 (CAHU-HGE1) (Chae et al., 2000; Rar and Golovljova, 2011). This *A. phagocytophilum* strain was originally detected in HGA patients.

5.1.4. Other genes

In addition to the sequencing of the *16S rRNA*, the partial citrate synthase gene (*gltA*) of *A. phagocytophilum* from rodents and sheep from China was investigated in order to further classify the pathogen (Cao et al., 2006; Zhan et al., 2010a). The detected nucleotide sequences clearly differed from US and Russian *gltA* sequences and therefore defined a separate Chinese cluster (Zhan et al., 2010a). Bovine samples of *A. phagocytophilum* were sequenced with a high degree of

specificity detecting three new partial *23S rRNA* gene variants of *A. phagocytophilum* (Dahmani et al., 2015). Similar to the partial *16S rRNA*, this gene supports the protein synthesis of bacteria and therefore represents a part of the larger subunit 50S of the prokaryotic ribosome (Lin et al., 2011). The *major surface protein 5* (*msp5*) gene is another highly conserved surface protein analyzed in various *A. phagocytophilum* isolates from the USA and Europe (Strik et al., 2007). Additionally to the *groEL*, other *heat shock protein* genes, like the *heat shock protein 60* gene, were analyzed (Kolbert et al., 1997). Principal distinction of *A. phagocytophilum* from heat shock protein genes of other pathogens like *E. chaffeensis* has been described, but intraspecifical discrimination seems to be limited (Kolbert et al., 1997).

Huhn et al. (2014) used multilocus sequencing to characterize human and animal strains of *A. phagocytophilum* mostly originating from Europe. Seven housekeeping genes were analyzed: the phenylalanyl-tRNA-synthetase alpha subunit (*pheS*) gene, NAD-dependant malate dehydrogenase (*glyA*) gene, fumarate hydratase class II (*fumC*) gene, DNA Polymerase III beta subunit (*dnaN*) gene, ATP synthase subunit alpha (*atpA*) gene, NAD-dependent malate dehydrogenase (*mdh*) gene and 2-oxoglutarate dehydrogenase E1 component (*sucA*) gene. A common clonal complex was assumed for *A. phagocytophilum* strains from humans, domestic animals, hedgehogs and wild boars. Only distant relatedness was shown for *A. phagocytophilum* from roe deer, voles and shrews (Huhn et al., 2014).

5.2. Strain diversity

Several vector-borne pathogens of humans and animals have developed different strategies to be successfully transmitted to their hosts and maintain their natural life cycle. Particularly arthropod-transmitted pathogens use different genetic mechanisms in order to prolong their presence in the blood circulation and therefore increase the likelihood of transmission (Barbour and Restrepo, 2000). Thus, *A. phagocytophilum* shows great genetic variety in strains infecting different susceptible hosts. Although currently being reorganized as one single species (Dumler et al., 2001), genetic studies have shown differences in nucleotide sequences of diverse genes of *A. phagocytophilum*. The genetic variability of the pathogen reflects the potential of *A. phagocytophilum* to develop new strategies of infection possibly resulting in altered host tropism,

pathogenicity or persistence of infection. Foley et al. (2008a) described a possible host transmission barrier by the lack of infection potential of *A. phagocytophilum* strains in different animal species. For example, the strain of a wood rat, a putative reservoir host of *A. phagocytophilum* in the USA, did not cause clinical disease in horses when experimentally infected. Besides, the genetic variability of *A. phagocytophilum* increases the capacity of the pathogen to persist in its host (Palmer et al., 2009). Especially the antigenic variation of immunodominant major surface proteins, like *msp2*, could possibly be the reason for persistent infection (Brown, 2012). By modifying immunodominant antigens, *A. phagocytophilum* has developed a way to circumvent the immune system of its hosts and thus allow its persistence. Persistence of *A. phagocytophilum* was confirmed in several animal species, including sheep, dogs and horses (Egenvall et al., 2000; Foggie, 1951; Franzen et al., 2009).

6. Enzootic life cycles of *A. phagocytophilum*

6.1. Possible endemic cycles of *A. phagocytophilum* occurring in nature

6.1.1. Endemic life cycles of *A. phagocytophilum* in Europe

In Europe, phylogenetic analysis on the basis of *groEL* sequences of *A. phagocytophilum* from different animal and tick samples revealed a clustering into four different ecotypes (Jahfari et al., 2014). Each ecotype represented a possible endemic life cycle of *A. phagocytophilum* sharing similar genetic variants of the bacterium (Fig. 5). The first ecotype included the broadest host range with wild (mouflons, red deer, roe deer, hedgehogs, wild boars, mice) and domesticated animals (cattle, sheep, goats, horses, dogs). Besides, all of the 34 European *A. phagocytophilum* strains originating from humans with HGA symptoms referred to the same ecotype. A second ecotype was represented by roe deer as hosts. Deer keds were also described as part of the same ecotype, possibly resulting from a blood meal infesting roe deer. Further, separate ecotypes were suggested for rodents and birds (Jahfari et al., 2014). As strains from rodents and birds lacked in ecotype I, Jahfari et al. (2014) concluded that these species might play a minor role in the zoonotic transmission cycle of *A. phagocytophilum* in Europe. The vector responsible for the transmission of *A. phagocytophilum* differed in each life cycle. The most common tick in Europe, *I. ricinus*, was suggested as vector of ecotype I. *I. persulcatus* was part of the second ecotype.

Vectors of ecotype III and IV are probably rodent and bird specific ticks, respectively, but were not investigated by Jahfari et al. (2014). Rodents and birds could therefore be involved in separate alternative endemic subcycles for specific *A. phagocytophilum* strains.

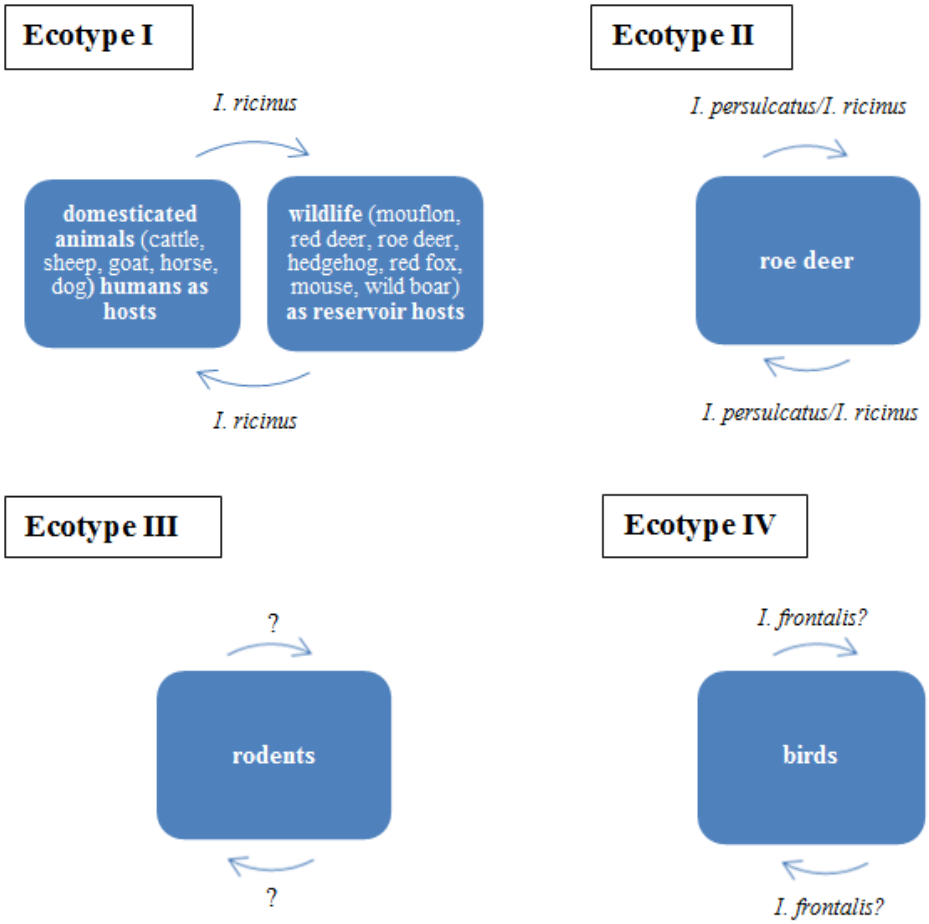

Figure 5: Four suggested ecotypes of *A. phagocytophilum* based on *groEL* strains from Europe (Jahfari et al., 2014)

Examining nine different gene loci, Chastagner et al. (2014) detected three possible clusters of *A. phagocytophilum* strains in France comparing cattle samples with horse, dog and roe deer samples (Fig. 6). One cluster comprised *A. phagocytophilum* strains of clinically manifest domestic animals, like cattle, horses and dogs (Cluster C). Huhn et al. (2014) confirmed this cluster and additionally proposed humans as hosts and wild boars, hedgehogs and red deer as reservoir hosts for the same cluster. A separate cluster included strains specific for cattle (Cluster B). According to similar *A. phagocytophilum* strains, other studies sporadically found sheep, red deer and roe deer in the same cluster (Majazki et al., 2013; Scharf et al., 2011a). Noticeably, most variants originating from roe deer and two cattle strains were part of another distinct cluster (Cluster A).

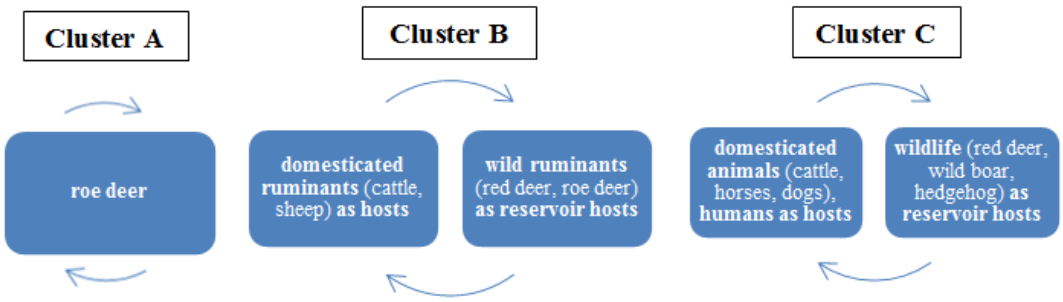

Figure 6: Three possible clusters of *A. phagocytophilum* strains in France [according to Chastagner et al. (2014) and confirmed by Huhn et al. (2014) and Scharf et al. (2011a)]

In Europe, Bown et al. (2009) proposed two distinct co-existing enzootic cycles: One consisted of roe deer as host and *I. ricinus* as vector. The other included field voles (*Microtus agrestis*) as host and assumed *I. trianguliceps* as vector, which was proposed as an alternative subcycle of *A. phagocytophilum* (Fig. 7). These two cycles are considered to co-exist independently of each other in nature.

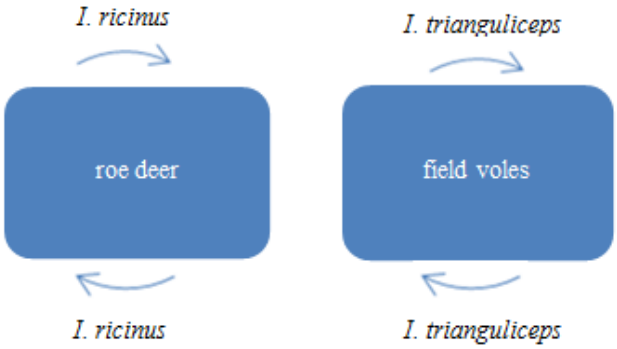

Figure 7: The two co-existing endemic cycles of *A. phagocytophilum* with roe deer and voles [according to Bown et al. (2009)]

Previously, European wild boars (*Sus scrofa*) were discussed as potential reservoirs for *A. phagocytophilum*. Prevalence data of *A. phagocytophilum* in wild boar samples and ticks (*I. ricinus*) infesting this animal species led to the assumption, that wild boars are part of the natural enzootic cycle (Michalik et al., 2012; Petrovec et al., 2003; Silaghi et al., 2014). As *A. phagocytophilum* strains detected in wild boars matched with variants causing HGA in humans, a potential zoonotic risk was assumed (Michalik et al., 2012; Petrovec et al., 2003). Huhn et al. (2014) showed a common cluster for wild boar, human and domestic animal

strains based on multilocus sequence typing of *A. phagocytophilum* (Fig. 6, Cluster C). Naturally and experimentally infected wild boars were susceptible for *A. phagocytophilum* infection, but also showed the ability to control the infection and thus raised the question of their role in the enzootic cycle of *A. phagocytophilum* (De la Fuente and Gortazar, 2012).

6.1.2. Endemic life cycles of *A. phagocytophilum* in the USA

In the USA, two main *A. phagocytophilum* life cycles are considered on the basis of two strains differentiated by the *16S rRNA* gene (Massung et al., 2002). Firstly, the AP-ha strain exists in nature infecting humans and white-footed mice as natural reservoir (Massung et al., 2003). The same strain was also detected in dogs (*C. lupus familiaris*) from the USA suggesting a common evolutionary origin (Fig. 8) (Morissette et al., 2009). Secondly, AP-V1 was shown in white-tailed deer as reservoir host, which was neither associated with human infection nor with rodents. White-tailed deer were experimentally confronted with *I. scapularis* transmitting both variants, AP-ha and AP-V1. Thereby, only the strain AP-V1 was successfully transmitted to white-tailed deer. Strain AP-ha could not be transmitted experimentally (Massung et al., 2005). As goats were infected by the same AP-V1 strain, Massung et al. (2006) assumed a possible discrimination between ruminant and non-ruminant strains. Besides, a Norwegian sheep was also infected with the AP-V1, wherefore the AP-V1 was believed to represent a common marker for ruminant tropism (Fig. 9) (Morissette et al., 2009).

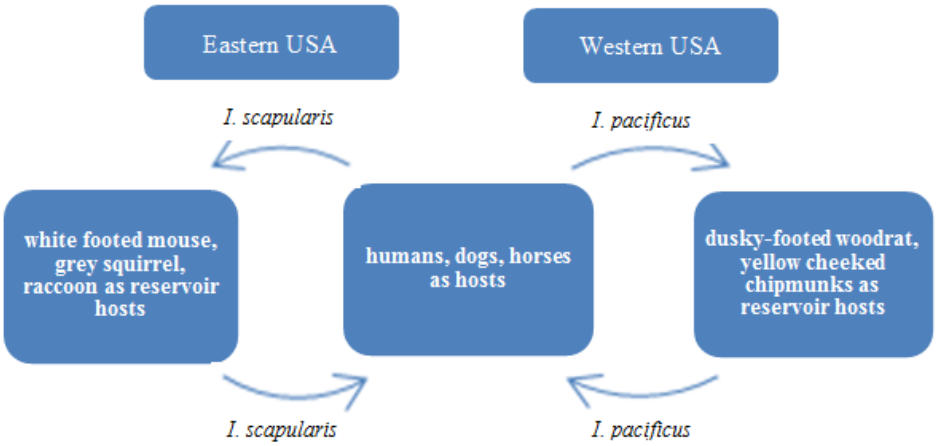

Figure 8: Animal and tick species possibly infected with the Ap-ha strain in the USA (Massung et al., 2003; Morissette et al., 2009)

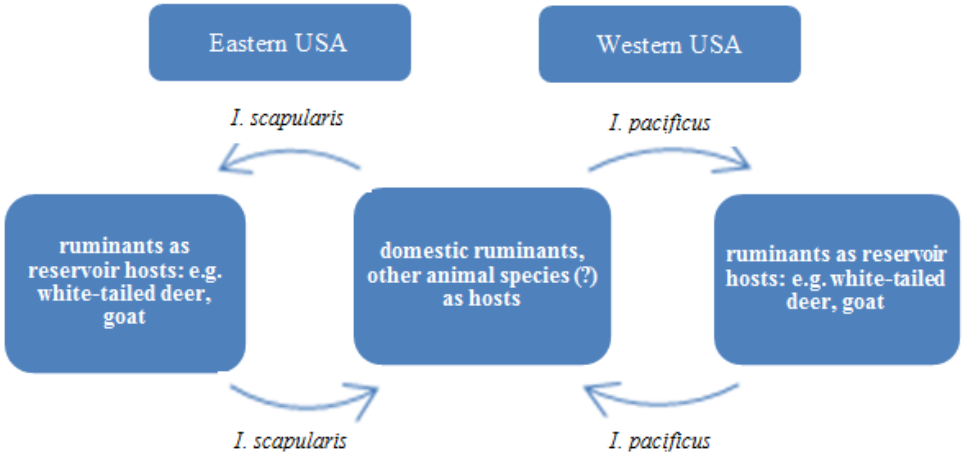

Figure 9: Animal and tick species possibly infected with the Ap-variant 1 in the USA (Massung et al., 2005; 2006)

6.1.3. Endemic life cycles of *A. phagocytophilum* in Asia

In Asia, rodents and wild ruminants like Korean water deer or sika deer are considered reservoir hosts for *A. phagocytophilum*. Nevertheless, experimental studies verifying distinct life cycles of *A. phagocytophilum* in Asia are lacking so far (Zhan et al., 2008). As livestock and small rodents were infected with the same *msp2/p44* variant of *A. phagocytophilum*, Asian life cycles of *A. phagocytophilum* may include both ruminants and rodents (Fig. 10) (Zhan et al., 2010b). In contrast to these findings, distinct life cycles of rodent and ruminant strains of *A. phagocytophilum* are considered in Europe (Bown et al., 2003).

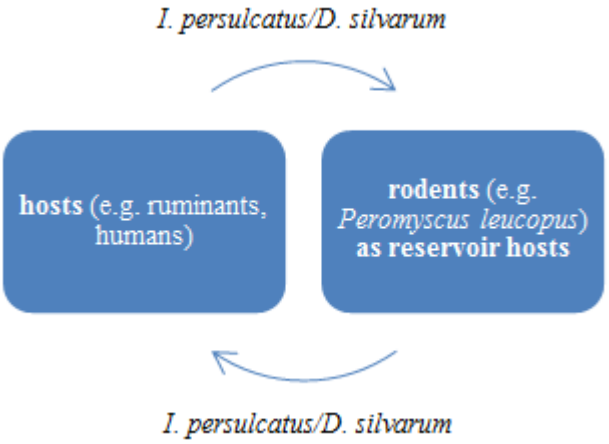

Figure 10: Example of a common enzootic cycle of *A. phagocytophilum* with hosts like ruminants and rodents in Asia [according to Zhan et al. (2010b) and Cao et al. (2006)]

6.2. Niche cycles of *A. phagocytophilum*

In Germany, the European hedgehog (*E. europaeus*) has been proposed as niche reservoir host. In combination with the vector tick *I. ricinus* and *I. hexagonus*, hedgehogs might form a subcycle of *A. phagocytophilum* transmitting variants even pathogenic for humans (Silaghi et al., 2012a) (Fig. 11). Another insectivore discussed as reservoir host in England is the common shrew (*Sorex araneus*) (Bown et al., 2011). Besides, only a further study in Switzerland detected *A. phagocytophilum* infection in common shrews (Liz et al., 2000).

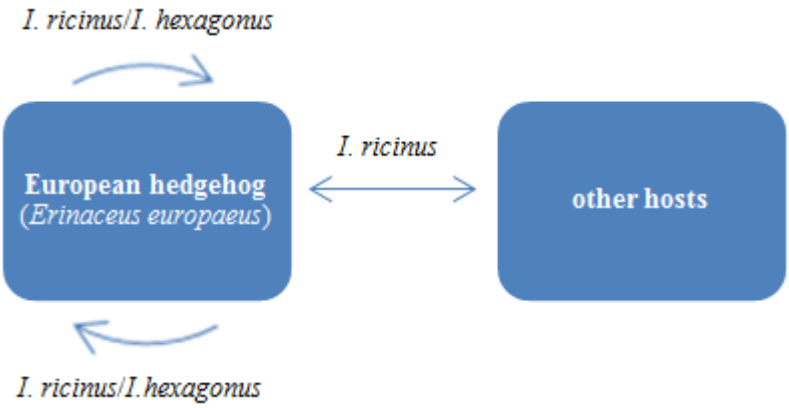

Figure 11: Enzootic subcycle of *A. phagocytophilum* with the European hedgehog and ticks as vectors (Silaghi et al., 2012a)

In Massachusetts, USA, the cottontail rabbit (*Sylvilagus floridanus*) is considered a reservoir host for the Ap-ha variant of *A. phagocytophilum*. With *I. dendatus* as vector, this rabbit species is suggested to be involved in an enzootic subcycle (Goethert and Telford, 2003) (Fig. 12).

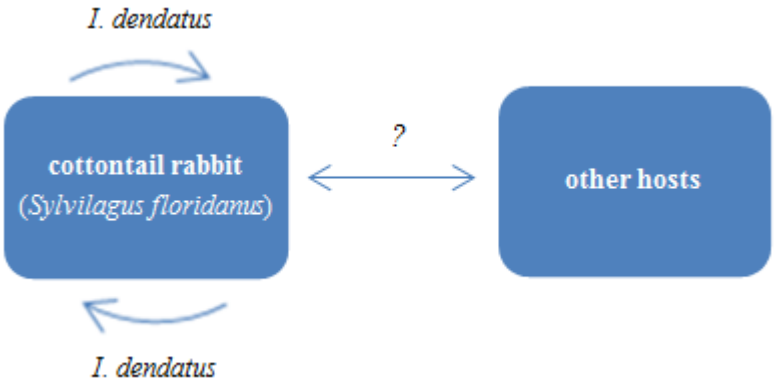

Figure 12: Enzootic subcycle of *A. phagocytophilum* with the cottontail rabbit and ticks as vectors (Goethert and Telford, 2003)

Moreover, high prevalence rates of *A. phagocytophilum* in deer mice (*Peromyscus maniculatus*) and Mexican woodrats (*Neotoma mexicana*) indicate reservoir competency of these rodents in Colorado, USA. A potential vector in an alternate natural cycle of *A. phagocytophilum* could be represented by *I. spinipalpis* (Zeidner et al., 2000) (Fig. 13).

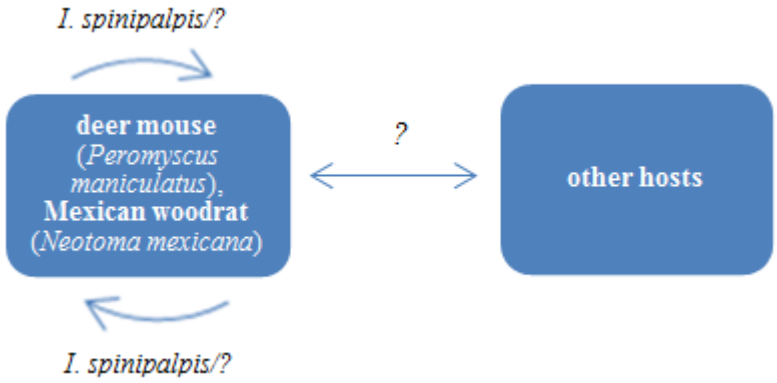

Figure 13: Enzootic subcycles of *A. phagocytophilum* with the deer mouse and the Mexican woodrats and ticks as vectors (Zeidner et al., 2000)

III. MATERIAL AND METHODS

1. Overview of the workflow

The concept of the study is shown in Fig. 14. Three strategies to obtain nucleotide sequences were used with the aim of a final dataset of sequences of all four genotyping genes for every sample.

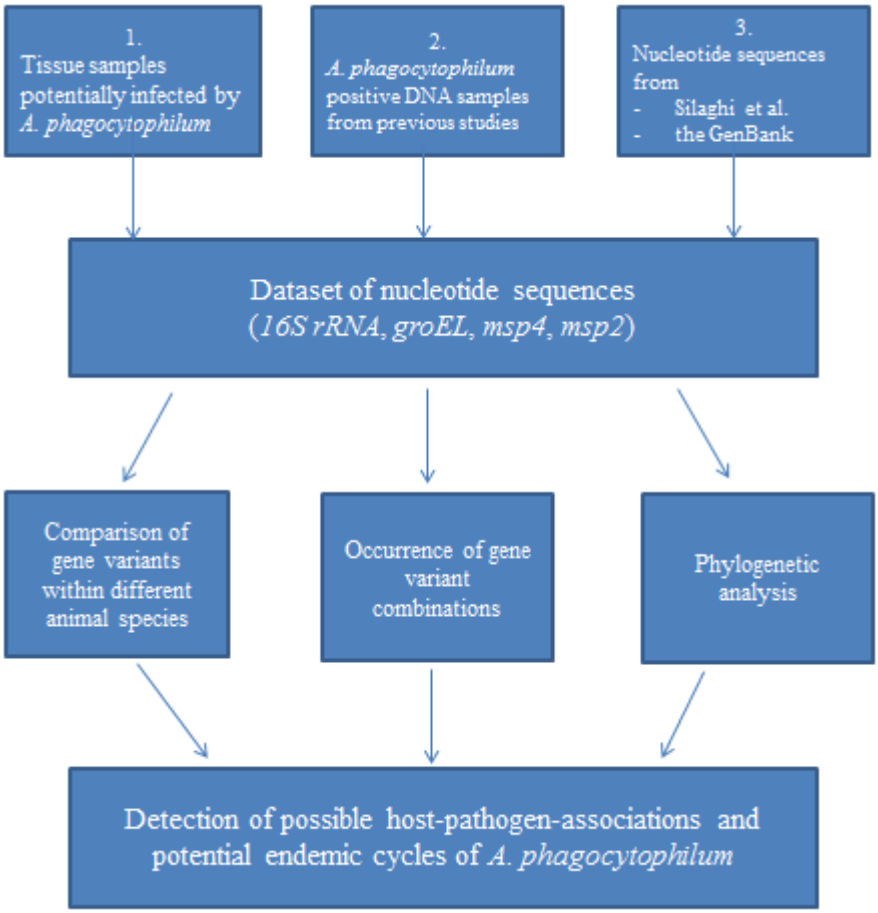

Figure 14: Concept of the study

Steps 1 – 3 describe the proceedings of samples according to the origin of the sample (Fig. 14). After tissue of diverse animal species had been collected (Fig. 14, step 1), DNA was extracted. DNA samples of *A. phagocytophilum*, which had already been screened positive in previous studies, were directly used for PCR amplification of the four partial genes (Fig. 14, step 2). The workflow in the laboratory is shown in Fig. 15. After sequencing, the nucleotide and amino acid sequences were compared to each other and to additional sequences from previous

studies and to sequences from the GenBank (Fig. 14, step 3).

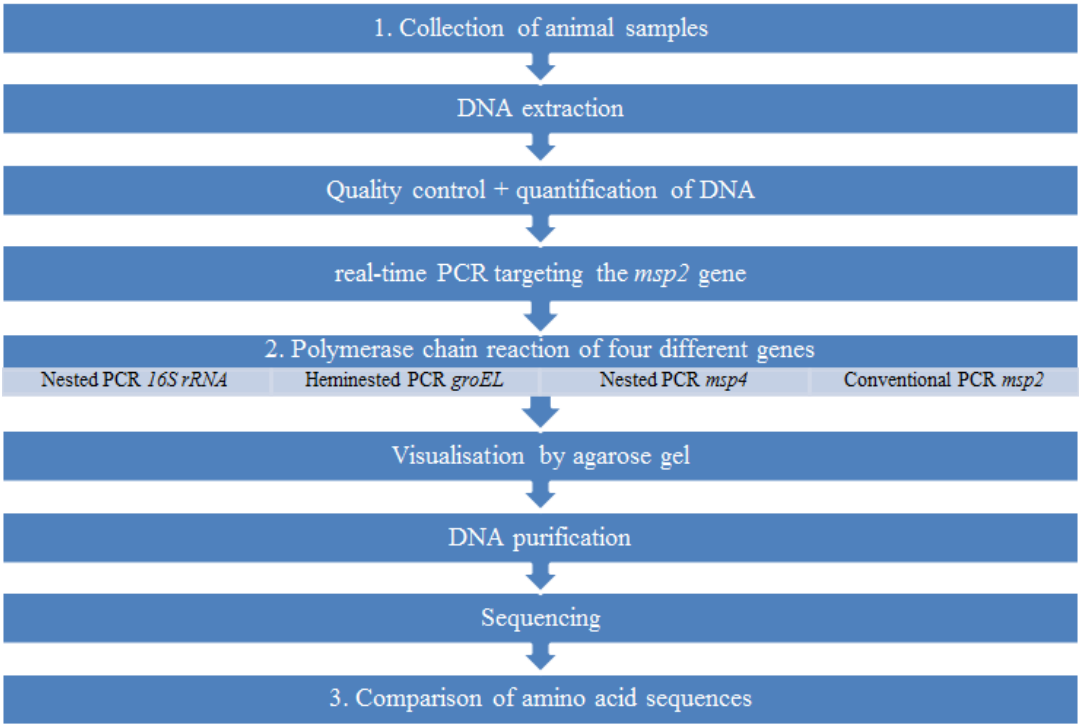

Figure 15: Laboratory workflow after sample collection

2. Animal samples

In total, 781 samples from 17 different animal species (dog, horse, cat, cattle, goat, hedgehog, red fox, roe deer, red deer, sika deer, fallow deer, mouflon, chamois, ibex, wild boar, bank vole, wood mouse) from Central Europe suspected to be hosts for *A. phagocytophilum* were included in this study (Table 1, Fig. 16). Of these, 356 samples were collected for the present study and used for DNA extraction. The additional 425 DNA samples were available from previous studies and had already been screened positive by real-time PCR for *A. phagocytophilum* (Tab. 1). Of those 425 samples, 263 had already been investigated at least in parts with genotyping PCRs for *A. phagocytophilum* in previous studies. In total, samples originated either from tissue (483 spleen, one liver and one ear sample) or from blood (296).

Figure 16: A map of Europe showing the origin of the animal samples included in this study

Table 1: Origin of the *A. phagocytophilum*-positive animal samples available for this study

Animal species	n	Origin	Material	Year of collection (n)	Reference (n[1])
Cat (*Felis silvestris catus*)	1	Finland	blood	2008	Heikkilä et al. (2010)
Cattle (*Bos primigenius taurus*)	28	Germany		2011	Nieder et al. (unpublished)
	8	Germany	blood	2002	Silaghi et al. (2011e) (8)
	9	Switzerland		2005	Silaghi et al. (2011e)
Chamois (*Rupicapra r.rupicapra*)	10	Austria	spleen	2008 (4) / 2009 (2), partially unknown (4)	Silaghi et al. (2011a) (11)
Dog (*Canis lupus familiaris*)	6	Spain	blood	2009	Hamel et al. (2013), sequences from this study
	3	Italy		2009	
	1	Hungary	blood	2009	Hamel et al. (2012), sequences from this study
	3	Romania		2009 (2) / 2010 (1)	
	4	Germany	blood	2009 (3) / 2010 (1)	Silaghi et al. (2011c) (55), sequences from this study
	16	Germany		2008 (3) / 2009 (13)	
	14	Germany		2008 (1) / 2009 (13)	
	5	Germany		2008 (1) / 2009 (2), partially unknown (2)	
	2	Slovenia		2009	
	1	Czech Republic		2009	
	1	Turkey		2009	
	9	Germany	blood	2009 (1) / 2010 (6), partially unknown (1	Sequences from this study
	1	Germany	blood	2010	Sequences from this study
	1	Switzerland	blood	2010	Sequences from this study
	7	Albania	blood	2009 (1) / 2010 (2), partially unknown (4)	Sequences from this study
Fallow deer (*Dama dama*)	5	Germany	spleen	2013	This study
	3	Germany		2009	

Red fox (*Vulpes vulpes*)	34	Germany	spleen	2009	Meyer-Kayser et al. (2012), sequences from this study
Goat *(Capra hircus)*	4	Switzerland	blood	2008	Silaghi et al. (2011e) (11)
Hedgehog (*Erinaceus europaeus*)	148	Germany	blood	2007/2008	Silaghi et al. (2012a) (7), sequences from this study
Horse (*Equus ferus caballus*)	14	Germany	blood	2004-2009	Silaghi et al. (2011d) (56)
Ibex (*Capra ibex*)	3	Austria	spleen	2008-2010	Silaghi et al. (2011b) (6), sequences from this study
Mouflon (*Ovis orientalis musimon*)	8	Austria	spleen	2008(2)-2009(16), partially unknown (5)	Silaghi et al. (2011b) (6), sequences from this study
	15	Germany		2009	Kauffmann et al. (2016), sequences from this study
Bank vole (*Myodes glareolus)*	1	Germany	skin	2012	Obiegala et al. (2014)
Wood mouse (*Apodemus sylvaticus*)	2	Germany	spleen / blood	2013	
	2	Germany	spleen	2012	
Red deer (*Cervus elaphus*)	19	Germany	spleen	2013	This study
	12	Austria	spleen / liver	2008 (5) – 2009 (1), partially unknown (6)	Silaghi et al. (2011b), partially unpublished
Roe deer (*Capreolus capreolus*)	29	Germany	spleen	2013	This study
	17	Germany	spleen (8), blood (9)	2010-2011	Overzier et al. (2013a) (9)
	22	Germany	spleen	2009	Kauffmann et al. (2016), sequences from this study
	10	Austria	spleen	2008-2009, partially unknown (1)	Silaghi et al. (2011b) (26)
	4	Germany	spleen (3), blood (1)	2011	Nieder et al. (unpublished)
Sika deer (*Cervus nippon*)	21	Germany	spleen	2013	This study
Wild boar (*Sus scrofa*)	4	Germany	spleen	2011	Silaghi et al. (2014) (1)
	1	Germany		2013	This study

[1] Number of sequences available from previous studies; Accession numbers are listed in the annex, Tab. 30

3. DNA extraction

Of the 356 animal samples collected for this study, approximately 10 – 15 g of spleen or liver preserved in 70% ethanol were available for DNA extraction. The samples originated from wild animals, including roe deer (16), red deer (37), fallow deer (7), sika deer (17), chamois (7), red fox (260) and wild boar (12). With a sterile scalpel blade, forceps and a mortar, 25 – 50 mg of tissue were separated of each sample and DNA was extracted using the High Pure PCR Template Preparation Kit® (Roche Diagnostics GmbH, Mannheim, Germany). At first 300 µl of Lysis Buffer and 30 µl of the enzyme Proteinase K was added for protein digestion and incubated in an Eppendorf Thermomixer® comfort over night at a temperature of 56°C and a speed of 500 rpm. After several washing steps according to the manufacturer`s instructions, the DNA was eluted with 60 µl of Elution Buffer.

Additionally, the QIAmp DNA Mini Kit (Qiagen, Hilden, Germany) was used for the DNA extraction of the spleen samples from red foxes following the manufacturer`s protocol.

4. Quality control of extraction and quantification of DNA

In order to evaluate the quality and quantity of the extracted DNA, a full-spectrum (220 – 750nm) spectrophotometer (NanoDrop®ND.1000, PeqLab, Erlangen, Germany) was used according to the manufacturer`s instructions (NanoDrop® 1000 Spectrophotometer User`s Manual, 2008). Samples with a very high amount of DNA, averaging 130 ng/µl or more, were diluted at a ratio of 1:10 using DNA free, sterile water in order to prevent false negative PCR results due to excessive amounts of starting template.

5. PCR amplification of DNA of *A. phagocytophilum*

5.1. Real-time PCR targeting a fragment of the *msp2* gene

The 356 DNA extracts from this study were screened for a 77 bp fragment of the *msp2* gene by real-time PCR modified after Courtney et al. (2004). Investigated samples with positive results of this real-time PCR were considered positive for *A. phagocytophilum*, since the *msp2* gene is specific for the bacterium. The Applied Biosystems® TaqMan® Gene Expression Master Mix (Life Technologies GmbH,

Darmstadt, Germany) was used with the primers ApMSP2f and ApMSP2r and the TaqMan® probe ApMSP2p (Eurofins Genomics, Ebersberg, Germany) (Tab. 2 – 4). The real-time PCR was performed using the Applied Biosystems® 7500 Fast Real-Time PCR System (Life Technologies GmbH, Darmstadt, Germany).

Table 2: Primers and probe for real-time PCR targeting the *msp2* gene of *A. phagocytophilum* modified after Courtney et al. (2004)

Primer	Oligonucleotide sequence
ApMSP2f	5'-ATG GAA GGT AGT GTT GGT TAT GGT ATT-3'
ApMSP2r	5'-TTG GTC TTG AAG CGC TCG TA-3'
ApMSP2p	5'-TGG TGC CAG GGT TGA GCT TGA GAT TG-3' labeled 5'-FAM, 3'-TAMRA

Table 3: PCR reaction mix of the real-time PCR targeting the *msp2* gene of *A. phagocytophilum*

Components	Amount
TaqMan Gene Expression Master Mix	15 µl
5'-Primer: *ApMSP2f* (10 µM)	2.25 µl
3'-Primer: *ApMSP2r* (10 µM)	2.25 µl
Probe: *ApMSP2p* (10 µM)	0.5 µl
Template DNA	5 µl
Total volume	**25 µl**

Table 4: Cycling conditions of the real-time PCR targeting the *msp2* gene of *A. phagocytophilum*

Step	Temperature	Time	Number of Cycles
Initial activation step	95°C	5 min	1
Denaturation	94°C	15 sec	50
Annealing-Extension	60°C	60 sec	

5.2. Nested PCR targeting the partial *16S rRNA*-gene

A nested PCR of the partial *16S rRNA* gene with primers described by Massung et al. (1998) was performed with all samples, which had no available partial *16S rRNA* gene sequences (Tab. 5 – 7). For both PCR runs the HotStarTaq Kit® (Qiagen, Hilden, Germany) and an Eppendorf Mastercycler® (Eppendorf,

Hamburg, Germany) was used. Primer, reaction mix and cycling conditions for the nested *16S rRNA* PCR are shown in tables 5 – 7. The *16S rRNA* is described as an approximately 1.5 kbp gene fragment (Chen et al., 1994). The amplicon of the first step of the nested PCR has a length of 932 bp. The final PCR product consists of 546 bp (Massung et al., 1998). The present study compared *16S rRNA* nucleotide sequences of 497 bp length. The exact PCR conditions are shown in Tab. 5 to 7.

Table 5: Primers for nested PCR targeting the *16S rRNA* gene of *A. phagocytophilum* (Massung et al., 1998)

Nested PCR	Primer	Oligonucleotide sequence
1st amplification	ge3a	5'-CAC ATG CAA GTC GAA CGG ATT ATT C 3'
	ge10r	5'-TTC CGT TAA GAA GGA TCT AAT CTC C-3'
2nd amplification	ge9f	5'-AAC GGA TTA TTC TTT ATA GCT TGC T-3'
	ge2	5'-GGC AGT ATT AAA AGC AGC TCC AGG-3'

Table 6: PCR reaction mix of the nested PCR targeting the *16S rRNA* gene of *A. phagocytophilum*

Nested PCR	Components	Amount
1st amplification	$MgCl_2$ (25 mM)	1 µl
	10x PCR Buffer	5 µl
	dNTP mixture (10 mM)	1 µl
	Autoclaved distilled water	36.5 µl
	5'-Primer: ge3a (100 µM)	0.5 µl
	3'-Primer: ge10r (100 µM)	0.5 µl
	HotStarTaq *Plus* DNA Polymerase (5U/µl)	0.5 µl
	Template DNA	5 µl
	Total volume	**50 µl**
2nd amplification	$MgCl_2$ (25 mM)	1 µl
	10x PCR Buffer	5 µl
	dNTP mixture (10 mM)	1 µl
	Autoclaved distilled water	40.5 µl
	5'-Primer: ge9f (100 µM)	0.5 µl
	3'-Primer: ge2 (100 µM)	0.5 µl
	HotStarTaq *Plus* DNA Polymerase (5 U/µl)	0.5 µl
	Template DNA (product of 1st amplification)	1µl
	Total volume	**50 µl**

Table 7: Cycling conditions of the nested PCR targeting the *16S rRNA* gene of *A. phagocytophilum*

Nested PCR	Step	Temperature	Time	Number of Cycles
1st amplification	Initial activation step	95°C	5 min	1
	Denaturation	94°C	30 sec	
	Annealing	55°C	30 sec	40
	Extension	72°C	1 min	
	Final extension	72°C	5 min	1
2nd amplification	Initial activation step	95°C	5 min	1
	Denaturation	94°C	30 sec	
	Annealing	55°C	30 sec	25
	Extension	72°C	1 min	
	Final extension	72°C	5 min	1

5.3. Heminested PCR targeting the *groEL* gene

In samples without available partial *groEL* nucleotide sequences, a heminested PCR was performed for detection of the partial *groEL* gene according to Alberti et al. (2005b) (Tab. 8 – 10). The primers AphplgroELF and AphplgroELR were applied for the first amplification step. The length of the PCR product was 624 bp. The second PCR run used the same forward primer as in the first step, but EphgroELR as reverse primer. The resulting amplicon was of 573 bp length. In the present study *groEL* nucleotide sequences of 530 bp were compared. The HotStar Taq Kit® (Qiagen, Hilden, Germany) and an Eppendorf Mastercycler® (Eppendorf, Hamburg, Germany) were used. The exact PCR conditions are shown in tables 8 to 10.

Table 8: Primers for heminested PCR targeting the *groEL* gene of *A. phagocytophilum* (Alberti et al., 2005a; 2005b)

Nested PCR	Primer	Oligonucleotide sequence
1st amplification	AphplgroELF	5'-ATG GTA TGC AGT TTG ATC GC-3'
	AphplgroELR	5'-TCT ACT CTG TCT TTG CGT TC-3'
2nd amplification	AphplgroELF	5'-ATG GTA TGC AGT TTG ATC GC-3'
	EphgroELR	5'-TTG AGT ACA GCA ACA CCA CCG GAA-3'

Table 9: PCR reaction mix of the heminested PCR targeting the *groEL* gene of *A. phagocytophilum*

Nested PCR	Components	Amount
1st amplification	$MgCl_2$ (25 mM)	1 µl
	10x PCR Buffer	5 µl
	dNTP mixture (10 mM)	1 µl
	Autoclaved distilled water	36.75 µl
	5'-Primer: AphplgroELF (100 µM)	0.5 µl
	3'-Primer: AphplgroELR (100 µM)	0.5 µl
	HotStarTaq *Plus* DNA Polymerase (5 U/µl)	0.25 µl
	Template DNA	5 µl
	Total volume	**50 µl**
2nd amplification	$MgCl_2$ (25 mM)	1 µl
	10x PCR Buffer	5 µl
	dNTP mixture (10 mM)	1 µl
	Autoclaved distilled water	36.75 µl
	5'-Primer: AphplgroELF (100 µM)	0.5 µl
	3'-Primer: EphgroELR (100 µM)	0.5 µl
	HotStarTaq *Plus* DNA Polymerase (5 U/µl)	0.25 µl
	Template DNA	5 µl
	Total volume	**50 µl**

Table 10: Cycling conditions of the heminested PCR targeting the *groEL* gene of *A. phagocytophilum*

Nested PCR	Step	Temperature	Time	Number of Cycles
1st amplification/ 2nd amplification	Initial activation step	95°C	5 min	1
	Denaturation	94°C	30 sec	
	Annealing	55°C	30 sec	40
	Extension	72°C	45 sec	
	Final extension	72°C	10 min	1

5.4. Nested PCR targeting the *msp4* gene

The nested PCR targeting the partial *msp4* gene was developed by De la Fuente et al. (2005a) and modified by Bown et al. (2007b) (Tab. 11 – 13). In order to amplify the 849 bp long partial *msp4* gene, De la Fuente et al. (2005a) first used a conventional PCR with the primer pair Msp4AP5 and Msp4AP3. In the present study, *msp4* nucleotide sequences of 340 bp were compared. Bown et al. (2007b) improved the sensitivity of the latter PCR protocol by developing a nested PCR adding the primers Msp4f and Msp4r for the second amplification. All DNA samples without available *msp4* sequences were used for amplification. The resulting PCR product was 301 bp of length.

The HotStar Taq Kit® (Qiagen, Hilden, Germany) and the Eppendorf Mastercycler® (Eppendorf, Hamburg, Germany) were used in both amplification steps of the *msp4*-PCR. PCR conditions are shown in Tab. 11 and 13.

Table 11: Primers for nested PCR targeting the *msp4* gene of *A. phagocytophilum* (De la Fuente et al., 2007)

Nested PCR	Primer	Oligonucleotide sequence
1st amplification	Msp4AP5	5'-ATG AAT TAC AGA GAA TTG CTT GTA GG-3'
	Msp4AP3	5'-TTA ATT GAA AGC AAA TCT TGC TCC TAT G-3'
2nd amplification	Msp4f	5'-CTA TTG GYG GNG CYA GAG T-3'
	Msp4r	5'-GTT CAT CGA AAA TTC CGT GGT A-3'

Table 12: PCR reaction mix of the nested PCR targeting the *msp4* gene of *A. phagocytophilum*

Nested PCR	Components	Amount
1st amplification	$MgCl_2$ (25 mM)	1 µl
	10x PCR Buffer	5 µl
	dNTP mixture (10 mM)	1 µl
	Autoclaved distilled water	36.75 µl
	5'-Primer: Msp4AP5 (100 µM)	0.5 µl
	3'-Primer: Msp4AP3 (100 µM)	0.5 µl
	HotStarTaq *Plus* DNA Polymerase (5 U/µl)	0.25 µl
	Template DNA	5 µl
	Total volume	**50 µl**
2nd amplification	$MgCl_2$ (25 mM)	1 µl
	10x PCR Buffer	5 µl
	dNTP mixture (10 mM)	1 µl
	Autoclaved distilled water	36.75 µl
	5'-Primer: Msp4f (100 µM)	0.5 µl
	3'-Primer: Msp4r (100 µM)	0.5 µl
	HotStarTaq *Plus* DNA Polymerase (5 U/µl)	0.25 µl
	Template DNA	5 µl
	Total volume	**50 µl**

Table 13: Cycling conditions of the nested PCR targeting the *msp4* gene of *A. phagocytophilum*

Nested PCR	Step	Temperature	Time	Number of Cycles
1st amplification/ 2nd amplification	Initial activation step	95°C	5 min	1
	Denaturation	94°C	30 sec	
	Annealing	54°C	45 sec	40
	Extension	72°C	1 min	
	Final extension	72°C	10 min	1

5.5. Conventional PCR targeting the whole *msp2* gene

A conventional PCR was used in order to target the partial *msp2* gene (Lin et al.,

2004b). Msp25 and Msp23 were used as primers (Tab. 14 – 16). The HotStar Taq Kit® (Qiagen, Hilden, Germany) and an Eppendorf Mastercycler® (Eppendorf, Hamburg, Germany) were used. In the present study, *msp2* nucleotide sequences of up to 893 bp were compared. PCR conditions are listed in tab. 14 to 16.

Table 14: Primers for the conventional PCR targeting the *msp2* gene of *A. phagocytophilum* (Lin et al., 2004a)

Primer	Oligonucleotide sequence
msp25	5'-TTA TGA TTA GGC CTT TGG GCA TG-3'
msp23	5'-TCA GAA AGA TAC ACG TGC GCC C-3'

Table 15: PCR reaction mix for the conventional PCR targeting the *msp2* gene of *A. phagocytophilum*

Components	Amount
$MgCl_2$ (25 mM)	1 µl
10x PCR Buffer	5 µl
dNTP mixture (10 mM)	1 µl
Autoclaved distilled water	36.75 µl
5'-Primer: msp25 (100 µM)	0.5 µl
3'-Primer: msp23 (100 µM)	0.5 µl
HotStarTaq *Plus* DNA Polymerase (5U/µl)	0.25 µl
Template DNA	5 µl
Total volume	**50 µl**

Table 16: Cycling conditions for the conventional PCR targeting the *msp2* gene of *A. phagocytophilum*

Step	Temperature	Time	Number of Cycles
Initial activation step	95°C	5 min	1
Denaturation	95°C	1 min	
Annealing	62°C	1 min	35
Extension	72°C	1 min 30 sec	
Final extension	72°C	10 min	1

6. Visualisation

The products of the conventional PCRs were analysed by gel electrophoresis using a 2% agarose (Top Vision Agarose®; Fermentas, St. Leon-Rot, Germany).

The gel was dyed with Gel Red [Gel Red™ Nucleic Acid stain, 10.0000x in water (Biotium Hayward, USA), end concentration 1x]. For comparison of the detected PCR products with the positive control, a standardized DNA ladder (GeneRuler 100 bp Plus DNA Ladder®, St. Leon-Rot, Germany) was used and visualized under UV-light (PeqLab, Erlangen, Germany).

7. DNA purification

PCR products were purified with the QIAquick PCR Purification Kit® (Qiagen, Hilden, Germany) according to the manufacturer`s instructions. Quality and quantity of the purified PCR products were determined with a spectrophotometer (NanoDrop®ND.1000, PeqLab, Erlangen, Germany).

8. Sequencing and sequence analysis

After purification, forward and reverse sequencing was performed by Eurofins Genomics (Ebersberg, Germany) with forward and reverse primers. In case of nested PCR protocols, the product and the primers of the second PCR run were chosen. Prior to sending the PCR products to Eurofins Genomics, the purified DNA was diluted with sterile PCR-water until a concentration of 5 ng/µl was obtained.

The chromatogrammes of the sequences were analyzed and evaluated with Chromas Lite® (www.technelysium.com.au). In order to compare forward and reverse sequences of the samples, the reverse complement of the reverse sequence was created (www.bioinformatics.org/sms/rev_comp.html). Clustal Omega (www.ebi.ac.uk/Tools/msa/clustalo/) was utilized for the comparison of the forward and the reverse sequences. Subsequently, multiple alignments of *A. phagocytophilum* sequences resulting from the present (217 sequences) and from previous (355 sequences) studies were done with ClustalW2 according to the respective gene (www.ebi.ac.uk/Tools/msa/clustalw2/). BLAST (Basic Local Alignment Search Tool) provided by the NCBI (National Center for Biotechnology Information, Rockville Pike, USA), was applied in order to find similar *A. phagocytophilum* sequences. For easier distinction of the different sequence types, an alternative name for each examined gene variant (*16S rRNA*, *groEL*, *msp4*, *msp2*) was determined. In a summary of sequences for each of the four genes, every single variation of the sequence occurring in the study was

collected. The nomenclature of these summarized sequences consisted of the abbreviation for the four partial genes (“16S-“, “g-“, “m2-“, “m4-“) and a numerical sequence. The letter in brackets describes the nomenclature created in previous studies of Silaghi et al. (Overzier et al., 2013b; Silaghi et al., 2011b; Silaghi et al., 2011c; Silaghi et al., 2011d; Silaghi et al., 2012a).

9. Nucleotide database (GenBank)

Corresponding sequences from the NCBI database GenBank (http://www.ncbi.nlm.nih.gov/) were chosen and downloaded by screening the database for “*Anaplasma phagocytophilum*” and the four genes examined in this study – partial *16S rRNA*, *msp2*, *msp4* and *groEL*. The selection of sequences available in the GenBank until September 2014 included all sequences in the same gene region and of the same length used in this study. If necessary, selected sequences were shortened to the corresponding part of the gene and aligned with the sequences from this study. Shorter sequences or sequences from a different gene region were excluded.

In total, 531 nucleotide sequences from the NCBI nucleotide database were used for comparison with the sequences obtained from this study. In the end 23 different host animal species were taken into consideration for the analysis of sequence variability of *A. phagocytophilum* (Tab. 17). Altogether 367 *16S rRNA* gene sequences, 90 *groEL* gene sequences, 55 *msp4* gene sequences and 19 *msp2* gene sequences were downloaded from GenBank and used for comparison. The sequences originated from Europe, the USA or Asia/Russia (Tab. 18).

All 531 database nucleotide sequences were aligned with the sequences for partial *16S rRNA*, *groEL*, *msp4* and *msp2* resulting from this study. For ease of analysis, nucleotide sequences that did not match with sequences from this study were given their own name (“16S-nm”, “g-nm”, “m4-nm”, “m2-nm”). The letters “nm” stand for “no match”. Besides, each sequence not matching with the sequences obtained from this study was enumerated (e.g. 16S-nm1).

Table 17: Number of *A. phagocytophilum* sequences from GenBank used for the comparative analysis in this study[1]

Host animal species	*16S rRNA* (n)	*Msp2* (n)	*Msp4* (n)	*groEL* (n)	In total (n)
Domestic animals					
Dog	66	4	0	8	**78**
Cat	4	0	0	1	**5**
Equids (horse, donkey)	34	1	1	11	**47**
Cattle	7	1	8	0	**16**
Goat	4	0	0	0	**4**
Sheep	15	3	8	4	**30**
Wild animals					
Bear	0	1	0	0	**0**
Bison	14	0	3	0	**0**
Cervid (deer)	0	0	0	4	**4**
Chamois	3	0	0	0	**0**
Fox	2	0	0	0	**0**
Moose	3	0	0	3	**3**
Mufflon	1	0	0	0	**4**
Red deer	19	0	6	0	**0**
Reindeer	0	0	1	0	**0**
Roe deer	53	0	11	26	**26**
Sika deer	0	0	0	2	**2**
Wild boar	21	0	0	1	**1**
Water deer	5	0	0	1	**1**
Small mammals					
Rodent	66	1	15	24	**106**
Rabbit	1	0	0	0	**1**
Human	49	8	2	5	**64**
TOTAL	**367**	**19**	**55**	**90**	**531**

[1] Acc. no. of the sequences from the GenBank are listed in annex, Tab. 53 – 56.

Table 18: The origin of the nucleotide sequences available from GenBank database selected for comparison[1]

Country/ Region	***16SrRNA*** **(n)**	***Msp2*** **(n)**	***Msp4*** **(n)**	***groEL*** **(n)**
Europe	305	2	49	53
USA	30	17	2	24
Asia & Russia	29	0	4	13
Other*	3	0	0	0
Total	**367**	**19**	**55**	**90**

*Two nucleotide sequences from Tunesia and one from Brazil

[1] Acc. no. of the sequences from the GenBank are listed in annex, Tab. 53 – 56.

10. Amino acid sequences

Divergent nucleotide sequences obtained in the present study and from GenBank of the three protein-coding genes (*groEL, msp4, msp2*) were translated into their putative amino acid protein sequence using the translation tool ExPASy (Expert Protein Analysis System) Bioinformatics Resource Portal (http://web.expasy.org/translate/). Subsequently, the amino acid sequences were aligned applying ClustalW2 (www.ebi.ac.uk/Tools/msa/clustalw2/).

11. Statistical analysis

The statistical analysis was performed in cooperation with Stablab (Institute of statistics, LMU Munich).

11.1. Heat Maps

The creation of a heat map for each gene was performed with the software directory R Project [R Core Team (2015). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. URL;https://www.R-project.org/]. The different numbers of *A. phagocytophilum* gene variants occurring in different animal species contained in the matrix are represented by colors from green to red. Green stands for a low or no occurrence of the strains and red for a high number of occurrences.

11.2. Variance calculation

The arithmetic mean was calculated using the following formula:

$$\bar{x} = \frac{1}{n} \sum_{i=1}^{n} x_i$$

The empirial variance was calculated using the following formula:

$$s^2 = \frac{1}{n} \cdot \sum_{i=1}^{n} (x_i - \bar{x})^2$$

11.3. Trend analysis

A trend analysis was performed in order to demonstrate the variation of the four detected genes within animal species. The number of different strains within an animal species depending on the number of nucleotide sequences obtained in total was demonstrated in a line with the aid of Microsoft Excel (Redmond, USA). A common trendline was also added by the same programme including the total number of different strains of all animal species in respect of the total number of nucleotide sequences obtained in the present study.

The slope of the lines was calculated by the same formula as for the arithmetic mean.

11.4. Odd`s ratio

In order to describe the statistical chances of certain animal species to be infected with the most frequently occurring *16S rRNA* strains of *A. phagocytophilum*, the odd`s ratio was calculated. For this calculation, the investigated animal species were clustered into different animal groups according to animal orders (ruminants vs. non-ruminants) and level of domestication (wild animals vs. domestic animals).

The following formula was applied:

	E	$\bar{E}$
I	a	b
$\bar{I}$	c	d

$$OR = \frac{a / b}{c / d} = \frac{a \times d}{b \times c}$$

12. Phylogenetic analysis

Phylogenetic analysis was performed with the protein encoding partial *groEL* gene, the partial *msp4* gene and the partial *msp2* gene obtained in the present study or selected from the NCBI nucleotide database using MEGA version 6 (Tamura, Stecher, Peterson, Filipski, and Kumar 2013). The phylogenetic trees were created using the neighbor joining method with a bootstrap value of 1.000 repeats. For the phylogenetic analysis of the *msp2* gene, only sequences of a length of 813 bp were considered, shorter *msp2* sequences were excluded.

IV. RESULTS

1. Real-time PCR

In total, 83 out of 356 samples (23.3%) were positive for *A. phagocytophilum* by real-time PCR (Tab. 19). High numbers of *A. phagocytophilum*-positive samples were detected in roe deer with 81.3%, in sika deer with 76.5% and in fallow deer with 71.4% of the investigated samples. In contrast, less samples from red fox or wild boar were positive for *A. phagocytophilum* (Tab. 19).

Table 19: *A. phagocytophilum*-positive samples detected by real time PCR

Animal species	Number of samples	Number of pos. samples (%)
Roe deer	16	13 (81.3)
Red deer	37	13 (35.1)
Sika deer	17	13 (76.5)
Fallow deer	7	5 (71.4)
Chamois	7	4 (57.1)
Red fox	260	34 (13.1)
Wild boar	12	1 (8.3)
Total	**356**	**83 (23.3)**

2. PCR and nucleotide sequencing

The 83 *A. phagocytophilum*-positive samples from the present study as well as 259 positive samples from previous studies were taken into consideration for nucleotide sequencing, i.e. altogether 342 *A. phagocytophilum* samples from 15 different animal species were analyzed. The PCR amplification of the partial *16S rRNA* gene yielded the highest number of readouts with 250 out of 342 (73%) possible sequences. *GroEL* sequences were obtained from 145 out of 342 (42%) samples and *msp4* sequences from 133 out of 342 (39%) samples. The PCR amplification of the *msp2* gene resulted in 44 sequences out of 342 (13%) samples (Tab. 20).

Table 20: Overview of the *A. phagocytophilum* sequences obtained in the present study

Host animal species	Pos. samples (n)	No. of successfully sequenced partial genes (%)							
		16S rRNA		*groEL*		*msp4*		*msp2*	
Dog (*Canis lupus familiaris*)	74	50	(68)	33	(45)	30	(41)	20	(27)
Cattle (*Bos primigenius taurus*)	17	15	(88)	16	(94)	12	(71)	9	(53)
Hedgehog (*Erinaceus europaeus*)	33	31	(94)	33	(100)	18	(55)	4	(12)
Red fox (*Vulpes vulpes*)	34	7	(21)	3	(9)	3	(9)	1	(3)
Roe deer (*Capreolus capreolus*)	82	72	(88)	25	(30)	25	(30)	6	(7)
Red deer (*Cervus elaphus*)	31	26	(84)	9	(29)	18	(58)	3	(10)
Sika deer (*Cervus nippon*)	21	15	(71)	5	(24)	11	(52)	0	(0)
Fallow deer (*Dama dama*)	8	6	(75)	4	(50)	3	(38)	1	(13)
Chamois (*Rupicapra r. rupicapra*)	10	9	(90)	5	(50)	4	(40)	0	(0)
Ibex (*Capra i. ibex*)	3	2	(67)	2	(67)	2	(67)	0	(0)
Mouflon (*Ovis orientalis musimon*)	23	19	(83)	10	(43)	7	(30)	0	(0)
Wild boar (*Sus scrofa*)	1	0	(0)	0	(0)	0	(0)	0	(0)
Bank vole (*Myodes glareolus*)	2	0	(0)	0	(0)	0	(0)	0	(0)
Wood mouse (*Apodemus sylvaticus*)	2	0	(0)	0	(0)	0	(0)	0	(0)
Yellow-necked mouse (*Apodemus flavicollis*)	1	0	(0)	0	(0)	0	(0)	0	(0)
Total	**342**	**252**	**(73)**	**145**	**(42)**	**133**	**(39)**	**44**	**(13)**

2.1. *16S rRNA* gene sequences

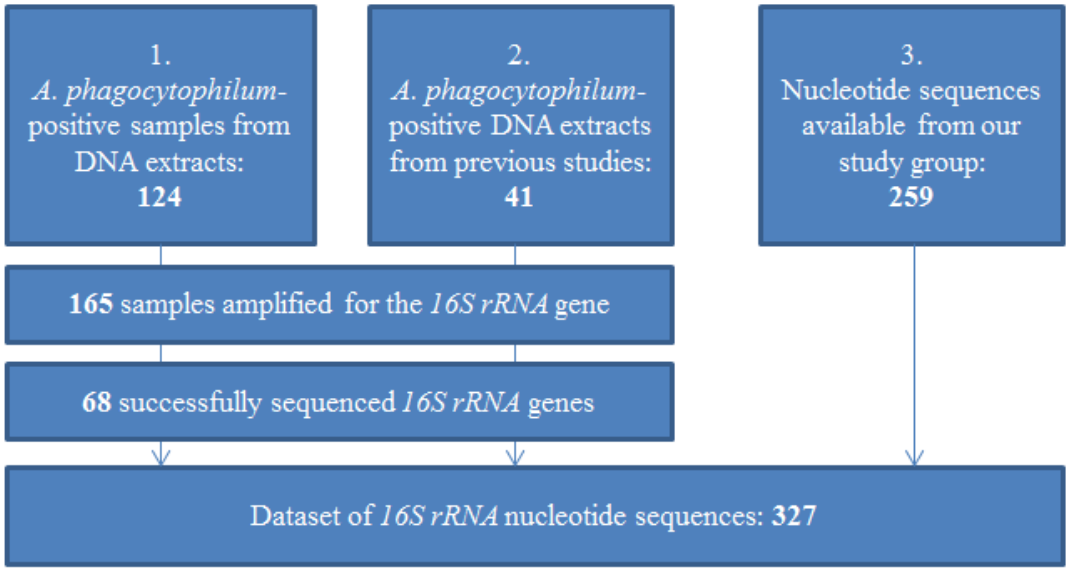

Figure 17: Origin of the nucleotide sequences of the *16S rRNA* gene. The enumerations 1., 2. and 3. relate to Fig. 14: Concept of the study.

In total, 68 partial *16S rRNA* (497 bp) sequences of the 165 analysed *A. phagocytophilum* DNA samples from nine of the fifteen different animal species (dog, red fox, roe deer, red deer, sika deer, fallow deer, mouflon, chamois, wild boar) were obtained in the present study (acc. nos.: annex Tab. 33). Of these 165 samples, 41 extracted DNA samples were already available from previous studies. Together with the sequences obtained from previous studies including further six animal species (cat, horse, cattle, goat, ibex, hedgehog), 327 *16S rRNA* nucleotide sequences were taken into consideration (Fig. 17). A multiple alignment of all obtained partial *16S rRNA* genes resulted in 23 different variants ("16S-1(A)" – "16S-30"). The heatmap of the *16S rRNA* shows the distribution of the different *A. phagocytophilum* variants in the animal species depending on the number of successfully sequenced genes (Fig. 18). According to the multiple sequence alignments, the strains differing most had a similarity score of 98.4%. The nucleotide sequences differed at altogether 27 of the 497 nucleotide positions (annex Tab. 32).

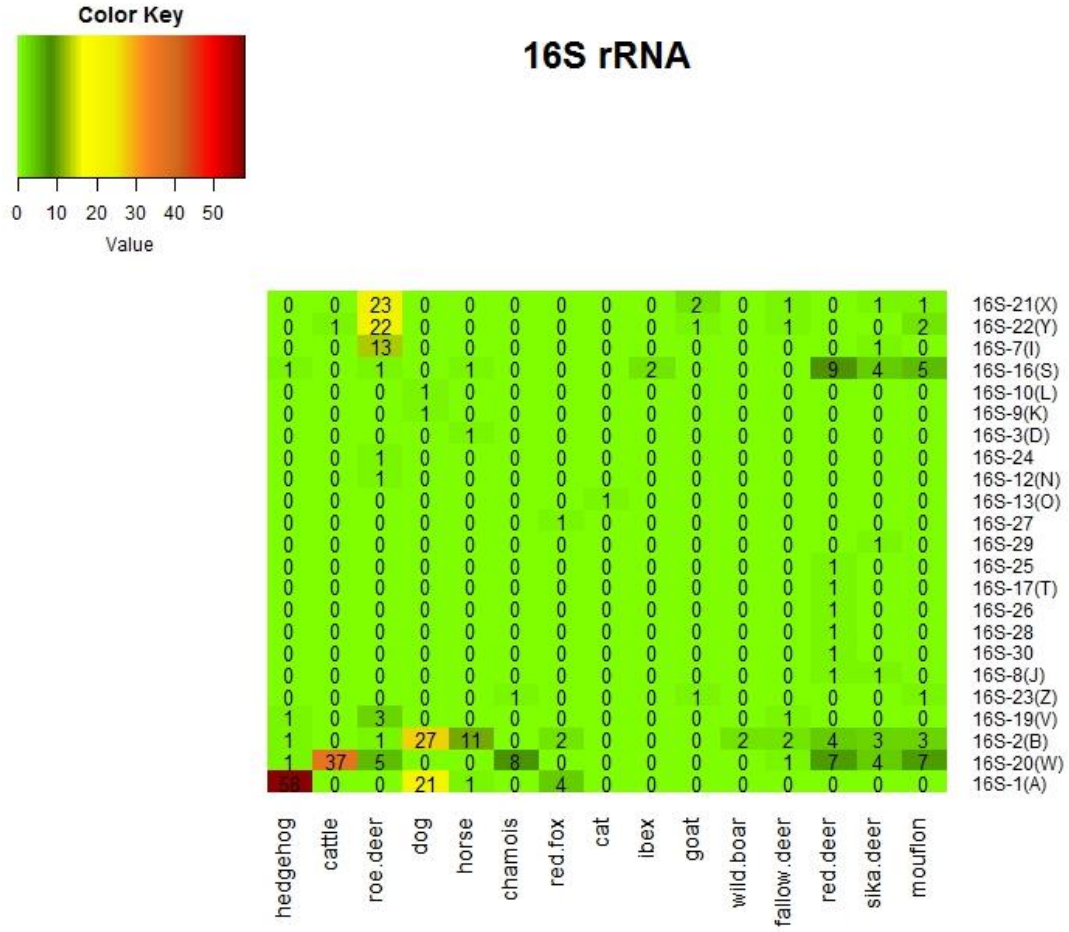

Figure 18: Heatmap of the *16S rRNA* variants of *A. phagocytophilum* occurring in the 15 different examined animal species. The numbers describe the number of occurrence of a certain variant.

Variant 16S-2(B) was 100% homologue to the prototype sequence of the HGA agent in the amplified region (acc. no.: U02521), which was detected in 10 of the 15 examined animal species in the present study (Fig. 18). Dogs and horses showed the highest presence of the variant 16S-2(B) reaching 54.0% (27/50) of the successfully sequenced dog samples and 78.6% (11/17) of the successfully sequenced horse samples, respectively. Although less often, wild ruminants also showed this prototype, for example 15.4% (4/26) in successfully sequenced red deer samples.

In general, variants 16S-1(A) and 16S-2(B) were detected more often in non-ruminant animals compared to ruminant species. 16S-1(A) was even detected in non-ruminants exclusively (100%), while variant 16S-2(B) reached 76.8% in non-ruminant species. Most 16S-1(A) variants were found in *A. phagocytophilum* from dog (24.7%) and hedgehog samples (68.2%).

Variant 16S-20(W) occurred in ruminants exclusively, with the exception of one

A. phagocytophilum sample from a hedgehog. Thereby, 52.9% of the 16S-20(W) variants were detected in cattle showing clinical signs of TBF. The other 45.7% of this variant were detected in *A. phagocytophilum* samples from wild ruminants (roe deer, red deer, sika deer, fallow deer, mouflon, chamois).

Variants 16S-21(X) and 16S-22(Y) dominated in wild ruminants and were detected especially in *A. phagocytophilum* samples from roe deer [16S-21(X): 32.9% (23/70), 16S-22(Y): 31.4% (22/70)]. In single animals of domestic (cattle, goat) and other wild ruminant species (sika deer, fallow deer, mouflon) variants 16S-21(X) and 16S-22(Y) were shown, while other domestic animals and humans did not show these variants.

Variant 16S-7(I) was also widely spread in roe deer (92.9%), making 18.6% of the *A. phagocytophilum* samples originating from roe deer.

Additionally, 13 further variants were detected only in single domestic and wild animal species (Fig. 18).

2.2. *groEL* gene sequences

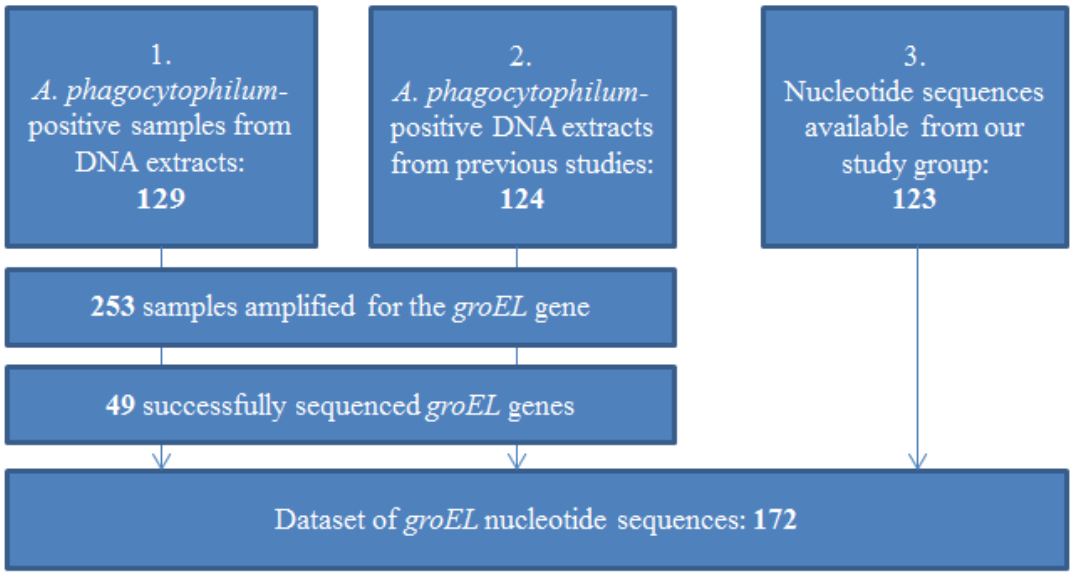

Figure 19: Origin of the nucleotide sequences of the *groEL* gene. The enumerations 1., 2. and 3. relate to the Fig. 14: Concept of the study.

The partial *groEL* gene (530 bp) was successfully sequenced in 49 animal samples of 253 *A. phagocytophilum* samples from ten of the fifteen different animal species (dog, hedgehog, roe deer, red deer, sika deer, fallow deer, mouflon, chamois, ibex, red fox) (acc. nos.: annex Tab. 37). Of these 253 samples, 124

extracted DNA samples were available from previous studies. Together with the sequences obtained from previous studies including further four animal species (cat, horse, cattle, goat), 172 *groEL* nucleotide sequences were taken into consideration (Fig. 19). An alignment of all *groEL* sequences occurring in this study resulted in 33 different variants ("g-1(A)"–"g-35"). The distribution of these *A. phagocytophilum* variants among the animal species is shown in the heatmap depending on the number of successfully sequenced *groEL* genes (Fig. 20). The variants differing most from the other strains had a similarity score of 95.9%. Variation was detected in 31 of the 530 nucleotide positions (annex Tab. 35).

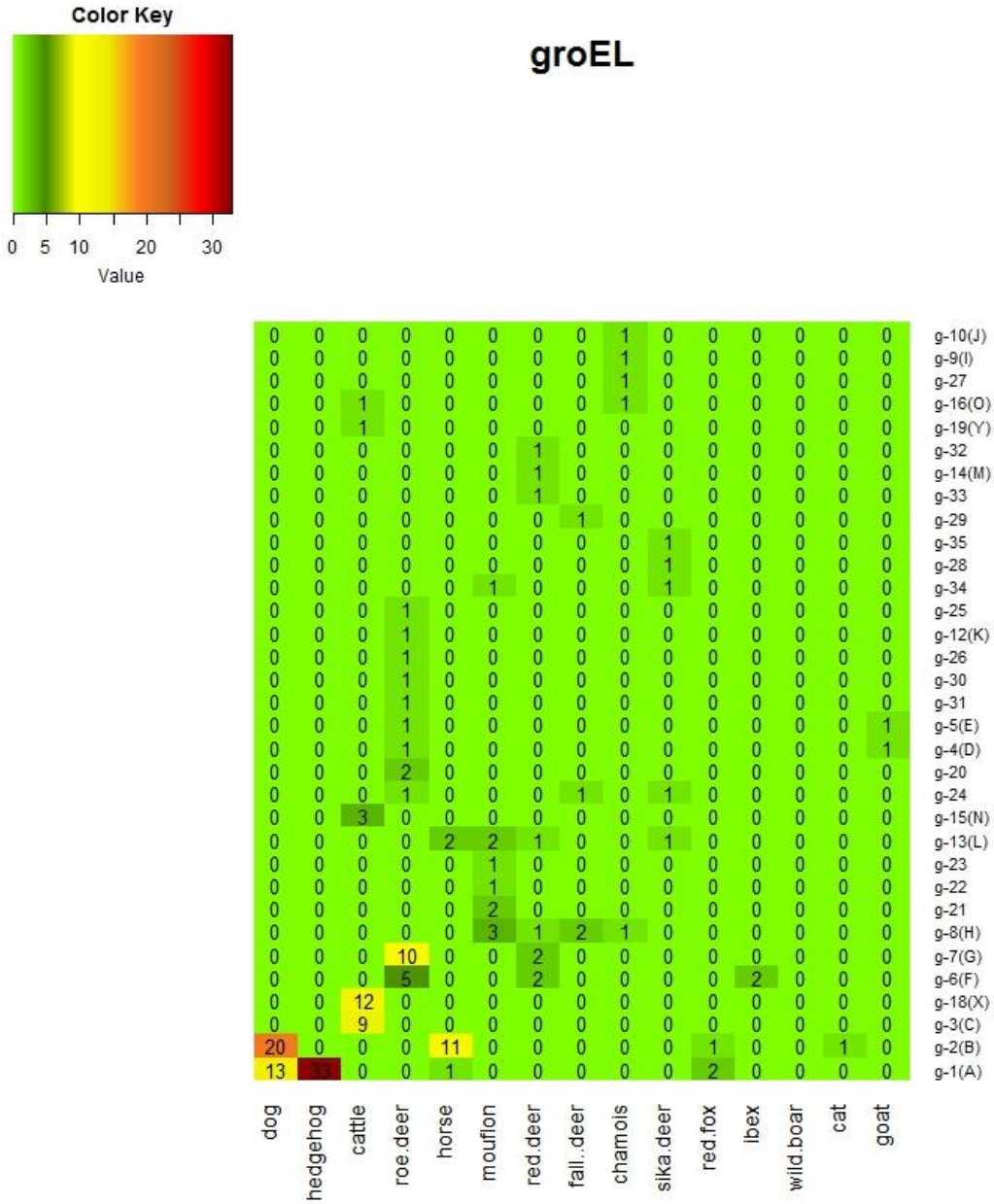

Figure 20: Heatmap of the *groEL* variants of *A. phagocytophilum* occurring in the 14 different examined animal species. The numbers describe the number of occurrence of a certain variant.

Analysis of the *groEL* gene showed variants occurring in ruminants exclusively [e.g. g-6(F), g-7(G)] and variants occurring exclusively in non-ruminants [e.g. g-1(A), g-2(B)]. The only exception was variant g-13(L), which was detected in both a horse and wild ruminants. Variants g-1(A) and g-2(B) were mainly found in dogs [g-1(A): 39.4%, g-2(B): 60.6%], horses [g-1(A): 78.6%, g-2(B): 7.1%] and cats [g-2(B): 100%]. In hedgehog samples, the *A. phagocytophilum* variant g-1(A) was detected exclusively, while fox samples showed both variants g-1(A) and g-2(B).

Four different variants occurred in cattle exclusively. Most of the *A. phagocytophilum* from cattle showed variant g-18(X) (46.2%). The other variants g-3(C) (34.6%), g-15(N) (11.5%) and g-19(Y) (3.9%) occurred less often, but also exclusively in cattle samples. Variant g-16(O) was detected in one cattle and one chamois sample. *Anaplasma phagocytophilum* from roe deer showed great diversity with eleven variants occurring in the 25 investigated roe deer samples. Thereby, 40.0% of these samples showed variant g-7(G). Sixteen additional variants occurred once in single wild ruminant species.

On amino acid level, seven expressed substitutions in nucleotide sequence resulted in changes of the translated amino acid sequences (annex, Tab. 36). The variation of the amino acid at position 48 divided all *groEL* sequences into two groups. The amino acid was either represented by alanine (9/35, 26.0%) or serine (26/35, 74.0%) (annex, Tab. 36). All *A. phagocytophilum* variants from the alanine group (g-4(D), g-5(E), g-6(F), g-7(G), g-12(K), g-20, g-26, g-30, g-31) originated either from goats or wild ruminants, predominantly from roe deer. The other nucleotide changes caused punctual mutations. Sequence variant g-15(N) showed a guanine nucleotide in position 29, resulting in an amino acid change from asparagine to serine (X29G → A position S). The same nucleotide and amino acid change was found in g-27 at nucleotide position 92 and amino acid 33, respectively. The nucleotide change at position 418 in g-28 caused an amino acid change from valine to phenylalanine at position 126. The last lysine was missing in g-25. The remaining nucleotide variations did not cause any amino acid changes.

2.3. *Msp4* gene sequences

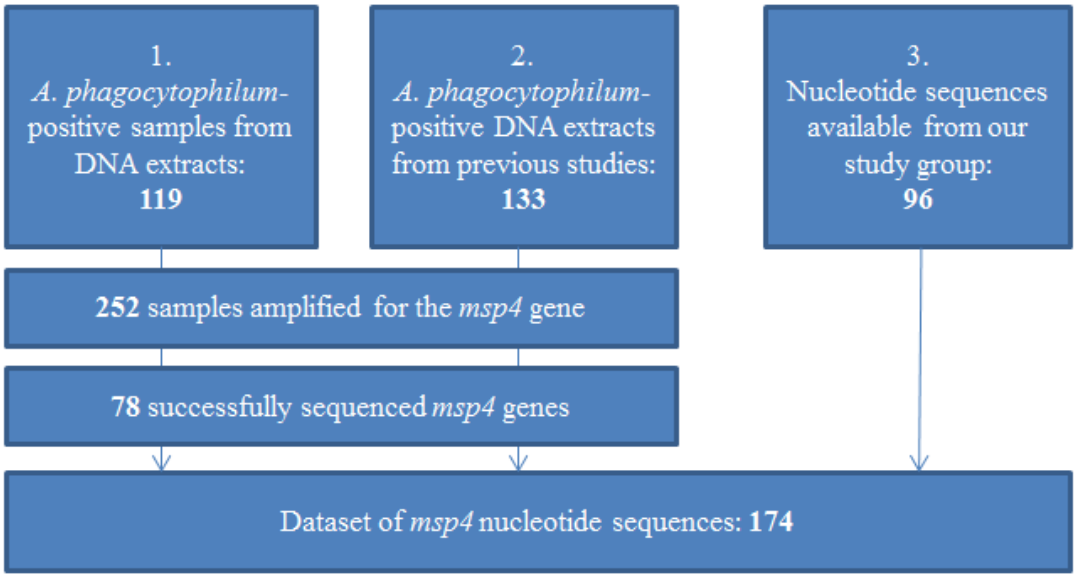

Figure 21: Origin of the nucleotide sequences of the *msp4* gene. The enumerations 1., 2. and 3. relate to the Fig. 14: Concept of the study.

Altogether, 78 partial *msp4* sequences (340 bp) of 252 *A. phagocytophilum* samples were obtained in the present study from eight of the fifteen different animal species (dog, hedgehog, roe deer, red deer, sika deer, fallow deer, mouflon, red deer) (acc. nos.: annex, Tab. 41). Of these 252 samples, 133 DNA samples were already available from previous studies. Together with the sequences obtained from previous studies including further six animal species (cat, horse, cattle, goat, ibex, red fox), 174 nucleotide sequences were taken into consideration (Fig. 21). The multiple alignment of all *msp4* nucleotide sequences obtained in the study showed 50 variants (here called "m4-1(A)" – "m4-51") differing in 56 of the 340 nucleotides. The heatmap demonstrates the distribution of the *msp4* variants among the infected animal samples (Fig. 22). The variant most different to the others had a similarity score of 87.9% (annex, Tab. 39).

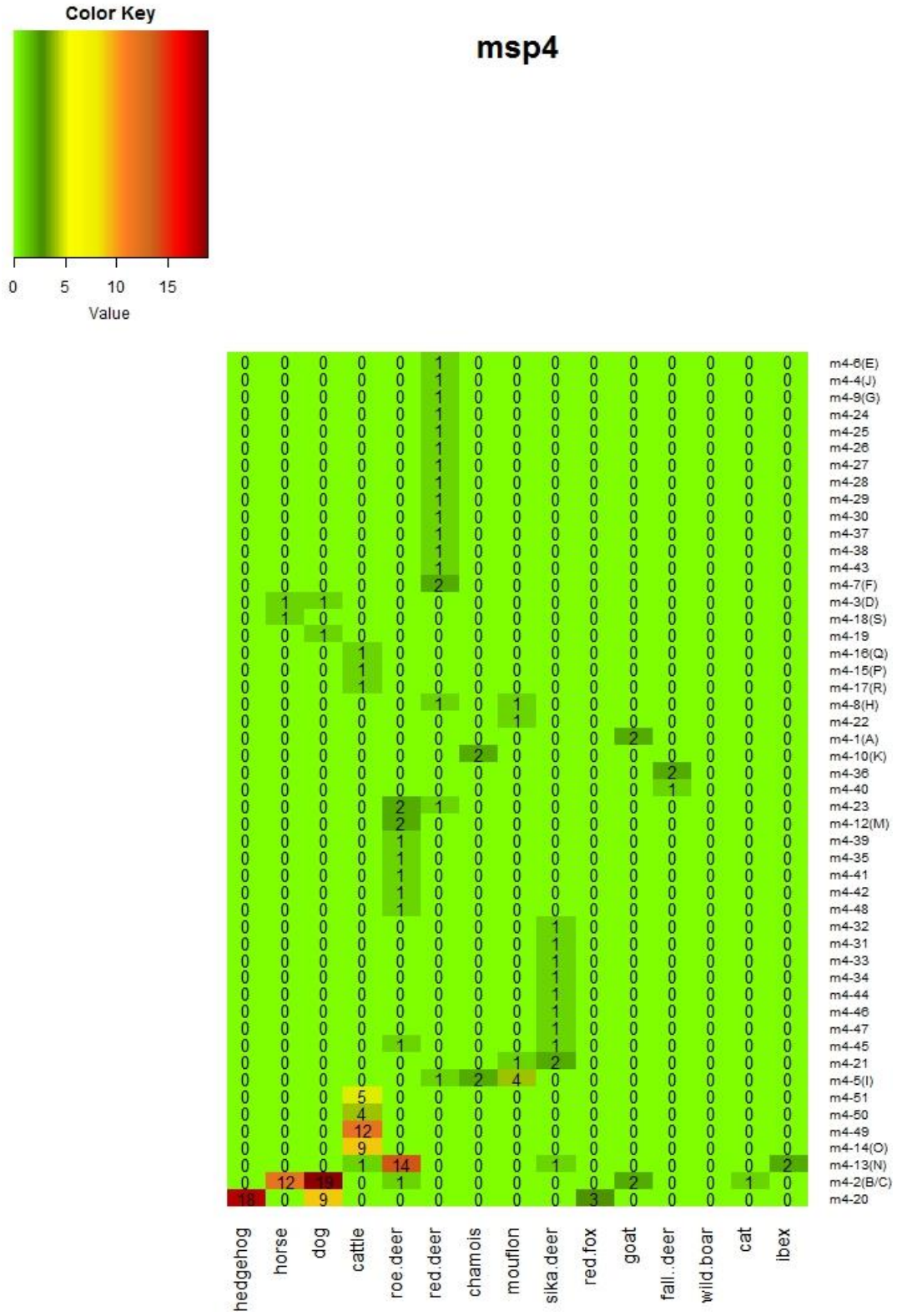

Figure 22: Heatmap of the *msp4* variants of *A. phagocytophilum* occurring in the 14 different examined animal species. The numbers describe the number of occurrence of a certain variant.

Variant m4-2(B/C) was the only one detected in both ruminant and non-ruminant strains. Thereby, the majority of the samples with this *A. phagocytophilum* variant were dogs (54.8%) and horses (34.3%). Additionally, two goats (5.7%) and one

roe deer (2.9%) showed this *A. phagocytophilum* variant. The other variants were found either in ruminants or in non-ruminants. Variant m4-20, for example, was detected in *A. phagocytophilum* infecting dogs (30.0%), hedgehogs (60.0%) and foxes (10.0%). Cattle samples revealed eight different *A. phagocytophilum* variants. Except from variant m4-13(N), all of them occurred in cattle exclusively. Variant m4-13(N) was revealed in *A. phagocytophilum* infecting cattle, roe deer, sika deer and ibex. Thirty-two single variants occurring in only one animal were detected in both wild and domestic animal species (annex, Tab. 38).

In total, the translation of the fifty *msp4* nucleotide sequences resulted in ten different amino acid sequences (annex, Tab. 40). Thereby, variants m4-23, m4-39 and m4-45 originated from wild ruminants exclusively and differed from the other variants in amino acid sequence and showed high similarity amongst each other, except from an amino acid mutation of m4-45 in position 34 (annex, Tab. 40). Six amino acids of these sequences (position 15, 18, 21, 51, 89 and 92) differed from all the other sequences. However, the two detected groups on basis of the *groEL* gene were not confirmed on basis of the *msp4* nucleotide sequence level. Additional point mutations induced amino acid changes. In variant m4-35, prolin was changed to leucine on position 39 (X108T → P position L) and in m4-1(A), the 50th amino acid was changed from aspartic acid to glutamic acid (X136A → A position G). Position 109 in m4-22, m4-26 and m4-44 showed the amino acid arginine instead of lysine (X303G → K position R).

2.4. *Msp2* gene sequences

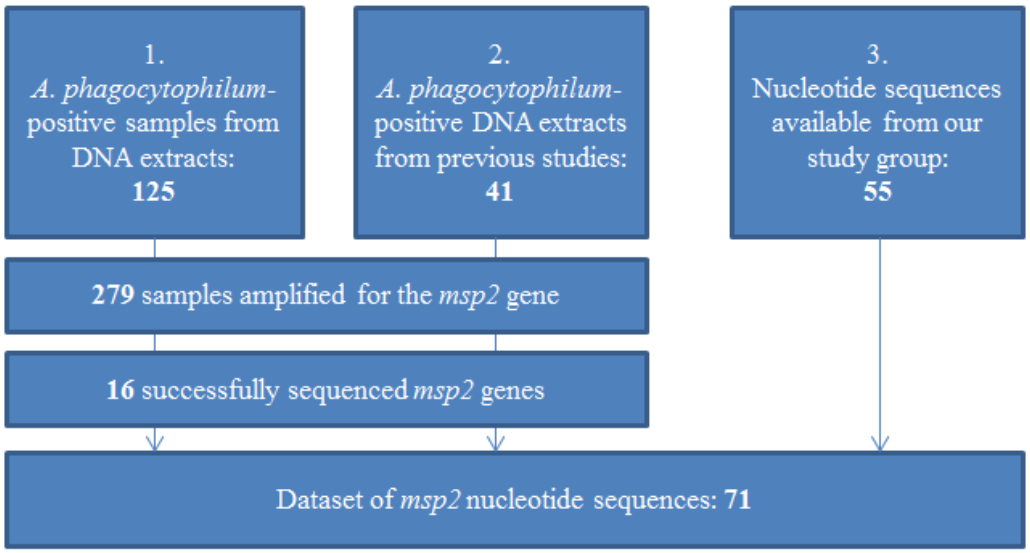

Figure 23: Origin of the nucleotide sequences of the *msp2* gene. The enumerations 1., 2. and 3. relate to the Fig. 14: Concept of the study.

Sixteen partial *msp2* genes of 279 *A. phagocytophilum* samples from six of the fifteen different animal species (dog, hedgehog, roe deer, red deer, fallow deer, red fox) were successfully sequenced (acc. nos.: annex, Tab. 45). Of these 279 samples, 154 extracted DNA samples were already available from previous studies. Together with the sequences obtained from previous studies including further four animal species (cat, horse, cattle, goat), 71 nucleotide sequences were taken into consideration (Fig. 23). The diversity concerning length and structure of the resulting sequences among each other was very high with the lowest similarity score of 67.1% compared to the other analyzed genes (annex, Tab. 43). The heatmap showed the distribution of the different *msp2* variants among *A. phagocytophilum* infecting different animals species (Fig. 24). The alignment of the obtained *msp2* sequences resulted in 22 different variants (here called "m2-2(a)" – "m2-27") (Fig. 24).

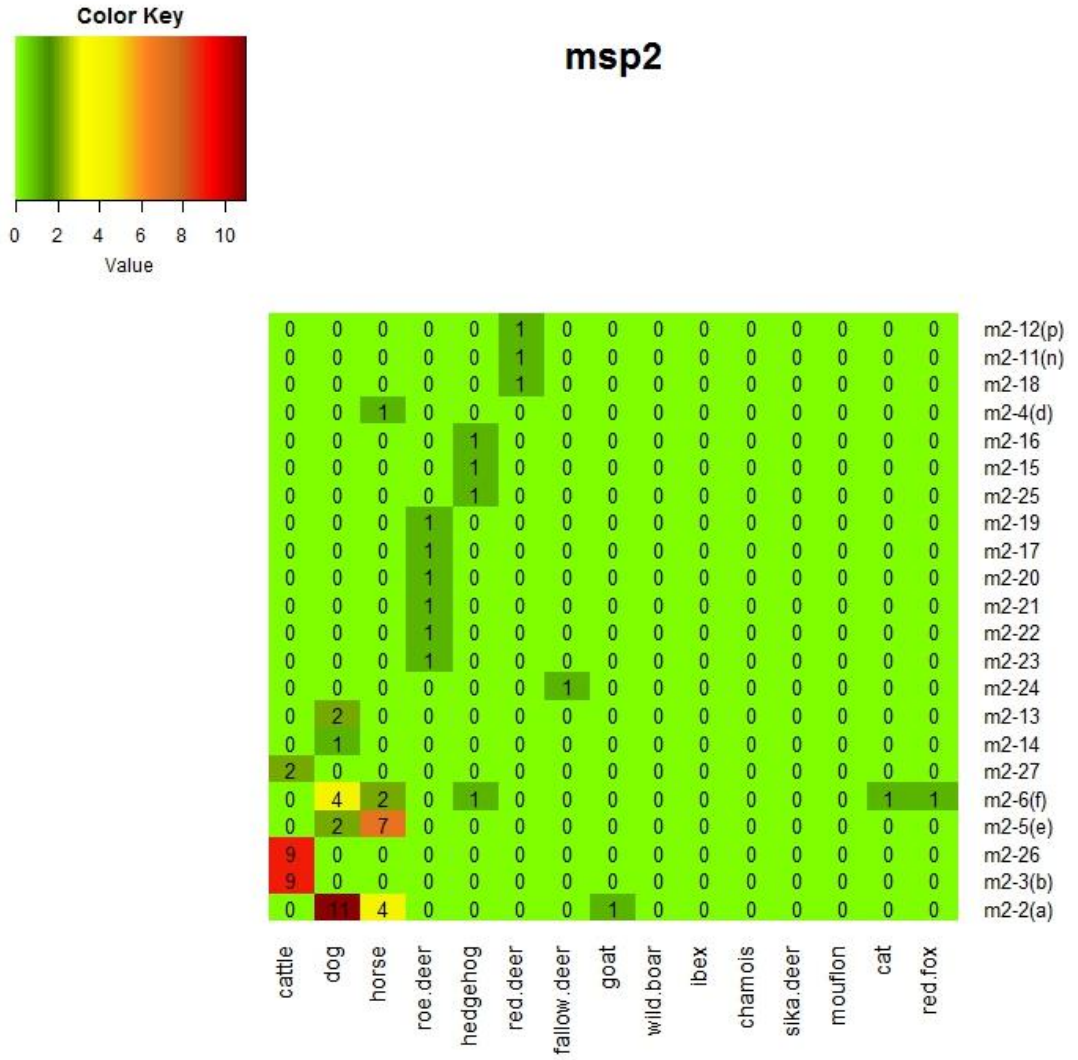

Figure 24: Heatmap of the *msp2* variants of *A. phagocytophilum* occurring in the ten different examined animal species. The numbers describe the number of occurrence of a certain variant.

The alignment of the *msp2* nucleotide sequences revealed three groups of sequences resembling each other strongly. The first group consisted of the variants m2-19, m2-20, m2-22, m2-23, m2-24 and m2-25. All of these variants were detected in wild cervids, except from the variant m2-25, which was found in a hedgehog sample. The second group was represented by the variants m2-26 and m2-27, which were both detected in cattle. Thereby, m2-26 and m2-27 were very homologue to each other (similarity score: 99.1). Additionally, cattle samples revealed *A. phagocytophilum* infections with m2-3(b). All of these variants were restricted to cattle exclusively (100%). The third group included all of the other variants. Thereby, the variant m2-6(f), for instance, was detected in wild and domestic non-ruminant animals. Domestic animals included dogs (44.4%), horses

(22.2%) and a cat (11.1%), while wild animals included a hedgehog (11.1%) and a fox (11.1%) sample. The only variant detected in domestic animals [dogs (68.8%) and horses (25.0%)] as well as in a goat (6.3%) sample was the variant m2-2(a). Of the 22 different *msp2* variants, 15 nucleotide sequences were detected in only one single animal affecting both domestic and wild animal species (68.2%).

The translation of the nucleotide sequence into the according amino acid structure reflected the high variability of the *msp2* gene with 22 variants (annex, Tab. 44). In total, the alignment of the amino acid sequences showed differences in 85 of the 308 (27.6%) positions. Interestingly, the amino acid structures of the two *A. phagocytophilum* strains from cattle, m2-26 and m2-27, differed clearly from the rest of the variants. Twenty-two of the 85 variating positions were shown solely in these two variants. The other two groups described on nucleotide sequence level were not confirmed on amino acid level, i. e. the changes on nucleotide level had no additional implications on the translation into amino acid sequences.

3. *A. phagocytophilum* variants within animal species

A. phagocytophilum derived from host animals potentially developing symptoms showed less variation in the four analyzed partial genes compared to strains infecting suspected reservoir host species. In the following, selected animal species are described.

Dog: The sequencing of the partial *16S rRNA*, the *groEL* and the *msp4* gene of the dog samples resulted in predominantly two different variants for each gene (Fig. 25). The *msp2* gene showed five different nucleotide sequences. In respect of the relatively large number of *msp2* strains of dog samples successfully sequenced (n=20), the diversity of the *msp2* was still rather small compared to other animal species.

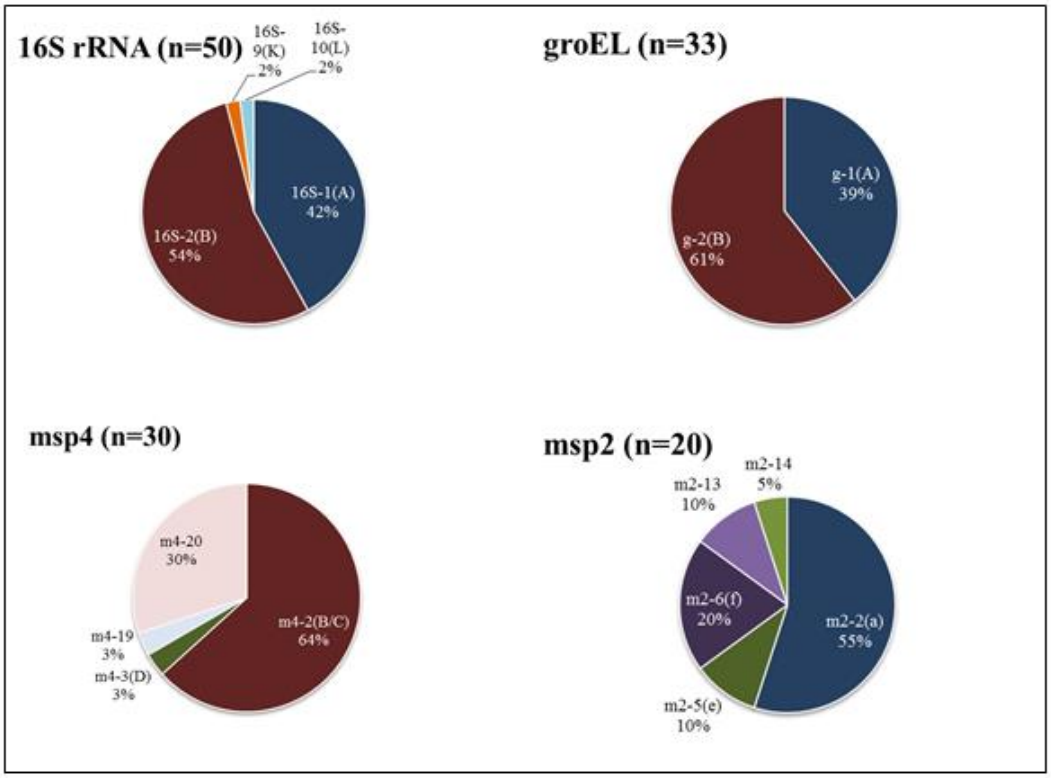

Figure 25: Partial *16S rRNA*, *groEL*, *msp4* and *msp2* gene sequences of all examined dog samples

Horse: The *A. phagocytophilum* nucleotide sequences from horse samples also showed little variation in the *16S rRNA*, *groEL* and *msp4*. One variant combination dominated in each of these genes [16S-2(B), g-2(B), m4-2(B/C)] (Fig. 26). The same *A. phagocytophilum* strains also dominated in *A. phagocytophilum* infecting dogs. Most of the *msp2* gene variants were

represented by the m2-5(e) (50.0%). The other variants m2-2(a) (29.0%), m2-4(d) (14.0%) and m2-6(f) (7.0%) were detected less often.

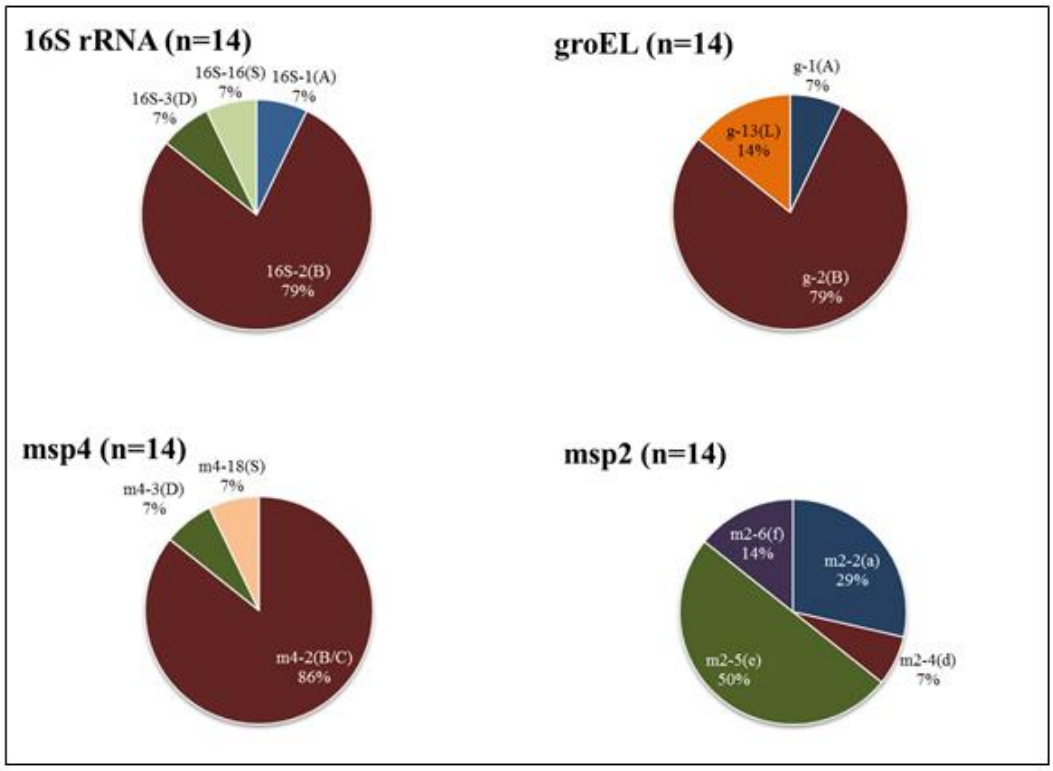

Figure 26: Partial *16S rRNA*, *groEL*, *msp4* and *msp2* gene sequences of all available horse samples

Cattle: In comparison to *A. phagocytophilum* infecting wild ruminants, the sequences revealed in cattle samples were less heterogenic. Especially the partial *16S rRNA* nucleotide sequences were very uniform with the variant 16S-20(W) representing 97.0% of the sequences from the cattle samples (Fig. 27). Similarly, the *msp2* gene revealed only three variants (m2-3(b), m2-26, m2-27). In contrast, the *msp4* and the *groEL* gene were more diverse.

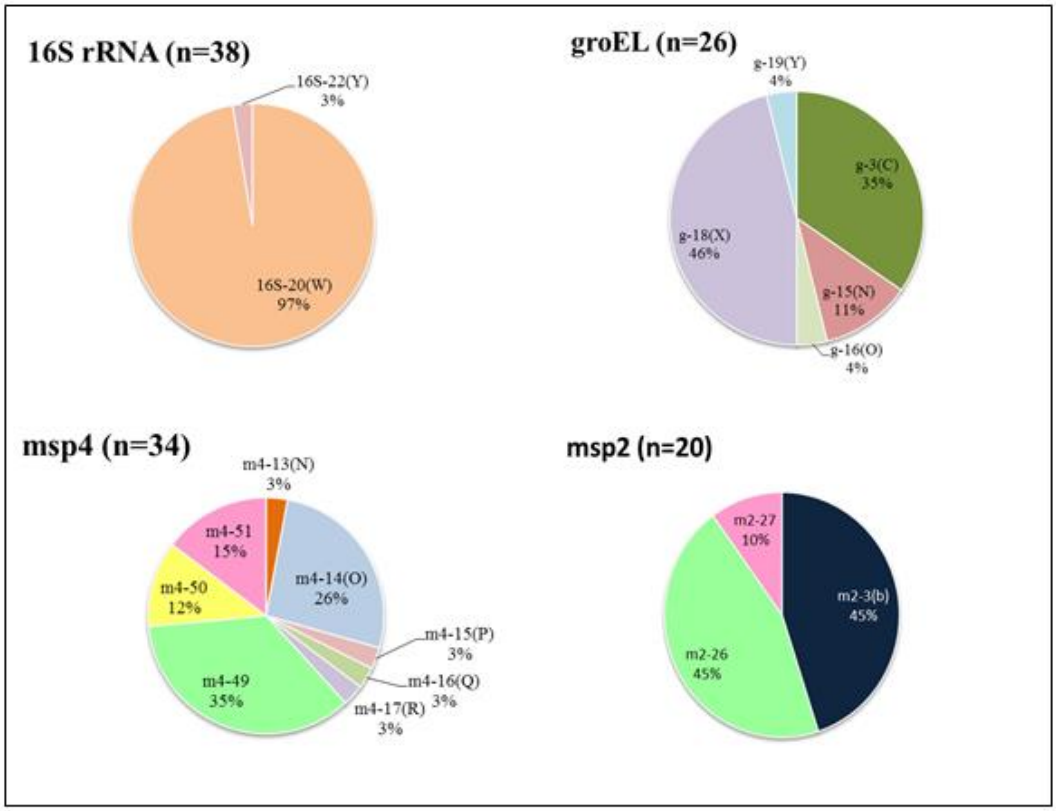

Figure 27: Partial *16S rRNA*, *groEL*, *msp4* and *msp2* gene sequences of all available and examined cattle samples

Hedgehog: A total of 93.0% of the partial *16S rRNA* variants of *A.phagocytophilum* detected in hedgehog samples represented the 16S-1(A) strain. Similarly, one single *groEL* [g-1(A)] and one *msp4* (m4-20) variant were detected in all successfully sequenced hedgehog samples. In four successfully sequenced partial *msp2* sequences, four different nucleotide sequences were revealed (Fig. 28).

Red fox: *A. phagocytophilum* variants occurring in fox samples were similar to the strains detected in dog samples with a different ratio: Variant 16S-1(A) (57%) and the g-1(A) (67%) dominated in fox samples, whereas 16S-2(B) and g-2(B) dominated in dog samples (Fig. 25/29). Variant m4-20 was detected in fox samples exclusively, which was also revealed in 30% of the dog samples (Fig. 29).

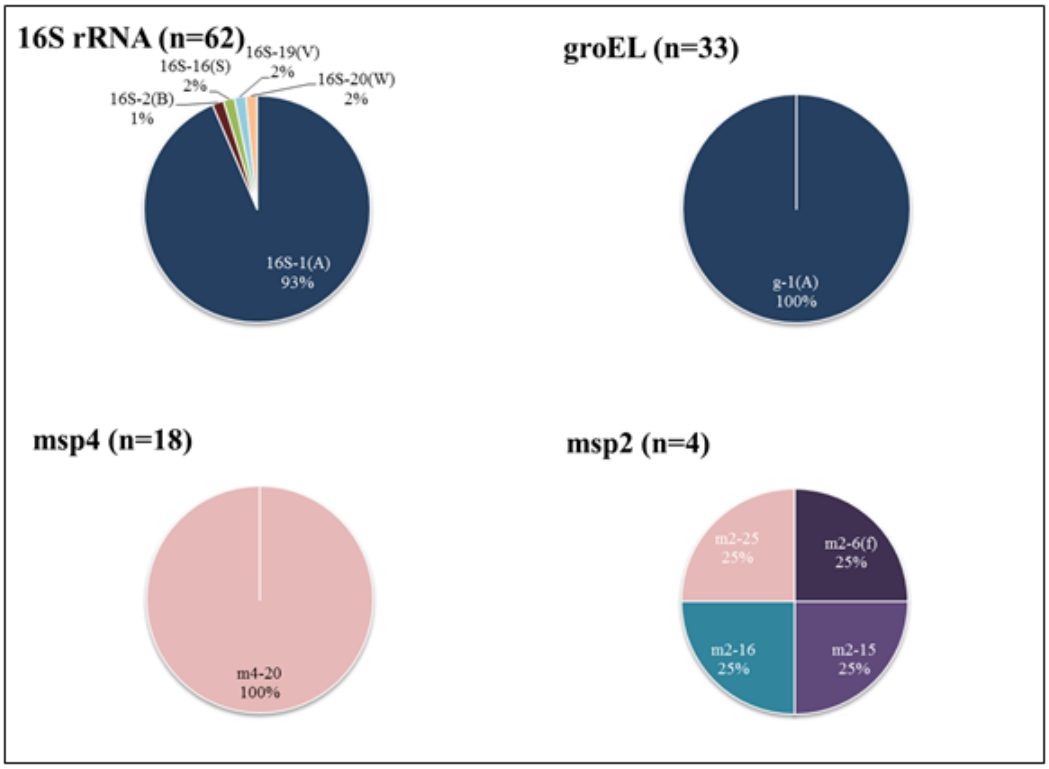

Figure 28: Partial *16S rRNA*, *groEL*, *msp4* and *msp2* gene sequences of all examined and available hedgehog samples

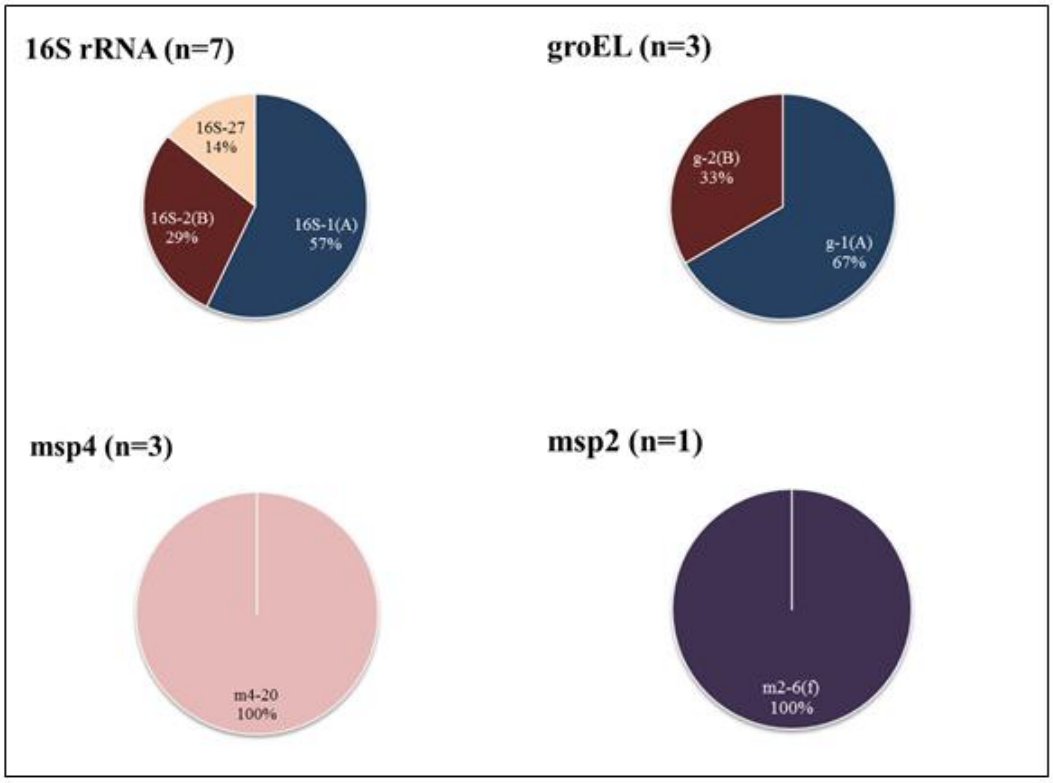

Figure 29: Partial *16S rRNA*, *groEL*, *msp4* and *msp2* gene sequences of all available red fox samples

Roe deer: *A. phagocytophilum* from roe deer samples showed great heterogeneity. The variants 16S-21(X), 16S-22(Y) and 16S-20(W), which mainly occurred in ruminants, were detected in 71% of all available *16S rRNA* sequences originating from roe deer. The potentially human pathogenic variant 16S-2(B) only occurred in 2% of the roe deer samples. Apart from the variants g-24 (4%) and g-25 (4%), all of the *groEL* variants (92%) belonged to the ruminant-only group characterized by an alanine amino acid on position 48 of the *groEL* heat shock protein. The *msp2* gene showed most variation, since all sequenced strains differed from each other (Fig. 30).

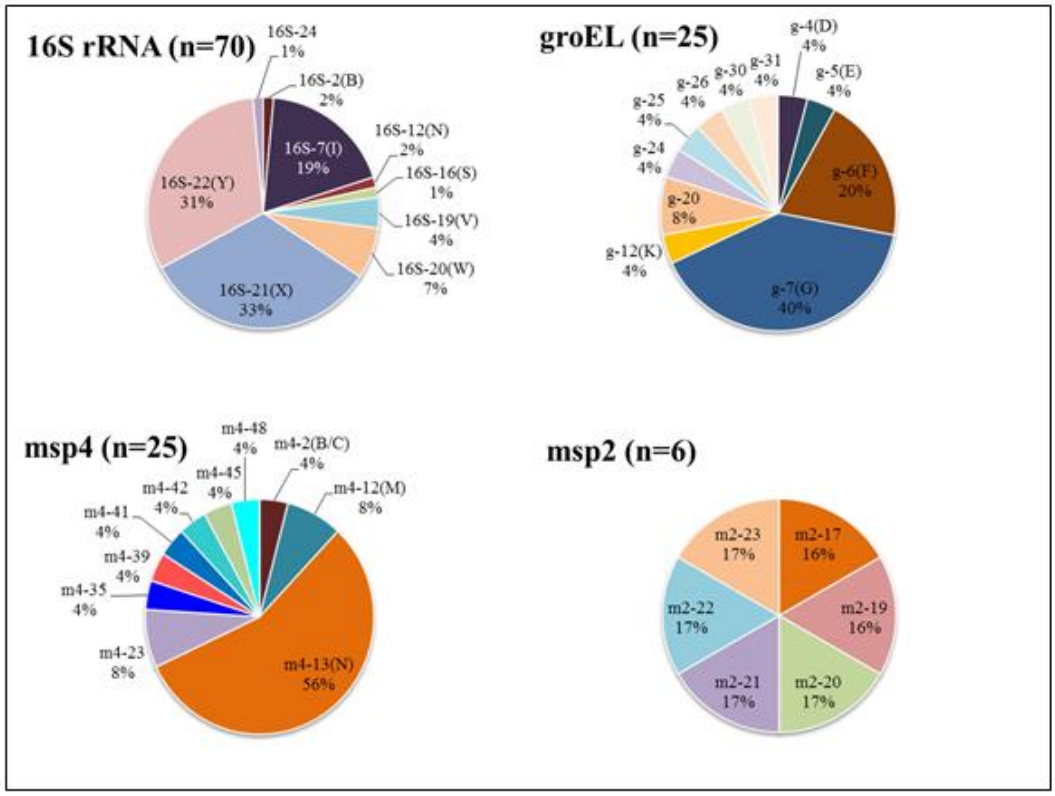

Figure 30: Partial *16S rRNA*, *groEL*, *msp4* and *msp2* gene sequences of all examined and available roe deer samples

Deer (red deer/fallow deer/sika deer): Similar to the *A. phagocytophilum* strains from roe deer samples, *A. phagocytophilum* strains originating from other deer species were very heterogenic (Figs. 31 – 33). Most variants were detected in the *msp4* and *groEL* gene. For example, 17 of the 18 successfully sequenced *msp4* genes from infected red deer differed in their nucleotide sequences (94.4%), most of them only occurring once (Fig. 33).

The prototype of *A. phagocytophilum*, the 16S-2(B) variant, was revealed in 33% of *A. phagocytophilum* from fallow deer samples, in 20% from sika deer samples and in 15% from red deer samples (Figs. 31 – 33). Similarly, the variant 16S-20(W), which potentially causes TBF in domestic ruminants, was represented in all three deer species (red deer: 27%, fallow deer: 17%, sika deer: 27%). Variants 16S-21(X) and 16S-22(Y) were detected in *A. phagocytophilum* from fallow deer (16S-21(X): 17%, 16S-22(Y): 17%) and sika deer (16S-21(X): 7%), but were lacking in red deer (Figs. 31 – 33).

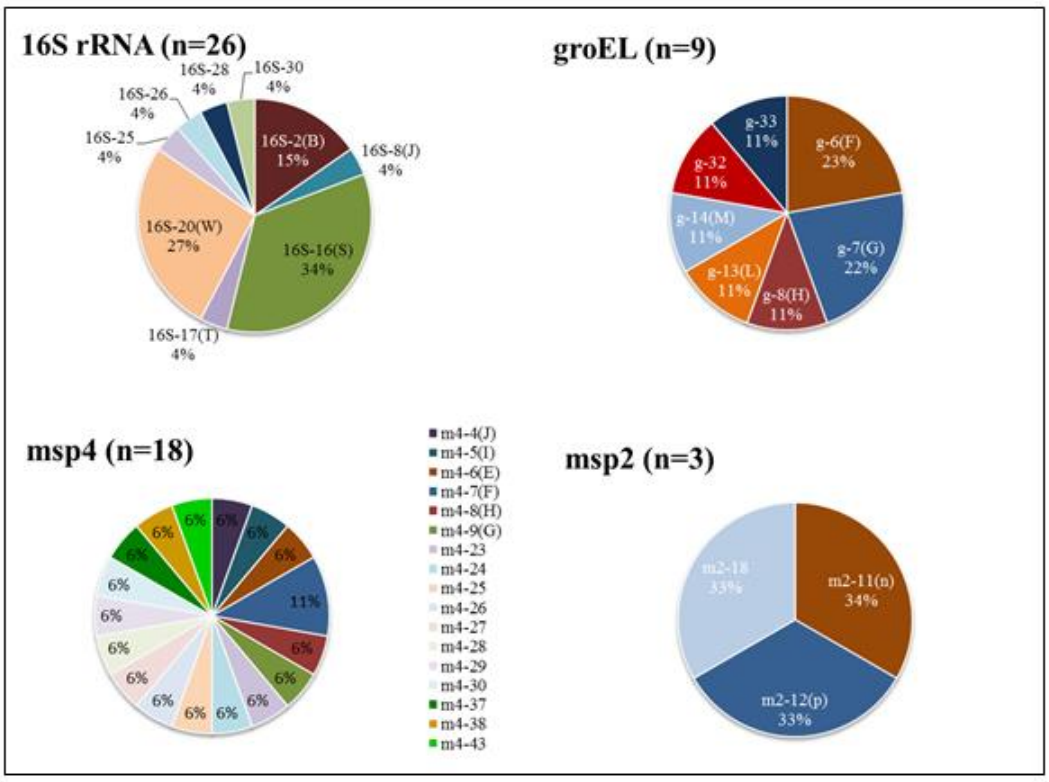

Figure 31: Partial *16S rRNA*, *groEL*, *msp4* and *msp2* gene sequences of all examined and available red deer samples

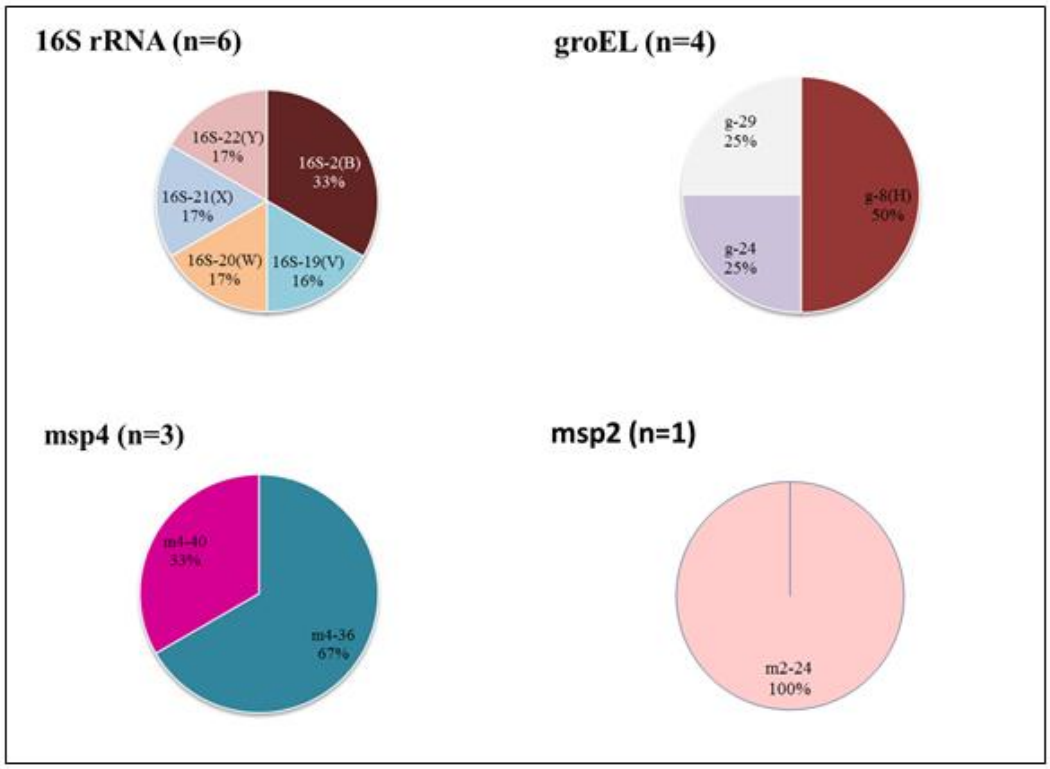

Figure 32: Partial *16S rRNA*, *groEL*, *msp4* and *msp2* gene sequences of all examined fallow deer samples

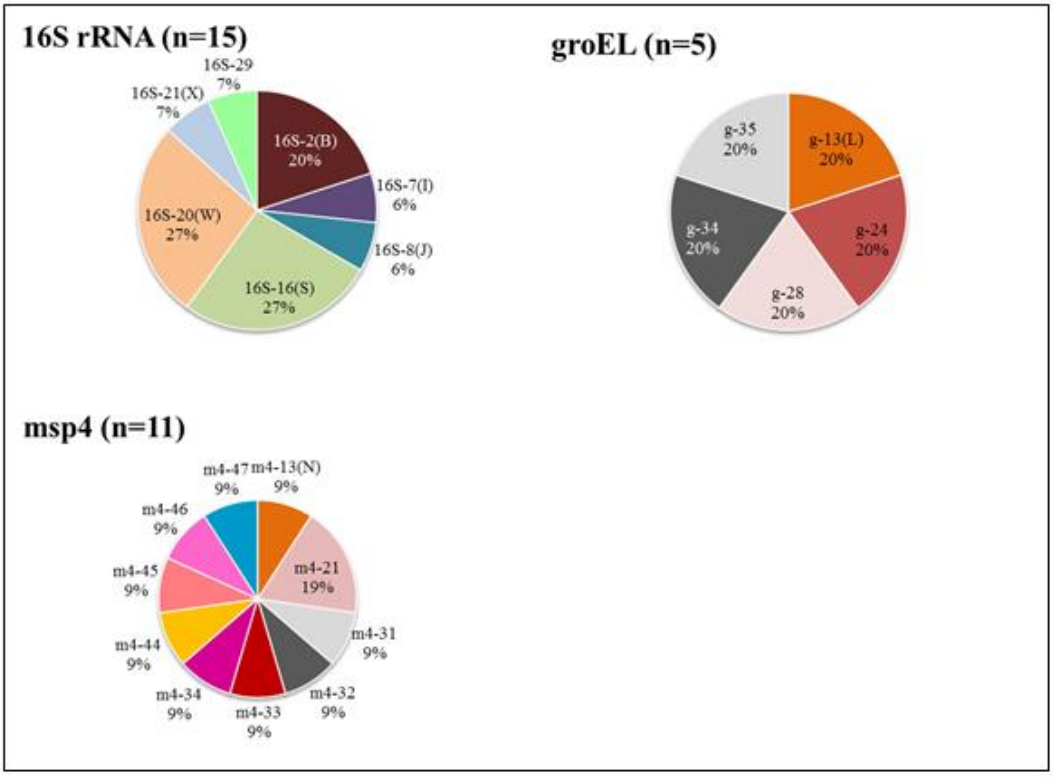

Figure 33: Partial *16S rRNA*, *groEL* and *msp4* gene sequences of all examined sika deer samples

Chamois: Altogether 89% of the *16S rRNA* sequences of *A. phagocytophilum* derived from chamois were variant 16S-20(W). The *groEL* nucleotide sequences of *A. phagocytophilum* infecting chamois reflected a higher variability with five different strains (Fig. 34).

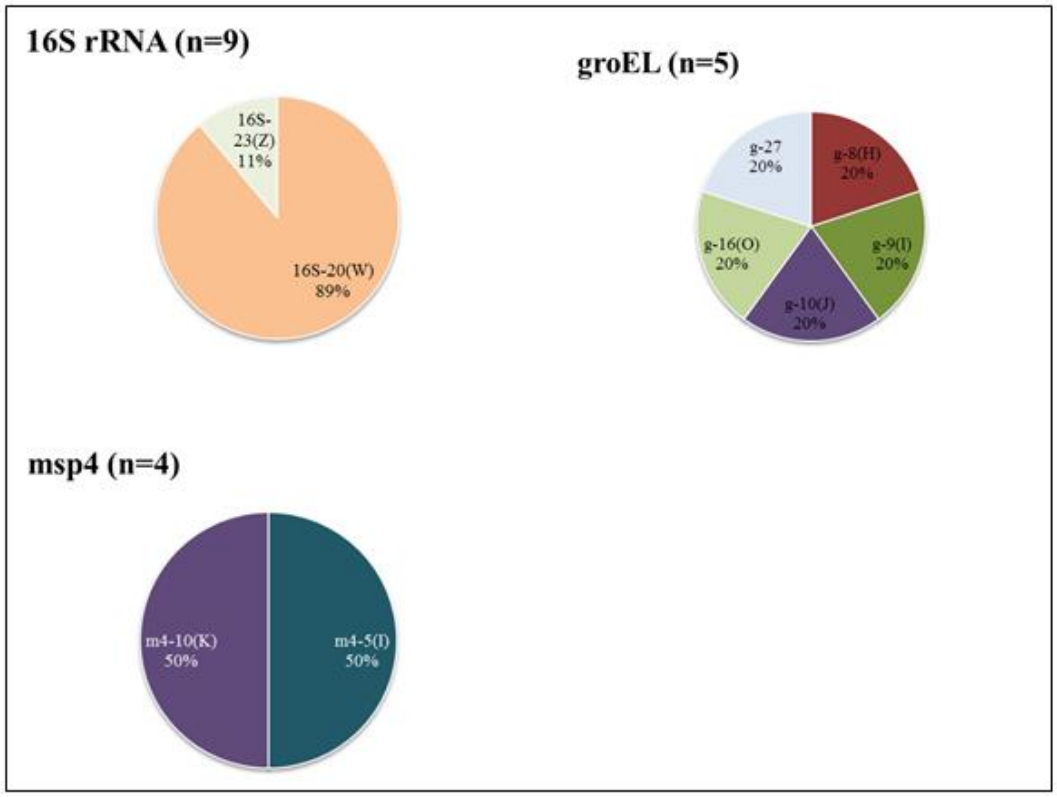

Figure 34: Partial *16S rRNA*, *groEL* and *msp4* gene sequences of all examined and available chamois samples

4. Statistical analysis of *A. phagocytophilum* strains

4.1. Empirical variance

In general, *A. phagocytophilum* from ruminants showed more variants than *A. phagocytophilum* from non-ruminants in the *16S rRNA*, the *groEL* and the *msp4* gene (Tab. 21). In each of these genes, the mean number of variants as well as the empirical variance was notably higher in *A. phagocytophilum* from ruminants compared to non-ruminants. The *msp4* gene in particular was highly variable with a mean number of variants of 6.2 and the empirical variance of 15.8 was the highest compared to the others. Solely, the *msp2* gene showed more variants in *A. phagocytophilum* from non-ruminants than from ruminants (Tab. 21). Nevertheless, the scattering of the variance was almost the same in ruminants and non-ruminants (3.80 vs. 3.58).

Table 21: Mean ând empirical variance of *A. phagocytophilum* occurring in ruminant[1] and non-ruminant[2] species investigated in the present study

	Ruminants[1]		Non-ruminants[2]	
	Mean no. of variants	Empirical variance	Mean no. of variants	Empirical variance
16S rRNA	4.89	8.32	3.00	2.33
groEL	5.00	7.78	1.50	0.92
msp4	6.22	15.76	1.67	1.89
msp2	1.56	3.80	2.50	3.58

[1]Ruminants: cattle, goat, roe deer, red deer, sika deer, fallow deer, chamois, mouflon, ibex
[2]Non-ruminants: dog, horse, cat, red fox, hedgehog, wild boar

Similar results were detected in the comparison between *A. phagocytophilum* from wild and domestic animal species (Tab. 22). The mean number of variants in the *16S rRNA*, the *groEL* and the *msp4* gene was higher in *A. phagocytophilum* from wild animals and amounted 4.10 – 4.80. The empirical variance scattered up to 15.10 in the *msp4* gene. In contrast, the *msp2* gene varied more in *A. phagocytophilum* from domestic animals. Nevertheless, the empirical variance was slightly higher in wild animals (4.05 vs. 2.56) (Tab. 22).

Table 22: Mean and empirical variance of *A. phagocytophilum* occurring in wild[1] and domestic[2] species investigated in the present study

	Wild animals[1]		Domestic animals[2]	
	Mean no. of variants	Empirical variance	Mean no. of variants	Empirical variance
16S rRNA	4.80	8.16	2.80	1.36
groEL	4.10	10.29	2.60	1.84
msp4	4.80	15.20	3.60	5.84
msp2	1.50	4.05	2.80	2.56

[1]Wild animals: hedgehog, red fox, wild boar, roe deer, red deer, sika deer, fallow deer, chamois, mouflon, ibex
[2]Domestic animals: dog, horse, cat, cattle, goat

4.2. Combination of sequences

In total, partial sequences of all four genes were obtained in 54 (12.7%) *A. phagocytophilum* isolates of the 425 animal samples, of three genes in 83 (19.5%) samples and of two genes in 71 (16.7%) samples (Tab. 23). There were no overlaps in the combinations between ruminants and non-ruminants as described in tab. 23. The variant combination 16S-1(A) and g-1(A) was detected in *A. phagocytophilum* from both domestic and wild animal samples, but was

lacking in ruminants. The variant combination with different major surface proteins [msp4: m4-2(B/C), m4-20; msp2: m2-6(f)] was shared by six dogs, one horse, 26 hedgehogs and two foxes. Samples from cattle were relatively uniform with two distinct sequence variant combinations in 15 examined animals (Tab. 23). Apart from a mouflon and a chamois detecting infections with *A. phagocytophilum* revealing the same gene combination [16S-20(W) g-8(H) m4-5(I)], wild ruminants showed similar gene combinations within the same animal species exclusively, i. e. intraspecific matches were detected more often in comparison to interspecific similarities.

Table 23: Combination of sequence variants including the combinations occurring at least twice

	Animal species	No. animal species	Combination of sequence variants			
			16S rRNA	***groEL***	***msp4***	***msp2***
4 / 4 partial genes successfully sequenced	dog	1	16S-1(A)	g-1(A)	m4-2(B/C)	m2-6(f)
	horse	1				
	fox	1	16S-1(A)	g-1(A)	m4-20	m2-6(f)
	hedgehog	1				
	dog	8	16S-2(B)	g-2(B)	m4-2(B/C)	m2-2(a)
	horse	4				
	horse	**7**	16S-2(B)	g-2(B)	m4-2(B/C)	m2-5(e)
	cattle	9	16S-20(W)	g-3(C)	m4-14(O)	m2-3(b)
	cattle	6	16S-20(W)	g-18(X)	m4-49	m2-26
3 / 4 partial genes sequenced	dog	2	16S-1(A)	g-1(A)		m2-6(f)
	dog	2	16S-1(A)	g-1(A)	m4-2(B/C)	
	dog	2	16S-1(A)	g-1(A)	m4-20	
	hedgehog	12				
	dog	2	16S-2(B)	g-2(B)		m2-2(a)
	dog	2	16S-2(B)	g-2(B)	m4-2(B/C)	

	dog	2	16S-2(B)	g-2(B)		m2-2(a)
	dog fox	1 1	16S-2(B)	g-2(B)	m4-20	
	ibex	2	16S-16(S)	g-6(F)	m4-13(N)	
	cattle	2	16S-20(W)		m4-49	m2-26
	cattle	3	16S-20(W)	g-15(N)	m4-51	
	chamois mouflon	1 1	16S-20(W)	g-8(H)	m4-5(I)	
	roe deer	3	16S-22(Y)	g-7(G)	m4-13(N)	
2 / 4 partial genes sequenced	dog	2	16S-1(A)		m4-2(B/C)	
	hedgehog dog fox	13 3 1	16S-1(A)	g-1(A)		
	dog	2	16S-2(B)	g-2(B)		
	roe deer	2	16S-7(I)		m4-23	
	cattle	4	16S-20(W)		m4-49	

4.3. Trend analysis

In order to allow a comparison between the different number of available *A. phagocytophilum* samples and nucleotide sequences, a trend analysis was performed. The analysis included strains obtained in the present study and strains from GenBank (Acc. nos. of strains from GenBank are provided in annex, Tab. 53 – 56). *A. phagocytophilum* variants originating from dogs, horses, cattle and hedgehog showed a low tendency of variation (Figs. 35 – 38). The line describing the *groEL* and the *msp4* gene of *A. phagocytophilum* from hedgehog samples represented the line with the lowest slope (Figs. 36 and 37).

All lines with a very steep line revealed a high amount of genetic diversity of *A. phagocytophilum*. Thus, the *A. phagocytophilum* strains originating from goat samples and wild ruminants showed a high tendency of variability (Figs. 35 – 38).

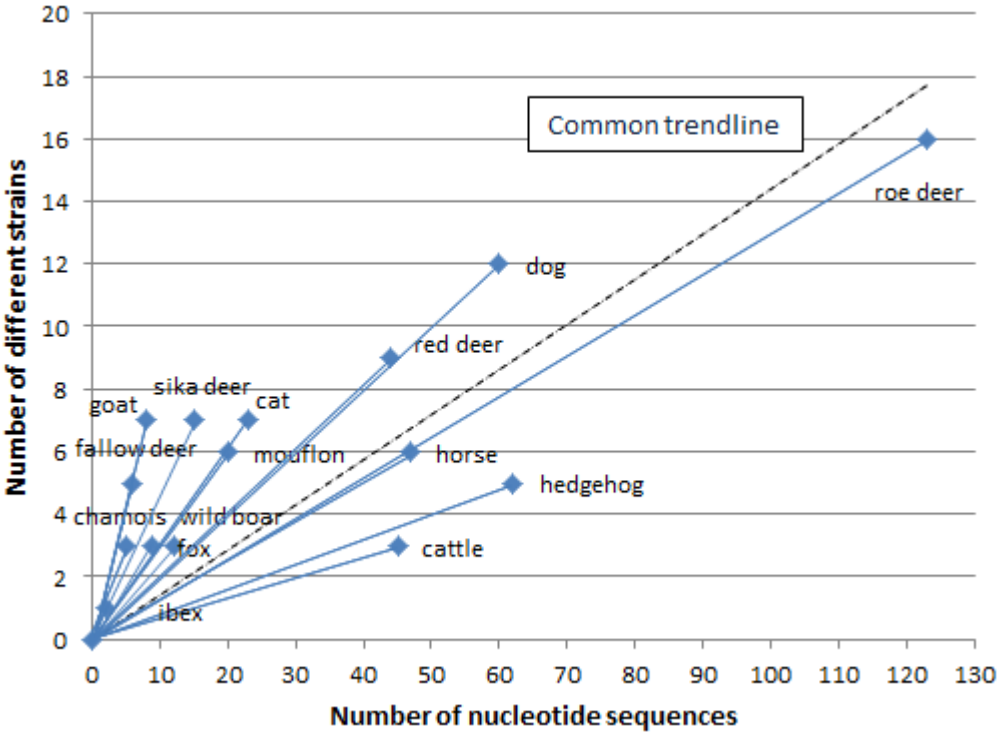

Figure 35: Trend analysis of the *16S rRNA* gene according to the sequences obtained from animal species investigated in this study in addition to available sequences from the GenBank

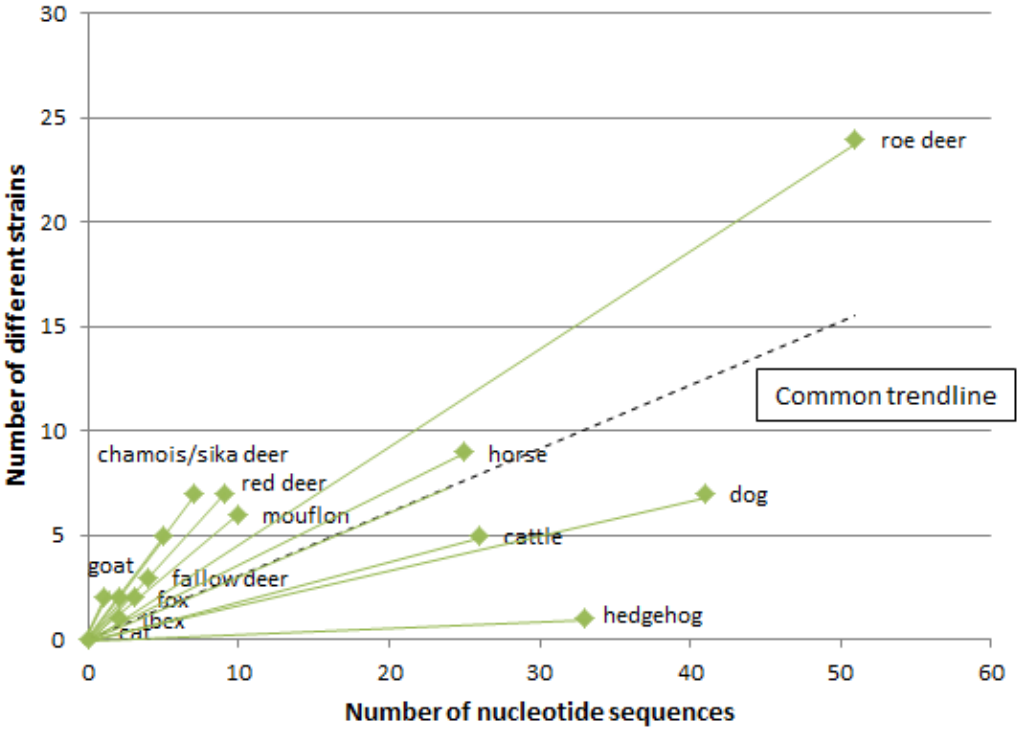

Figure 36: Trend analysis of the *groEL* gene according to the sequences obtained from animal species investigated in this study in addition to available sequences from the GenBank

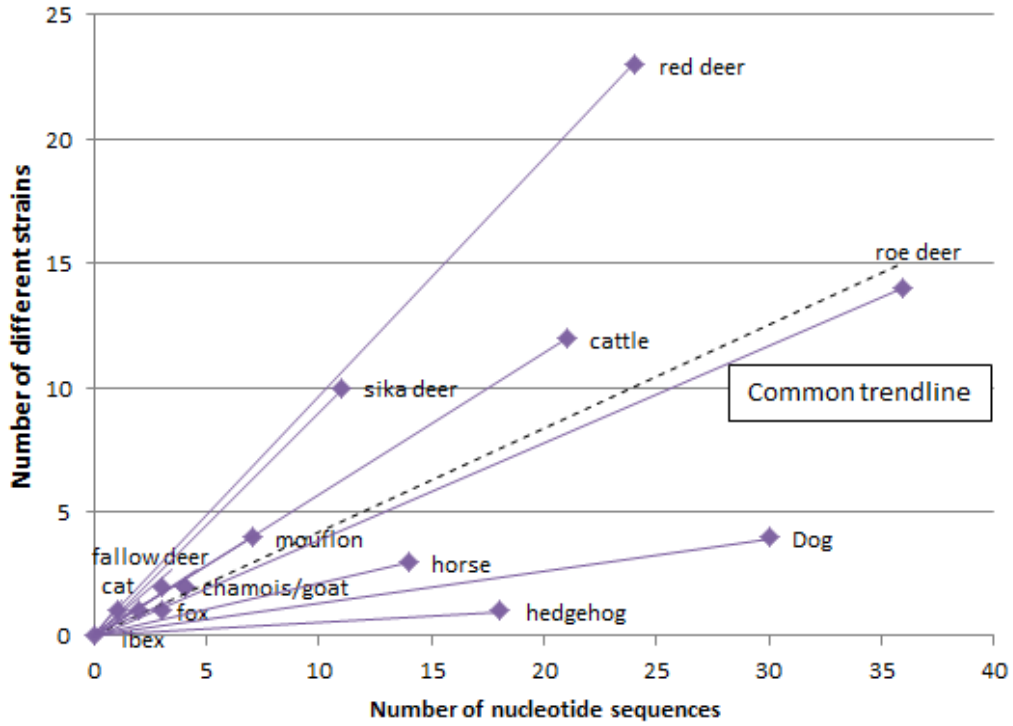

Figure 37: Trend analysis of the *msp4* gene according to the sequences obtained from animal species investigated in this study in addition to available sequences from the GenBank

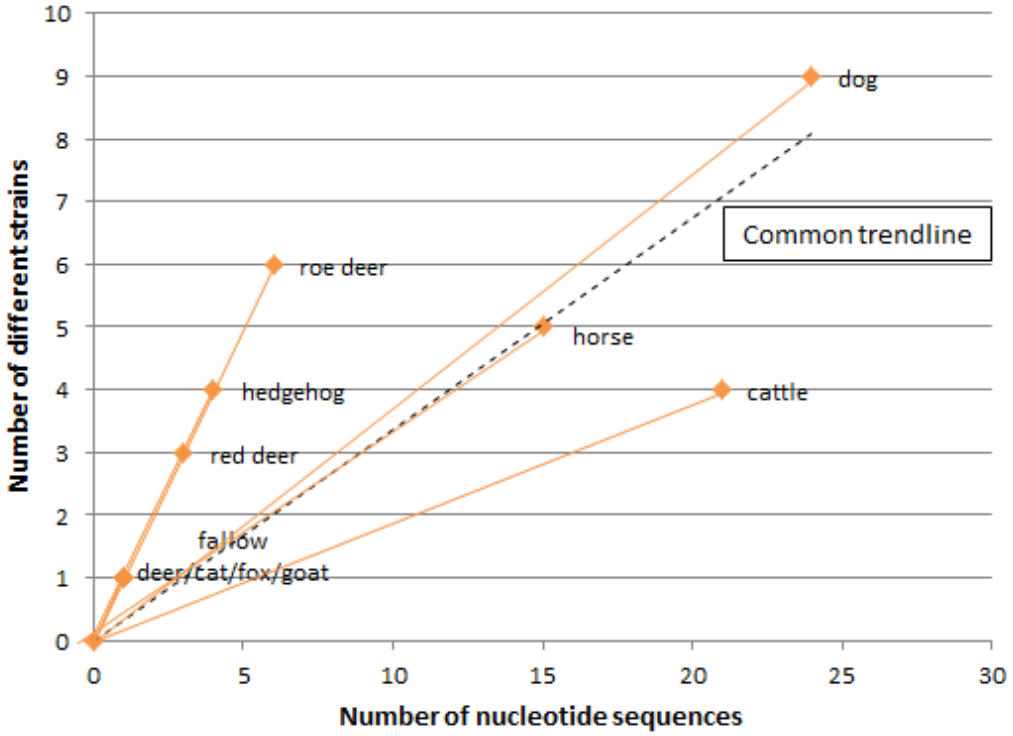

Figure 38: Trend analysis of the *msp2* gene according to the sequences obtained from animal species investigated in this study in addition to available sequences from the GenBank

4.4. Odd`s ratio

In order to statistically evaluate the chance of ruminants and non-ruminants being infected with the most common *16S rRNA* variants [16S-1(A), 16S-2(B), 16S-20(W)], the odd`s ratio was determined (Tab. 24). Thereby, the chance of variant 16S-20(W) infecting ruminants was 77.6 times higher than infecting non-ruminants. In contrast, the chance of an infection with variant 16S-2(B) was smaller in ruminants compared to non-ruminants (odd`s ratio: 0.6).

Besides the results of the odd`s ratio confirmed the higher chance of variants 16S-21(X) and 16S-22(Y) to infect wild animals compared to domestic animals (odd`s ratio: 7.1 and 6.8, respectively). Among wild animals, the chance of an infection with 16S-21(X) and 16S-22(Y) was about seven times higher than in domestic animals (Tab. 25).

Table 24: Odd`s ratio for ruminants and non-ruminants being infected with the variants 16S-1(A), 16S-2(B) and 16S-20(W)

	16S-1(A)	**Non 16S-1(A)**	**Odd`s ratio (OR)**
Ruminant[1]	0	186	0.0
Non-ruminant[2]	84	52	
	16S-2(B)	**Non 16S-2(B)**	
Ruminant[1]	13	93	0.6
Non-ruminant[2]	43	176	
	16S-20(W)	**Nicht 16S-20(W)**	
Ruminant[1]	69	120	77.6
Non-ruminant[2]	1	135	

[1] Ruminants: cattle, goat, roe deer, red deer, sika deer, fallow deer, mouflon, chamois, ibex
[2] Non-ruminants: dog, horse, cat, hedgehog, red fox, wild boar

Table 25: Odd`s ratio for wild and domestic animals being infected with the variant 16S-21(X) and 16S-22(Y)

	16S-21(X)	**Nicht 16S-21(X)**	**Odd`s ratio (OR)**
Wild animals[1]	26	192	7.1
Domestic animals[2]	2	105	
	16S-22(Y)	**Nicht 16S-22(Y)**	
Wild animals[1]	25	193	6.8
Domestic animals[2]	2	105	

[1] Wild animals: hedgehog, red fox, red deer, roe deer, fallow deer, sika deer, chamois, ibex, wild boar
[2] Domestic animals: dog, horse, cat, cattle, goat

5. Comparison with sequences from the NCBI GenBank

Eleven different ***16S rRNA*** variants of the consensus matched with diverse *A. phagocytophilum* strains obtained from the GenBank (Tab. 26). Noticeably, additional mammalian species including humans (*Homo sapiens*) from Eastern Europe (Slovenia, Czech Republic) and the USA, bisons (*B. bonasus*) from Europe (Poland), moose (*A. alces*) from Europe (Sweden), Korean water deer (*H. inermis argyropus*) from Asia and rodents from all investigated continents showed identical *16S rRNA* sequences of *A. phagocytophilum* to strains obtained from the present study (Tab. 26). Domestic animals mainly represented the variants 16S-1(A) and 16S-2(B). More heterogeneity was found in *16S rRNA* strains of *A. phagocytophilum* originating from wild ruminants. *A. phagocytophilum* strains from roe deer, for example, matched with six different *16S rRNA* strains from this study. The 16S-20(W), the variant mainly occurring in ruminants (cattle), was detected in bisons (n=14), water deer (n=1) and rodents (n=29). In the USA, this variant was also detected in a dog (n=1), in horses (n=2) and in humans (n=5). Other human strains matched with the variants 16S-1(A) and 16S-2(B). Variants 16S-1(A) and 16S-2(B) and 16S-20(W) matched with strains of *A. phagocytophilum* in humans from the USA (12 samples from Genbank) and from parts of Eastern European countries (31 samples from Genbank).

Fewer similar nucleotide sequences for the ***groEL*** and the ***msp4*** gene were available in the GenBank compared to the *16S rRNA* gene. The ***groEL*** sequences matched with twelve different *groEL* variants obtained in the present study (Tab. 27). In addition to roe deer from the present study, *A. phagocytophilum* strains from two moose from Sweden showed the *groEL* sequence g-20.

Twelve ***msp4*** variants were identical to sequences of European origin from the GenBank (Tab. 28) and included apart from animal species examined in this study, two *A. phagocytophilum* strains of bisons from Poland and one of a reindeer (*Rangifer tarandus*) from France. The strains originating from bisons were identical to variant m4-51, which was detected in five cattle in the present study (Tab. 28).

Sequences differing from sequences from the present study and therefore representing new variants of the four analyzed genes, were lowest in European samples compared to samples from the other continents (*16S rRNA*: 9.0%, *groEL*: 55.0%, *msp4*: 36.0%). Asian and Russian samples varied most, with the rate of

"no matches" reaching 72% for the *16S rRNA* gene and 100% for the *groEL* and *msp4* gene, respectively. The *msp2* sequences available in the Genbank did not match with any sequences obtained in the present study (Tab. 29).

Table 26: Distribution of the partial *16S rRNA* variants in sequences obtained from the GenBank

	Human	Dog	Horse	Cat	Hedgehog	Red fox	Cattle	Small ruminants	Bison	Moose	Roe deer	Red deer	Sika deer	Fallow deer	Water deer	Mouflon	Chamois	Ibex	Wild boar	Rodent
16S-1(A)	1	21		3		1														
16S-2(B)	43	35	31	1		1		4			1	4							14	3
16S-7(I)											9									
16S-13(O)											2									
16S-16(S)												12					2		1	3
16S-19(V)								1			2								2	2
16S-20(W)	5	1	2				6	9	14			1			1	1	1		1	29
16S-21(X)							1				20								1	
16S-22(Y)										2	13									
16S-24		1																		
16S-30								1												3
16S-nm1								2												13
16S-nm2															2					
16S-nm3		2																		
16S-nm4		2																		
16S-nm5																				5
16S-nm6																				2
16S-nm7											2									
16S-nm8											2									
16S-nm9												1								
16S-nm10																1				
16S-nm11																			1	
16S-nm12																			1	
16S-nm13															1					
16S-nm14															1					
16S-nm15								1												
16S-nm16																				1
16S-nm17		1																		
16S-nm18		1																		
16S-nm19			1																	
16S-nm20		1																		
16S-nm21		1																		
16S-nm22								1												
16S-nm23																				1
16S-nm24																				1
16S-nm25																				1
16S-nm26											1									
16S-nm27											1									
16S-nm28											1									

The abbreviation "nm" stands for "no matches".

Table 27: Distribution of the *groEL* variants in sequences obtained from the GenBank

	Human	Dog	Horse	Cat	Hedgehog	Red fox	Cattle	Small ruminants	Moose	Cervid	Roe deer	Red deer	Sika deer	Fallow deer	Water deer	Mouflon	Chamois	Ibex	Wild boar	Rodent
g-1(A)		2									1									
g-2(B)	1	1	1																1	
g-3(C)											1									
g-4(D)											1									
g-5(E)										1	1									
g-6(F)											1									
g-7(G)										1	5									
g-8(H)								1			1									
g-11/12(K)			1								1									
g-18(X)								1												
g-20									2		1									
g-35											1									
g-nm1																				6
g-nm2																				3
g-nm3	4	1	5	1																5
g-nm4																				2
g-nm5										1	1									
g-nm6										3										
g-nm7																				2
g-nm8																				2
g-nm9											2									
g-nm10									1											
g-nm11																				1
g-nm12																				1
g-nm13																				1
g-nm14								1												
g-nm15																				1
g-nm16		1																		
g-nm17															1					
g-nm18		1																		
g-nm19		1																		
g-nm20		1																		
g-nm21			1																	
g-nm22			1																	
g-nm23			1																	
g-nm24										1										
g-nm25								1												
g-nm26			1																	
g-nm27											1									
g-nm28											1									
g-nm29											1									
g-nm30											1									
g-nm31											1									
g-nm32													1							
g-nm33													1							
g-nm34											1									

The abbreviation “nm” stands for “no matches”.

Table 28: Distribution of the *msp4* variants in sequences obtained from the GenBank

	Human	Dog	Horse/Donkey	Cat	Hedgehog	Red fox	Cattle	Small ruminants	Bison	Roe deer	Red deer	Sika deer	Reindeer	Mouflon	Chamois	Ibex	Rodent
m4-2(B/C)										1							
m4-8(H)								1									
m4-5(I)								2									
m4-13(N)										4			1				
m4-15(P)							1										
m4-16(Q)							2										
m4-20								1									
m4-23										1							
m4-35										1							
m4-37								1									
m4-44											1						
m4-51							1		2								
m4-nm1																	11
m4-nm2	2																
m4-nm3																	3
m4-nm4							1										
m4-nm5							1										
m4-nm6							1										
m4-nm7										1							
m4-nm8										1							
m4-nm9																	1
m4-nm10								1									
m4-nm11									1								
m4-nm12				1													
m4-nm13								1									
m4-nm14											1						
m4-nm15											1						
m4-nm16											1						
m4-nm17										1							
m4-nm18											1						
m4-nm19											1						
m4-nm20										1							
m4-nm21							1										
m4-nm22								1									

The abbreviation “nm” stands for “no matches”.

Table 29: Distribution of the *msp2* consensus variants in sequences obtained from the GenBank

	Human	Donkey	Cattle	Sheep	Bison	Roe deer	Red deer	Rodent
m4-nm1								11
m4-nm2	2							
m4-nm3								3
m4-nm4			1					
m4-nm5			1					
m4-nm6			1					
m4-nm7						1		
m4-nm8						1		
m4-nm9								1
m4-nm10				1				
m4-nm11					1			
m4-nm12		1						
m4-nm13				1				
m4-nm14							1	
m4-nm15							1	
m4-nm16							1	
m4-nm17						1		
m4-nm18							1	
m4-nm19							1	
m4-nm20						1		
m4-nm21			1					
m4-nm22				1				

The abbreviation “nm” stands for “no matches”.

6. Phylogenetic analysis

The partial sequences of the protein-coding genes *groEL* (530 bp), *msp4* (340 bp) and *msp2* (813 bp) were phylogenetically analyzed (Figs. 39 – 41).

In general, separate clusters for European and Asian strains were described by the phylogenetic tree of the *groEL* gene with relatively high bootstrap values (91 – 99) (Fig. 39). Within the European cluster two lineages of European samples were revealed, although bootstrap values were distinctly lower for these nodes (<55) (Fig. 39). One lineage included primarily ruminants, especially roe deer (71.0%). The other lineage showed diverse animal species, among others also a human strain [g-2(B); acc. no.: AF033101]. The separate Asian cluster included primarily Asian strains of rodent and wild ruminant origin. The strains from the USA were divided between the European clusters, but were distinct from the Asian cluster (Fig. 39).

The phylogenetic tree of the *msp4* gene showed high heterogeneity between the analyzed strains (Fig. 40). Most of the *A. phagocytophilum* from European samples were part of one cluster with very low bootrap values (0 – 54), lowering the actual existence of this European cluster in nature. Noticeably, a roe deer strain from Poland differed from all the other variants (acc. no.: JN005726). High

bootstrap values of 36 – 98 were detected for a small cluster including wild ruminants from Europe and rodents from Europe and Asia.

The phylogenetic tree resulting from the partial *msp2* sequences showed a clear distinction between a European and a USA cluster (Fig. 41). No outgroup was available for the *msp2*-gene to root the phylogenetic tree, which is why deviations of the evolutionary relationships are possible. The European cluster included ten different animal species, but revealed rather low bootstrap values. Noticeably, three human strains from the USA (acc. no.: CP000235.1, AY568558.1, CP006616.1) occurred in the European cluster, very closely related to the variants m2-26 and m2-27 from cattle and a strain from a Swiss cow (acc. no.: AY706392). The cluster from the USA revealed higher bootstrap values and included *A. phagocytophilum* from dogs and a horse. A strain from a rodent (acc. no.: DQ519570) and a bear (acc. no.: DQ519567) from the USA occurred in the European cluster resembling variants from a European sheep (acc. no.: DQ519569.1) and a dog (acc. no.: DQ519568.1) strain. No Asian *msp2* sequences of the same length or the same gene region were available from GenBank.

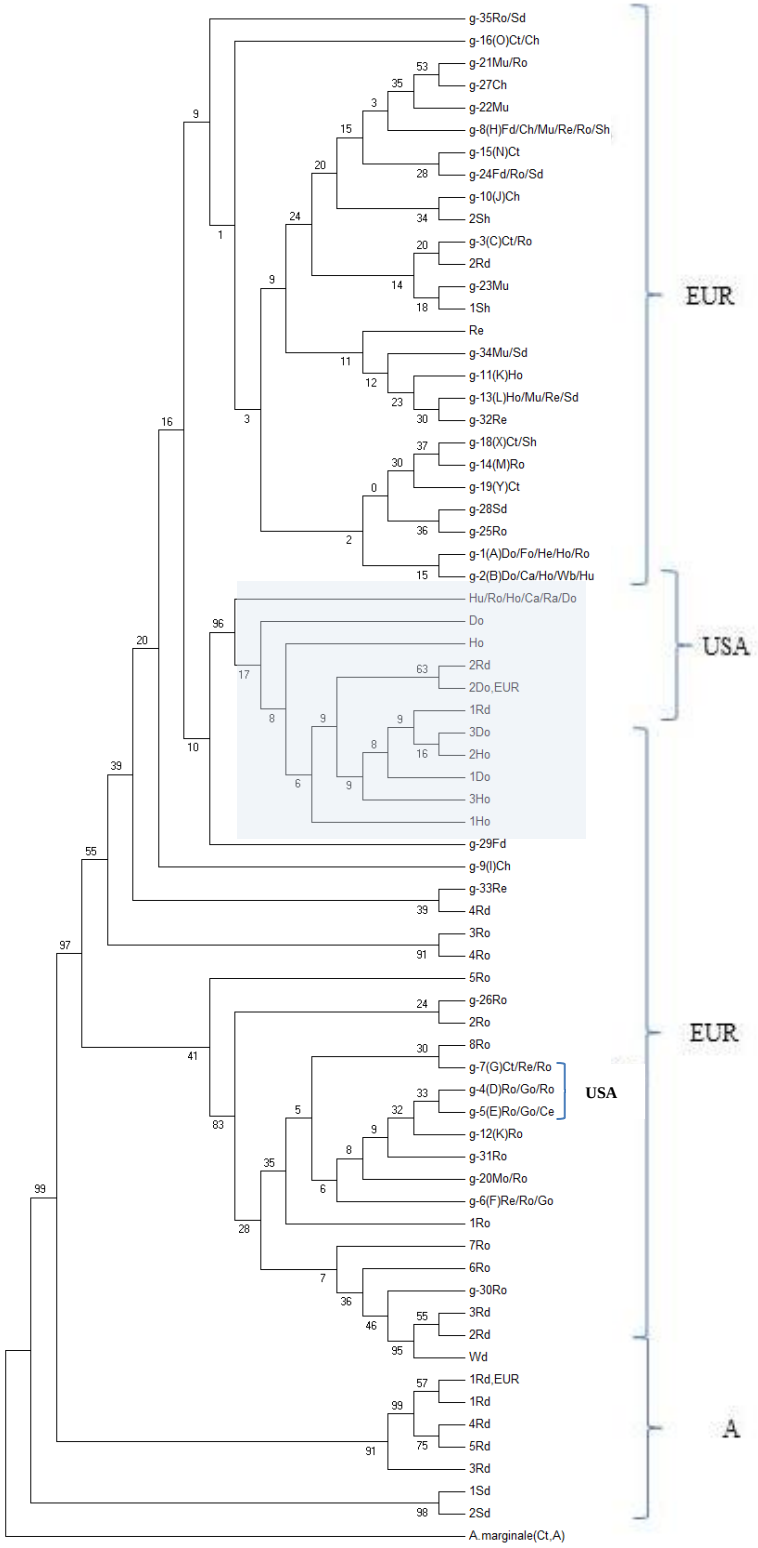
g-35Ro/Sd
g-16(O)Ct/Ch
g-21Mu/Ro
g-27Ch
g-22Mu
g-8(H)Fd/Ch/Mu/Re/Ro/Sh
g-15(N)Ct
g-24Fd/Ro/Sd
g-10(J)Ch
2Sh
g-3(C)Ct/Ro
2Rd
g-23Mu
1Sh
Re
g-34Mu/Sd
g-11(K)Ho
g-13(L)Ho/Mu/Re/Sd
g-32Re
g-18(X)Ct/Sh
g-14(M)Ro
g-19(Y)Ct
g-28Sd
g-25Ro
g-1(A)Do/Fo/He/Ho/Ro
g-2(B)Do/Ca/Ho/Wb/Hu
EUR
Hu/Ro/Ho/Ca/Ra/Do
Do
Ho
2Rd
2Do,EUR
1Rd
3Do
2Ho
1Do
3Ho
1Ho
USA
g-29Fd
g-9(I)Ch
g-33Re
4Rd
3Ro
4Ro
5Ro
g-26Ro
2Ro
8Ro
g-7(G)Ct/Re/Ro
g-4(D)Ro/Go/Ro
g-5(E)Ro/Go/Ce
g-12(K)Ro
g-31Ro
g-20Mo/Ro
g-6(F)Re/Ro/Go
1Ro
7Ro
6Ro
g-30Ro
3Rd
2Rd
Wd
EUR
USA
1Rd,EUR
1Rd
4Rd
5Rd
3Rd
1Sd
2Sd
A
A.marginale(Ct,A)

Figure 39: Phylogenetic tree of the *groEL* gene (530 bp). The nucleotide sequences were obtained in the present and previous studies (x=33) of the same study group (“g-“) and selected from GenBank (x=36). Bootstrap values >50% (1000 repeats) are presented at the respective nodes. The light blue square marks the node with a very high bootstrap value of 96 in comparison to the other nodes on the same level. (Accession nos. of nucleotide sequences from GenBank are provided in the annex, Tab. 54.)

Abbreviations: Be-Bear, Bi-Bison, Ca-Cat, Ct-Cattle, Ce -Cervid (deer), Ch-Chamois, Do-Dog, Dn-Donkey, Fd-Fallow deer, Fo-Fox, Go-Goat, He-Hedgehog, Ho-Horse, Hu-Human, Ib-Ibex, Mo- Moose, Mu-Mufflon, Ra-Rabbit, Re-Red deer, Rn-Reindeer, Rd-Rodent, Ro-Roe deer, Sd-Sika deer, Sh-Sheep, Wd-Water deer, Wb-Wild boar

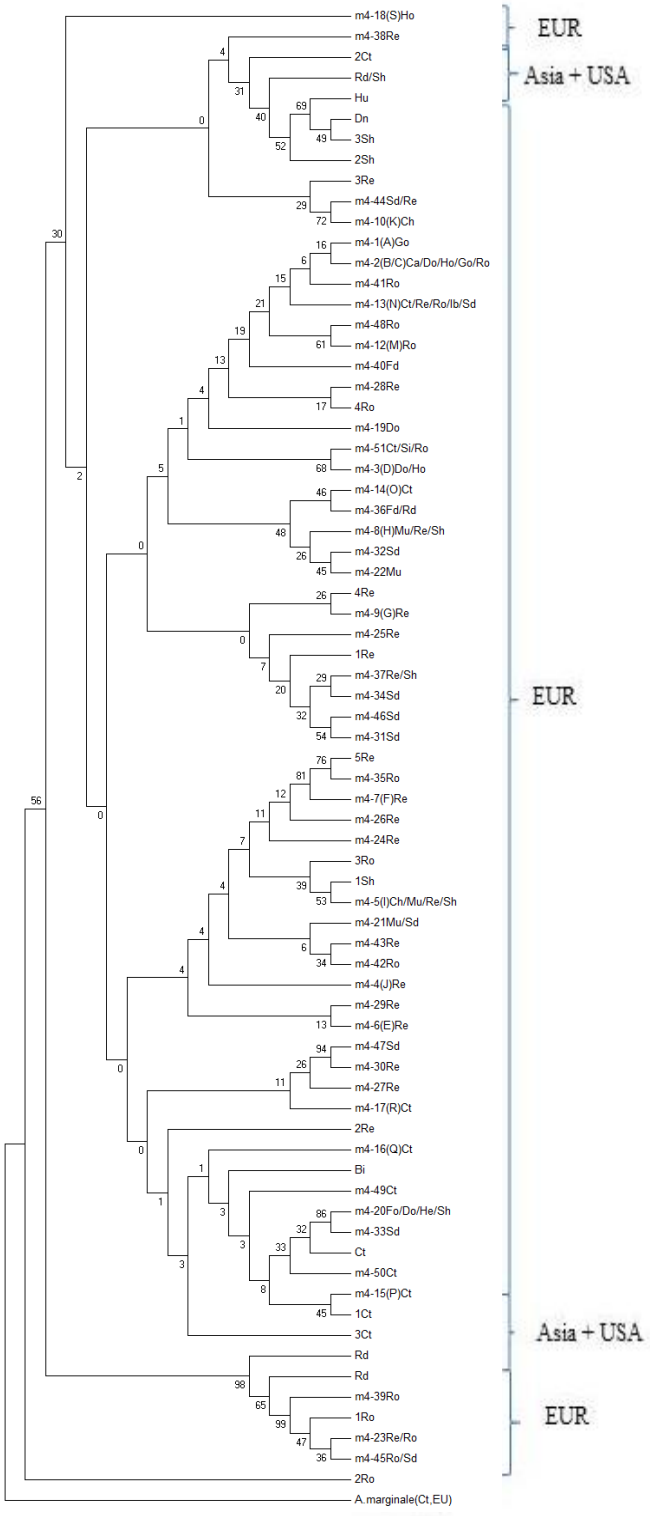
m4-18(S)Ho
m4-38Re
2Ct
Rd/Sh
Hu
Dn
3Sh
2Sh
3Re
m4-44Sd/Re
m4-10(K)Ch
m4-1(A)Go
m4-2(B/C)Ca/Do/Ho/Go/Ro
m4-41Ro
m4-13(N)Ct/Re/Ro/Ib/Sd
m4-48Ro
m4-12(M)Ro
m4-40Fd
m4-28Re
4Ro
m4-19Do
m4-51Ct/Si/Ro
m4-3(D)Do/Ho
m4-14(O)Ct
m4-36Fd/Rd
m4-8(H)Mu/Re/Sh
m4-32Sd
m4-22Mu
4Re
m4-9(G)Re
m4-25Re
1Re
m4-37Re/Sh
m4-34Sd
m4-46Sd
m4-31Sd
5Re
m4-35Ro
m4-7(F)Re
m4-26Re
m4-24Re
3Ro
1Sh
m4-5(I)Ch/Mu/Re/Sh
m4-21Mu/Sd
m4-43Re
m4-42Ro
m4-4(J)Re
m4-29Re
m4-6(E)Re
m4-47Sd
m4-30Re
m4-27Re
m4-17(R)Ct
2Re
m4-16(Q)Ct
Bi
m4-49Ct
m4-20Fo/Do/He/Sh
m4-33Sd
Ct
m4-50Ct
m4-15(P)Ct
1Ct
3Ct
Rd
Rd
m4-39Ro
1Ro
m4-23Re/Ro
m4-45Ro/Sd
2Ro
A.marginale(Ct,EU)
EUR
Asia + USA
EUR
Asia + USA
EUR

Figure 40: Phylogenetic tree of the *msp4* gene (340 bp). The nucleotide sequences were obtained in the present and previous studies (x=50) of the same study group (“m4-“) and nucleotide sequences selected from GenBank (x=22). Bootstrap values >50% (1000 repeats) are presented at the respective nodes. (Accession nos. of nucleotide sequences from GenBank are provided in the annex, Tab. 55.)

Abbreviations: Be-Bear, Bi-Bison, Ca-Cat, Ct-Cattle, Ce -Cervid (deer), Ch-Chamois, Do-Dog, Dn-Donkey, Fd-Fallow deer, Fo-Fox, Go-Goat, He-Hedgehog, Ho-Horse, Hu-Human, Ib-Ibex, Mo- Moose, Mu-Mufflon, Ra-Rabbit, Re-Red deer, Rn-Reindeer, Rd-Rodent, Ro-Roe deer, Sd-Sika deer, Sh-Sheep, Wd-Water deer, Wb-Wild boar

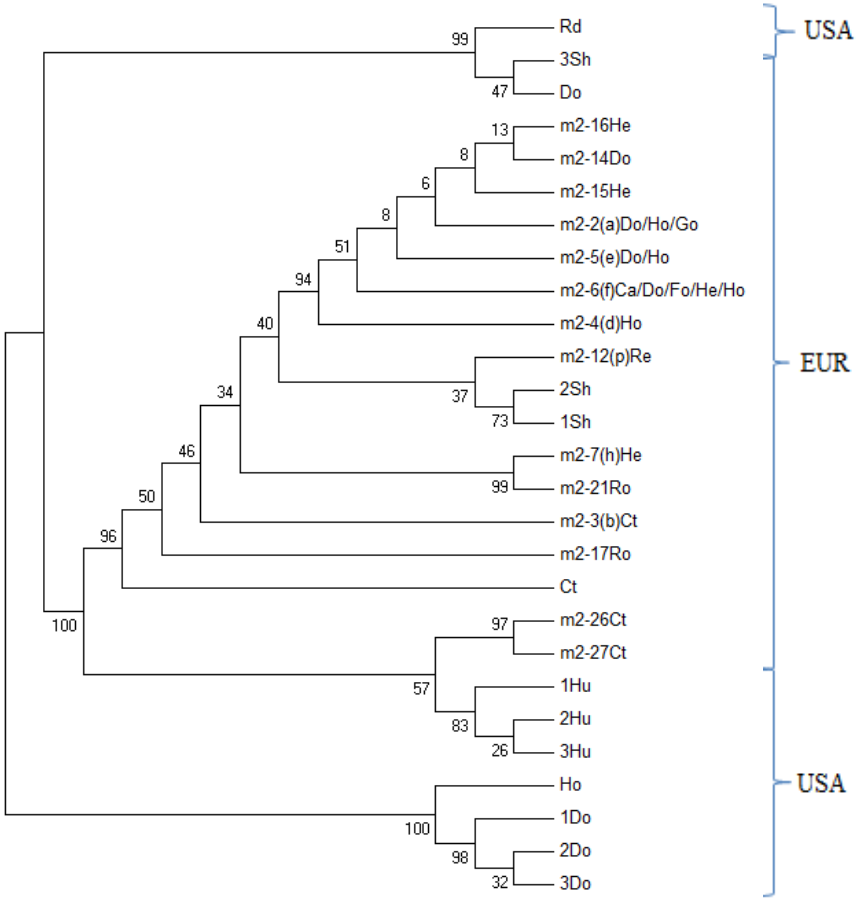

Figure 41: Phylogenetic tree of the *msp2* gene (app. 813 bp). The nucleotide sequences were obtained in the present and previous studies (x=11) of the same study group ("m2-") and nucleotide sequences selected from GenBank (x=9). Bootstrap values >50% (1.000 repeats) are presented at the respective nodes. (Accession nos. of nucleotide sequences from GenBank are provided in the annex, Tab. 56.)

Abbreviations: Be-Bear, Bi-Bison, Ca-Cat, Ct-Cattle, Ce -Cervid (deer), Ch-Chamois, Do-Dog, Dn-Donkey, Fd-Fallow deer, Fo-Fox, Go-Goat, He-Hedgehog, Ho-Horse, Hu-Human, Ib-Ibex, Mo- Moose, Mu-Mufflon, Ra-Rabbit, Re-Red deer, Rn-Reindeer, Rd-Rodent, Ro-Roe deer, Sd-Sika deer, Sh-Sheep, Wd-Water deer, Wb-Wild boar

V. DISCUSSION

1. Sequence comparison

1.1. Prevalences of *A. phagocytophilum*

The overall molecular prevalence of *A. phagocytophilum* in the investigated animal samples in this study amounted to 23% with roe deer (81%), sika deer (76%) and fallow deer (71%) reaching highest prevalence rates. Similar high molecular prevalence rates for *A. phagocytophilum* in roe deer was shown by Overzier et al. (2013a) in Southern Germany and in fallow deer by Ebani et al. (2007) in Italy. Nevertheless, the prevalence rates in European wild ruminants varied extensively. For instance, only 9.6% of investigated roe deer and 1.5% of fallow deer were infected by *A. phagocytophilum* in Poland (Hapunik et al., 2011; Michalik et al., 2009). Sika deer showed a molecular prevalence of *A. phagocytophilum* in about 40% of samples (Robinson et al., 2009; Zeman and Pecha, 2008).

In contrast, red foxes (13%) and wild boar (8%) had the lowest molecular prevalences in the present study. Similar results were detected previously in red foxes (Ebani et al., 2011), and in wild boars with a prevalence of 3.6% in Japan and 12.5% in Europe (Masuzawa et al., 2011; Michalik et al., 2012).

1.2. The *16S rRNA* variants

Altogether 23 partial *16S rRNA* variants were detected in *A. phagocytophilum* in 15 different animal species. Variant 16S-2(B) was 100% homologue to the "prototype" (acc. no.: U02521) of *A. phagocytophilum* in the amplified region and was previously described to infect both humans and diverse animal species including dogs, horses, sheep, red deer and roe deer (Chen et al., 1994; Huhn et al., 2014; Scharf et al., 2011a; Zeman and Jahn, 2009). In the USA, the white-footed mouse, humans and other mammals were infected with the identical variant of *A. phagocytophilum*, named "Ap-ha" (Massung et al., 2006). In this study, the prototype was detected in ten of the 15 examined animal species and therefore revealed high variability in its host tropism. Dogs and horses had the highest rate of infections with the 16S-2(B) variant. The same partial *16S rRNA* variant of *A. phagocytophilum* was also confirmed in several other studies in dogs and

horses (Scharf et al., 2011a; Von Loewenich et al., 2003b; Zeman and Jahn, 2009). Although less often, the present study detected the prototype in samples from wild animals including wild ruminants, red foxes and wild boars. The calculation of the odd`s ratio also confirmed that the chance of non-ruminants being infected with variant 16S-2(B) was higher than in ruminants. Nevertheless, some reports previously detected this variant in wild ruminant and in wild boar samples (Huhn et al., 2014; Kang et al., 2011; Michalik et al., 2012). This high variability of host tropism of the 16S-2(B) strain might reflect the highly developed adaptation skills of this *A. phagocytophilum* variant. It may not discriminate between its host species and could potentially infect even more host species than investigated so far. Therefore, the question arises which conditions enabled such a development of variability. Wild ruminants, which were sporadically infected with 16S-2(B) in the present study, ticks and *A. phagocytophilum* variant 16S-2(B) might form an endemic life cycle. A spillover of *A. phagocytophilum* could contribute to the distribution of the infection to other susceptible hosts such as domestic animals and humans (Silaghi et al., 2011d). As accidental dead-end hosts, humans, dogs and horses possibly develop severe disease (Parola et al., 2005; Silaghi et al., 2011c; Silaghi et al., 2011d). Silaghi et al. (2011c) confirmed symptoms of CGA in dogs infected with the prototype of *A. phagocytophilum*. Clinical symptoms were even more severe than in dogs infected with 16S-1(A) (Silaghi et al., 2011c). The wildlife animals infected with the 16S-2(B) variant from this study showed no apparent clinical symptoms of granulocytic anaplasmosis according to the hunters. However, unspecific symptoms or subclinical infection with *A. phagocytophilum* might have been overseen and cannot be fully excluded. In case the investigated wild cervids indeed lacked symptoms, the role of wild animals as reservoir hosts in the natural life cycle of *A. phagocytophilum* is supported.

Variant 16S-1(A) was detected in non-ruminants exclusively, most frequently in dogs and hedghogs. Wild ungulates might therefore play a very minor role in the distribution of this *A. phagocytophilum* variant. Although potentially less pathogenic than the prototype of *A. phagocytophilum*, the 16S-1(A) variant was mainly associated with clinically symptomatic dogs, cats and *I. ricinus* from Central Europe so far (Huhn et al., 2014; Overzier et al., 2013b; Scharf et al., 2011a). Red foxes were also infected with this variant. This was previously reported from a red fox from Germany (Hartwig et al., 2014). According to

Genbank information, the 16S-1(A) was detected in a human from Slovenia developing symptoms of HGA (Scharf et al., 2011a). Nevertheless, the role of this variant in HGA cases, and therefore the zoonotic potential of this variant, remains to be elucidated, since few cases of HGA were reported in Europe despite of high prevalence rates of this variant in ticks (Silaghi et al., 2011c). Intense contact between domestic animals and humans and increasing outdoor activities, with the subsequently higher potential of contact between humans and *I.ricinus,* might trigger the risk of transmission of the pathogen *A. phagocytophilum.* Thereby, the risk of an HGA infection with variant 16S-1(A) is considered high due to high occurrence of this variant in *I. ricinus.*

Variant 16S-20(W) occurred in wild and domestic ruminants and in one hedgehog. The calculation of the odd`s ratio also supported a higher chance of ruminants being infected with this variant than non-ruminants. Previous studies have reported the same variant in several ruminant species such as cattle, sheep, goat, bison, red deer, chamois and mouflon as well as in wild boar (Huhn et al., 2014; Michalik et al., 2012; Scharf et al., 2011a; Stuen et al., 2003). While several cases of TBF were reported in Europe, no cases of TBF in ruminants are known in the USA so far (Gokce and Woldehiwet, 1999; Nieder et al., 2012; Stuen et al., 2006). This fact indicates the evolution of a more pathogenic ruminant lineage of *A. phagocytophilum* in domestic ruminants in Europe, whereby variant 16S-(W) might play an important role. Besides, variant 16S-20(W) was also previously detected in small mammals (shrews, voles and European hedgehog) (Rar et al., 2011; Silaghi et al., 2012a). This supports the hypothesis of a common cycle of ruminants and small mammals with 16S-20(W) as proposed by Bown et al. (2009). Nevertheless, generally very low prevalences were found for small mammals among other rodents in Central Europe, which may indicate a rather minor role of these species in the natural cycle of *A. phagocytophilum* (Obiegala et al., 2014). Besides, the analysis of strains from GenBank, revealed several new variants distinct from the sequences from the present study (e.g. 16S-nm1, 16S-nm5). These results could indicate that rodents harbor several *A. phagocytophilum* strains specific for the respective rodent species. In accordance with this hypothesis, several previous reports claimed a distinct subcycle of rodent species and different ticks as vector (Bown et al., 2009; Zeidner et al., 2000). The variant 16S-20(W) was sporadically detected in humans and dogs from the USA, also known as the variant CAHU-HGE1 (acc. no.: AF093789) (Andersson and

Kurland, 1998; Chae et al., 2000; Poitout et al., 2005), but not yet in Europe. Variant 16S-20(W) seems, at least to some extent, adaptable in its host tropism, since a number of different animal species were affected.

The variants 16S-1(A) and 16S-2(B) and 16S-20(W) of the partial *16S rRNA* gene might play a role for human disease risk assessment, since these variants matched with strains of *A. phagocytophilum* infecting humans from the USA and Eastern European countries (Slovenia, Czech Republic) (Chen et al., 1994; Huhn et al., 2014). Geographical regions with high prevalence rates of these variants in domestic animals and ticks might indicate a higher risk for humans getting in touch with potential HGA strains.

Variants 16S-21(X) and 16S-22(Y) dominated in wild ruminants. Identical to the 16S-21(X) in the amplified part, Massung et al. (2006) described the putatively less pathogenic "Ap-Variant 1" of *A. phagocytophilum*, which occurred in white tailed deer (*O. virginianus*) and in goats. Another variant identical to the 16S-21(X) was detected in roe deer from the Czech Republic (corresponding Czech strain: "RV") (Zeman and Pecha, 2008). In previous studies, variant 16S-22(Y) occurred in further wild cervids, for instance in roe deer or European moose (Huhn et al., 2014; Malmsten et al., 2014; Scharf et al., 2011a). In respect of domestic ruminants, variant 16S-21(X) only occurred in two goats and variant 16S-22(Y) in a cow and a goat without apparent clinical symptoms. Since variants 16S-21(X) and 16S-22(Y) primarily occurred in wild ruminants, *A. phagocytophilum* might have developed in different lineages specializing in distinct host animal species (Massung et al., 2005). Other mammals and humans might not be susceptible for these lineages, as random screenings lacked the detection of variant 16S-21(X) and 16S-22(Y) in these species (Huhn et al., 2014; Poitout et al., 2005; Scharf et al., 2011a).

The other 18 *16S rRNA* variants only occurred scatterly in diverse animal species making a systematic characterization for these strains difficult. Nevertheless, the variation of the *16S rRNA* seemed limited compared to the other analyzed genes because of its high degree of conservation. A previous study showed eleven different *16S rRNA* variants in 195 *A. phagocytophilum* positive samples from eight animal species (horse, dog, cat, sheep, cow, roe deer, European bison, red deer) and humans, but claimed limited phylogenetic predictions according to these results (Scharf et al., 2011b).

Nevertheless, the decisive question is, whether the diversity on nucleotide

sequence level has an impact on hosts of *A. phagocytophilum*, i.e. on amino acid level. Therefore, different susceptibility of potential hosts might initiate a different virulence of diverse *A. phagocytophilum* strains, which would support the emergence of *A. phagocytophilum* variants. Most domestic animals included in the present study showed clinical symptoms, like fever or lameness. In contrast, most of the putative reservoir hosts revealed no obvious symptoms of anaplasmosis, although infected with a wide range of *A. phagocytophilum* variants, among them even strains known to cause HGA in humans [e.g. 16S-2(B)] or TBF in ruminants [e.g. 16S-20(W)]. This observation was supported by the experimental infection of lambs with *A. phagocytophilum* from a sheep and a red deer without obvious clinical symptoms. Both strains were able to cause disease in the experimentally infected sheep (Stuen et al., 2010). Besides, the experimental infection of a white-tailed deer with the human pathogenic Ap-ha strain caused a seroconversion of the affected animal, but did not cause clinical symptoms of anaplasmosis (Tate et al., 2005). These results underline the fact, that different strains of *A. phagocytophilum* can evoke different reactions in diverse animal hosts.

1.3. The *groEL* variants

The *groEL* sequence analysis revealed 33 different sequence types. In comparison to the *16S rRNA*, the *groEL* gene showed higher heterogeneity in the investigated samples.

Except from the variant g-13(L), a clear distinction between *A. phagocytophilum* strains infecting ruminants and non-ruminants was observed in the present study. This distinction was confirmed in previous studies based on other genes of *A. phagocytophilum* than the *groEL*, for instance the *16S rRNA* (Massung et al., 2005). In contrast, Von Loewenich et al. (2003a) and Petrovec et al. (2003) detected two European lineages of the *groEL* gene, whereby one of them included strains from wild ruminants and non-ruminants: One lineage primarily included *A. phagocytophilum* sequences from roe deer and the other represented *groEL* sequences isolated from red deer and other species including humans. These two lineages were also confirmed in Sardinia/Italy, Austria and Poland (Alberti et al., 2005b; Rymaszewska, 2008; Silaghi et al., 2011b).

Dog and horse samples showed the variants g-1(A) and g-2(B) exclusively. These variants were previously also revealed in dog and horse samples from other

European countries (Italy, Sweden, Slovenia) (acc. no.: EU381151, EU381151, AF478558, EU982549, AY529490) and a human HGA patient (acc. no.: AF033101) from Slovenia (Petrovec et al., 1999). Besides, wildlife, for instance hedgehog and red fox samples, revealed these variants. These results were consistent with the results of Jahfari et al. (2014). Especially variant g-1(A) also occurred in Northern whitebreasted hedgehogs (*Erinaceus roumanicus*) in Hungary (acc. no.: KF803998) (Földvári et al., 2014). Wild boar (acc. no.: EU184703) and a roe deer (acc. no.: AF478558) sample also showed these variants (Petrovec et al., 2002). Variants g-1(A) and g-2(B) showed broad host tropism including wild and domestic animals, but the majority was detected in domestic animals. Variant g-2(B) showed zoonotic potential as it matched with an *A. phagocytophilum* strain infecting Slovenian HGA patients with a history of tickbite, and it was also shown in ticks from Slovenia (Petrovec et al., 1999).

Variant g-13(L) was detected in ruminant and non-ruminant species (two horses, red deer, sika deer, mouflon). Additionally, a mouflon from the Netherlands also showed this variant (Jahfari et al., 2014). Therefore, variant g-13(L) seems more diverse in its host tropism compared to most of the other *groEL* variants detected in the present study.

Most *groEL* variants from cattle samples were restricted to cattle. Variant g-18(X), which was detected most frequently in cattle, matched with *A. phagocytophilum* from a sheep from Norway (AF548385) (Stuen et al., 2003). The variant g-3(C) was 100% homologue to a strain from roe deer from Slovenia (Petrovec et al., 2002). Both of these variants caused typical symptoms of TBF in cattle, like apathy, fever and a sudden drop of milk production. In contrast, neither *A. phagocytophilum* infections in cattle nor TBF cases were so far detected in the USA, although high seroprevalence rates were found in Connecticut for instance (Magnarelli et al., 2002). These findings might indicate that TBF is caused by very distinct *groEL* variants of *A. phagocytophilum* specialized on European ruminants, since most cattle samples originated from clinically manifest animals. Cattle without obvious clinical symptoms showed several diverse variants [e.g. g-16(O), g-18(X), g-19(Y)], whereas variants of clinically manifest cattle were rather uniform [g-3(C) and g-18(X)].

Roe deer represented the animal species with greatest heterogeneity in the *groEL* gene variants. Except from variants g-6(F), g-7(G) and g-20, all *groEL* variants (eight variants) occurred only once in a single animal sample confirming the high

variability of *A. phagocytophilum* in roe deer. Additionally, eight variants deriving from GenBank were also restricted to roe deer. This great heterogeneity of *A. phagocytophilum* in roe deer samples was also detected in previous studies, whereby only few *groEL* variants matched with the variants occurring in the present study (Liz et al., 2002; Petrovec et al., 2002; Petrovec et al., 2003). Variants g-20 or g-7(G) and g-4(D), for instance, occurred in roe deer samples from Switzerland and Slovenia (Liz et al., 2002; Petrovec et al., 2002). Nine additional variants (acc. nos.: JN005748, JN005747, AF478561, AY220469, AY220468, AF478563, AF478556, AF478554, AF478553) were revealed analyzing GenBank information. According to the striking amount of variable strains of *A. phagocytophilum* restricted to roe deer, most of these variants might circulate between ticks and roe deer exclusively. Few might be transmitted to domestic animals and potentially humans by accident (Silaghi et al., 2011d).

However, the zoonotic impact of the *A. phagocytophilum* variants occurring in wild ruminants is still unclear. Only variant g-24 was 100% homologue to a strain previously detected in a HGA patient from Belgium (Jahfari et al., 2014). In the present study, this variant showed a broader range of susceptible hosts by infecting a roe deer, a sika deer and a fallow deer. Besides, Liz et al. (2002) described three *groEL* strains isolated from Swiss roe deer showing strong homology to human strains, but were not completely identical (≥99.7 and ≥98.6%, respectively). Two of these variants matched with g-7(G) and g-4(D), also detected in roe deer in the present study. These results indicate that some *A. phagocytophilum* variants circulating in wild cervids and ticks are indeed capable of infecting humans. Nevertheless, the low number of matches between strains from wild cervids and humans might suggest humans as accidential hosts.

Most changes on nucleotide level were silent on amino acid level. Therefore, the structure of the resulting protein might not be influenced by most of the point mutations in the nucleotide sequences and the pathogenicity of these *A. phagocytophilum* strains might remain similar.

Remarkably, a division of all *groEL* sequences into two groups was observed depending on the amino acid in position 48. *A. phagocytophilum* variants with the amino acid alanine in this position originated either from goats or wild ruminants, predominantly from roe deer, whereas all of the other variants showed serine. The exchange from a polar amino acid (serine) to the non-polar amino acid (alanine) in

this position might lead to a distinct protein folding of the heat shock protein *groEL* and could potentially influence pathogenicity and host tropism of the particular *A. phagocytophilum* strain. A modification of the expression of the *groEL* protein might enable *A. phagocytophilum* to circumvent the immune system of potential hosts and therefore find access to new hosts. A clear host tropism for ruminants, for example, was confirmed by the alanine group, which was also suggested on nucleotide level. The same amino acid change was detected previously in a study comparing *A. phagocytophilum* isolates from roe deer and red deer (Rymaszewska, 2008). In contrast, two *A. phagocytophilum* strains with the *groEL* gene differing in three nucleotides caused a different clinical manifestation in experimentally infected sheep, although no changes on amino acid level were detected (Stuen et al., 2003). One of these strains was identical to the g-18(X), which was also revealed previously in TBF manifest cattle (Silaghi et al., 2011e). In the present study, variant g-18(X) was often associated with variant 16S-20(W) in a single *A. phagocytophilum* strain, which was also shown to cause TBF in cattle. This result might indicate a distinct host tropism for special *A. phagocytophilum* strains in cattle.

1.4. The *msp4* variants

Altogether 50 different variants with an identity score of >87.9% were detected with the *msp4* PCR assay. High heterogeneity of the *msp4* gene was also shown by Bown et al. (2007a) with eleven variants occurring in 20 examined samples from six different animal species (dog, cow, sheep, goat, roe deer, boar) and humans. Unfortunately, the eleven detected *A. phagocytophilum* variants could not be used for comparison in the present study due to short sequence lengths (301 bp vs. 340 bp in the present study). Besides, sheep and red deer from Norway revealed nine different *msp4* variants, of which five had not been described in the GenBank before (Stuen et al., 2013b). These sequences were also too short for comparison (287 bp vs. 340 bp in the present study). It seems, however, that there are still multiple *msp4* variants in nature remaining to be detected, as additional 22 sequences were obtained from the GenBank (m4-nm1 – m4-nm22) (Stuen et al., 2013b). Since *major surface proteins* are in contact with the immune systeme of the different (reservoir) hosts, extensive variability might result as natural selection. Therefore, the number of *msp4* variants was clearly higher than the number of variants of the other analyzed genes. *A. phagocytophilum* has

developed a strategy to survive and even expand in nature with diverse strains infecting different animal species.

Except from variant m4-2(B/C), the *msp4* variants were clearly differentiated between ruminants and non-ruminants. Thus, the present study confirmed the study by De la Fuente et al. (2005a) with over 50 animal (dogs, horses, donkeys, white-tailed deer, roe deer, sheep, European bison and cattle) and human samples from Europe (Germany, Poland, Norway, Italy and Switzerland) and the USA. Clearly, *A. phagocytophilum* strains from humans, dogs and horses were distinguished from ruminant strains (De la Fuente et al., 2005a). Of these detected *msp4* strains, *A. phagocytophilum* from a cow (acc. no.: AY530198) and two bison (acc. no.: AY706387, AY706389) samples matched with variant m4-51. This strain was also confirmed to infect cattle in the present study. Additionally, De la Fuente et al. (2005a) detected variant m4-23 in a roe deer originating from Germany (acc. no.: AY706386), which also occurred in roe deer and red deer samples in the present study. Therefore, *A. phagocytophilum* strains might have coevolved from different development of lineages specializing on ruminants and non-ruminants as hosts. Consequently, most variants originating from ruminants might not play a role in the infection of humans and domestic animals.

Variant m4-2(B/C) was the only strain to infect both ruminants and non-ruminants. This variant occurred primarily in dogs and horses, but was also shown in roe deer and goats. Similar to the prototype [16S-2(B)] of *A. phagocytophilum*, this variant might be an indicator for broad host tropism. Indeed, variant m4-2(B/C) was detected together with variant 16S-2(B) and 16S-1(A) in several dogs and horses. The respective gene combinations m4-2(B/C) / 16S-2(B) and m4-2(B/C) / 16S-1(A) might be concentrated on the infection of domestic animals. Nevertheless, the possibility of *A. phagocytophilum* strains with m4-2(B/C) to infect further animal species is probably higher compared to other strains, which seem more specialized.

Twenty-seven of the fifty *msp4* variants were detected in a wild cervid exclusively. Additionally, nine variants also infecting wild cervids, were obtained from GenBank. Especially red deer, sika deer and roe deer variants revealed high heterogeneity. Similarly, the *A. phagocytophilum* strain from a German roe deer (acc. no.: AY706386), which was 100% homologue to variant m4-23, also showed most variation in comparison to the other investigated *A. phagocytophilum* strains in the study of De la Fuente et al. (2005a).

Nevertheless, it remains unclear whether this is a result of genomic diversity or environmental pressure. Since major surface proteins, like the *msp4* gene, are in steady interactions with the immune system of hosts and vectors, selective pressure could result in faster evolution of new *A. phagocytophilum* strains (De la Fuente et al., 2001). Wild ruminants as reservoir hosts are inapparent permanent carriers of the pathogen. Therefore, antigen variation of *A. phagocytophilum* might be essential in order to allow its persistence in the organism and therefore in nature. Noticeably, the amplification of the *msp4* gene was less successful compared to the *16S rRNA* protocol. This finding is in line with the results of De la Fuente et al. (2005a), where also less *A.phagocytophilum* samples were amplified with the *msp4* PCR protocol, than with the *16S rRNA* protocol.

The translation into amino acid sequences did not reflect the high variability detected on nucleotide level with only ten different variants of the 50 potentially different strains. Accordingly, De la Fuente et al. (2005a) revealed a higher identity level among the analyzed *A. phagocytophilum* samples on protein level compared to the nucleotide level. The variants m4-23, m4-39 and m4-45 were very similar to each other and differed from the other strains remarkably (variation in six amino acid positions). All of these strains originated from wild cervids. Strains derived from ruminants have also previously shown a greater divergence than non-ruminants in comparison to the amino acid sequence of the HZ strain (acc. no.: AY530194). Thereby, equidae (horses and donkeys) and dogs differed in not more than one amino acid, whereas roe deer varied most with 23 amino acid substitutions (De la Fuente et al., 2005a). Therefore, a coevolution of different strains, especially in wild cervids, seems likely.

1.5. The *msp2* variants

The *msp2* strains of *A. phagocytophilum* obtained from animal samples were remarkably diverse. Only 71 *msp2* gene sequences were successfully sequenced. Among them, 22 different variants were detected, in parts varying extensively in nucleotide sequences with a similarity score of only >67.1%. Three groups of variants resembling each other were detected on nucleotide level. Except from one strain from a hedgehog, the first group of *A. phagocytophilum* variants originated from wild cervids (roe deer, fallow deer) exclusively. Similarly, a study from Poland revealed strains of deer ("genotype 3": acc. no.: DQ105671) resembling each other in contrast to *A. phagocytophilum* from Polish tick and dog samples

and HGE strains from the USA (Rymaszewska, 2010). Unfortunately, the length of the Polish nucleotide sequences was too short (334 bp) for comparison with strains from the present study. Besides, Lin et al. (2004a) discriminated a *msp2* variant from a ruminant (sheep) from the United Kingdom (acc. no.: AY541004) distinct from the other variants causing clinical disease in domestic animals (dogs, horses) and humans from the USA. Similarities were also shown for variant Ap-V1 infecting ruminants in the USA and a sheep strain from Norway, whereas a discrimination of strains infecting humans and dogs was confirmed (Morissette et al., 2009). The distinct ruminant cluster revealed in the present study only included wild ruminants without obvious clinical symptoms. In contrast, two *A. phagocytophilum* strains from clinical symptomatic cattle (m2-26 and m2-27) were part of the second group. These two *msp2* variants resembled each other strongly and were never described in GenBank before (Nieder et al., unpublished). An *msp2* sequence originating from Swiss cattle (acc. no.: AY706392) was available in the GenBank with a similarity score of 94.9% compared to m2-26 and m2-27. These Swiss cattle samples were 100% homologue to each other, whereas the two strains differed from an ovine strain (acc. no.: A706393) (De la Fuente et al., 2005a). Therefore, European variants of *A. phagocytophilum* causing TBF in cattle might differ from the strains occurring in the USA and in European sheep.

Accordingly, all of the strains from GenBank not matching with the sequences from the present study (m2-nm1 – m2-nm18) originated from the USA, except from *A. phagocytophilum* from European ruminants (acc. nos.: AY706393, AY706392, AY706393). Previous reports support this mismatch between European and US strains (Barbet et al., 2006; Morissette et al., 2009; Rymaszewska, 2010). Barbet et al. (2006), for example, revealed a low similarity score of less than 90% in dog and sheep strains from Norway and Sweden compared to US strains. Since the present study confirmed the high rate of heterogeneity of *msp2* strains detected in European animals versus US strains from GenBank and no clinical case of TBF was reported in the USA so far, the risk of an infection with a TBF causing *A. phagocytophilum* strain might be higher in Europe indicating a higher importance of prevention programmes for cattle in Europe. Nevertheless, short termed or subclinical fever periods of TBF might be overseen easily due to extensive cattle farming in the USA.

The third group consisted of all the remaining *A. phagocytophilum* strains mainly originating from non-ruminants. Thereby, variants m2-6(f) and m2-2(a) were the

only variants detected in symptomatic domestic animals (dogs, horses, cats) and scatterly also in ruminants. Accordingly, *A. phagocytophilum* from a dog and a roe deer from Poland were 100% homologue to each other ("genotype 2": acc. no.: DQ105670) (Rymaszewska, 2010). These findings might imply the potential of flexibility of the immunodominant *msp2* gene, although the other genes analyzed in the present study showed a broader host range of some variants [e.g. 16S-2(B), m4-2(B/C), g-1(A)].

More than half of the *msp2* variants occurred only once in domestic and wild animals alike [e.g. m2-11(n), m2-14, m2-25]. Such a striking variability of the *msp2* gene could probably be the result of interactions with the adaptive immune system of the animal or human host or of environmental pressure. Since the *msp2* gene is considered as immunodominant, the occurrence of different *msp2* variants might facilitate the evasion of specific immune responses of its host by antigenic variation (Brown, 2012). This adaptive function of *A. phagocytophilum* was confirmed by the detection of cyclic waves of bacteraemia in experimentally infected lambs during a persistent infection (Granquist et al., 2010). Noticeably, especially putative reservoir hosts showed single *msp2* genes, which might be a way to circumvent the reservoir host`s immune system. Nevertheless, a confirmation of this hypothesis was not possible due to limited numbers of available sequences. Of course, the environmental pressure exerting on *A. phagocytophilum* in different habitats might also be a key driver in the development of new *msp2* variants. Barbet et al. (2003) described a predominance of special *msp2* variants when conditions of *in vitro* cultured *A. phagocytophilum* strains in human (HL-60) and tick (ISE6) cell lines were changed. Further studies including cell lines from different hosts and reservoir hosts might be useful to understand the change of the *msp2* expression in the different susceptible hosts as shown in the present study.

The translation of the *msp2* sequences into the amino acid sequences confirmed two of the three groups, detected on nucleotide level. Thereby, the two variants from cattle, m2-26 and m2-27, differed clearly from all of the other variants. Several changes of polar and unpolar amino acids (e.g. alanine to asparagine) likely result in a change of protein folding. These findings are consistent with the results of De la Fuente et al. (2005a) differentiating *A. phagocytophilum* from ruminant strains and domestic animal and human strains. Apart from these variants, most of the nucleotide changes of the *msp2* sequences were silent. The

msp2 sequences of roe deer, dog and tick samples from Poland only showed four amino acid changes compared to a HGA strain from the USA (acc. no.: AY164490). According to Rymaszewska (2010), none of these changes would have an impact on the structure of *Anaplasma* strains. Surprisingly, the high diversity shown for the nucleotide sequences was not confirmed on amino acid level, questioning the clinical impact of these *A. phagocytophilum* strains. Nevertheless, the *msp2* gene indeed confirmed a high potential of variance and therefore the possibility of the development of pathogenic strains affecting domestic animals and humans exists.

In summary, the *msp2* gene might be an important tool for the bacterium *A. phagocytophilum* enabling flexible adaptation in different stages and environments of its life cycle. Nevertheless, the comparison of *msp2* nucleotide and amino acid sequences revealed poor additional information concerning epidemiological questions of *A. phagocytophilum* in the present study. Its structure of a genomic expression site with ≈100 functional pseudogenes including conserved and hypervariable regions resulting in a high diversity of *msp2* variants might make an interpretation of the strains as evolutionary divergence or spontaneous recombination difficult (Barbet et al., 2003).

2. Evaluation of the PCR assays

In order to gain a general overview of *A. phagocytophilum* variants and their distribution among the investigated animal species, the sequence of the partial *16S rRNA* gene was determined. Thereby, the *16S rRNA* PCR protocol designed by Massung et al. (1998) was most successful with respect to the amount of sequences obtained [325 of the 781 samples (41.6%) nucleotide sequences] in comparison to the amplification assays of the other genes analyzed in the present study. A similar success rate was previously shown in a study investigating 346 ruminant samples, which revealed 194 *16S rRNA* sequences (Scharf et al., 2011a). The sequencing of the *16S rRNA* was a very reliable tool for a first classification of the involved *A. phagocytophilum* strain into ruminant and non-ruminant strains. Nevertheless, this gene seemed too conserved for the analysis of genetic variation in *A. phagocytophilum* strains with only 23 different *16S rRNA* variants detected in this study. In previous reports, the identification of *16S rRNA* variants for

molecular characterization was therefore discussed controversially. Bown et al. (2009), for instance, claimed that the sequencing of the partial *16S rRNA* gene does not provide enough information for the detection of distinct ecotypes of *A. phagocytophilum* in Europe. Since diverse *A. phagocytophilum* strains were detected in this and other European studies from wild ruminants, the *16S rRNA* was questioned as marker for similar life cycles of *A. phagocytophilum* as shown in the USA with only two confirmed *16S rRNA* variants (Ap-ha and Ap-V1) (Bown et al., 2009). To date, other studies confirmed 15 different *16S rRNA* variants in the USA among these also the two variants Ap-ha and Ap-V1 (Rar and Golovljova, 2011). These reservations concerning the *16S rRNA* were supported by the present study as it was impossible to reflect the real potential of diversity provided by *A. phagocytophilum.* Therefore, the investigation of additional genes was useful in order to gain further information concerning the evolution of different *A. phagocytophilum* variants.

The success rate of the sequencing of the *groEL* gene was comparable to the sequencing of the *msp4* gene with 172 (22.0%) and 174 sequences (22.3%), respectively. Even more variability was detected in the alignment of all *msp4* nucleotide sequences with 50 variants versus 33 different *groEL*variants. This analysis of the *groEL* and the *msp4* was essential in order to further genetically characterize the *A. phagocytophilum* strains of the investigated animal samples. In the alignments of the variants, a clear differentiation of ruminant and non-ruminant strains and therefore a possible coevolution of *A. phagocytophilum* strains was indicated by the *16S rRNA* amplification, but was clearly confirmed by the *msp4* and the *groEL* sequences. This definition of different lineages, especially between ruminant versus human strains, was consistent with the findings of De la Fuente et al. (2007) and Bown et al. (2007b). Several single variants occurring in wild cervids exclusively were shown by the sequencing of these two genes. Red deer, for instance, showed five single *groEL* variants and even 13 single *msp4* variants only infecting one wild cervid in the present study. The special role of European roe deer in the maintenance of *A. phagocytophilum* in nature became obvious by the sequencing of the *msp4* and the *groEL* gene, although its role is not yet fully explored. Due to a sporadical matching of *msp4* and *groEL* strains of *A. phagocytophilum* from roe deer and both ruminants and non-ruminants, roe deer as reservoir host might form a bridge host for these strains between the other hosts. This result was also supported by Rymaszewska

(2008) in the comparison of *groEL* nucleotide sequences from roe deer and red deer. In contrast, several other studies suggested a separate life cycle with roe deer specific *A. phagocytophilum* strains excluding variants infecting humans and domestic animals (Chastagner et al., 2014; Dugat et al., 2014a; Scharf et al., 2011a).

The *msp2* gene was successfully sequenced in only 9.1% of the samples. The amplification resulted in 22 different variants and the lowest similarity score of 67.1% compared to the other analyzed genes. Therefore, the great diversity of European *msp2* nucleotide sequences was confirmed in accordance with previous reports (Lin et al., 2004a; Morissette et al., 2009). The success rate of the *msp2* PCR in previous studies was similarly low with only 4.6% in Korean Water Deer (Kang et al., 2011). In contrast, *msp2* of *A. phagocytophilum* originating from the USA was said to be highly conserved, at least in some parts, and even comparable to the degree of conservation of the *16S rRNA* gene (Lin et al., 2004a). Accordingly, a similar conventional *msp2* PCR assay was claimed to perform with an excellent specificity and sensitivity comparable to the *16S rRNA* assay used in the present study (Massung and Slater, 2003). However, limited sequence identity was found between US strains and a sheep and a dog strain from Norway and Sweden, respectively (Barbet et al., 2006). The great variability of European *A. phagocytophilum* strains might have been the reason for a limited outcome rate of the *msp2* amplification in the present study. Apart from the differentiation of *msp2* sequences from different continents, the gain of new information resulting from the *msp2* PCR assay concerning host associations was limited.

3. Statistical analysis of *A. phagocytophilum* strains

Although Dunning Hotopp et al. (2006) supported a reorganization aiming for one common species for *A. phagocytophilum*, there are differences between strains infecting different animal species supported by the statistical analysis of the present study. In general, the variance (mean and empirical variance) of *A. phagocytophilum* strains from ruminants was higher than in non-ruminants, at least in terms of the *16S rRNA*, the *groEL* and the *msp4* gene. In conclusion, ruminants showed a tendency for greater genetical variation than non-ruminants, in particular wild ruminants. This finding was also confirmed by the trend analysis. In contrast, a rather uniform range of *A. phagocytophilum* strains in dogs, horses and especially in hedgehogs was shown with a low mean and

empirical variance. Hedgehogs in particular revealed the lowest slope in the trend analysis regarding the *groEL* and the *msp4* gene. The trend analysis was a way to demonstrate and compare the variability of the different number of available *A. phagocytophilum* strains from different animal species. Most other studies rather concentrate on the investigation of few animal species infected with *A. phagocytophilum*, wherefore this study enabled a broad insight of *A. phagocytophilum* in nature (Stuen et al., 2013b). Nevertheless, the continuation of the line with an increasing number of investigated *A. phagocytophilum* strains in different animal species remains to be elucidated, e.g. the variability of the *16S rRNA* gene in domestic animals might be limited, whereby the straight line of the trend analysis would flatten. Possible reasons for different levels of diversity of *A. phagocytophilum* in different animal species could either relate to the divergence of evolving strains with a special host tropism in distinct epidemiological cycles or to a different response of animal species to certain *A. phagocytophilum* strains.

In respect of gene combinations, gene variants 16S-1(A) and g-1(A) in combination with different major surface protein sequences [m4-2(B/C), m4-20, m2-6(f)] occurred in dogs, horses, foxes and hedgehogs and thereby revealed great uniformity. Especially the strain combination 16S-1(A), g-1(A), m4-20 and m2-6(f) was detected in hedgehogs very often. Therefore, the high level of conservation of the *16S rRNA* and the house-keeping gene *groEL* was confirmed, whereas the major surface proteins varied remarkably (Dasch et al., 1990; Jahfari et al., 2014; Silaghi et al., 2011d; Woese et al., 1975). In contrast, wild cervids showed great differences in comparison to other animal species, but same gene combinations within the same animal species. Thereby, strains with more than one successfully sequenced gene did not match with ruminants and non-ruminants in respect of gene combinations. Although less diversity of *A. phagocytophilum* was detected in the USA, a similar situation was described. The co-existence of the Ap-ha strain infecting domestic animals and humans and the Ap-V1 infecting ruminants was shown (Massung et al., 2005; 2006). Foley et al. (2008a) even described a mismatch between strains infecting domestic animals and humans and strains from wildlife reservoir hosts, like the dusky-footed woodrat in Western US. With many potential reservoir hosts living in a common habitat, the transmission of different strains by hard ticks and the co-evolution of different *A. phagocytophilum* genotypes could potentially be enforced.

4. Phylogenetic analysis

The evolutionary relationship between *A. phagocytophilum* strains can provide information about geographic distribution patterns and the development of host-vector-parasite interactions (Brooks and McLennan, 1991; Weller et al., 1998).

Two distinct European lineages were detected based on the phylogenetic analysis of the *groEL* gene. Roe deer was the dominating animal species in one of the lineages representing over 71%. Strains with zoonotic potential occurred in the second lineage exclusively. Our study therefore confirms previous reports of these two main lineages in Europe (Alberti et al., 2005a; Petrovec et al., 2002; Silaghi et al., 2011d; Von Loewenich et al., 2003a). To date, variants involved in human infection in Europe clustered in the second lineage (Petrovec et al., 2002). Although these strains were not detected in red deer directly, *A. phagocytophilum* from red deer were also closely related to human HGA strains as they were part of the same cluster. A similar clade including wild ungulates, humans and horses was also shown in a Slovenian study, whereas rodents grouped in another cluster (Vichova et al., 2014). In the present study, Asian *groEL* variants originating from *A. phagocytophilum* of rodent and deer samples formed a separate cluster distinct from other continents. Such a separate Asian clade distinct from other continents was also described by Zhan et al. (2010b) analyzing *A. phagocytophilum* from clinical symptomatic livestock and rodents. This result might confirm the hypothesis of a common natural life cycle of *A. phagocytophilum* including rodents and ungulates at least in Asia (Telford et al., 1996; Yang et al., 2013).

The phylogenetic analysis of the *msp4* sequences was unable to confirm the clear clustering of *A. phagocytophilum* strains originating from same continents. Instead, the *msp4* variants clustered according to their host animal species as ruminant (wild and domestic ruminants) variants were distinguished from human, dog and horse variants. This finding is consistent with previous reports analyzing the *msp4* gene (Bown et al., 2007b; De la Fuente et al., 2005a). Besides, a separate cluster of *msp4* strains included strains of some of the European wild ruminants (roe deer, red deer and sika deer) and also rodent strains from China (acc. no.: EU008082) and Slovakia (acc. no.: KF420102). This result clearly supports the phylogenetic analysis of the *groEL* gene in the present study. Thus, a common endemic life cycle for ruminants and rodents in Asia might be possible. Noticably, a roe deer from Poland (acc. no.: JN005726) clearly diverged from all

the other *msp4* strains in the phylogenetic analysis, although the strain showed no striking abnormalities on nucleotide level with a similarity score of 91.5 – 97.4%. Similar observations were detected in previous reports with an *A. phagocytophilum* strain from roe deer from Germany (acc.no.: AY706386) (Bown et al., 2007b; De la Fuente et al., 2005a). This strong diversity might have evolved either from coevolutionary divergence or from selective pressure as immunodominant protein. In any case, the phylogenetic analysis of the *msp4* gene supports the suggestion of an own subcycle of roe deer specific *A. phagocytophilum* strains.

The phylogenetic tree of the partial *msp2* gene revealed a distinction of a European and a US cluster. Accordingly, a previous study comparing strains of *A. phagocytophilum* from the USA and Poland also supported a low level of congruency between the different genotypes (Rymaszewska, 2010). The high variability of European *msp2* sequences confirmed a high evolutionary flexibility and might therefore enable the pathogen to quickly adapt and distribute to new habitats and animal hosts. Additionally, Kang et al. (2011) also described a clustering of European versus Asian and American strains, when examining *msp2* sequences of Korean water deer as potential reservoir host for *A. phagocytophilum*. Thereby, *A. phagocytophilum* from a cow in Switzerland (acc. no.: AAW32620) and a red deer from Austria (acc. no.: ACO40509) differed from human and equine strains from the USA (acc. no., AAT51861, AAT08123, AAX07946) and Korean water deer (acc. no.: ADO34909). This finding might indicate a variable conservation rate of the *msp2* gene of *A. phagocytophilum* of different origin. Unfortunately, no *msp2* sequences of compatible length were available from Asian countries. Nevertheless, partial *msp2* gene sequences of *A. phagocytophilum* in Korean water deer or Japanese sika deer showed very uniform sequences and therefore resembled US strains (Gaowa et al., 2012; Kang et al., 2011). The US cluster of the present study consisted of dog and horse samples. Such a distinct cluster was also shown by Morissette et al. (2009), additionally also including human strains from the USA. Human strains from the USA clustered in the European branch in the present study showing close relationship to European variant m2-26 and m2-27 from cattle. These variants were remarkably different from the other strains on nucleotide and amino acid level. The resemblance to human pathogenic strains from the USA might indicate the emergence of new pathogenic strains of *A. phagocytophilum* different from the

strains in Europe detected so far. Similarly, a *msp2* variant of a rodent and a bear sample occurred in the American cluster. As incidental dead-end host, humans might not be infected by most of the *A. phagocytophilum* strains circulating in European wildlife.

5. Possible natural life cycles of *A. phagocytophilum*

Based on analysis of the nucleotide sequences *16S rRNA*, the *groEL*, the *msp4* and the *msp2* originating from the present study and the GenBank, two life cycles could be suggested for *A. phagocytophilum* in nature. Considering these suggestions, a certain kind of sampling bias due to availability of samples has to be taken into account. For example, *A. phagocytophilum* from horses and several dogs showed clinical signs of anaplasmosis, whereas other animal species (hedgehogs, roe deer, red deer, sika deer, fallow deer, mouflon, chamois, red fox, ibex, wild boar) were free of obvious clinical symptoms. Besides, most samples of an animal species originated from few different or the same places in Central Europe. In order to compensate this sampling bias, corresponding *A. phagocytophilum* genes available from GenBank were taken into consideration for sequence variability, host spectrum and phylogenetic analysis and consequently for this evaluation of potential endemic life cycles of *A. phagocytophilum.*

5.1. A possible life cycle with domestic animals

The first cycle includes hedgehogs, foxes and red deer as reservoir hosts and domestic animals and humans as hosts of *A. phagocytophilum* (Fig. 42). Ticks as vectors spread the pathogen to other animal species. Especially *I. ricinus* with its broad host range transmitts the bacterium to several host animals. All animal species included in this potential life cycle had variant 16S-2(B) in common. With the exception of red deer, the variant 16S-1(A) was also shared by all animal species and occurred as predominant *16S rRNA* variant in hedgehogs. On the basis of the *msp4* gene, domestic animals shared the variant m4-20 with hedghogs and red foxes as reservoir hosts, and the m4-2(B/C) variant with red deer as reservoir host. Variant g-13(L) was the only *groEL* variant including red deer and domestic animals (horses). Variants g-1(A) and g-2(B) were shared by the rest of the animal species in this potential *A. phagocytophilum* life cycle.

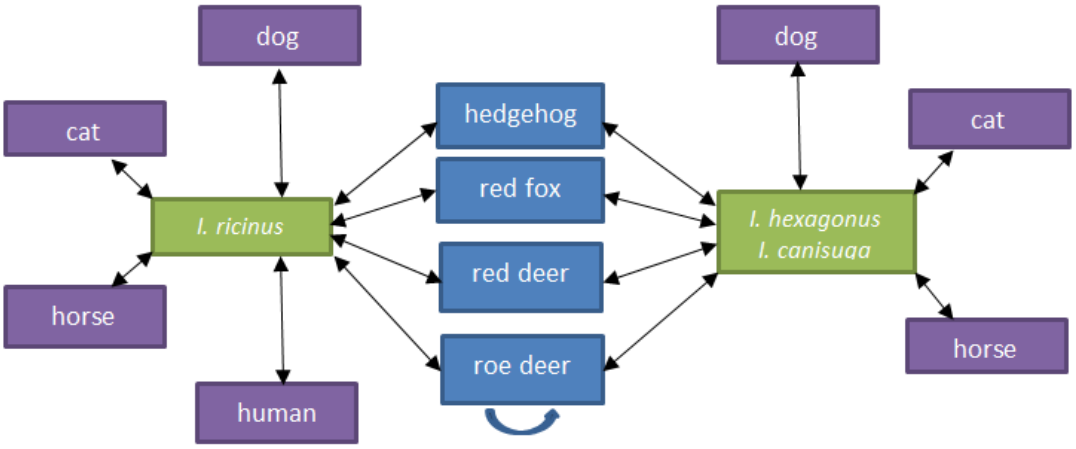

Figure 42: Possible natural life cycle of *A. phagocytophilum* with hedgehogs, red foxes and red deer as reservoir hosts. The blue arrow describes the possible subcycle of roe deer specific *A. phagocytophilum* variants.

A similar life cycle was suggested previously on the basis of *groEL* sequencing of *A. phagocytophilum*. Thereby, Jahfari et al. (2014) described the ecotype I involving domestic animals and humans and proposed red deer, hedgehogs and mouflons as reservoir hosts in Europe. Red deer and hedgehogs as possible reservoir hosts were confirmed by the present study. Hedgehogs showed high uniformity of infectious *A. phagocytophilum* strains, also occurring in domestic animals and humans and therefore supporting their role as reservoir hosts. For instance, variant 16S-1(A), putatively less pathogenic for humans compared to the prototype of *A. phagocytophilum*, occurred predominantly in hedgehogs. Since hedgehogs are heavily parasitized animals, other hedgehog specific tick species, like *I. hexagonus*, might form a subcycle of *A. phagocytophilum* as proposed by Silaghi et al. (2012a). However, due to uniform strains of *A. phagocytophilum* all matching with strains from other domestic animals the present study was unable to confirm such a subcycle for hedgehogs. Red deer as potential reservoir hosts of *A. phagocytophilum* harboured strains from domestic ruminants [e.g. 16S-20(W)] as well as from non-ruminants [e.g. 16S-2(B), 16S-16(S)]. Rymaszewska (2008) also supported red deer as reservoir for human pathogenic *A. phagocytophilum* strains, whereas other previous reports strongly questioned this hypothesis (Chastagner et al., 2014; Dugat et al., 2014a). Although red deer might predominantly be a reservoir for *A. phagocytophilum* from ruminants, the fact that red deer showed the prototype [16S-2(B)] should not be underestimated as this strain represents a potential threat for humans. In respect of the *A. phagocytophilum* from mouflons, the *groEL* variants from the present study

rather clustered with other wild ruminants than domestic animals. The only exception was the variant g-16(O), which occurred in a mouflon and a cattle sample. Several *16S rRNA* variants matched with sequences from mouflons and domestic animals [16S-2(B), 16S-16(S), 16S-20(W)]. In summary, mouflons could not be allocated to one specific life cycle due to a limited number of available nucleotide sequences. Red foxes were also proposed as possible reservoir hosts for *A. phagocytophilum* in the present study. Most variants of the four analyzed genes detected in red foxes also occurred in domestic animals like dogs and horses, in the present study [16S-1(A), 16S-2(B), g-1(A), g-2(B), m4-20, m2-6(f)]. Variants 16S-1(A) and 16S-2(B) also occurred in the fox samples investigated by Hartwig et al. (2014). Since *A. phagocytophilum* has been frequently detected in European red foxes, experimental studies could be useful to finally confirm the role of red foxes as reservoir hosts (Ebani et al., 2011; Hartwig et al., 2014; Hulinska et al., 2004).

In contrast to the proposed life cycle in the present study, Chastagner et al. (2014) clustered dog, horse and cattle strains of *A. phagocytophilum* (Cluster C) based on a supertree, which consolidated the results of a multilocus sequence typing approach. The same cluster included *A. phagocytophilum* from red deer, hedghogs and wild boars as reservoir hosts and humans as hosts (Chastagner et al., 2014; Huhn et al., 2014). This finding was inconsistent with the present study, since *A. phagocytophilum* from cattle samples clearly differed from variants infecting other domestic animals. Apart from the supertree, Chastagner et al. (2014) also described a separate cluster including cattle strains and roe deer strains, respectively. Therefore, European variants might have evolved from different evolutionary lineages infecting different animals or vectors. Based on the supertree of the multilocus sequencing analysis, Chastagner et al. (2014) assumed that the cause for the divergence of French strains might not be due to diverse geographical origin. Since similar clusters were detected in this study for European strains, the development of *A. phagocytophilum* strains independently of their geographical origin was confirmed. Nevertheless, differences between continents might still exist. In the same study, Chastagner et al. (2014) proposed wild boars as reservoir hosts. Since two prototype sequences [16S-2(B)] were also detected in the present study, wild boars might represent a potential reservoir host for *A. phagocytophilum* in Europe for human pathogenic strains as was also suggested previously (De la Fuente and Gortazar, 2012; Michalik et al., 2012;

Petrovec et al., 2003; Silaghi et al., 2014). However, due to a very limited number of wild boar samples, the role of wild boars remains uncertain to date.

The suggested life cycle of *A. phagocytophilum* showed similarities to the proposed hosts and dissimilarities to the proposed reservoir hosts of the Ap-ha strain from the USA. Since the Ap-ha is identical to the variant 16S-2(B) from this study, similar hosts were confirmed: dogs, horses and humans (Massung et al., 2003; Morissette et al., 2009). Nevertheless, mainly rodents, like the white-footed mouse (*P.leucopus*), were suggested as reservoir of Ap-ha, (Massung et al., 2005; Massung et al., 2003). Variant 16S-2(B) was also detected in rodents from Slovakia (acc. no.: KF481940, KF481945), which was identical to the strain causing HGA in Slovenian patients (Huhn et al., 2014; Scharf et al., 2011a). In contrast, the rodent samples examined in the present study originated from a study site with a very low prevalence of *A. phagocytophilum* (0 – 5.6%). It was impossible to obtain any nucleotide sequences. The reason for low prevalences in rodents might be the fact that mainly larval and nymphal tick stages with lower prevalence rates of *A. phagocytophilum* target rodents as hosts. Nevertheless, European rodents clustered in clades distinct from other mammals in previous studies concluding that rodents are unlikely reservoir hosts for *A. phagocytophilum* from domestic animals and humans (Jahfari et al., 2014; Majazki et al., 2013).

In the present study, *16S rRNA* strains of *A. phagocytophilum* from Slovenian and American HGA patients matched predominantly with the 16S-2(B). Since this was the *A. phagocytophilum* strain with the broadest host tropism in this study, the risk for HGA in humans in Europe might be higher than expected. Especially domestic animals, like dogs (48.2%) and horses (19.6%), were infected with this *A. phagocytophilum* strain. Due to close and intense contact between dogs and their owners, the distribution of the human pathogenic variant might be facilitated.

5.2. A possible life cycle with ruminants

The second proposed life cycle of *A. phagocytophilum* included wild cervids as reservoir hosts and domestic ruminants as hosts (Fig. 43). Similar to *A. phagocytophilum* infecting roe deer, other ruminant variants were highly diverse and very often unique to a single infected animal. Variant 16S-20(W) was primarily detected in cattle (52.9%), but was also revealed in wild cervids like roe deer, red deer, sika deer or fallow deer. Additionally, the 16S-21(X) and the

m4-13(N) occurred in *A. phagocytophilum* infected cattle, roe deer, sika deer or fallow deer samples.

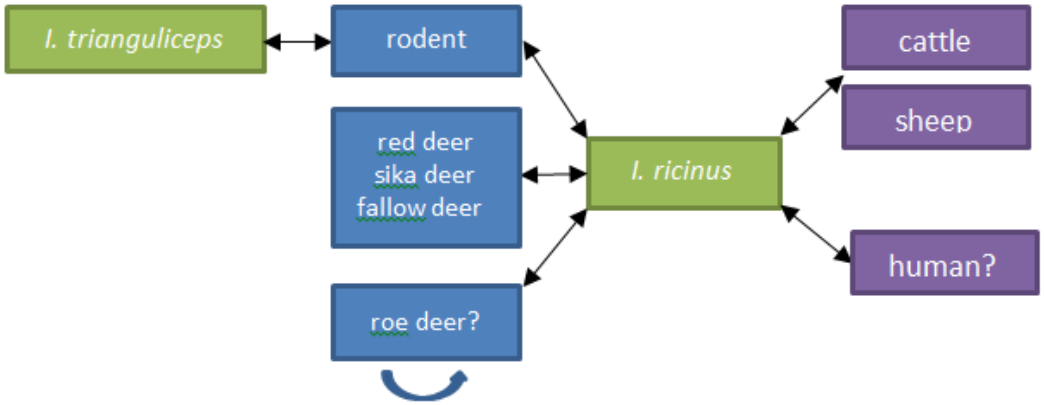

Figure 43: Possible natural life cycle of *A. phagocytophilum* including wild ruminants and rodents as reservoir hosts. The blue arrow describes the possible subcycle of roe deer specific *A. phagocytophilum* variants.

A. phagocytophilum strains from red deer in particular showed identical variants [especially the variant 16S-20(W)] to domestic ruminant strains, suggesting red deer as reservoir hosts in the epidemiological cycle of *A. phagocytophilum* with ruminants. In a Multiple Locus VNTR (Variable-Number Tandem Repeat) Analysis (MLVA), Dugat et al. (2014a) confirmed red deer as reservoir hosts of domestic ruminant strains, representing either accidental or long-term hosts. Besides, the role of red deer as reservoir hosts for ruminant pathogenic *A. phagocytophilum* strains was experimentally confirmed by the infection of a sheep with a red deer strain, developing symptoms of TBF (Stuen et al., 2010). Other wild cervids, like sika deer and fallow deer, were also suggested as potential wildlife reservoir of *A. phagocytophilum* in the present study. Both of these species showed the ruminant specific strains 16S-20(W) and 16S-21(X). Previous reports have supported these species as reservoir for *A. phagocytophilum* (Michalik et al., 2009; Veronesi et al., 2011; Wu et al., 2015). Unfortunately, only two comparable *groEL* strains from Asian sika deer (acc. nos.: JN055360, JN055359) are available from the GenBank, which showed no similarity to the variants detected in the present study. The number of available fallow deer and sika deer samples was limited, making additional epidemiological and experimental studies inevitable in order to confirm their role as reservoir hosts.

A life cycle including *A. phagocytophilum* from French cattle was described by Chastagner et al. (2014) in a multilocus sequence typing approach based on a

supertree consolidating all the investigated sequences (Cluster B). In contrast to the present study, this cluster included cattle samples exclusively (Chastagner et al., 2014). The same genotypes as in cluster B were also detected in sheep and sporadically wild deer like red deer and roe deer, wherefore these animal species were suggested as part of the same phylogenetic cluster (Huhn et al., 2014; Scharf et al., 2011a).

It was not possible to successfully sequence any of the *A. phagocytophilum* strains from rodents in the present study. Nevertheless, rodents and ticks infesting rodents, like *I. triangularis*, might form an own subcycle in nature. Numerous *A. phagocytophilum* sequences of rodents derived from GenBank were 100% homologue to variant 16S-20(W), which was previously detected in several ruminant species. Consequently, rodents might play a role in the distribution of this variant. In contrast to the present study, a distinct ecotype (ecotype III) involving rodents and rodent specific vectors of *A. phagocytophilum*, like *I. trianguliceps*, from the Netherlands and Belgium was proposed by Jahfari et al. (2014) on the basis of the *groEL* gene. This life cycle differed clearly from the ecotype I related to *A. phagocytophilum* strains infecting domestic animals and humans. As a consequence a direct contribution of rodents in the transmission cycle of *A. phagocytophilum* was denied (Jahfari et al., 2014). The *A. phagocytophilum* strains from the Dutch and Belgian rodents did not match with the variants of the present study. Nevertheless, the phylogenetic analysis of the *groEL* gene clustered *A. phagocytophilum* from rodents primarily into the "ruminant" clade. Thus, an own subcycle of rodents, as described in the present study as well as in several previously performed studies, might also exist additionally in nature (Bown et al., 2009; Majazki et al., 2013).

Interestingly, five US patients showed infections with *A. phagocytophilum* variant 16S-20(W). Whether humans play a role in the proposed ruminant cycle, remains to be elucidated, since most of the other strains causing HGA clustered into the other proposed life cycle.

5.3. The role of roe deer in *both* life cycles

The role of roe deer as reservoir hosts in the suggested life cycles of *A. phagocytophilum* appeared ambiguous. On the one hand roe deer represented a wildlife reservoir for several *A. phagocytophilum* strains and might therefore function as bridge for an exchange of strains between animal species. On the other

hand multiple variants of *A. phagocytophilum* occurred in roe deer exclusively and might thus suggest an own subcycle of roe deer specific *A. phagocytophilum* strains.

In the present study, roe deer were infected with *A. phagocytophilum* variants occurring in other ruminant species [e.g. 16S-20(W), 16S-21(X), 16S-22(Y), m4-13(N)] and non-ruminant species [e.g. 16S-2(B), m4-2(B/C)]. Therefore, roe deer might represent reservoir hosts for both suggested life cycles of *A. phagocytophilum*, whereby a systematic occurrence of distinct variants was not identifiable in the present study. According to Overzier et al. (2013a), the occurrence of special variants might be dependant on habitat structure and the existence of different potential reservoir hosts in this habitat. Since all of the roe deer samples originated from similar habitats, it was not possible to evaluate the latter hypothesis in the present study. The majority of the analyzed roe deer samples from both the present study and GenBank were the variants 16S-21(X) and 16S-22(Y). These two variants were mostly associated with roe deer in previous reports (Overzier et al., 2013a; Silaghi et al., 2011b). Additionally, variant 16S-21(X) was detected in goats and 16S-22(Y) in a goat and a cow sample. Interestingly, the affected cow did not show any symptoms of TBF in comparison to other cattle from the same pasture, which were primarily infected with the 16S-20(W) (Nieder et al., unpublished). In return, roe deer from the present study were also infected with variant 16S-20(W), which might indicate the transmission of ruminant pathogen and non-pathogen strains of *A. phagocytophilum* by roe deer. Previous reports also revealed infections of domestic and wild ruminants as well as small mammals with the variant 16S-20(W) (Rar et al., 2011; Scharf et al., 2011a; Silaghi et al., 2011b). In the USA, variant Ap-V1, which was confirmed to infect white-tailed deer and goats as reservoir hosts, was 100% homologue to variant 16S-21(X) (Massung et al., 2005; 2006). The Ap-V1 was recognized as a putatively less pathogenic strain in the USA, at least regarding infections of dogs and humans (Morissette et al., 2009). Stuen et al. (2006) described the case of a paretic roe deer calf infected with variant 16S-21(X) (acc. no.: AJ242784). Nevertheless, TBF symptomatic roe deer seem to represent an exception and depend on tick burden and health condition of the affected animal. In respect of the *groEL* gene, Rar et al. (2011) detected the variants g-4(D) and g-7(G) in roe deer, which also matched with *A. phagocytophilum* from a goat and a red deer, respectively. Sporadically, roe

deer samples revealed *A. phagocytophilum* variants causing disease in non-ruminant domestic animals and humans, for example variant 16S-2(B) or 16S-16(S). Especially variant 16S-2(B) as the prototype of *A. phagocytophilum* showed great potential to infect humans and domestic animals with a broad host topism. Since roe deer were regularly detected to be infected with this strain, the potential threat for humans may rise if ticks transmit this variant in a roe deer dense habitat (Silaghi et al., 2011b; Zeman and Pecha, 2008).

In contrast, several variants (e.g. 16S-24, m2-19, m4-24, g-25) isolated from roe deer samples were unique to one single animal and restricted to roe deer as hosts. This finding might indicate the existence of a subcycle in nature primarily including *A. phagocytophilum* variants from roe deer. Such a separate cluster of roe deer associated *A. phagocytophilum* strains was strongly supported by previous studies. For instance, Rymaszewska (2008) revealed specific *A. phagocytophilum* strains from roe deer distinct from red deer, livestock and humans. Except from one variant [g-7(G)], all of the *groEL* variants showed no similarity to the variants detected in the present study. Since numerable *A. phagocytophilum* strains detected in the present study were specific for roe deer, a cycle for these *A. phagocytophilum* strains was clearly supported. Besides, Jahfari et al. (2014) described a distinct ecotype (ecotype II) for *A. phagocytophilum* infecting roe deer on the basis of the *groEL* gene. Most of these roe deer samples originating from the Netherlands and Belgium showed variant g-24. The latter variant was also found in sika deer and fallow deer in the present study, possibly indicating a subcycle for wild cervids in general. The differentiation between a cluster for *A. phagocytophilum* strains from roe deer and domestic ruminants was supported by Dugat et al. (2014a) with a multilocus sequencing approach. Similarly, Chastagner et al. (2014) proposed a distinct cluster for roe deer analyzing French cattle samples (cluster A). According to the results of the present study and previous reports, roe deer might participate in both the transmission cycle of domestic animals and humans and a roe deer specific subcycle of *A. phagocytophilum*, whereby the variants differ from each other in each cycle.

VI. CONCLUSION

A. phagocytophilum in Europe was found to be an emerging pathogen with strong genetic diversity infecting different animal species. Analysis of four well established PCR protocols targeting the partial *16S rRNA*, *groEL*, *msp4* and *msp2* gene demonstrated great heterogeneity within these four different genes of *A. phagocytophilum* strains. Thereby, the *msp4* gene was most variable and the partial *16S rRNA* revealed the greatest level of conservation. Based on the partial *16S rRNA*, the *groEL* and the *msp* genes, a genetic classification of the investigated *A. phagocytophilum* strains into ruminant and non-ruminant strains could be shown, indicating an evolutionary divergence of the different strains. Depending on reservoir hosts rather than on geographic origin, two endemic life cycles of *A. phagocytophilum* were suggested ensuring its maintenance in nature. The first endemic cycle was characterized by a broad range of species as both hosts (horse, dog, cat, humans, roe deer) and reservoir hosts (hedgehog, red fox, red deer), possibly causing disease in domestic animals and humans. The second endemic cycle of *A. phagocytophilum* included wild cervids and rodents as reservoir hosts and domestic ruminants as hosts. Due to the detection of *A. phagocytophilum* variants of roe deer circulating in both life cycles, the role of roe deer as reservoir host seemed ambiguous. Roe deer might participate in the two life cycles as reservoir hosts, but might also form an own subcycle with roe deer specific *A. phagocytophilum* strains circulating between roe deer and their vectors in nature. According to the phylogenetic analysis, *A. phagocytophilum* harbouring zoonotic potential might resemble strains from domestic animals.

VII. SUMMARY

A. phagocytophilum is the causative agent for granulocytic anaplasmosis, which is a tick-transmitted emerging disease in humans and domestic animals. Genetic diversity in its genome could be a reason for a diverging pathogenicity in Europe compared to the USA. In order to maintain in nature, the bacterium has developed a survival strategy of different enzootic life cycles by circulating between hosts potentially developing clinical symptoms, reservoir hosts and ticks as vectors. In order to enhance the knowledge of possible endemic life cycles of *A. phagocytophilum* and its host-pathogen-associations, strains of *A. phagocytophilum* from different animal species were genetically characterized and phylogenetically classified on the basis of four partial genes (*16S rRNA*, *groEL*, *msp4*, *msp2*) within the present study.

In total, 781 samples of 17 different animal species (dog, horse, cat, cattle, goat, hedgehog, red fox, roe deer, red deer, sika deer, fallow deer, mouflon, chamois, ibex, wild boar, bank vole, wood mouse), which were screened positive for *A. phagocytophilum* by real-time PCR, were considered in the present study. The animal samples included 425 samples already tested positive available from previous studies. Subsequently, all positive samples were investigated using conventional PCR and nested PCR protocols targeting the partial *16S rRNA*, *msp4*, *groEL* and *msp2* gene. Forward and reverse sequencing was performed with all amplified samples. The obtained nucleotide sequences were compared to each other and to corresponding sequences from the GenBank, including a phylogenetic analysis (neighbor joining method; bootstrap value: 1.000 repeats) of the *groEL* and the *msp* genes. Statistical analysis concentrated on the evaluation of diversity occurring in *A. phagocytophilum* of different animal species including the empirical variance of special animal groups, a trend analysis of the four analyzed genes and the odd`s ratio of common *16S rRNA* variants.

Altogether, 327 *16S rRNA* sequences, 172 *groEL* sequences, 174 *msp4* sequences and 71 *msp2* sequences from the present and previous studies were taken into consideration. The amplification of the partial *16S rRNA* gene resulted in 23 variants, the *groEL* gene in 33 variants, the *msp4* gene in 50 variants and the *msp2* gene in 22 variants. *A. phagocytophilum* from wild cervids revealed a statistically

higher mean and empirical variance of the analyzed genes in comparison to strains from domestic animals. Accordingly, domestic animals, red foxes and hedgehogs showed rather uniform nucleotide sequences of *A. phagocytophilum*. The calculation of the odd`s ratio of the most common *16S rRNA* strains confirmed the preference of special variants in wild ruminants and domestic animal species, respectively. A preliminary genetic classification of the *A. phagocytophilum* strains into ruminant and non-ruminant variants was possible based on the partial *16S rRNA* variants. The phylogenetic analysis of the *groEL* and the *msp2* gene showed a clustering according to the three continents Europe, USA and Asia. On the contrary, the *msp4* gene rather clustered strains according to host animal species of *A. phagocytophilum*, discriminating clusters with wild and domestic ruminant variants from clusters with human, dog and horse variants.

Based on the genetic classification two endemic life cycles of *A. phagocytophilum* were proposed in the present study. The first life cycle includes hedgehogs, red foxes and red deer as reservoir hosts and domestic animals and humans as potential hosts developing clinical symptoms of the disease. The second endemic cycle of *A. phagocytophilum* possibly involves wild cervids and rodents as reservoir hosts and domestic ruminants as hosts. Since roe deer showed a high number of different variants of *A. phagocytophilum*, it might play a role in both suggested life cycles as reservoir host. However, the existence of an own subcycle of roe deer specific *A. phagocytophilum* strains is also possible.

VIII. ZUSAMMENFASSUNG

A. phagocytophilum ist der Erreger für die Granulozytäre Anaplasmose, einer immer häufiger auftretende Krankheit bei Mensch und Tier. Ein Grund für die unterschiedliche Pathogenität des Erregers in Europa im Vergleich zu Fällen aus den USA könnte an der genetischen Diversität in dessen Genom liegen. Das Bakterium hat eine Strategie von unterschiedlichen endemischen Lebenszyklen entwickelt, um seine Existenz in der Natur sicherzustellen. Dabei zirkuliert es zwischen Wirten, die potenziell Krankheitssymptome entwickeln, Reservoirwirten und Zecken als Vektoren. Zur Identifikation möglicher endemischer Lebenszyklen von *A. phagocytophilum* unter natürlichen Lebensbedingungen und seiner Wirt-Erreger-Assoziationen, wurden Proben von *A. phagocytophilum* unterschiedlicher Tierspezies auf Grundlage von vier partiellen Genen (*16S rRNA, groEL, msp4, msp2*) genetisch charakterisiert und phylogenetisch klassifiziert.

Insgesamt wurden 781 Proben von 17 unterschiedlichen Tierspezies (Hund, Pferd, Katze, Rind, Ziege, Igel, Rotfuchs, Reh, Rotwild, Sikawild, Dammwild, Mufflon, Gams, Steinbock, Wildschwein, Rötelmaus, Waldmaus), die durch real-time PCR *A. phagocytophilum*-positiv getestet wurden, in diese Studie einbezogen. Von den Tierproben standen bereits 425 positiv getestete Proben von vorheriger Studien zur Verfügung. Anschließend wurde bei positiv getesteten Proben das partielle *16S rRNA-*, *msp4-*, *groEL-* und *msp2*-Gen durch konventionelle, nested und hemi-nested PCR amplifiziert und sequenziert. Die daraus resultierenden Nukleotidsequenzen wurden sowohl untereinander als auch mit entsprechenden Sequenzen aus der GenBank verglichen und hinsichtlich der *groEL*- und der *msp*-Gene phylogenetisch (Methode: Neighbor Joining; Bootstrap Value: 1.000 Wiederholungen) untersucht. Die Diversität von *A. phagocytophilum* in verschiedenen Tierspezies wurde statistisch ausgewertet. Diese beinhaltete die Berechnung der empirischen Varianz spezieller Tierspeziesgruppen, eine Trend-Analyse der vier untersuchten Gene und der Berechnung des *Odd`s ratio* in besonders häufig auftretenden *16S rRNA* Varianten.

Insgesamt wurden 327 *16S rRNA*, 172 *groEL*, 174 *msp4* und 71 *msp2* Sequenzen aus dieser und vorheriger Studien berücksichtigt. Die Amplifizierung des partiellen *16S rRNA*-Gens resultierte in 23 Varianten, des *groEL*-Gens in 33

Varianten, des *msp4*-Gens in 50 Varianten und des *msp2*-Gens in 22 Varianten. *A. phagocytophilum* von wild lebenden Hirscharten zeigte eine signifikant höhere mittlere und empirische Varianz als von Haustieren. Dagegen wiesen *A. phagocytophilum* Proben von Haustieren, Rotfüchsen und Igeln sehr einheitliche Nukleotidsequenzen auf. Die Berechnung des *Odd`s ratio* der besonders häufig auftretenden *16S rRNA*-Gene bestätigten die statistisch signifikante Präferenz von bestimmten, sich unterscheidenden *A. phagocytophilum* Varianten in Wildwiederkäuern bzw. in Haustieren. Eine vorläufige Klassifizierung der *A. phagocytophilum* in Wiederkäuer und Nicht-Wiederkäuer-Varianten wurde auf Grundlage der partiellen *16S rRNA* Varianten möglich. Die phylogenetische Analyse des *groEL*- und des *msp2*-Genes zeigte eine Gruppierung entsprechend der Herkunft der drei Kontinente Europa, USA und Asien. Im Gegensatz dazu zeigten die *msp4*-Gene eine Gruppierung entsprechend der unterschiedlichen Wirtstiere von *A. phagocytophilum*, wobei zwischen Sequenzen von Wild- und Hauswiederkäuern und Varianten von Mensch, Hund und Pferd differenziert wurde.

Auf Grundlage der genetischen Klassifizierung wurden zwei endemische Lebenszyklen von *A. phagocytophilum* vorgeschlagen. Der erste Lebenszyklus umfasste Igel, Rotfüchse und Rotwild als Reservoirwirte und Haustiere und Menschen als Wirte, die potenziell Krankheitssymptome entwickeln könnten. Der zweite endemische Lebenszyklus von *A. phagocytophilum* schloss wild lebende Hirscharten und Nagetiere als Reservoirwirte und Hauswiederkäuer als Wirtstiere ein. Da eine Vielzahl an unterschiedlichen *A. phagocytophilum* Varianten in Rehen vorkam, wurden diese als Reservoirtiere in beiden Lebenszyklen vorgeschlagen, wobei auch ein eigener Subzyklus von Reh-spezifischen *A. phagocytophilum* Varianten existieren könnte.

IX. REFERENCES

Adaszek, L., Gorna, M., Skrzypczak, M., Buczek, K., Balicki, I., Winiarczyk, S., 2013. Three clinical cases of *Anaplasma phagocytophilum* infection in cats in Poland. J Feline Med Surg 15, 333-337.

Alberti, A., Addis, M.F., Sparagano, O., Zobba, R., Chessa, B., Cubeddu, T., Parpaglia, M.L., Ardu, M., Pittau, M., 2005a. *Anaplasma phagocytophilum*, Sardinia, Italy. Emerging infectious diseases 11, 1322-1324.

Alberti, A., Zobba, R., Chessa, B., Addis, M.F., Sparagano, O., Pinna Parpaglia, M.L., Cubeddu, T., Pintori, G., Pittau, M., 2005b. Equine and canine *Anaplasma phagocytophilum* strains isolated on the island of Sardinia (Italy) are phylogenetically related to pathogenic strains from the United States. Applied and environmental microbiology 71, 6418-6422.

Andersson, S.G., Kurland, C.G., 1998. Reductive evolution of resident genomes. Trends in microbiology 6, 263-268.

Baldridge, G.D., Scoles, G.A., Burkhardt, N.Y., Schloeder, B., Kurtti, T.J., Munderloh, U.G., 2009. Transovarial transmission of *Francisella*-like endosymbionts and *Anaplasma phagocytophilum* variants in *Dermacentor albipictus* (*Acari*: *Ixodidae*). Journal of medical entomology 46, 625-632.

Barbet, A., Al-Khedery, B., Stuen, S., Granquist, E., Felsheim, R., Munderloh, U., 2013. An Emerging Tick-Borne Disease of Humans Is Caused by a Subset of Strains with Conserved Genome Structure. Pathogens 2, 544-555.

Barbet, A.F., Lundgren, A.M., Alleman, A.R., Stuen, S., Bjoersdorff, A., Brown, R.N., Drazenovich, N.L., Foley, J.E., 2006. Structure of the expression site reveals global diversity in MSP2 (P44) variants in *Anaplasma phagocytophilum*. Infection and immunity 74, 6429-6437.

Barbet, A.F., Meeus, P.F., Belanger, M., Bowie, M.V., Yi, J., Lundgren, A.M., Alleman, A.R., Wong, S.J., Chu, F.K., Munderloh, U.G., Jauron, S.D., 2003. Expression of multiple outer membrane protein sequence variants from a single genomic locus of *Anaplasma phagocytophilum*. Infection and immunity 71, 1706-1718.

Barbour, A.G., Restrepo, B.I., 2000. Antigenic variation in vector-borne pathogens. Emerging infectious diseases 6, 449-457.

Beall, M.J., Chandrashekar, R., Eberts, M.D., Cyr, K.E., Diniz, P.P., Mainville, C., Hegarty, B.C., Crawford, J.M., Breitschwerdt, E.B., 2008. Serological and molecular prevalence of *Borrelia burgdorferi*, *Anaplasma phagocytophilum*, and *Ehrlichia* species in dogs from Minnesota. Vector borne and zoonotic diseases (Larchmont, N.Y.) 8, 455-464.

Bjoersdorff, A., Bagert, B., Massung, R.F., Gusa, A., Eliasson, I., 2002. Isolation and characterization of two European strains of *Ehrlichia phagocytophila* of equine origin. Clinical and diagnostic laboratory immunology 9, 341-343.

Bjoersdorff, A., Bergstrom, S., Massung, R.F., Haemig, P.D., Olsen, B., 2001. Ehrlichia-infected ticks on migrating birds. Emerging infectious diseases 7, 877-879.

Bjoersdorff, A., Svendenius, L., Owens, J.H., Massung, R.F., 1999. Feline granulocytic ehrlichiosis--a report of a new clinical entity and characterisation of the infectious agent. J Small Anim Pract 40, 20-24.

Bown, K.J., Begon, M., Bennett, M., Woldehiwet, Z., Ogden, N.H., 2003. Seasonal dynamics of *Anaplasma phagocytophila* in a rodent-tick (*Ixodes trianguliceps*) system, United Kingdom. Emerging infectious diseases 9, 63-70.

Bown, K.J., Lambin, X., Ogden, N.H., Begon, M., Telford, G., Woldehiwet, Z., Birtles, R.J., 2009. Delineating *Anaplasma phagocytophilum* ecotypes in coexisting, discrete enzootic cycles. Emerging infectious diseases 15, 1948-1954.

Bown, K.J., Lambin, X., Ogden, N.H., Petrovec, M., Shaw, S.E., Woldehiwet, Z., Birtles, R.J., 2007a. High-resolution genetic fingerprinting of European strains of Anaplasma phagocytophilum by use of multilocus variable-number tandem-repeat analysis. Journal of clinical microbiology 45.

Bown, K.J., Lambin, X., Ogden, N.H., Petrovec, M., Shaw, S.E., Woldehiwet, Z., Birtles, R.J., 2007b. High-resolution genetic fingerprinting of European strains of *Anaplasma phagocytophilum* by use of multilocus variable-number tandem-repeat analysis. Journal of clinical microbiology 45, 1771-1776.

Bown, K.J., Lambin, X., Telford, G., Heyder-Bruckner, D., Ogden, N.H., Birtles, R.J., 2011. The common shrew (*Sorex araneus*): a neglected host of tick-borne infections? Vector borne and zoonotic diseases (Larchmont, N.Y.) 11, 947-953.

Brayton, K.A., Knowles, D.P., McGuire, T.C., Palmer, G.H., 2001. Efficient use of a small genome to generate antigenic diversity in tick-borne ehrlichial pathogens. Proceedings of the National Academy of Sciences of the United States of America 98, 4130-4135.

Brayton, K.A., Palmer, G.H., Brown, W.C., 2006. Genomic and proteomic approaches to vaccine candidate identification for *Anaplasma marginale*. Expert review of vaccines 5, 95-101.

Brooks, D.R., McLennan, D.A., 1991. Phylogeny, ecology, and behavior: a research program in comparative biology. University of Chicago press.

Brown, W.C., 2012. Adaptive immunity to *Anaplasma* pathogens and immune dysregulation: implications for bacterial persistence. Comparative immunology, microbiology and infectious diseases 35, 241-252.

Butler, C.M., Nijhof, A.M., Jongejan, F., van der Kolk, J.H., 2008. *Anaplasma phagocytophilum* infection in horses in the Netherlands. The Veterinary record 162, 216-217.

Cao, W.C., Zhan, L., He, J., Foley, J.E., SJ, D.E.V., Wu, X.M., Yang, H., Richardus, J.H., Habbema, J.D., 2006. Natural *Anaplasma phagocytophilum* infection of ticks and rodents from a forest area of Jilin Province, China. The American journal of tropical medicine and hygiene 75, 664-668.

Carrade, D.D., Foley, J.E., Borjesson, D.L., Sykes, J.E., 2009. Canine granulocytic anaplasmosis: a review. Journal of veterinary internal medicine / American College of Veterinary Internal Medicine 23, 1129-1141.

Casey, A.N., Birtles, R.J., Radford, A.D., Bown, K.J., French, N.P., Woldehiwet, Z., Ogden, N.H., 2004. Groupings of highly similar major surface protein (p44)-encoding paralogues: a potential index of genetic diversity amongst isolates of *Anaplasma phagocytophilum*. Microbiology (Reading, England) 150, 727-734.

Castro, M.B., Wright, S.A., 2007. Vertebrate hosts of *Ixodes pacificus* (*Acari*: *Ixodidae*) in California. J Vector Ecol 32, 140-149.

CDC, C.o.D.C.a.P. 2013. Anaplasmosis - Statistics and Epidemiology (Atlanta, USA).

Centers for Disease Control and Prevention, C., 1997. Case definitions for infectious conditions under public health surveillance. Centers for Disease Control and Prevention. MMWR. Recommendations and reports : Morbidity and mortality weekly report. Recommendations and reports / Centers for Disease Control 46, 1-55.

Chae, J.S., Foley, J.E., Dumler, J.S., Madigan, J.E., 2000. Comparison of the nucleotide sequences of 16S rRNA, 444 Ep-ank, and groESL heat shock operon genes in naturally occurring *Ehrlichia equi* and human granulocytic ehrlichiosis agent isolates from Northern California. Journal of clinical microbiology 38, 1364-1369.

Chae, J.S., Yu do, H., Shringi, S., Klein, T.A., Kim, H.C., Chong, S.T., Lee, I.Y., Foley, J., 2008. Microbial pathogens in ticks, rodents and a shrew in northern

Gyeonggi-do near the DMZ, Korea. J Vet Sci 9, 285-293.

Chapman, A.S., Bakken, J.S., Folk, S.M., Paddock, C.D., Bloch, K.C., Krusell, A., Sexton, D.J., Buckingham, S.C., Marshall, G.S., Storch, G.A., Dasch, G.A., McQuiston, J.H., Swerdlow, D.L., Dumler, S.J., Nicholson, W.L., Walker, D.H., Eremeeva, M.E., Ohl, C.A., 2006. Diagnosis and management of tickborne *rickettsial* diseases: Rocky Mountain spotted fever, ehrlichioses, and anaplasmosis--United States: a practical guide for physicians and other health-care and public health professionals. MMWR. Recommendations and reports : Morbidity and mortality weekly report. Recommendations and reports / Centers for Disease Control 55, 1-27.

Chastagner, A., Dugat, T., Vourc, H.G., Verheyden, H., Legrand, L., Bachy, V., Chabanne, L., Joncour, G., Maillard, R., Boulouis, H.J., Haddad, N., Bailly, X., Leblond, A., 2014. Multilocus sequence analysis of *Anaplasma phagocytophilum* reveals three distinct lineages with different host ranges in clinically ill French cattle. Veterinary research 45, 114.

Chen, S.M., Dumler, J.S., Bakken, J.S., Walker, D.H., 1994. Identification of a granulocytotropic *Ehrlichia* species as the etiologic agent of human disease. Journal of clinical microbiology 32, 589-595.

Clark, K.L., 2012. *Anaplasma phagocytophilum* in small mammals and ticks in northeast Florida. J Vector Ecol 37, 262-268.

Cochez, C., Ducoffre, G., Vandenvelde, C., Luyasu, V., Heyman, P., 2011. Human anaplasmosis in Belgium: a 10-year seroepidemiological study. Ticks and tick-borne diseases 2, 156-159.

Cotte, V., Bonnet, S., Cote, M., Vayssier-Taussat, M., 2010. Prevalence of five pathogenic agents in questing *Ixodes ricinus* ticks from western France. Vector borne and zoonotic diseases (Larchmont, N.Y.) 10, 723-730.

Courtney, J.W., Dryden, R.L., Montgomery, J., Schneider, B.S., Smith, G.,

Massung, R.F., 2003. Molecular characterization of *Anaplasma phagocytophilum* and *Borrelia burgdorferi* in *Ixodes scapularis* ticks from Pennsylvania. Journal of clinical microbiology 41, 1569-1573.

Courtney, J.W., Kostelnik, L.M., Zeidner, N.S., Massung, R.F., 2004. Multiplex real-time PCR for detection of *Anaplasma phagocytophilum* and *Borrelia burgdorferi*. Journal of clinical microbiology 42, 3164-3168.

Dahlgren, F.S., Mandel, E.J., Krebs, J.W., Massung, R.F., McQuiston, J.H., 2011. Increasing incidence of *Ehrlichia chaffeensis* and *Anaplasma phagocytophilum* in the United States, 2000-2007. The American journal of tropical medicine and hygiene 85, 124-131.

Dahmani, M., Davoust, B., Benterki, M.S., Fenollar, F., Raoult, D., Mediannikov, O., 2015. Development of a new PCR-based assay to detect *Anaplasmataceae* and the first report of *Anaplasma phagocytophilum* and *Anaplasma platys* in cattle from Algeria. Comparative immunology, microbiology and infectious diseases 39, 39-45.

Daniels, T.J., Battaly, G.R., Liveris, D., Falco, R.C., Schwartz, I., 2002. Avian reservoirs of the agent of human granulocytic ehrlichiosis? Emerging infectious diseases 8, 1524-1525.

Dasch, G.A., Ching, W.M., Kim, P.Y., Pham, H., Stover, C.K., Oaks, E.V., Dobson, M.E., Weiss, E., 1990. A structural and immunological comparison of rickettsial HSP60 antigens with those of other species. Annals of the New York Academy of Sciences 590, 352-369.

De la Fuente, J., Atkinson, M.W., Naranjo, V., Fernandez de Mera, I.G., Mangold, A.J., Keating, K.A., Kocan, K.M., 2007. Sequence analysis of the msp4 gene of *Anaplasma ovis* strains. Veterinary microbiology 119, 375-381.

De La Fuente, J., Golsteyn Thomas, E.J., van den Bussche, R.A., Hamilton, R.G., Tanaka, E.E., Druhan, S.E., Kocan, K.M., 2003. Characterization of *Anaplasma*

marginale isolated from North American bison. Applied and environmental microbiology 69, 5001-5005.

De la Fuente, J., Gortazar, C., 2012. Wild Boars as Hosts of Human-Pathogenic *Anaplasma phagocytophilum* Variants. Emerging infectious diseases 18, 2094-2095.

De la Fuente, J., Massung, R.F., Wong, S.J., Chu, F.K., Lutz, H., Meli, M., von Loewenich, F.D., Grzeszczuk, A., Torina, A., Caracappa, S., Mangold, A.J., Naranjo, V., Stuen, S., Kocan, K.M., 2005a. Sequence analysis of the msp4 gene of *Anaplasma phagocytophilum* strains. Journal of clinical microbiology 43, 1309-1317.

De La Fuente, J., Naranjo, V., Ruiz-Fons, F., Hofle, U., Fernandez De Mera, I.G., Villanua, D., Almazan, C., Torina, A., Caracappa, S., Kocan, K.M., Gortazar, C., 2005b. Potential vertebrate reservoir hosts and invertebrate vectors of *Anaplasma marginale* and *A. phagocytophilum* in central Spain. Vector borne and zoonotic diseases (Larchmont, N.Y.) 5, 390-401.

De la Fuente, J., Van Den Bussche, R.A., Garcia-Garcia, J.C., Rodriguez, S.D., Garcia, M.A., Guglielmone, A.A., Mangold, A.J., Friche Passos, L.M., Barbosa Ribeiro, M.F., Blouin, E.F., Kocan, K.M., 2002. Phylogeography of New World isolates of *Anaplasma marginale* based on major surface protein sequences. Veterinary microbiology 88, 275-285.

De la Fuente, J., Van Den Bussche, R.A., Kocan, K.M., 2001. Molecular phylogeny and biogeography of North American isolates of *Anaplasma marginale* (*Rickettsiaceae*: *Ehrlichieae*). Veterinary parasitology 97, 65-76.

Demma, L.J., Holman, R.C., McQuiston, J.H., Krebs, J.W., Swerdlow, D.L., 2005. Epidemiology of human ehrlichiosis and anaplasmosis in the United States, 2001-2002. The American journal of tropical medicine and hygiene 73, 400-409.

Deplazes, P., Eckert, J., von Samson-Himmelstjerna, G., Zahner, H., 2013.

Lehrbuch der Parasitologie für die Tiermedizin. Eke Vera, 656 p.

Des Vignes, F., Piesman, J., Heffernan, R., Schulze, T.L., Stafford, K.C., 3rd, Fish, D., 2001. Effect of tick removal on transmission of *Borrelia burgdorferi* and *Ehrlichia phagocytophila* by *Ixodes scapularis* nymphs. The Journal of infectious diseases 183, 773-778.

Doudier, B., Olano, J., Parola, P., Brouqui, P., 2010. Factors contributing to emergence of *Ehrlichia* and *Anaplasma spp.* as human pathogens. Veterinary parasitology 167, 149-154.

Drazenovich, N., Foley, J., Brown, R.N., 2006. Use of real-time quantitative PCR targeting the msp2 protein gene to identify cryptic *Anaplasma phagocytophilum* infections in wildlife and domestic animals. Vector borne and zoonotic diseases (Larchmont, N.Y.) 6, 83-90.

Dugan, V.G., Yabsley, M.J., Tate, C.M., Mead, D.G., Munderloh, U.G., Herron, M.J., Stallknecht, D.E., Little, S.E., Davidson, W.R., 2006. Evaluation of white-tailed deer (*Odocoileus virginianus*) as natural sentinels for *Anaplasma phagocytophilum*. Vector borne and zoonotic diseases (Larchmont, N.Y.) 6, 192-207.

Dugat, T., Chastagner, A., Lagrée, A.-C., Petit, E., Durand, B., Thierry, S., Corbière, F., Verheyden, H., Chabanne, L., Bailly, X., Leblond, A., Vourc'h, G., Boulouis, H.-J., Maillard, R., Haddad, N., 2014a. A new multiple-locus variable-number tandem repeat analysis reveals different clusters for *Anaplasma phagocytophilum* circulating in domestic and wild ruminants. Parasites & vectors 7, 1-11.

Dugat, T., Lagrée, A.-C., Maillard, R., Boulouis, H.-J., Haddad, N., 2015. Opening the black box of *Anaplasma phagocytophilum* diversity: current situation and future perspectives. Frontiers in cellular and infection microbiology 5.

Dugat, T., Loux, V., Marthey, S., Moroldo, M., Lagree, A.C., Boulouis, H.J.,

Haddad, N., Maillard, R., 2014b. Comparative genomics of first available bovine *Anaplasma phagocytophilum* genome obtained with targeted sequence capture. BMC genomics 15, 973.

Dumler, J.S., 2005. *Anaplasma* and *Ehrlichia* infection. Annals of the New York Academy of Sciences 1063, 361-373.

Dumler, J.S., 2012. The biological basis of severe outcomes in *Anaplasma phagocytophilum* infection. FEMS immunology and medical microbiology 64, 13-20.

Dumler, J.S., Barbet, A.F., Bekker, C.P., Dasch, G.A., Palmer, G.H., Ray, S.C., Rikihisa, Y., Rurangirwa, F.R., 2001. Reorganization of genera in the families *Rickettsiaceae* and *Anaplasmataceae* in the order *Rickettsiales*: unification of some species of *Ehrlichia* with *Anaplasma*, *Cowdria* with *Ehrlichia* and *Ehrlichia* with *Neorickettsia*, descriptions of six new species combinations and designation of *Ehrlichia equi* and 'HGE agent' as subjective synonyms of *Ehrlichia phagocytophila*. International journal of systematic and evolutionary microbiology 51, 2145-2165.

Dumler, J.S., Choi, K.S., Garcia-Garcia, J.C., Barat, N.S., Scorpio, D.G., Garyu, J.W., Grab, D.J., Bakken, J.S., 2005. Human granulocytic anaplasmosis and *Anaplasma phagocytophilum*. Emerging infectious diseases 11, 1828-1834.

Dunning Hotopp, J.C., Lin, M., Madupu, R., Crabtree, J., Angiuoli, S.V., Eisen, J.A., Seshadri, R., Ren, Q., Wu, M., Utterback, T.R., Smith, S., Lewis, M., Khouri, H., Zhang, C., Niu, H., Lin, Q., Ohashi, N., Zhi, N., Nelson, W., Brinkac, L.M., Dodson, R.J., Rosovitz, M.J., Sundaram, J., Daugherty, S.C., Davidsen, T., Durkin, A.S., Gwinn, M., Haft, D.H., Selengut, J.D., Sullivan, S.A., Zafar, N., Zhou, L., Benahmed, F., Forberger, H., Halpin, R., Mulligan, S., Robinson, J., White, O., Rikihisa, Y., Tettelin, H., 2006. Comparative genomics of emerging human ehrlichiosis agents. PLoS genetics 2, e21.

Dziegiel, B., Adaszek, L., Kalinowski, M., Winiarczyk, S., 2013. Equine

granulocytic anaplasmosis. Research in veterinary science 95, 316-320.

Ebani, V.V., Cerri, D., Fratini, F., Ampola, M., Andreani, E., 2007. *Anaplasma phagocytophilum* infection in a fallow deer (*Dama dama*) population in a preserve of central Italy. New Microbiol 30, 161-165.

Ebani, V.V., Verin, R., Fratini, F., Poli, A., Cerri, D., 2011. Molecular survey of *Anaplasma phagocytophilum* and *Ehrlichia canis* in red foxes (*Vulpes vulpes*) from central Italy. Journal of wildlife diseases 47, 699-703.

Edouard, S., Koebel, C., Goehringer, F., Socolovschi, C., Jaulhac, B., Raoult, D., Brouqui, P., 2012. Emergence of human granulocytic anaplasmosis in France. Ticks and tick-borne diseases 3, 403-405.

Egenvall, A., Lilliehook, I., Bjoersdorff, A., Engvall, E.O., Karlstam, E., Artursson, K., Heldtander, M., Gunnarsson, A., 2000. Detection of granulocytic *Ehrlichia* species DNA by PCR in persistently infected dogs. The Veterinary record 146, 186-190.

Foggie, A., 1951. STUDIES ON THE INFECTIOUS AGENT OF TICK-BORNE FEVER IN SHEEP. Journal of Pathology and Bacteriology 63, 1-&.

Földvári, G., Jahfari, S., Rigó, K., Jablonszky, M., Szekeres, S., Majoros, G., Tóth, M., Molnár, V., Coipan, E.C., Sprong, H., 2014. *Candidatus Neoehrlichia mikurensis* and *Anaplasma phagocytophilum* in urban hedgehogs. Emerging infectious diseases 20, 496-498.

Foley, J., Nieto, N.C., Madigan, J., Sykes, J., 2008a. Possible differential host tropism in *Anaplasma phagocytophilum* strains in the Western United States. Annals of the New York Academy of Sciences 1149, 94-97.

Foley, J.E., Barlough, J.E., Kimsey, R.B., Madigan, J.E., DeRock, E., Poland, A., 1998. *Ehrlichia spp*. in cervids from California. Journal of wildlife diseases 34,

731-737.

Foley, J.E., Nieto, N.C., Adjemian, J., Dabritz, H., Brown, R.N., 2008b. *Anaplasma phagocytophilum* infection in small mammal hosts of Ixodes ticks, western United States. Emerging infectious diseases 14, 1147-1150.

Franzen, P., Aspan, A., Egenvall, A., Gunnarsson, A., Karlstam, E., Pringle, J., 2009. Molecular evidence for persistence of *Anaplasma phagocytophilum* in the absence of clinical abnormalities in horses after recovery from acute experimental infection. Journal of veterinary internal medicine / American College of Veterinary Internal Medicine 23, 636-642.

Franzen, P., Berg, A.L., Aspan, A., Gunnarsson, A., Pringle, J., 2007. Death of a horse infected experimentally with *Anaplasma phagocytophilum*. The Veterinary record 160, 122-125.

Gabriel, M.W., Brown, R.N., Foley, J.E., Higley, J.M., Botzler, R.G., 2009. Ecology of *Anaplasma phagocytophilum* infection in gray foxes (*Urocyon cinereoargenteus*) in northwestern California. Journal of wildlife diseases 45, 344-354.

Gaowa, Wuritu, Wu, D., Yoshikawa, Y., Ohashi, N., Kawamori, F., Sugiyama, K., Ohtake, M., Ohashi, M., Yamamoto, S., Kitano, T., Takada, N., Kawabata, H., 2012. Detection and characterization of p44/msp2 transcript variants of *Anaplasma phagocytophilum* from naturally infected ticks and wild deer in Japan. Japanese journal of infectious diseases 65, 79-83.

Garcia-Garcia, J.C., Rennoll-Bankert, K.E., Pelly, S., Milstone, A.M., Dumler, J.S., 2009. Silencing of host cell CYBB gene expression by the nuclear effector AnkA of the intracellular pathogen *Anaplasma phagocytophilum*. Infection and immunity 77, 2385-2391.

Georgopoulos, C., Welch, W.J., 1993. Role of the major heat shock proteins as molecular chaperones. Annual review of cell biology 9, 601-634.

Goethert, H.K., Telford, S.R., 3rd, 2003. Enzootic transmission of the agent of human granulocytic ehrlichiosis among cottontail rabbits. The American journal of tropical medicine and hygiene 68, 633-637.

Gokce, H.I., Woldehiwet, Z., 1999. Differential haematological effects of tick-borne fever in sheep and goats. Zentralblatt fur Veterinarmedizin. Reihe B. Journal of veterinary medicine. Series B 46, 105-115.

Gordon, W.S., Brownlee, A., Wilson, D.R., 1940. Studies on louping ill, tickborne fever and scrapie. Proceedings of the Third International Congress on Microbiology, New York.

Granquist, E.G., Bardsen, K., Bergstrom, K., Stuen, S., 2010. Variant -and individual dependent nature of persistent *Anaplasma phagocytophilum* infection. Acta veterinaria Scandinavica 52, 25.

Granquist, E.G., Stuen, S., Lundgren, A.M., Braten, M., Barbet, A.F., 2008. Outer membrane protein sequence variation in lambs experimentally infected with *Anaplasma phagocytophilum*. Infection and immunity 76, 120-126.

Gribble, D.H., 1969. Equine ehrlichiosis. Journal of the American Veterinary Medical Association 155, 462-469.

Hagedorn, P., Imhoff, M., Fischer, C., Domingo, C., Niedrig, M., 2014. Human granulocytic anaplasmosis acquired in Scotland, 2013. Emerging infectious diseases 20, 1079-1081.

Hamel, D., Silaghi, C., Lescai, D., Pfister, K., 2012. Epidemiological aspects on vector-borne infections in stray and pet dogs from Romania and Hungary with focus on *Babesia spp*. Parasitology research 110, 1537-1545.

Hamel, D., Silaghi, C., Pfister, K., 2013. Arthropod-borne infections in travelled dogs in Europe. Parasite (Paris, France) 20, 9.

Hapunik, J., Vichova, B., Karbowiak, G., Wita, I., Bogdaszewski, M., Pet'ko, B., 2011. Wild and farm breeding cervids infections with *Anaplasma phagocytophilum*. Annals of agricultural and environmental medicine : AAEM 18, 73-77.

Hartwig, V., von Loewenich, F.D., Schulze, C., Straubinger, R.K., Daugschies, A., Dyachenko, V., 2014. Detection of *Anaplasma phagocytophilum* in red foxes (*Vulpes vulpes*) and raccoon dogs (*Nyctereutes procyonoides*) from Brandenburg, Germany. Ticks and tick-borne diseases 5, 277-280.

Haschke-Becher, E., Bernauer, R., Walleczek, A.M., Apfalter, P., Afazel-Saeedi, S., Kraus, J., Ladurner, G., Strasser, P., 2010. First detection of the *Anaplasma phagocytophilum* groEL-A genotype in man. The Journal of infection 60, 300-305.

Heikkilä, H.M., Bondarenko, A., Mihalkov, A., Pfister, K., Spillmann, T., 2010. *Anaplasma phagocytophilum* infection in a domestic cat in Finland: Case report. Acta veterinaria Scandinavica 52, 62.

Henn, J.B., Gabriel, M.W., Kasten, R.W., Brown, R.N., Theis, J.H., Foley, J.E., Chomel, B.B., 2007. Gray foxes (*Urocyon cinereoargenteus*) as a potential reservoir of a *Bartonella clarridgeiae*-like bacterium and domestic dogs as part of a sentinel system for surveillance of zoonotic arthropod-borne pathogens in northern California. Journal of clinical microbiology 45, 2411-2418.

Heo, E.J., Park, J.H., Koo, J.R., Park, M.S., Park, M.Y., Dumler, J.S., Chae, J.S., 2002. Serologic and molecular detection of *Ehrlichia chaffeensis* and *Anaplasma phagocytophila* (human granulocytic ehrlichiosis agent) in Korean patients. Journal of clinical microbiology 40, 3082-3085.

Holden, K., Boothby, J.T., Anand, S., Massung, R.F., 2003. Detection of *Borrelia burgdorferi, Ehrlichia chaffeensis,* and *Anaplasma phagocytophilum* in ticks (*Acari*: *Ixodidae*) from a coastal region of California. Journal of medical entomology 40, 534-539.

Hudson, J.R., 1950. The recognition of tick-borne fever as a disease of cattle. British Veterinary Journal 106, 317.

Huhn, C., Winter, C., Wolfsperger, T., Wuppenhorst, N., Strasek Smrdel, K., Skuballa, J., Pfaffle, M., Petney, T., Silaghi, C., Dyachenko, V., Pantchev, N., Straubinger, R.K., Schaarschmidt-Kiener, D., Ganter, M., Aardema, M.L., von Loewenich, F.D., 2014. Analysis of the population structure of *Anaplasma phagocytophilum* using multilocus sequence typing. PloS one 9, e93725.

Hulinska, D., Langrova, K., Pejcoch, M., Pavlasek, I., 2004. Detection of *Anaplasma phagocytophilum* in animals by real-time polymerase chain reaction. APMIS : acta pathologica, microbiologica, et immunologica Scandinavica 112, 239-247.

Ismail, N., Bloch, K.C., McBride, J.W., 2010. Human ehrlichiosis and anaplasmosis. Clinics in laboratory medicine 30, 261-292.

Jahfari, S., Coipan, E.C., Fonville, M., van Leeuwen, A.D., Hengeveld, P., Heylen, D., Heyman, P., van Maanen, C., Butler, C.M., Foldvari, G., Szekeres, S., van Duijvendijk, G., Tack, W., Rijks, J.M., van der Giessen, J., Takken, W., van Wieren, S.E., Takumi, K., Sprong, H., 2014. Circulation of four *Anaplasma phagocytophilum* ecotypes in Europe. Parasites & vectors 7, 365.

Jahn, P., Zeman, P., Bezdekova, B., Praskova, I., 2010. Equine granulocytic anaplasmosis in the Czech Republic. The Veterinary record 166, 646-649.

Jilintai, Seino, N., Hayakawa, D., Suzuki, M., Hata, H., Kondo, S., Matsumoto, K., Yokoyama, N., Inokuma, H., 2009. Molecular survey for *Anaplasma bovis* and *Anaplasma phagocytophilum* infection in cattle in a pastureland where sika deer appear in Hokkaido, Japan. Japanese journal of infectious diseases 62, 73-75.

Johnson, R.C., Kodner, C., Jarnefeld, J., Eck, D.K., Xu, Y., 2011. Agents of human anaplasmosis and Lyme disease at Camp Ripley, Minnesota. Vector borne and zoonotic diseases (Larchmont, N.Y.) 11, 1529-1534.

Kang, J.G., Ko, S., Kim, Y.J., Yang, H.J., Lee, H., Shin, N.S., Choi, K.S., Chae, J.S., 2011. New genetic variants of *Anaplasma phagocytophilum* and *Anaplasma bovis* from Korean water deer (*Hydropotes inermis argyropus*). Vector borne and zoonotic diseases (Larchmont, N.Y.) 11, 929-938.

Karbowiak, G., Víchová, B., Majláthová, V., Hapunik, J., Peťko, B., 2009. *Anaplasma phagocytophilum* infection of red foxes (*Vulpes vulpes*). Annals of agricultural and environmental medicine : AAEM 16, 299-300.

Karlsson, U., Bjoersdorff, A., Massung, R.F., Christensson, B., 2001. Human granulocytic ehrlichiosis--a clinical case in Scandinavia. Scandinavian journal of infectious diseases 33, 73-74.

Kauffmann, M., Rehbein, S., Hamel, D., Lutz, W., Heddergott, M., Pfister, K., Silaghi, C., 2016. Anaplasma phagocytophilum and Babesia spp. in roe deer (Capreolus capreolus), fallow deer (Dama dama) and mouflon (Ovis musimon) in Germany. Mol Cell Probes.

Kawahara, M., Rikihisa, Y., Lin, Q., Isogai, E., Tahara, K., Itagaki, A., Hiramitsu, Y., Tajima, T., 2006. Novel genetic variants of *Anaplasma phagocytophilum*, *Anaplasma bovis*, *Anaplasma centrale*, and a novel *Ehrlichia sp.* in wild deer and ticks on two major islands in Japan. Applied and environmental microbiology 72, 1102-1109.

Keesing, F., Hersh, M.H., Tibbetts, M., McHenry, D.J., Duerr, S., Brunner, J., Killilea, M., LoGiudice, K., Schmidt, K.A., Ostfeld, R.S., 2012. Reservoir competence of vertebrate hosts for *Anaplasma phagocytophilum*. Emerging infectious diseases 18, 2013-2016.

Kim, C.M., Yi, Y.H., Yu, D.H., Lee, M.J., Cho, M.R., Desai, A.R., Shringi, S., Klein, T.A., Kim, H.C., Song, J.W., Baek, L.J., Chong, S.T., O'Guinn M, L., Lee, J.S., Lee, I.Y., Park, J.H., Foley, J., Chae, J.S., 2006. Tick-borne rickettsial pathogens in ticks and small mammals in Korea. Applied and environmental microbiology 72, 5766-5776.

Kim, K.H., Yi, J., Oh, W.S., Kim, N.H., Choi, S.J., Choe, P.G., Kim, N.J., Lee, J.K., Oh, M.D., 2014. Human granulocytic anaplasmosis, South Korea, 2013. Emerging infectious diseases 20, 1708-1711.

Kocan, K.M., Busby, A.T., Allison, R.W., Breshears, M.A., Coburn, L., Galindo, R.C., Ayllon, N., Blouin, E.F., de la Fuente, J., 2012. Sheep experimentally infected with a human isolate of *Anaplasma phagocytophilum* serve as a host for infection of *Ixodes scapularis* ticks. Ticks and tick-borne diseases 3, 147-153.

Koebel, C., Kern, A., Edouard, S., Hoang, A.T., Celestin, N., Hansmann, Y., Jaulhac, B., Brouqui, P., De Martino, S.J., 2012. Human granulocytic anaplasmosis in eastern France: clinical presentation and laboratory diagnosis. Diagnostic microbiology and infectious disease 72, 214-218.

Kohn, B., Galke, D., Beelitz, P., Pfister, K., 2008. Clinical features of canine granulocytic anaplasmosis in 18 naturally infected dogs. Journal of veterinary internal medicine / American College of Veterinary Internal Medicine 22, 1289-1295.

Kohn, B., Silaghi, C., Galke, D., Arndt, G., Pfister, K., 2011. Infections with *Anaplasma phagocytophilum* in dogs in Germany. Research in veterinary science 91, 71-76.

Kolbert, C.P., Bruinsma, E.S., Abdulkarim, A.S., Hofmeister, E.K., Tompkins, R.B., Telford, S.R., 3rd, Mitchell, P.D., Adams-Stich, J., Persing, D.H., 1997. Characterization of an immunoreactive protein from the agent of human granulocytic ehrlichiosis. Journal of clinical microbiology 35, 1172-1178.

Kybicova, K., Schanilec, P., Hulinska, D., Uherkova, L., Kurzova, Z., Spejchalova, S., 2009. Detection of *Anaplasma phagocytophilum* and *Borrelia burgdorferi sensu lato* in dogs in the Czech Republic. Vector borne and zoonotic diseases (Larchmont, N.Y.) 9, 655-661.

Lappin, M.R., Breitschwerdt, E.B., Jensen, W.A., Dunnigan, B., Rha, J.Y.,

Williams, C.R., Brewer, M., Fall, M., 2004. Molecular and serologic evidence of *Anaplasma phagocytophilum* infection in cats in North America. Journal of the American Veterinary Medical Association 225, 893-896, 879.

Laus, F., Veronesi, F., Passamonti, F., Paggi, E., Cerquetella, M., Hyatt, D., Tesei, B., Fioretti, D.P., 2013. Prevalence of tick borne pathogens in horses from Italy. The Journal of veterinary medical science / the Japanese Society of Veterinary Science 75, 715-720.

Li, H., Zhou, Y., Wang, W., Guo, D., Huang, S., Jie, S., 2011. The clinical characteristics and outcomes of patients with human granulocytic anaplasmosis in China. Int J Infect Dis 15, e859-866.

Lin, M., Kikuchi, T., Brewer, H.M., Norbeck, A.D., Rikihisa, Y., 2011. Global proteomic analysis of two tick-borne emerging zoonotic agents: *Anaplasma phagocytophilum* and *Ehrlichia chaffeensis*. Frontiers in microbiology 2, 24.

Lin, Q., Rikihisa, Y., Felek, S., Wang, X., Massung, R.F., Woldehiwet, Z., 2004a. *Anaplasma phagocytophilum* has a functional msp2 gene that is distinct from p44. Infection and immunity 72, 3883-3889.

Lin, Q., Rikihisa, Y., Massung, R.F., Woldehiwet, Z., Falco, R.C., 2004b. Polymorphism and transcription at the p44-1/p44-18 genomic locus in *Anaplasma phagocytophilum* strains from diverse geographic regions. Infection and immunity 72, 5574-5581.

Littlejohn, A.I., 1950. Tick-borne fever as a cause of abortion in sheep. The Veterinary record 62, 577-578.

Liz, J.S., Anderes, L., Sumner, J.W., Massung, R.F., Gern, L., Rutti, B., Brossard, M., 2000. PCR detection of granulocytic ehrlichiae in *Ixodes ricinus* ticks and wild small mammals in western Switzerland. Journal of clinical microbiology 38, 1002-1007.

Liz, J.S., Sumner, J.W., Pfister, K., Brossard, M., 2002. PCR detection and serological evidence of granulocytic ehrlichial infection in roe deer (*Capreolus capreolus*) and chamois (*Rupicapra rupicapra*). Journal of clinical microbiology 40, 892-897.

M'Ghirbi, Y., Ghorbel, A., Amouri, M., Nebaoui, A., Haddad, S., Bouattour, A., 2009. Clinical, serological, and molecular evidence of ehrlichiosis and anaplasmosis in dogs in Tunisia. Parasitology research 104, 767-774.

M'Ghirbi, Y., Yaich, H., Ghorbel, A., Bouattour, A., 2012. *Anaplasma phagocytophilum* in horses and ticks in Tunisia. Parasites & vectors 5, 180.

MacLeod, J., 1932. Preliminary studies in tick transmission of louping ill. II. A study of the reaction of sheep to tick infestation. Vet. J. 88, 276–284.

Madewell, B.R., Gribble, D.H., 1982. Infection in two dogs with an agent resembling *Ehrlichia equi*. Journal of the American Veterinary Medical Association 180, 512-514.

Magnarelli, L.A., Ijdo, J.W., Shermant, B.A., Bushmich, S.L., Levy, S.A., Fikrig, E., 2002. Antibodies to granulocytic ehrlichiae in cattle from Connecticut. Journal of medical microbiology 51, 326-331.

Majazki, J., Wuppenhorst, N., Hartelt, K., Birtles, R., von Loewenich, F.D., 2013. *Anaplasma phagocytophilum* strains from voles and shrews exhibit specific ankA gene sequences. BMC veterinary research 9, 235.

Malmsten, J., Widen, D.G., Rydevik, G., Yon, L., Hutchings, M.R., Thulin, C.G., Soderquist, L., Aspan, A., Stuen, S., Dalin, A.M., 2014. Temporal and spatial variation in *Anaplasma phagocytophilum* infection in Swedish moose (*Alces alces*). Epidemiology and infection 142, 1205-1213.

Massung, R.F., Courtney, J.W., Hiratzka, S.L., Pitzer, V.E., Smith, G., Dryden,

R.L., 2005. *Anaplasma phagocytophilum* in white-tailed deer. Emerging infectious diseases 11, 1604-1606.

Massung, R.F., Mather, T.N., Levin, M.L., 2006. Reservoir competency of goats for the Ap-variant 1 strain of *Anaplasma phagocytophilum*. Infection and immunity 74, 1373-1375.

Massung, R.F., Mauel, M.J., Owens, J.H., Allan, N., Courtney, J.W., Stafford, K.C., 3rd, Mather, T.N., 2002. Genetic variants of *Ehrlichia phagocytophila*, Rhode Island and Connecticut. Emerging infectious diseases 8, 467-472.

Massung, R.F., Owens, J.H., Ross, D., Reed, K.D., Petrovec, M., Bjoersdorff, A., Coughlin, R.T., Beltz, G.A., Murphy, C.I., 2000. Sequence analysis of the ank gene of granulocytic ehrlichiae. Journal of clinical microbiology 38, 2917-2922.

Massung, R.F., Priestley, R.A., Miller, N.J., Mather, T.N., Levin, M.L., 2003. Inability of a variant strain of *Anaplasma phagocytophilum* to infect mice. The Journal of infectious diseases 188, 1757-1763.

Massung, R.F., Slater, K., Owens, J.H., Nicholson, W.L., Mather, T.N., Solberg, V.B., Olson, J.G., 1998. Nested PCR assay for detection of granulocytic ehrlichiae. Journal of clinical microbiology 36, 1090-1095.

Massung, R.F., Slater, K.G., 2003. Comparison of PCR assays for detection of the agent of human granulocytic ehrlichiosis, *Anaplasma phagocytophilum*. Journal of clinical microbiology 41, 717-722.

Masuzawa, T., Uchishima, Y., Fukui, T., Okamoto, Y., Muto, M., Koizumi, N., Yamada, A., 2011. Detection of *Anaplasma phagocytophilum* from Wild Boars and Deer in Japan. Japanese journal of infectious diseases 64, 333-336.

McEwen, A., 1947. Tick-borne fever in young lambs. The Veterinary record 59, 198-201.

Medlock, J.M., Hansford, K.M., Bormane, A., Derdakova, M., Estrada-Pena, A., George, J.C., Golovljova, I., Jaenson, T.G., Jensen, J.K., Jensen, P.M., Kazimirova, M., Oteo, J.A., Papa, A., Pfister, K., Plantard, O., Randolph, S.E., Rizzoli, A., Santos-Silva, M.M., Sprong, H., Vial, L., Hendrickx, G., Zeller, H., Van Bortel, W., 2013. Driving forces for changes in geographical distribution of *Ixodes ricinus* ticks in Europe. Parasites & vectors 6, 1.

Meyer-Kayser, E., Hoffmann, L., Silaghi, C., Pfister, K., Mahling, M., Passos, L.M., 2012. Dynamics of tick infestations in foxes in Thuringia, Germany. Ticks and tick-borne diseases 3, 232-239.

Michalik, J., Stanczak, J., Cieniuch, S., Racewicz, M., Sikora, B., Dabert, M., 2012. Wild boars as hosts of human-pathogenic *Anaplasma phagocytophilum* variants. Emerging infectious diseases 18, 998-1001.

Michalik, J., Stanczak, J., Racewicz, M., Cieniuch, S., Sikora, B., Szubert-Kruszynska, A., Grochowalska, R., 2009. Molecular evidence of *Anaplasma phagocytophilum* infection in wild cervids and feeding *Ixodes ricinus* ticks from west-central Poland. Clinical microbiology and infection : the official publication of the European Society of Clinical Microbiology and Infectious Diseases 15 Suppl 2, 81-83.

Moreno, C.X., Moy, F., Daniels, T.J., Godfrey, H.P., Cabello, F.C., 2006. Molecular analysis of microbial communities identified in different developmental stages of *Ixodes scapularis* ticks from Westchester and Dutchess Counties, New York. Environmental microbiology 8, 761-772.

Morgenthal, D., Hamel, D., Arndt, G., Silaghi, C., Pfister, K., Kempf, V.A., Kohn, B., 2012. [Prevalence of haemotropic *Mycoplasma spp*., *Bartonella spp*. and *Anaplasma phagocytophilum* in cats in Berlin/Brandenburg (Northeast Germany)]. Berliner und Munchener tierarztliche Wochenschrift 125, 418-427.

Morissette, E., Massung, R.F., Foley, J.E., Alleman, A.R., Foley, P., Barbet, A.F., 2009. Diversity of *Anaplasma phagocytophilum* strains, USA. Emerging

infectious diseases 15, 928-931.

Mosavi, L.K., Cammett, T.J., Desrosiers, D.C., Peng, Z.Y., 2004. The ankyrin repeat as molecular architecture for protein recognition. Protein science : a publication of the Protein Society 13, 1435-1448.

Murase, Y., Konnai, S., Hidano, A., Githaka, N.W., Ito, T., Takano, A., Kawabata, H., Ato, M., Tajima, T., Tajima, M., Onuma, M., Murata, S., Ohashi, K., 2011. Molecular detection of *Anaplasma phagocytophilum* in cattle and *Ixodes persulcatus* ticks. Veterinary microbiology 149, 504-507.

Murphy, C.I., Storey, J.R., Recchia, J., Doros-Richert, L.A., Gingrich-Baker, C., Munroe, K., Bakken, J.S., Coughlin, R.T., Beltz, G.A., 1998. Major antigenic proteins of the agent of human granulocytic ehrlichiosis are encoded by members of a multigene family. Infection and immunity 66, 3711-3718.

Nieder, M., Silaghi, C., Hamel, D., Pfister, K., Schmaschke, R., Pfeffer, M., 2012. Tick-borne fever caused by *Anaplasma phagocytophilum* in Germany: first laboratory confirmed case in a dairy cattle herd. Tierarztliche Praxis. Ausgabe G, Grosstiere/Nutztiere 40, 101-106.

Novakova, M., Vichova, B., Majlathova, V., Lesnakova, A., Pochybova, M., Petko, B., 2010. First case of human granulocytic anaplasmosis from Slovakia. Annals of agricultural and environmental medicine : AAEM 17, 173-175.

Obiegala, A., Pfeffer, M., Pfister, K., Tiedemann, T., Thiel, C., Balling, A., Karnath, C., Woll, D., Silaghi, C., 2014. *Candidatus Neoehrlichia mikurensis* and *Anaplasma phagocytophilum*: prevalences and investigations on a new transmission path in small mammals and ixodid ticks. Parasites & vectors 7, 563.

Oteo, J.A., Blanco, J.R., Martinez de Artola, V., Ibarra, V., 2000. First report of human granulocytic ehrlichiosis from southern Europe (Spain). Emerging infectious diseases 6, 430-432.

Overzier, E., Pfister, K., Herb, I., Mahling, M., Bock, G., Jr., Silaghi, C., 2013a. Detection of tick-borne pathogens in roe deer (*Capreolus capreolus*), in questing ticks (*Ixodes ricinus*), and in ticks infesting roe deer in southern Germany. Ticks and tick-borne diseases 4, 320-328.

Overzier, E., Pfister, K., Thiel, C., Herb, I., Mahling, M., Silaghi, C., 2013b. *Anaplasma phagocytophilum* in questing *Ixodes ricinus* ticks: comparison of prevalences and partial 16S rRNA gene variants in urban, pasture, and natural habitats. Applied and environmental microbiology 79, 1730-1734.

Palmer, G.H., Bankhead, T., Lukehart, S.A., 2009. 'Nothing is permanent but change'- antigenic variation in persistent bacterial pathogens. Cellular microbiology 11, 1697-1705.

Parola, P., Davoust, B., Raoult, D., 2005. Tick- and flea-borne rickettsial emerging zoonoses. Veterinary research 36, 469-492.

Paulauskas, A., Radzijevskaja, J., Rosef, O., 2012. Molecular detection and characterization of *Anaplasma phagocytophilum* strains. Comparative immunology, microbiology and infectious diseases 35, 187-195.

Petrovec, M., Bidovec, A., Sumner, J.W., Nicholson, W.L., Childs, J.E., Avsic-Zupanc, T., 2002. Infection with *Anaplasma phagocytophila* in cervids from Slovenia: evidence of two genotypic lineages. Wiener klinische Wochenschrift 114, 641-647.

Petrovec, M., Lotric Furlan, S., Zupanc, T.A., Strle, F., Brouqui, P., Roux, V., Dumler, J.S., 1997. Human disease in Europe caused by a granulocytic *Ehrlichia* species. Journal of clinical microbiology 35, 1556-1559.

Petrovec, M., Sixl, W., Schweiger, R., Mikulasek, S., Elke, L., Wust, G., Marth, E., Strasek, K., Stunzner, D., Avsic-Zupanc, T., 2003. Infections of wild animals with *Anaplasma phagocytophila* in Austria and the Czech Republic. Annals of the New York Academy of Sciences 990, 103-106.

Petrovec, M., Sumner, J.W., Nicholson, W.L., Childs, J.E., Strle, F., Barlic, J., Lotric-Furlan, S., Avsic Zupanc, T., 1999. Identity of ehrlichial DNA sequences derived from *Ixodes ricinus* ticks with those obtained from patients with human granulocytic ehrlichiosis in Slovenia. Journal of clinical microbiology 37, 209-210.

Poitout, F.M., Shinozaki, J.K., Stockwell, P.J., Holland, C.J., Shukla, S.K., 2005. Genetic variants of *Anaplasma phagocytophilum* infecting dogs in Western Washington State. Journal of clinical microbiology 43, 796-801.

Prukk, T., Ainsalu, K., Laja, E., Aigro, A., 2003. Human granulocytic ehrlichiosis in Estonia. Emerging infectious diseases 9, 1499-1500.

Rar, V., Golovljova, I., 2011. *Anaplasma*, *Ehrlichia*, and "*Candidatus Neoehrlichia*" bacteria: pathogenicity, biodiversity, and molecular genetic characteristics, a review. Infection, genetics and evolution : journal of molecular epidemiology and evolutionary genetics in infectious diseases 11, 1842-1861.

Rar, V.A., Epikhina, T.I., Livanova, N.N., Panov, V.V., Doroschenko, E.K., Pukhovskaya, N.M., Vysochina, N.P., Ivanov, L.I., 2011. Genetic variability of *Anaplasma phagocytophilum* in *Ixodes persulcatus* ticks and small mammals in the Asian part of Russia. Vector borne and zoonotic diseases (Larchmont, N.Y.) 11, 1013-1021.

Rejmanek, D., Foley, P., Barbet, A., Foley, J., 2012. Antigen variability in *Anaplasma phagocytophilum* during chronic infection of a reservoir host. Microbiology (Reading, England) 158, 2632-2641.

Rikihisa, Y., 1991. The tribe *Ehrlichieae* and ehrlichial diseases. Clinical microbiology reviews 4, 286-308.

Rikihisa, Y., Lin, M., 2010. *Anaplasma phagocytophilum* and *Ehrlichia chaffeensis* type IV secretion and Ank proteins. Current opinion in microbiology 13, 59-66.

Robinson, M., Shaw, S., Morgan, E., 2009. *Anaplasma phagocytophilum* infection in a multi-species deer community in the New Forest, England. Eur J Wildl Res 55, 439-442.

Roux, S., Enault, F., Bronner, G., Debroas, D., 2011. Comparison of 16S rRNA and protein-coding genes as molecular markers for assessing microbial diversity (*Bacteria* and *Archaea*) in ecosystems. FEMS microbiology ecology 78, 617-628.

Rurangirwa, F.R., Stiller, D., French, D.M., Palmer, G.H., 1999. Restriction of major surface protein 2 (MSP2) variants during tick transmission of the ehrlichia *Anaplasma marginale*. Proceedings of the National Academy of Sciences of the United States of America 96, 3171-3176.

Ruscio, M., Cinco, M., 2003. Human granulocytic ehrlichiosis in Italy: first report on two confirmed cases. Annals of the New York Academy of Sciences 990, 350-352.

Rymaszewska, A., 2008. Divergence within the marker region of the groESL operon in *Anaplasma phagocytophilum*. European journal of clinical microbiology & infectious diseases : official publication of the European Society of Clinical Microbiology 27, 1025-1036.

Rymaszewska, A., 2010. Variability within the msp2 gene in populations of *Anaplasma phagocythopilum*. Folia biologica 56, 269-275.

Santos, A.S., Santos-Silva, M.M., Almeida, V.C., Bacellar, F., Dumler, J.S., 2004. Detection of *Anaplasma phagocytophilum* DNA in *Ixodes ticks* (*Acari*: *Ixodidae*) from Madeira Island and Setubal District, mainland Portugal. Emerging infectious diseases 10, 1643-1648.

Savidge, C., Ewing, P., Andrews, J., Aucoin, D., Lappin, M.R., Moroff, S., 2015. *Anaplasma phagocytophilum* infection of domestic cats: 16 cases from the northeastern USA. J Feline Med Surg.

Scharf, W., Schauer, S., Freyburger, F., Petrovec, M., Schaarschmidt-Kiener, D., Liebisch, G., Runge, M., Ganter, M., Kehl, A., Dumler, J.S., Garcia-Perez, A.L., Jensen, J., Fingerle, V., Meli, M.L., Ensser, A., Stuen, S., von Loewenich, F.D., 2011a. Distinct host species correlate with *Anaplasma phagocytophilum* ankA gene clusters. Journal of clinical microbiology 49, 790-796.

Scharf, W., Schauer, S., Freyburger, F., Petrovec, M., Schaarschmidt-Kiener, D., Liebisch, G., Runge, M., Ganter, M., Kehl, A., Dumler, J.S., Garcia-Perez, A.L., Jensen, J., Fingerle, V., Meli, M.L., Ensser, A., Stuen, S., von Loewenich, F.D., 2011b. Distinct host species correlate with Anaplasma phagocytophilum ankA gene clusters▽. Journal of clinical microbiology 49.

Shaw, S.E., Binns, S.H., Birtles, R.J., Day, M.J., Smithson, R., Kenny, M.J., 2005. Molecular evidence of tick-transmitted infections in dogs and cats in the United Kingdom. The Veterinary record 157, 645-648.

Sidel'nikov Iu, N., Mediannikov, O., Ivanov, L.I., Zdanovskaia, N.I., 2003. [The first case of granulocytic ehrlichiosis in the Far East of the Russian Federation]. Klinicheskaia meditsina 81, 67-68.

Silaghi, C., Hamel, D., Pfister, K., Rehbein, S., 2011a. *Babesia* species and co-infection with *Anaplasma phagocytophilum* in free-ranging ungulates from Tyrol (Austria). Wiener Tierarztliche Monatsschrift 98, 268-274.

Silaghi, C., Hamel, D., Thiel, C., Pfister, K., Friche Passos, L.M., Rehbein, S., 2011b. Genetic Variants of *Anaplasma phagocytophilum* in Wild Caprine and Cervid Ungulates from the Alps in Tyrol, Austria. Vector-Borne and Zoonotic Diseases 11, 355-362.

Silaghi, C., Kohn, B., Chirek, A., Thiel, C., Nolte, I., Liebisch, G., Pfister, K., 2011c. Relationship of molecular and clinical findings on *Anaplasma phagocytophilum* involved in natural infections of dogs. Journal of clinical microbiology 49, 4413-4414.

Silaghi, C., Liebisch, G., Pfister, K., 2011d. Genetic variants of *Anaplasma phagocytophilum* from 14 equine granulocytic anaplasmosis cases. Parasites & vectors 4, 161.

Silaghi, C., Pfister, K., Overzier, E., 2014. Molecular investigation for bacterial and protozoan tick-borne pathogens in wild boars (*Sus scrofa*) from southern Germany. Vector borne and zoonotic diseases (Larchmont, N.Y.) 14, 371-373.

Silaghi, C., Scheuerle, M.C., Friche Passos, L.M., Thiel, C., Pfister, K., 2011e. PCR detection of *Anaplasma phagocytophilum* in goat flocks in an area endemic for tick-borne fever in Switzerland. Parasite (Paris, France) 18, 57-62.

Silaghi, C., Skuballa, J., Thiel, C., Pfister, K., Petney, T., Pfaffle, M., Taraschewski, H., Passos, L.M., 2012a. The European hedgehog *(Erinaceus europaeus*)--a suitable reservoir for variants of *Anaplasma phagocytophilum*? Ticks and tick-borne diseases 3, 49-54.

Silaghi, C., Woll, D., Hamel, D., Pfister, K., Mahling, M., Pfeffer, M., 2012b. *Babesia spp.* and *Anaplasma phagocytophilum* in questing ticks, ticks parasitizing rodents and the parasitized rodents--analyzing the host-pathogen-vector interface in a metropolitan area. Parasites & vectors 5, 191.

Sonenshine, D.E., 1991. Biology of Ticks, Vol 1. Oxford University Press, New York, 472 p.

Stamp, J.T., Watt, J.A., Jamieson, S., 1950. Tick-borne fever as a cause of abortion in sheep. The Veterinary record 62, 465-470.

Stanczak, J., Racewicz, M., Kruminis-Lozowska, W., Kubica-Biernat, B., 2002. Coinfection of *Ixodes ricinus* (*Acari*: *Ixodidae*) in northern Poland with the agents of Lyme borreliosis (LB) and human granulocytic ehrlichiosis (HGE). International journal of medical microbiology : IJMM 291 Suppl 33, 198-201.

Strik, N.I., Alleman, A.R., Barbet, A.F., Sorenson, H.L., Wamsley, H.L., Gaschen, F.P., Luckschander, N., Wong, S., Chu, F., Foley, J.E., Bjoersdorff, A., Stuen, S., Knowles, D.P., 2007. Characterization of *Anaplasma phagocytophilum* major surface protein 5 and the extent of its cross-reactivity with *A. marginale*. Clinical and vaccine immunology : CVI 14, 262-268.

Stuen, S., 1996. Experimental tick-borne fever infection in reindeer (*Rangifer tarandus tarandus*). The Veterinary record 138, 595-596.

Stuen, S., Bergstrom, K., Petrovec, M., Van de Pol, I., Schouls, L.M., 2003. Differences in clinical manifestations and hematological and serological responses after experimental infection with genetic variants of *Anaplasma phagocytophilum* in sheep. Clinical and diagnostic laboratory immunology 10, 692-695.

Stuen, S., Engvall, E.O., van de Poll, I., Schouls, L.M., 2001. Granulocytic ehrlichiosis in a roe deer calf in Norway. Journal of wildlife diseases 37, 614-616.

Stuen, S., Granquist, E.G., Silaghi, C., 2013a. *Anaplasma phagocytophilum*--a widespread multi-host pathogen with highly adaptive strategies. Frontiers in cellular and infection microbiology 3, 31.

Stuen, S., Moum, T., Bernhoft, A., Vene, S., 2006. A paretic condition in an *Anaplasma phagocytophilum* infected roe deer calf. Journal of wildlife diseases 42, 170-174.

Stuen, S., Pettersen, K.S., Granquist, E.G., Bergstrom, K., Bown, K.J., Birtles, R.J., 2013b. *Anaplasma phagocytophilum* variants in sympatric red deer (*Cervus elaphus*) and sheep in southern Norway. Ticks and tick-borne diseases 4, 197-201.

Stuen, S., Scharf, W., Schauer, S., Freyburger, F., Bergstrom, K., von Loewenich, F.D., 2010. Experimental infection in lambs with a red deer (*Cervus elaphus*) isolate of *Anaplasma phagocytophilum*. Journal of wildlife diseases 46, 803-809.

Sumner, J.W., Nicholson, W.L., Massung, R.F., 1997. PCR amplification and comparison of nucleotide sequences from the groESL heat shock operon of *Ehrlichia* species. Journal of clinical microbiology 35, 2087-2092.

Swanson, S.J., Neitzel, D., Reed, K.D., Belongia, E.A., 2006. Coinfections acquired from *Ixodes* ticks. Clinical microbiology reviews 19, 708-727.

Tate, C.M., Mead, D.G., Luttrell, M.P., Howerth, E.W., Dugan, V.G., Munderloh, U.G., Davidson, W.R., 2005. Experimental infection of white-tailed deer with *Anaplasma phagocytophilum*, etiologic agent of human granulocytic anaplasmosis. Journal of clinical microbiology 43.

Telford, S.R., 3rd, Dawson, J.E., Katavolos, P., Warner, C.K., Kolbert, C.P., Persing, D.H., 1996. Perpetuation of the agent of human granulocytic ehrlichiosis in a deer tick-rodent cycle. Proceedings of the National Academy of Sciences of the United States of America 93, 6209-6214.

Tomanovic, S., Chochlakis, D., Radulovic, Z., Milutinovic, M., Cakic, S., Mihaljica, D., Tselentis, Y., Psaroulaki, A., 2013. Analysis of pathogen co-occurrence in host-seeking adult hard ticks from Serbia. Exp Appl Acarol 59, 367-376.

Torina, A., Alongi, A., Naranjo, V., Estrada-Pena, A., Vicente, J., Scimeca, S., Marino, A.M., Salina, F., Caracappa, S., de la Fuente, J., 2008a. Prevalence and genotypes of *Anaplasma* species and habitat suitability for ticks in a Mediterranean ecosystem. Applied and environmental microbiology 74, 7578-7584.

Torina, A., Alongi, A., Naranjo, V., Scimeca, S., Nicosia, S., Di Marco, V., Caracappa, S., Kocan, K.M., de la Fuente, J., 2008b. Characterization of *Anaplasma* infections in Sicily, Italy. Annals of the New York Academy of Sciences 1149, 90-93.

Tylewska-Wierzbanowska, S., Chmielewski, T., Kondrusik, M., Hermanowska-

Szpakowicz, T., Sawicki, W., Sułek, K., 2001. First cases of acute human granulocytic ehrlichiosis in Poland. European journal of clinical microbiology & infectious diseases : official publication of the European Society of Clinical Microbiology 20, 196-198.

Van Dobbenburgh, A., Van Dam, A.P., Fikrig, E., 1999. Human granulocytic ehrlichiosis in western Europe. The New England journal of medicine 340, 1214-1216.

Veronesi, F., Galuppi, R., Tampieri, M.P., Bonoli, C., Mammoli, R., Piergili Fioretti, D., 2011. Prevalence of *Anaplasma phagocytophilum* in fallow deer (*Dama dama*) and feeding ticks from an Italy preserve. Research in veterinary science 90, 40-43.

Vichova, B., Majlathova, V., Novakova, M., Stanko, M., Hviscova, I., Pangracova, L., Chrudimsky, T., Curlik, J., Petko, B., 2014. *Anaplasma* infections in ticks and reservoir host from Slovakia. Infection, genetics and evolution : journal of molecular epidemiology and evolutionary genetics in infectious diseases 22, 265-272.

Vichova, B., Majlathova, V., Novakova, M., Straka, M., Pet'ko, B., 2010. First molecular detection of *Anaplasma phagocytophilum* in European brown bear (*Ursus arctos*). Vector borne and zoonotic diseases (Larchmont, N.Y.) 10, 543-545.

Vogl, U.M., Presterl, E., Stanek, G., Ramharter, M., Gattringer, K.B., Graninger, W., 2010. First described case of human granulocytic anaplasmosis in a patient in Eastern Austria. Wiener medizinische Wochenschrift (1946) 160, 91-93.

Von Loewenich, F.D., Baumgarten, B.U., Schroppel, K., Geissdorfer, W., Rollinghoff, M., Bogdan, C., 2003a. High diversity of ankA sequences of *Anaplasma phagocytophilum* among *Ixodes ricinus* ticks in Germany. Journal of clinical microbiology 41, 5033-5040.

Von Loewenich, F.D., Stumpf, G., Baumgarten, B.U., Rollinghoff, M., Dumler, J.S., Bogdan, C., 2003b. A case of equine granulocytic ehrlichiosis provides molecular evidence for the presence of pathogenic *Anaplasma phagocytophilum* (HGE agent) in Germany. European journal of clinical microbiology & infectious diseases : official publication of the European Society of Clinical Microbiology 22, 303-305.

Walder, G., Falkensammer, B., Aigner, J., Tiwald, G., Dierich, M.P., Wurzner, R., Lechleitner, P., 2003. First documented case of human granulocytic ehrlichiosis in Austria. Wiener klinische Wochenschrift 115, 263-266.

Walls, J.J., Greig, B., Neitzel, D.F., Dumler, J.S., 1997. Natural infection of small mammal species in Minnesota with the agent of human granulocytic ehrlichiosis. Journal of clinical microbiology 35, 853-855.

Wang, Y., Chen, C., Zhang, L., 2013. Molecular characterization of Msp2/P44 of *Anaplasma phagocytophilum* isolated from infected patients and *Haemaphysalis longicornis* in Laizhou Bay, Shandong Province, China. PloS one 8, e78189.

Welc-Faleciak, R., Kowalec, M., Zajkowska, J., Pancewicz, S.A., Sinski, E., 2015. Clinical and molecular features of one case of human infection with *Anaplasma phagocytophilum* from Podlaskie Province in eastern Poland. Annals of agricultural and environmental medicine : AAEM 22, 414-417.

Weller, S.J., Baldridge, G.D., Munderloh, U.G., Noda, H., Simser, J., Kurtti, T.J., 1998. Phylogenetic placement of rickettsiae from the ticks *Amblyomma americanum* and *Ixodes scapularis*. Journal of clinical microbiology 36, 1305-1317.

Woese, C.R., Fox, G.E., 1977. Phylogenetic structure of the prokaryotic domain: the primary kingdoms. Proceedings of the National Academy of Sciences of the United States of America 74, 5088-5090.

Woese, C.R., Fox, G.E., Zablen, L., Uchida, T., Bonen, L., Pechman, K., Lewis,

B.J., Stahl, D., 1975. Conservation of primary structure in 16S ribosomal RNA. Nature 254, 83-86.

Woldehiwet, Z., 2006. *Anaplasma phagocytophilum* in ruminants in Europe. Annals of the New York Academy of Sciences 1078, 446-460.

Woldehiwet, Z., 2010. The natural history of *Anaplasma phagocytophilum*. Veterinary parasitology 167, 108-122.

Wu, D., Wuritu, Yoshikawa, Y., Gaowa, Kawamori, F., Ikegaya, A., Ohtake, M., Ohashi, M., Shimada, M., Takada, A., Iwai, K., Ohashi, N., 2015. Molecular and Serological Survey of *Rickettsiales* Bacteria in Wild Sika Deer (*Cervus nippon nippon*) in Shizuoka Prefecture, Japan: High Prevalence of *Anaplasma* Species. Japanese journal of infectious diseases.

Yang, J., Liu, Z., Guan, G., Liu, Q., Li, Y., Chen, Z., Ma, M., Liu, A., Ren, Q., Luo, J., Yin, H., 2013. Prevalence of *Anaplasma phagocytophilum* in ruminants, rodents and ticks in Gansu, north-western China. Journal of medical microbiology 62, 254-258.

Yoshiie, K., Kim, H.Y., Mott, J., Rikihisa, Y., 2000. Intracellular infection by the human granulocytic ehrlichiosis agent inhibits human neutrophil apoptosis. Infection and immunity 68, 1125-1133.

Yoshikawa, Y., Ohashi, N., Wu, D., Kawamori, F., Ikegaya, A., Watanabe, T., Saitoh, K., Takechi, D., Murakami, Y., Shichi, D., Aso, K., Ando, S., 2014. *Anaplasma phagocytophilum* antibodies in humans, Japan, 2010-2011. Emerging infectious diseases 20, 508-509.

Yoshimoto, K., Matsuyama, Y., Matsuda, H., Sakamoto, L., Matsumoto, K., Yokoyama, N., Inokuma, H., 2010. Detection of *Anaplasma bovis* and *Anaplasma phagocytophilum* DNA from *Haemaphysalis megaspinosa* in Hokkaido, Japan. Veterinary parasitology 168, 170-172.

Zeidner, N.S., Burkot, T.R., Massung, R., Nicholson, W.L., Dolan, M.C., Rutherford, J.S., Biggerstaff, B.J., Maupin, G.O., 2000. Transmission of the agent of human granulocytic ehrlichiosis by *Ixodes spinipalpis* ticks: evidence of an enzootic cycle of dual infection with *Borrelia burgdorferi* in Northern Colorado. The Journal of infectious diseases 182, 616-619.

Zeman, P., Jahn, P., 2009. An entropy-optimized multilocus approach for characterizing the strains of *Anaplasma phagocytophilum* infecting horses in the Czech Republic. Journal of medical microbiology 58, 423-429.

Zeman, P., Pecha, M., 2008. Segregation of genetic variants of *Anaplasma phagocytophilum* circulating among wild ruminants within a Bohemian forest (Czech Republic). Int. J. Med. Microbiol. 298, 203-210.

Zhan, L., Cao, W.C., de Vlas, S., Xie, S.Y., Zhang, P.H., Wu, X.M., Dumler, J.S., Yang, H., Richardus, J.H., Habbema, J.D., 2008. A newly discovered *Anaplasma phagocytophilum* variant in rodents from southeastern China. Vector borne and zoonotic diseases (Larchmont, N.Y.) 8, 369-380.

Zhan, L., Cao, W.C., Jiang, J.F., Zhang, X.A., Liu, Y.X., Wu, X.M., Zhang, W.Y., Zhang, P.H., Bian, C.L., Dumler, J.S., Yang, H., Zuo, S.Q., Chu, C.Y., Liu, W., Richardus, J.H., Habbema, J.D., 2010a. *Anaplasma phagocytophilum* from Rodents and Sheep, China. Emerging infectious diseases 16, 764-768.

Zhan, L., Cao, W.C., Jiang, J.F., Zhang, X.A., Wu, X.M., Zhang, W.Y., Liu, W., Zuo, S.Q., Cao, Z.W., Yang, H., Richardus, J.H., Habbema, J.D., 2010b. *Anaplasma phagocytophilum* in livestock and small rodents. Veterinary microbiology 144, 405-408.

Zhang, L., Liu, H., Xu, B., Lu, Q., Li, L., Chang, L., Zhang, X., Fan, D., Li, G., Jin, Y., Cui, F., Shi, Y., Li, W., Xu, J., Yu, X.J., 2012. *Anaplasma phagocytophilum* infection in domestic animals in ten provinces/cities of China. The American journal of tropical medicine and hygiene 87, 185-189.

Zhang, L., Shan, A., Mathew, B., Yin, J., Fu, X., Zhang, J., Lu, J., Xu, J., Dumler, J.S., 2008. Rickettsial Seroepidemiology among farm workers, Tianjin, People's Republic of China. Emerging infectious diseases 14, 938-940.

1. Websites

http://www.ncbi.nlm.nih.gov/genome/genomes/1083? (date: 30.11.2014)

http://www.stepmap.de/profile.php#1 (date: 26.12.2014)

http://www.cdc.gov/anaplasmosis/stats/index.html (date: 12.10.2014)

X. TABLES

XI. FIGURES

XII. ANNEX

1. Animal samples

Table 30: Accession numbers of nucleotide sequences from the GenBank

Animal species	Available sequences (acc. no.) from previous studies				Reference
	16S rRNA	***groEL***	***msp4***	***msp2***	
cattle (*Bos primigenius taurus*)	FJ538291	GQ452228, GQ452230, GQ452232	HM028677-HM028680		Silaghi et al. (2011e)
chamois (*Rupicapra r. rupicapra*)	FJ812399, FJ812400, FJ812401, GU265827	GQ988769, GQ988768, GQ988756, GQ988774	GU265837, GU265834, GU265838		Silaghi et al. (2011a)
dog (*Canis lupus familiaris*)	FJ829748-FJ829750, FJ829752-FJ829753, FJ829755-FJ829756, FJ829758-FJ829762, JN656336-JN656340, JN656342-JN656346. FJ829764-FJ829766, FJ829768, FJ829770, FJ829772-FJ829774, FJ829777-FJ829779, FJ829781-FJ829784, FJ829786, JN656352-JN656371, JN656378, JN656380				Silaghi et al. (2011c)
goat *(Capra hircus)*	FJ538288, FJ538289, FJ538290	GQ452225, GQ452226, GQ452227, GQ452229, GQ452231	HM028674, HM028675, HM028676		Silaghi et al. (2011e)
hedgehog (*Erinaceus europaeus*)	JN571156, JN571159, JN571160, JN571161, JN571162,				Silaghi et al. (2012a)

	JN571163, JN571164				
horse (*Equus ferus caballus*)	JF893926-JF893936, JF893937-JF893938, JF893940	JF893913-JF893925	JF893886-JF893898, JF907577	JF893899, JF893912, JF893900-JF893911	Silaghi et al. (2011d)
ibex (*Capra ibex*)	FJ812395, FJ812396	GQ988773, GQ988772	GU265847, GU265846		Silaghi et al. (2011b)
mouflon (*Ovis orientalis musimon*)	FJ812402, FJ812409	GQ988755, GQ988759	GU265839, GU265840		Silaghi et al. (2011b), partially unpublished
	KU510417, KU510420, KU510428, KU510428, KU510433				Kauffmann et al. (2016)
fallow deer (*Dama dama*)	KU510418, KU510429				Kauffmann et al. (2016)
red deer (*Cervus elaphus*)	FJ812388, FJ812390-FJ812394, FJ812397-FJ812398,	GQ988757-GQ988758, GQ988765-GQ988767, GQ988771	GU265828-GU265833	FJ812384-FJ812385	Silaghi et al. (2011b), partially unpublished
roe deer (*Capreolus capreolus*)	FJ812389, FJ812391, FJ812403-FJ812408	GQ988753-GQ988754, GQ988760-GQ988764, GQ988770	GU265848, GU265844, GU265835, GU265842, GU265836, GU265841, GU265843. GU265845	FJ812386-FJ812387	Silaghi et al. (2011b), partially unpublished
	JX627360-JX627368				Overzier et al. (2013a)
	KU510422, KU510423, KU510425, KU510426, KU510427, KU510432, KU510434, KU510435				Kauffmann et al. (2016)
wild boar (*Sus scrofa*)	KC833754				Silaghi et al. (2014)

1.1. Gene-variant distributions and multiple alignments

1.1.1. *16S rRNA* gene

Table 31: Distribution of the *16S rRNA* variants in the different examined animal species

	Dog	Horse	Cat	Hedgehog	Red fox	Cattle	Goat	Roe deer	Red deer	Sika deer	Fallow deer	Mouflon	Chamois	Ibex	Wild boar
16S-1(A)	21	1		58	4										
16S-2(B)	27	11		1	2			1	4	3	2	3			2
16S-3(D)		1													
16S-7(I)								13		1					
16S-8(J)									1	1					
16S-9(K)	1														
16S-10(L)	1														
16S-12(N)								1							
16S-13(O)			1												
16S-16(S)		1		1	2			1	9	4		5		2	
16S-17(T)									1						
16S-19(V)				1				3			1				
16S-20(W)				1		37		5	7	4	1	7	8		
16S-21(X)							2	23		1	1	1			
16S-22(Y)						1	1	22			1	2			
16S-23(Z)							1					1	1		
16S-24								1							
16S-25									1						
16S-26									1						
16S-27					1										
16S-28									1						
16S-29										1					
16S-30									1						
Total	**50**	**14**	**1**	**62**	**7**	**38**	**4**	**70**	**26**	**15**	**6**	**19**	**9**	**2**	**2**

Table 32: Alignment of the *16S rRNA* nucleotide sequences - highlighted nucleotide positions are differing within the variants

```
CLUSTAL 2.1 multiple sequence alignment

16S-14_P_      GTAAAGAATAG TTAGTGGCAGACGGGTGA GTA ATGCATAGGAATCTA CCTAG TAGTATGG 60
16S-23_Z_      ATAAAGAATAG TTAGTGGCAGACGGGTGA GTA ATGCATAGGAATCTA CCTAG TAGTATGG 60
16S-18_U_      ATAAAGAATAG TTAGTGGCAGACGGGTGA GTA ATGCATAGGAATCTA CCTAG TAGTATGG 60
16S-10_L_      ATAAAGAATAG TTAGTGGCAGACGGGTGA GTA ATGCATAGGAATCTA CCTAG TAGTATGG 60
16S-5_G_       ATAAAGAATAG TTAGTGGCAGACGGGTGA GTA ATGCATAGGAATCTA CCTAG TAGTATGG 60
16S-1_A_       ATAAAGAATAG TTAGTGGCAGACGGGTGA GTA ATGCATAGGAATCTA CCTAG TAGTATGG 60
16S-27         ATAAAGAATAG TTAGTGGCAGACGGGTGA GTT ATGCATAGGAATCTA CCTAG TAGTATGG 60
16S-24         ATAAAGAATAG TTAGTGGCAGACGGGTGA GTA ATGCATAGGAATCTG CCTAG TAGTATGG 60
16S-15_R_      ATAAAGAATAG TTAGTGGCAGACGGGTGA GTA ATGCATAGGAATCTA CCTAG TAGTATGG 60
16S-4_E_       ATAAAGAATAG TTAGTGGCAGACGGGTGA GTA ATGCATAGGAATCTA CCTAG TAGTATGG 60
```

```
16S-2_B_      ATAAAGAATAG TTAGTGGCAGACGGGTGA GTA ATGCATAGGAATCTA CCTAG TAGTATGG 60
16S-6_H_      ATAAAGAATAA TTAGTGGCAGACGGGTGA GTA ATGCATAGGAATCTA CCTAG TAGTATGG 60
16S-19_V_     ATAAAGGATAA TTAGTGGCAGACGGGTGA GTA ATGCATAGGAATCTA CCTAG TAGTATGG 60
16S-29        ATAAAGAATAA TTAGTGGCAGACGGGTGA GTA ATGCATAGGAATCTA CCTAG TAGTATGG 60
16S-20_W_     ATAAAGAATAA TTAGTGGCAGACGGGTGA GTA ATGCATAGGAATCTA CCTAG TAGTATGG 60
16S-30        ATAGAGAATAA TTAGTGGCAGACGGGTGA GTA ATGCATAGGAATCTA CCTAG TAGTATGG 60
16S-11_M_     ATAGAGAATAG TTAGTGGCAGACGGGTGA GTA ATGCATAGGAATCTA CCTAG TAGTATGG 60
16S-25        ATAGAGAATAG TTAGTGGCAGACGGGTGA GTA ATGCATAGGAATCTA CCTAG TAGTATGG 60
16S-28        ATAGAGAATAG TTAGTGGCAGACGGGTGA GTA ATGCATAGGAATCTA CCTAG TAGTATGG 60
16S-16_S_     ATAGAGAATAG TTAGTGGCAGACGGGTGA GTA ATGCATAGGAATCTA CCTAG TAGTATGG 60
16S-26        ATAGAGAATAG TTAGTGGCAGACGGGTGA GTA ATGCATAGGAATCTA CCTAG TAGTATGG 60
16S-3_D_      ATGAAGAATAA TTAGTGGCAGACGGGTGA GTA ATGCATAGGAATCTA CCTAA TAGTATGG 60
16S-9_K_      ATGAAGAATAA TTAGTGGCAGACGGGTGG GTA ATGCATAGGAATCTA CCTAA TAGTATGG 60
16S-8_J_      ATGAAGAATAA TTAGTGGCAGACGGGTGA GTA ATGCATAGGAATCTA CCTAA TAGTATGG 60
16S-7_I_      ATGAAGAATAA TTAGTGGCAGACGGGTGA GTA ATGCATAGGAATCTA CCTAG TAGTATGG 60
16S-12_N_     ATGAAGAATAA TTAGTGGCAGACGGGTGA GTA ATGCATAGGAATCTA CCTAG TAGTATGG 60
16S-21_X_     ATGAAGAATAA TTAGTGGCAGACGGGTGA GTA ATGCATAGGAATCTA CCTAG TAGTATGG 60
16S-13_O_     ATGAGGAATAG TTAGTGGCAGACGGGTGA GTA ATGCATAGGAATCTA CCTAG TAGTATGG 60
16S-22_Y_     ATGAAGAATAG TTAGTGGCAGACGGGTGA GTA ATGCATAGGAATCTA CCTAG TAGTATGG 60
16S-17_T_     ATGGAGAATAG TTAGTGGCAGACGGGTGA GTA ATGCATAGGAATCTA CCTAG TAGTATGG 60
Position      1*345*7***11*****************29**32**************47****52********

16S-14_P_     GATAGCCACTAGA AATGGTGGGTAAT ACTGTATAATC CCTGC  GGGGGAAAGATTTATCGC 120
16S-23_Z_     GATAGCCACTAGA AATGGTGGGTAAT ACTGTATAATT CCTGC  GGGGGAAAGATTTATCGC 120
16S-18_U_     GATAGCCACTAGA AATGGTGGGTAAT ACTGTATAATC CCTGC  GGGGGAAAGATTTATCGC 120
16S-10_L_     GATAGCCACTAGA AATGGTGGGTAAT ACTGTATAATC CCTGC  GGGGGAAAGATTTATCGC 120
16S-5_G_      GATAGCCACTAGA AATGGTGGGTAAT ACTGTATAATC CCTGC  GGGGGAAAGATTTATCGC 120
16S-1_A_      GATAGCCACTAGA AATGGTGGGTAAT ACTGTATAATC CCTGC  GGGGGAAAGATTTATCGC 120
16S-27        GATAGCCACTAGA AATGGTGGGTAAT ACTGTATAATC CCTGC  GGGGGAAAGATTTATCGC 120
16S-24        GATAGCCACTAGA AATGGTGGGTAAT ACTGTATAATC CCTGC  GGGGGAAAGATTTATCGC 120
16S-15_R_     GATAGCCACTAGA AATGGTGGGTAAT ACTGTATAATC CCTGC  GGGGGAAAGATTTATCGC 120
16S-4_E_      GATAGCCACTAGG AATGGTGGGTAAT ACTGTATAATC CCTGC  GGGGGAAAGATTTATCGC 120
16S-2_B_      GATAGCCACTAGA AATGGTGGGTAAT ACTGTATAATC CCTGC  GGGGGAAAGATTTATCGC 120
16S-6_H_      GATAGCCACTAGA AATGGTGGGTAAT ACTGTATAATC CCTGC  GGGGGAAAGATTTATCGC 120
16S-19_V_     GATAGCCACTAGA AATGGTGGGTAAT ACTGTATAATC CCTGC  GGGGGAAAGATTTATCGC 120
16S-29        GATAGCCACTAGA AATGGTGGGTAAT ACTGTATAATC CCTGC  GGGGGAAAGATTTATCGC 120
16S-20_W_     GATAGCCACTAGA AATGGTGGGTAAT ACTGTATAATC CCTGC  GGGGGAAAGATTTATCGC 120
16S-30        GATAGCCACTAGA AATGGTGGGTAAT ACTGTATAATC CCTGC  GGGGGAAAGATTTATCGC 120
16S-11_M_     GATAGCCACTAGA AATGGTGGGTAAT ACTGTATAATC CCTGC  GGGGGAAAGATTTATCGC 120
16S-25        GATAGCCACTAGA AATGGTGGGTAAT ACTGTATAATC CCTGC  GGGGGAAAGATTTATCGC 120
16S-28        GATAGCCACTAGA AATGGTGGGTAAT ACTGTATAATC CCTGC  GGGGGAAAGATTTATCGC 120
16S-16_S_     GATAGCCACTAGA AATGGTGGGTAAT ACTGTATAATC CCTGC  GGGGGAAAGATTTATCGC 120
16S-26        GATAGCCACTAGA AATGGTGGGTAAT ACTGTATAATC CCTGC  GGGGGAAAGATTTATCGC 120
16S-3_D_      GATAGCCACTAGA AATGGTGGGTAAT ACTGTATAATC CCTGC  GGGGGAAAGATTTATCGC 120
16S-9_K_      GATAGCCACTAGA AATGGTGGGTAAC ACTGTATAATC CCTGC  GGGGGAAAGATTTATCGC 120
16S-8_J_      GATAGCCACTAGA AATGGTGGGTAAT ACTGTATAATC CCTGC  GGGGGAAAGATTTATCGC 120
16S-7_I_      GATAGCCACTAGA AATGGTGGGTAAT ACTGTATAATC CCTGA  GGGGGAAAGATTTATCGC 120
16S-12_N_     GATAGCCACTAGA AATGGTGGGTAAT ACTGTATAATC CCTGC  GGGGGAAAGATTTATCGC 120
16S-21_X_     GATAGCCACTAGA AATGGTGGGTAAT ACTGTATAATC CCTGC  GGGGGAAAGATTTATCGC 120
16S-13_O_     GATAGCCACTAGA AATGGTGGGTAAT ACTGTATAATC CCTGC  GGGGGAAAGATTTATCGC 120
16S-22_Y_     GATAGCCACTAGA AATGGTGGGTAAT ACTGTATAATC CCTGC  GGGGGAAAGATTTATCGC 120
16S-17_T_     GATAGCCACTAGA AATGGTGGGTAAT ACTGTATAATC CCTGC  GGGGGAAAGATTTATCGC 120
Position      ************73************86**********97****102******************

16S-14_P_      TATTAGATGAGCCTATGTTAGATTAGCTA  GTTGGTAGGGTAAAGGCCTACCA  AGGCGATG 180
16S-23_Z_      TATTAGATGAGCCTATGTTAGATTAGCTA  GTTGGTAGGGTAAAGGCCTACCA  AGGCGATG 180
16S-18_U_      TATTAGATGAGCCTATGTTAGATTAGCTA  GTTGGTAGGGTAAAGGCCTACCA  AGGCGATG 180
16S-10_L_      TATTAGATGAGCCTATGTTAGATTAGCTA  GTTGGTAGGGTAAAGGCCTACCA  AGGCGATG 180
16S-5_G_       TATTAGATGAGCCTATGTTAGATTAGCTA  GTTGGTAGGGTAAAGGCCTACCG  AGGCGATG 180
16S-1_A_       TATTAGATGAGCCTATGTTAGATTAGCTA  GTTGGTAGGGTAAAGGCCTACCA  AGGCGATG 180
16S-27         TATTAGATGAGCCTATGTTAGATTAGCTA  GTTGGTAGGGTAAAGGCCTACCA  AGGCGATG 180
16S-24         TATTAGATGAGCCTATGTTAGATTAGCTA  GTTGGTAGGGTAAAGGCCTACCA  AGGCGATG 180
16S-15_R_      TATTAGATGAGCCTATGTTAGATTAGCTA  GTTGGTAGGGTAAAGGCCTACCA  AGGCGATG 180
16S-4_E_       TATTAGATGAGCCTATGTTAGATTAGCTA  GTTGGTAGGGTAAAGGCCTACCA  AGGCGATG 180
16S-2_B_       TATTAGATGAGCCTATGTTAGATTAGCTA  GTTGGTAGGGTAAAGGCCTACCA  AGGCGATG 180
16S-6_H_       TATTAGATGAGCCTATGTTAGATTAGCTA  GTTGGTAGGGTAAAGGCCTACCA  AGGCGATG 180
16S-19_V_      TATTAGATGAGCCTATGTTAGATTAGCTA  GTTGGTAGGGTAAAGGCCTACCA  AGGCGATG 180
16S-29         TATTAGATGAGCCTATGTTAGATTAGCTA  GTTGGTAGGGTAAAGGCCTACCA  AGGCGATG 180
16S-20_W_      TATTAGATGAGCCTATGTTAGATTAGCTA  GTTGGTAGGGTAAAGGCCTACCA  AGGCGATG 180
16S-30         TATTAGATGAGCCTATGTTAGATTAGCTA  GTTGGTAGGGTAAAGGCCTACCA  AGGCGATG 180
16S-11_M_      TATTAGATGAGCCTATGTTAGATTAGCTA  GTTGGTAGGGTAAAGGCCTACCA  AGGCGATG 180
16S-25         TATTAGATGAGCCTATGTTAGATTAGCTA  GTTGGTAGGGTAAAGGCCTACCA  AGGCGATG 180
16S-28         TATTAGATGAGCCTATGTTAGATTAGCTA  GTTGGTAGGGTAAAGGCCTACCA  AGGCGATG 180
16S-16_S_      TATTAGATGAGCCTATGTTAGATTAGCTA  GTTGGTAGGGTAAAGGCCTACCA  AGGCGATG 180
16S-26         TATTAGATGAGCCTATGTTAGATTAGCTG  GTTGGTAGGGTAAAGGCCTACCA  AGGCGATG 180
16S-3_D_       TATTAGATGAGCCTATGTTAGATTAGCTG  GTTGGTAGGGTAAAGGCCTACCA  AGGCGATG 180
16S-9_K_       TATTAGATGAGCCTATGTTAGATTAGCTG  GTTGGTAGGGTAAAGGCCTACCA  AGGCGATG 180
16S-8_J_       TATTAGATGAGCCTATGTTAGATTAGCTA  GTTGGTAGGGTAAAGGCCTACCA  AGGCGATG 180
16S-7_I_       TATTAGATGAGCCTATGTTAGATTAGCTA  GTTGGTAGGGTAAAGGCCTACCA  AGGCGATG 180
16S-12_N_      TATTAGATGAGCCTATGTTAGATTAGCTA  GTTGGTAGGGTAAAGGCCTACCA  AGGCGATG 180
16S-21_X_      TATTAGATGAGCCTATGTTAGATTAGCTA  GTTGGTAGGGTAAAGGCCTACCA  AGGCGATG 180
16S-13_O_      TATTAGATGAGCCTATGTTAGATTAGCTA  GTTGGTAGGGTAAAGGCCTACCA  AGGCGATG 180
16S-22_Y_      TATTAGATGAGCCTATGTTAGATTAGCTA  GTTGGTAGGGTAAAGGCCTACCA  AGGCGATG 180
16S-17_T_      TATTAGATGAGCCTATGTTAGATTAGCTA  GTTGGTAGGGTAAAGGCCTACCA  AGGCGATG 180
Position       ****************************149**********************172********

16S-14_P_    ATCTATAGCTGGTCTGAGAGGATGA   T  CAGCCACACTGGAACTGAGATACGGTC  CAGACTC 240
16S-23_Z_    ATCTATAGCTGGTCTGAGAGGATGA   T  CAGCCACACTGGAACTGAGATACGGTC  CAGACTC 240
16S-18_U_    ATCTATAGCTGGTCTGAGAGGATGA   T  CAGCCACACTGGAACTGAGATACGGTC  CAGACTC 240
16S-10_L_    ATCTATAGCTGGTCTGAGAGGATGG   T  CAGCCACACTGGAACTGAGATACGGTC  CAGACTC 240
16S-5_G_     ATCTATAGCTGGTCTGAGAGGATGA   T  CAGCCACACTGGAACTGAGATACGGTC  CAGACTC 240
```

```
16S-1_A_     ATCTATAGCTGGTCTGAGAGGATGA   T  CAGCCACACTGGAACTGAGATACGGTC  CAGACTC 240
16S-27       ATCTATAGCTGGTCTGAGAGGATGA   T  CAGCCACACTGGAACTGAGATACGGTC  CAGACTC 240
16S-24       ATCTATAGCTGGTCTGAGAGGATGA   T  CAGCCACACTGGAACTGAGATACGGTC  CAGACTC 240
16S-15_R_    ATCTATAGCTGGTCTGAGAGGATGA   T  CAGCCACACTGGAACTGAGATACGGTC  CAGACTC 240
16S-4_E_     ATCTATAGCTGGTCTGAGAGGATGA   C  CAGCCACACTGGAACTGAGATACGGTC  CAGACTC 240
16S-2_B_     ATCTATAGCTGGTCTGAGAGGATGA   T  CAGCCACACTGGAACTGAGATACGGTC  CAGACTC 240
16S-6_H_     ATCTATAGCTGGTCTGAGAGGATGA   T  CAGCCACACTGGAACTGAGATACGGTC  CAGACTC 240
16S-19_V_    ATCTATAGCTGGTCTGAGAGGATGA   T  CAGCCACACTGGAACTGAGATACGGTC  CAGACTC 240
16S-29       ATCTATAGCTGGTCTGAGAGGATGA   T  CAGCCACACTGGAACTGAGATACGGTC  CAGACTC 240
16S-20_W_    ATCTATAGCTGGTCTGAGAGGATGA   T  CAGCCACACTGGAACTGAGATACGGTC  CAGACTC 240
16S-30       ATCTATAGCTGGTCTGAGAGGATGA   T  CAGCCACACTGGAACTGAGATACGGTC  CAGACTC 240
16S-11_M_    ATCTATAGCTGGTCTGAGAGGATGA   T  CAGCCACACTGGAACTGAGATACGGTC  CAGACTC 240
16S-25       ATCTATAGCTGGTCTGAGAGGATGA   T  CAGCCACACTGGAACTGAGATACGGTC  CAGACTC 240
16S-28       ATCTATAGCTGGTCTGAGAGGATGA   T  CAGCCACACTGGAACTGAGATACGGTT  CAGACTC 240
16S-16_S_    ATCTATAGCTGGTCTGAGAGGATGA   T  CAGCCACACTGGAACTGAGATACGGTC  CAGACTC 240
16S-26       ATCTATAGCTGGTCTGAGAGGATGA   T  CAGCCACACTGGAACTGAGATACGGTC  CAGACTC 240
16S-3_D_     ATCTATAGCTGGTCTGAGAGGATGA   T  CAGCCACACTGGAACTGAGATACGGTC  CAGACTC 240
16S-9_K_     ATCTATAGCTGGTCTGAGAGGATGA   T  CAGCCACACTGGAACTGAGATACGGTC  CAGACTC 240
16S-8_J_     ATCTATAGCTGGTCTGAGAGGATGA   T  CAGCCACACTGGAACTGAGATACGGTC  CAGACTC 240
16S-7_I_     ATCTATAGCTGGTCTGAGAGGATGA   T  CAGCCACACTGGAACTGAGATACGGTC  CAGACTC 240
16S-12_N_    ATCTATAGCTGGTCTGAGAGGATGA   T  CAGCCACACTGGAACTGAGATACGGTC  CAGACTC 240
16S-21_X_    ATCTATAGCTGGTCTGAGAGGATGA   T  CAGCCACACTGGAACTGAGATACGGTC  CAGACTC 240
16S-13_O_    ATCTATAGCTGGTCTGAGAGGATGA   T  CAGCCACACTGGAACTGAGATACGGTC  CAGACTC 240
16S-22_Y_    ATCTATAGCTGGTCTGAGAGGATGA   T  CAGCCACACTGGAACTGAGATACGGTC  CAGACTC 240
16S-17_T_    ATCTATAGCTGGTCTGAGAGGATGA   T  CAGCCACACTGGAACTGAGATACGGTC  CAGACTC 240
Position     ************************205 206***************************225*******

16S-14_P_          CTACGGGAGGCA  GCAGTGGGGAATATTGGACAATGGGCGCAAGCCTGATCCAGCTATGCC 300
16S-23_Z_          CTACGGGAGGCA  GCAGTGGGGAATATTGGACAATGGGCGCAAGCCTGATCCAGCTATGCC 300
16S-18_U_          CTACGGGAGGCA  GCAGTGGGGAATATTGGACAATGGGCGCAAGCCTGATCCAGCTATGCC 300
16S-10_L_          CTACGGGAGGCA  GCAGTGGGGAATATTGGACAATGGGCGCAAGCCTGATCCAGCTATGCC 300
16S-5_G_           CTACGGGAGGCA  GCAGTGGGGAATATTGGACAATGGGCGCAAGCCTGATCCAGCTATGCC 300
16S-1_A_           CTACGGGAGGCA  GCAGTGGGGAATATTGGACAATGGGCGCAAGCCTGATCCAGCTATGCC 300
16S-27             CTACGGGAGGCA  GCAGTGGGGAATATTGGACAATGGGCGCAAGCCTGATCCAGCTATGCC 300
16S-24             CTACGGGAGGCA  GCAGTGGGGAATATTGGACAATGGGCGCAAGCCTGATCCAGCTATGCC 300
16S-15_R_          CTACGGGAGGCA  GCAGTGGGGAATATTGGACAATGGGCGCAAGCCTGATCCAGCTATGCC 300
16S-4_E_           CTACGGGAGGCA  GCAGTGGGGAATATTGGACAATGGGCGCAAGCCTGATCCAGCTATGCC 300
16S-2_B_           CTACGGGAGGCA  GCAGTGGGGAATATTGGACAATGGGCGCAAGCCTGATCCAGCTATGCC 300
16S-6_H_           CTACGGGAGGCG  GCAGTGGGGAATATTGGACAATGGGCGCAAGCCTGATCCAGCTATGCC 300
16S-19_V_          CTACGGGAGGCA  GCAGTGGGGAATATTGGACAATGGGCGCAAGCCTGATCCAGCTATGCC 300
16S-29             CTACGGGAGGCA  GCAGTGGGGAATATTGGACAATGGGCGCAAGCCTGATCCAGCTATGCC 300
16S-20_W_          CTACGGGAGGCA  GCAGTGGGGAATATTGGACAATGGGCGCAAGCCTGATCCAGCTATGCC 300
16S-30             CTACGGGAGGCA  GCAGTGGGGAATATTGGACAATGGGCGCAAGCCTGATCCAGCTATGCC 300
16S-11_M_          CTACGGGAGGCA  GCAGTGGGGAATATTGGACAATGGGCGCAAGCCTGATCCAGCTATGCC 300
16S-25             CTACGGGAGGCA  GCAGTGGGGAATATTGGACAATGGGCGCAAGCCTGATCCAGCTATGCC 300
16S-28             CTACGGGAGGCA  GCAGTGGGGAATATTGGACAATGGGCGCAAGCCTGATCCAGCTATGCC 300
16S-16_S_          CTACGGGAGGCA  GCAGTGGGGAATATTGGACAATGGGCGCAAGCCTGATCCAGCTATGCC 300
16S-26             CTACGGGAGGCA  GCAGTGGGGAATATTGGACAATGGGCGCAAGCCTGATCCAGCTATGCC 300
16S-3_D_           CTACGGGAGGCA  GCAGTGGGGAATATTGGACAATGGGCGCAAGCCTGATCCAGCTATGCC 300
16S-9_K_           CTACGGGAGGCA  GCAGTGGGGAATATTGGACAATGGGCGCAAGCCTGATCCAGCTATGCC 300
16S-8_J_           CTACGGGAGGCA  GCAGTGGGGAATATTGGACAATGGGCGCAAGCCTGATCCAGCTATGCC 300
16S-7_I_           CTACGGGAGGCA  GCAGTGGGGAATATTGGACAATGGGCGCAAGCCTGATCCAGCTATGCC 300
16S-12_N_          CTACGGGAGGCA  GCAGTGGGGAATATTGGACAATGGGCGCAAGCCTGATCCAGCTATGCC 300
16S-21_X_          CTACGGGAGGCA  GCAGTGGGGAATATTGGACAATGGGCGCAAGCCTGATCCAGCTATGCC 300
16S-13_O_          CTACGGGAGGCA  GCAGTGGGGAATATTGGACAATGGGCGCAAGCCTGATCCAGCTATGCC 300
16S-22_Y_          CTACGGGAGGCA  GCAGTGGGGAATATTGGACAATGGGCGCAAGCCTGATCCAGCTATGCC 300
16S-17_T_          CTACGGGAGGCA  GCAGTGGGGAATATTGGACAATGGGCGCAAGCCTGATCCAGCTATGCC 300
Position           ***********233***************************************************

16S-14_P_  GCA  TGAGTGAGGAAGGCCTTAGG  GTTGTAAAA  CTCTTTCAGTAGGGAAGATAATGA  CGGT 360
16S-23_Z_  GCA  TGAGTGAGGAAGGCCTTAGG  GTTGTAAAA  CTCTTTCAGTAGGGAAGATAATGA  CGGT 360
16S-18_U_  GCA  TGAGTGAGGAAGGCCTTAGG  GTTGTAAAA  CTCTTTCAGTAGGGAAGATAATGA  CGGT 360
16S-10_L_  GCA  TGAGTGAGGAAGGCCTTAGG  GTTGTAAAA  CTCTTTCAGTAGGGAAGATAATGA  CGGT 360
16S-5_G_   GCA  TGAGTGAGGAAGGCCTTAGG  GTTGTAAAA  CTCTTTCAGTAGGGAAGATAATGA  CGGT 360
16S-1_A_   GCA  TGAGTGAGGAAGGCCTTAGG  GTTGTAAAA  CTCTTTCAGTAGGGAAGATAATGA  CGGT 360
16S-27     GCG  TGAGTGAGGAAGGCCTTAGG  GTTGTAAAA  CTCTTTCAGTAGGGAAGATAATGA  CGGT 360
16S-24     GCG  TGAGTGAGGAAGGCCTTAGG  GTTGTAAAA  CTCTTTCAGTAGGGAAGATAATGA  CGGT 360
16S-15_R_  GCG  TGAGTGAGGAAGGCCTTAGG  GTTGTAAAA  CTCTTTCAGTAGGGAAGATAATGA  CGGT 360
16S-4_E_   GCG  TGAGTGAGGAAGGCCTTAGG  GTTGTAAAA  CTCTTTCAGTAGGGAAGATAATGA  CGGT 360
16S-2_B_   GCG  TGAGTGAGGAAGGCCTTAGG  GTTGTAAAA  CTCTTTCAGTAGGGAAGATAATGA  CGGT 360
16S-6_H_   GCG  TGAGTGAGGAAGGCCTTAGG  GTTGTAAAA  CTCTTTCAGTAGGGAAGATAATGA  CGGT 360
16S-19_V_  GCG  TGAGTGAGGAAGGCCTTAGG  GTTGTAAAA  CTCTTTCAGTAGGGAAGATAATGA  CGGT 360
16S-29     GCG  TGAGTGAGGAAGGCCTTAGA  GTTGTAAAA  CTCTTTCAGTAGGGAAGATAATGA  CGGT 360
16S-20_W_  GCG  TGAGTGAGGAAGGCCTTAGG  GTTGTAAAA  CTCTTTCAGTAGGGAAGATAATGA  CGGT 360
16S-30     GCG  TGAGTGAGGAAGGCCTTAGG  GTTGTAAAA  CTCTTTCAGTAGGGAAGATAATGA  CGGT 360
16S-11_M_  GCG  TGAGTGAGGAAGGCCTTAGG  GTTGTAAAA  CTCTTTCAGTAGGGAAGATAATGG  CGGT 360
16S-25     GCG  TGAGTGAGGAAGGCCTTAGG  GTTGTAAAA  CTCTTTCAGTAGGGAAGATAATGA  CGGT 360
16S-28     GCG  TGAGTGAGGAAGGCCTTAGG  GTTGTAAAA  CTCTTTCAGTAGGGAAGATAATGA  CGGT 360
16S-16_S_  GCG  TGAGTGAGGAAGGCCTTAGG  GTTGTAAAA  CTCTTTCAGTAGGGAAGATAATGA  CGGT 360
16S-26     GCG  TGAGTGAGGAAGGCCTTAGG  GTTGTAAAA  CTCTTTCAGTAGGGAAGATAATGA  CGGT 360
16S-3_D_   GCG  TGAGTGAGGAAGGCCTTAGG  GTTGTAAAA  CTCTTTCAGTAGGGAAGATAATGA  CGGT 360
16S-9_K_   GCG  TGAGTGAGGAAGGCCTTAGG  GTTGTAAAA  CTCTTTCAGTAGGGAAGATAATGA  CGGT 360
16S-8_J_   GCG  TGAGTGAGGAAGGCCTTAGG  GTTGTAAAA  CTCTTTCAGTAGGGAAGATAATGA  CGGT 360
16S-7_I_   GCG  TGAGTGAGGAAGGCCTTAGG  GTTGTAAAA  CTCTTTCAGTAGGGAAGATAATGA  CGGT 360
16S-12_N_  GCG  TGAGTGAGGAAGGCCTTAGG  GTTGTAAAG  CTCTTTCAGTAGGGAAGATAATGA  CGGT 360
16S-21_X_  GCG  TGAGTGAGGAAGGCCTTAGG  GTTGTAAAA  CTCTTTCAGTAGGGAAGATAATGA  CGGT 360
16S-13_O_  GCG  TGAGTGAGGAAGGCCTTAGG  GTTGTAAAA  CTCTTTCAGTAGGGAAGATAATGA  CGGT 360
16S-22_Y_  GCG  TGAGTGAGGAAGGCCTTAGG  GTTGTAAAA  CTCTTTCAGTAGGGAAGATAATGA  CGGT 360
16S-17_T_  GCG  TGAGTGAGGAAGGCCTTAGG  GTTGTAAAA  CTCTTTCAGTAGGGAAGATAATGA  CGGT 360
Position   **303******************323********332************************356****
```

```
16S-14_P_       ACCTACAGAAGAAGTCCCGGCAAACTCCGTGCCAGCAGCCGCGGTAATACGGAGGGGGCA  420
16S-23_Z_       ACCTACAGAAGAAGTCCCGGCAAACTCCGTGCCAGCAGCCGCGGTAATACGGAGGGGGCA  420
16S-18_U_       ACCTACAGAAGAAGTCCCGGCAAACTCCGTGCCAGCAGCCGCGGTAATACGGAGGGGGCA  420
16S-10_L_       ACCTACAGAAGAAGTCCCGGCAAACTCCGTGCCAGCAGCCGCGGTAATACGGAGGGGGCA  420
16S-5_G_        ACCTACAGAAGAAGTCCCGGCAAACTCCGTGCCAGCAGCCGCGGTAATACGGAGGGGGCA  420
16S-1_A_        ACCTACAGAAGAAGTCCCGGCAAACTCCGTGCCAGCAGCCGCGGTAATACGGAGGGGGCA  420
16S-27          ACCTACAGAAGAAGTCCCGGCAAACTCCGTGCCAGCAGCCGCGGTAATACGGAGGGGGCA  420
16S-24          ACCTACAGAAGAAGTCCCGGCAAACTCCGTGCCAGCAGCCGCGGTAATACGGAGGGGGCA  420
16S-15_R_       ACCTACAGAAGAAGTCCCGGCAAACTCCGTGCCAGCAGCCGCGGTAATACGGAGGGGGCG  420
16S-4_E_        ACCTACAGAAGAAGTCCCGGCAAACTCCGTGCCAGCAGCCGCGGTAATACGGAGGGGGCA  420
16S-2_B_        ACCTACAGAAGAAGTCCCGGCAAACTCCGTGCCAGCAGCCGCGGTAATACGGAGGGGGCA  420
16S-6_H_        ACCTACAGAAGAAGTCCCGGCAAACTCCGTGCCAGCAGCCGCGGTAATACGGAGGGGGCA  420
16S-19_V_       ACCTACAGAAGAAGTCCCGGCAAACTCCGTGCCAGCAGCCGCGGTAATACGGAGGGGGCA  420
16S-29          ACCTACAGAAGAAGTCCCGGCAAACTCCGTGCCAGCAGCCGCGGTAATACGGAGGGGGCA  420
16S-20_W_       ACCTACAGAAGAAGTCCCGGCAAACTCCGTGCCAGCAGCCGCGGTAATACGGAGGGGGCA  420
16S-30          ACCTACAGAAGAAGTCCCGGCAAACTCCGTGCCAGCAGCCGCGGTAATACGGAGGGGGCA  420
16S-11_M_       ACCTACAGAAGAAGTCCCGGCAAACTCCGTGCCAGCAGCCGCGGTAATACGGAGGGGGCA  420
16S-25          ACCTACAGAAGAAGTCCCGGCAAACTCCGTGCCAGCAGCCGCGGTAATACGGAGGGGGCA  420
16S-28          ACCTACAGAAGAAGTCCCGGCAAACTCCGTGCCAGCAGCCGCGGTAATACGGAGGGGGCA  420
16S-16_S_       ACCTACAGAAGAAGTCCCGGCAAACTCCGTGCCAGCAGCCGCGGTAATACGGAGGGGGCA  420
16S-26          ACCTACAGAAGAAGTCCCGGCAAACTCCGTGCCAGCAGCCGCGGTAATACGGAGGGGGCA  420
16S-3_D_        ACCTACAGAAGAAGTCCCGGCAAACTCCGTGCCAGCAGCCGCGGTAATACGGAGGGGGCA  420
16S-9_K_        ACCTACAGAAGAAGTCCCGGCAAACTCCGTGCCAGCAGCCGCGGTAATACGGAGGGGGCA  420
16S-8_J_        ACCTACAGAAGAAGTCCCGGCAAACTCCGTGCCAGCAGCCGCGGTAATACGGAGGGGGCA  420
16S-7_I_        ACCTACAGAAGAAGTCCCGGCAAACTCCGTGCCAGCAGCCGCGGTAATACGGAGGGGGCA  420
16S-12_N_       ACCTACAGAAGAAGTCCCGGCAAACTCCGTGCCAGCAGCCGCGGTAATACGGAGGGGGCA  420
16S-21_X_       ACCTACAGAAGAAGTCCCGGCAAACTCCGTGCCAGCAGCCGCGGTAATACGGAGGGGGCA  420
16S-13_O_       ACCTACAGAAGAAGTCCCGGCAAACTCCGTGCCAGCAGCCGCGGTAATACGGAGGGGGCA  420
16S-22_Y_       ACCTACAGAAGAAGTCCCGGCAAACTCCGTGCCAGCAGCCGCGGTAATACGGAGGGGGCA  420
16S-17_T_       ACCTACAGAAGAAGTCCCGGCAAACTCCGTGCCAGCAGCCGCGGTAATACGGAGGGGGCA  420
Position        ************************************************************420

16S-14_P_       AGCGTTGTTCGGAATTATTGGGCGTAAAGGGCATGTAGGCGGTTCGGTAAGTT  AAAGGTG 480
16S-23_Z_       AGCGTTGTTCGGAATTATTGGGCGTAAAGGGCATGTAGGCGGTTCGGTAAGTT  AAAGGTG 480
16S-18_U_       AGCGTTGTTCGGAATTATTGGGCGTAAAGGGCATGTAGGCGGTTCGGTAAGTT  AAAGGTG 480
16S-10_L_       AGCGTTGTTCGGAATTATTGGGCGTAAAGGGCATGTAGGCGGTTCGGTAAGTT  AAAGGTG 480
16S-5_G_        AGCGTTGTTCGGAATTATTGGGCGTAAAGGGCATGTAGGCGGTTCGGTAAGTT  AAAGGTG 480
16S-1_A_        AGCGTTGTTCGGAATTATTGGGCGTAAAGGGCATGTAGGCGGTTCGGTAAGTT  AAAGGTG 480
16S-27          AGCGTTGTTCGGAATTATTGGGCGTAAAGGGCATGTAGGCGGTTCGGTAAGTT  AAAGGTG 480
16S-24          AGCGTTGTTCGGAATTATTGGGCGTAAAGGGCATGTAGGCGGTTCGGTAAGTT  AAAGGTG 480
16S-15_R_       AGCGTTGTTCGGAATTATTGGGCGTAAAGGGCATGTAGGCGGTTCGGTAAGTT  AAAGGTG 480
16S-4_E_        AGCGTTGTTCGGAATTATTGGGCGTAAAGGGCATGTAGGCGGTTCGGTAAGTT  AAAGGTG 480
16S-2_B_        AGCGTTGTTCGGAATTATTGGGCGTAAAGGGCATGTAGGCGGTTCGGTAAGTT  AAAGGTG 480
16S-6_H_        AGCGTTGTTCGGAATTATTGGGCGTAAAGGGCATGTAGGCGGTTCGGTAAGTT  AAAGGTG 480
16S-19_V_       AGCGTTGTTCGGAATTATTGGGCGTAAAGGGCATGTAGGCGGTTCGGTAAGTT  AAAGGTG 480
16S-29          AGCGTTGTTCGGAATTATTGGGCGTAAAGGGCATGTAGGCGGTTCGGTAAGTT  AAAGGTG 480
16S-20_W_       AGCGTTGTTCGGAATTATTGGGCGTAAAGGGCATGTAGGCGGTTCGGTAAGTT  AAAGGTG 480
16S-30          AGCGTTGTTCGGAATTATTGGGCGTAAAGGGCATGTAGGCGGTTCGGTAAGTT  AAAGGTG 480
16S-11_M_       AGCGTTGTTCGGAATTATTGGGCGTAAAGGGCATGTAGGCGGTTCGGTAAGTT  AAAGGTG 480
16S-25          AGCGTTGTTCGGAATTATTGGGCGTAAAGGGCATGTAGGCGGTTCGGTAAGTA  AAAGGTG 480
16S-28          AGCGTTGTTCGGAATTATTGGGCGTAAAGGGCATGTAGGCGGTTCGGTAAGTT  AAAGGTG 480
16S-16_S_       AGCGTTGTTCGGAATTATTGGGCGTAAAGGGCATGTAGGCGGTTCGGTAAGTT  AAAGGTG 480
16S-26          AGCGTTGTTCGGAATTATTGGGCGTAAAGGGCATGTAGGCGGTTCGGTAAGTT  AAAGGTG 480
16S-3_D_        AGCGTTGTTCGGAATTATTGGGCGTAAAGGGCATGTAGGCGGTTCGGTAAGTT  AAAGGTG 480
16S-9_K_        AGCGTTGTTCGGAATTATTGGGCGTAAAGGGCATGTAGGCGGTTCGGTAAGTT  AAAGGTG 480
16S-8_J_        AGCGTTGTTCGGAATTATTGGGCGTAAAGGGCATGTAGGCGGTTCGGTAAGTT  AAAGGTG 480
16S-7_I_        AGCGTTGTTCGGAATTATTGGGCGTAAAGGGCATGTAGGCGGTTCGGTAAGTT  AAAGGTG 480
16S-12_N_       AGCGTTGTTCGGAATTATTGGGCGTAAAGGGCATGTAGGCGGTTCGGTAAGTT  AAAGGTG 480
16S-21_X_       AGCGTTGTTCGGAATTATTGGGCGTAAAGGGCATGTAGGCGGTTCGGTAAGTT  AAAGGTG 480
16S-13_O_       AGCGTTGTTCGGAATTATTGGGCGTAAAGGGCATGTAGGCGGTTCGGTAAGTT  AAAGGTG 480
16S-22_Y_       AGCGTTGTTCGGAATTATTGGGCGTAAAGGGCATGTAGGCGGTTCGGTAAGTT  AAAGGTG 480
16S-17_T_       AGCGTTGTTCGGAATTATTGGGCGTAAAGGGCATGTAGGCGGTTCGGTAAGTT  AAAGGTG 480
Position        ********************************************************473*******

16S-14_P_       AAATGCCAGG  GCTTAAC 497
16S-23_Z_       AAATGCCAGG  GCTTAAC 497
16S-18_U_       AAATGCCAGA  GCTTAAC 497
16S-10_L_       AAATGCCAGG  GCTTAAC 497
16S-5_G_        AAATGCCAGG  GCTTAAC 497
16S-1_A_        AAATGCCAGG  GCTTAAC 497
16S-27          AAATGCCAGG  GCTTAAC 497
16S-24          AAATGCCAGG  GCTTAAC 497
16S-15_R_       AAATGCCAGG  GCTTAAC 497
16S-4_E_        AAATGCCAGG  GCTTAAC 497
16S-2_B_        AAATGCCAGG  GCTTAAC 497
16S-6_H_        AAATGCCAGG  GCTTAAC 497
16S-19_V_       AAATGCCAGG  GCTTAAC 497
16S-29          AAATGCCAGG  GCTTAAC 497
16S-20_W_       AAATGCCAGG  GCTTAAC 497
16S-30          AAATGCCAGG  GCTTAAC 497
16S-11_M_       AAATGCCAGG  GCTTAAC 497
16S-25          AAATGCCAGG  GCTTAAC 497
16S-28          AAATGCCAGG  GCTTAAC 497
16S-16_S_       AAATGCCAGG  GCTTAAC 497
16S-26          AAATGCCAGG  GCTTAAC 497
16S-3_D_        AAATGCCAGG  GCTTAAC 497
16S-9_K_        AAATGCCAGG  GCTTAAC 497
16S-8_J_        AAATGCCAGG  GCTTAAC 497
16S-7_I_        AAATGCCAGG  GCTTAAC 497
16S-12_N_       AAATGCCAGG  GCTTAAC 497
16S-21_X_       AAATGCCAGG  GCTTAAC 497
```

```
16S-13_O_          AAATGCCAGG  GCTTAAC 497
16S-22_Y_          AAATGCCAGG  GCTTAAC 497
16S-17_T_          AAATGCCAGG  GCTTAAC 497
Position           *********490*******
```

Table 33: Assigned accession numbers of the new *16S rRNA* sequences

Accession number	*16S rRNA* variant	Sample	Animal species
KU705116	16S-13(O)	971	cat
KU705117	16S-16(S)	Mufflon12	mouflon
KU705118	16S-2(B)	B17	mouflon
KU705119	16S-20(W)	G04	mouflon
KU705120	16S-16(S)	H24	mouflon
KU705121	16S-20(W)	Rd21	red deer
KU705122	16S-2(B)	Rd22	red deer
KU705123	16S-28	Rd23	red deer
KU705124	16S-16(S)	BwRo1.1	red deer
KU705125	16S-20(W)	BwRo2	red deer
KU705126	16S-30	BwRo3	red deer
KU705127	16S-20(W)	BwRo4	red deer
KU705128	16S-16(S)	BwRo5	red deer
KU705129	16S-20(W)	BwRo6	red deer
KU705130	16S-25	BwRo9	red deer
KU705131	16S-16(S)	BwRo10	red deer
KU705132	16S-20(W)	BwRo11	red deer
KU705133	16S-2(B)	BwRo13	red deer
KU705134	16S-16(S)	BwRo14	red deer
KU705135	16S-20(W)	BwRo15	red deer
KU705136	16S-2(B)	BwRo16	red deer
KU705137	16S-20(W)	BwRo18	red deer
KU705138	16S-16(S)	BwRo1.2	red deer
KU705139	16S-26	BwRo21	red deer
KU705140	16S-7(I)_	BwReh1	roe deer
KU705141	16S-7(I)	BwReh6	roe deer
KU705142	16S-21(X)	BWReh3	roe deer
KU705143	16S-22(Y)	BwReh7	roe deer
KU705144	16S-22(Y)	BwReh8	roe deer
KU705145	16S-20(W)	BwReh9	roe deer
KU705146	16S-21(X)	BwReh10	roe deer
KU705147	16S-24	BwReh11	roe deer
KU705148	16S-22(Y)	BwReh12	roe deer
KU705149	16S-21(X)	BwReh13	roe deer
KU705150	16S-21(X)	BwReh14	roe deer
KU705151	16S-22(Y)	BwReh15	roe deer
KU705152	16S-7(I)	BwReh16	roe deer

KU705153	16S-19(V)	BwReh18	roe deer
KU705154	16S-7(I)	BwReh21	roe deer
KU705155	16S-21(X)	BwReh22	roe deer
KU705156	16S-20(W)_	BwReh23	roe deer
KU705157	16S-22(Y)	BwReh24	roe deer
KU705158	16S-7(I)	BwReh31	roe deer
KU705159	16S-21(X)	BwReh32	roe deer
KU705160	16S-7(I)	BwReh33	roe deer
KU705161	16S-22(Y)	A04	roe deer
KU705162	16S-22(Y)	C34	roe deer
KU705163	16S-22(Y)	C35	roe deer
KU705164	16S-22(Y)	G14	roe deer
KU705165	16S-21(X)	K34	roe deer
KU705166	16S-20(W)	K41	roe deer
KU705167	16S-20(X)	K91	roe deer
KU705168	16S-22(Y)	L12	roe deer
KU705169	16S-22(Y)	L45	roe deer
KU705170	16S-22(Y)	N13	roe deer
KU705171	16S-22(Y)	N49	roe deer
KU705172	16S-21(X)	O08	roe deer
KU705173	16S-22(Y)	O11	roe deer
KU705174	16S-22(Y)	P10	roe deer
KU705175	16S-21(X)	P14	roe deer
KU705176	16S-20(W)	BwSi1	sika deer
KU705177	16S-20(W)	BwSi2	sika deer
KU705178	16S-16(S)	BwSi3	sika deer
KU705179	16S-20(W)	BwSi4	sika deer
KU705180	16S-20(B)	BwSi5	sika deer
KU705181	16S-2(B)	BwSi6	sika deer
KU705182	16S-16(S)	BwSi7	sika deer
KU705183	16S-29	BwSi11	sika deer
KU705184	16S-7(I)	BwSi11.2	sika deer
KU705185	16S-20(W)	BwSi12	sika deer
KU705186	16S-16(S)	BwSi13	sika deer
KU705187	16S-8(J)	BwSi14	sika deer
KU705188	16S-21(X)	BwSi17	sika deer
KU705189	16S-16(S)	BwSi19	sika deer
KU705190	16S-2(B)	S109	red fox
KU705191	16S-1(A)	S135	red fox
KU705192	16S-1(A)	S133	red fox
KU705193	16S-1(A)	S146	red fox
KU705194	16S-1(A)	S152	red fox
KU705195	16S-27	S220	red fox
KU705196	16S-2(B)	S226	red fox

KU705197	16S-21(X)	BwDa1	fallow deer
KU705198	16S-20(W)	BwDa3	fallow deer
KU705199	16S-19(V)	BwDa8	fallow deer
KU705200	16S-1(A)	69	dog
KU705201	16S-1(A)	413	dog
KU705202	16S-1(A)	438	dog
KU705203	16S-2(B)	Probe E	dog

1.1.2. *groEL* gene

Table 34: Distribution of the *groEL* variants in the different examined animal species

	Dog	Horse	Cat	Hedgehog	Red fox	Cattle	Goat	Roe deer	Red deer	Sika deer	Fallow deer	Mouflon	Chamois	Ibex
g-1(A)	13	1		33	2									
g-2(B)	20	11	1		1									
g-3(C)						9								
g-4(D)							1	1						
g-5(E)							1	1						
g-6(F)								5	2					2
g-7(G)								10	2					
g-8(H)									1		2	3	1	
g-9(I)													1	
g-10(J)													1	
g-12(K)								1						
g-13(L)		2							1	1		2		
g-14(M)									1					
g-15(N)						3								
g-16(O)						1							1	
g-18(X)						12								
g-19(Y)						1								
g-20								2						
g-21												2		
g-22												1		
g-23												1		
g-24								1		1	1			
g-25								1						
g-26								1						
g-27													1	
g-28										1				
g-29											1			
g-30								1						

g-31								1						
g-32									1					
g-33									1					
g-34										1		1		
g-35										1				
Total	**33**	**14**	**1**	**33**	**3**	**26**	**2**	**25**	**9**	**5**	**4**	**10**	**5**	**2**

Table 35: Alignment of the *groEL* nucleotide sequences – highlighted nucleotide positions are differing within the variants

```
CLUSTAL 2.1 multiple sequence alignment


g-1_A_          GGATATCTTTCGCCTTACTTTGTTACAAA TGCTGAAAAAATGCTGGTGGAATTTGAAAAT 60
g-2_B_          GGATATCTTTCGCCTTACTTTGTTACAAA TGCTGAAAAAATGCTGGTGGAATTTGAAAAT 60
g-3_C_          GGATATCTTTCGCCTTACTTTGTTACAAA TGCTGAAAAAATGCTGGTGGAATTTGAAAAT 60
g-14_M_         GGATATCTTTCGCCTTACTTTGTTACAAA TGCTGAAAAAATGCTGGTGGAATTTGAAAAT 60
g-18_X_         GGATATCTTTCGCCTTACTTTGTTACAAA TGCTGAAAAAATGCTGGTGGAATTTGAAAAT 60
g-19_Y_         GGATATCTTTCGCCTTACTTTGTTACAAA TGCTGAAAAAATGCTGGTGGAATTTGAAAAT 60
g-25            GGATATCTTTCGCCTTACTTTGTTACAAA TGCTGAAAAAATGCTGGTGGAATTTGAAAAT 60
g-28            GGATATCTTTCGCCTTACTTTGTTACAAA TGCTGAAAAAATGCTGGTGGAATTTGAAAAT 60
g-17_P_         GGATATCTTTCGCCTTACTTTGTTACAAA TGCTGAAAAAATGCTGGTGGAATTTGAAAAT 60
g-32            GGATATCTTTCGCCTTACTTTGTTACAAA TGCTGAAAAAATGCTGGTGGAATTTGAAAAT 60
g-13_L_         GGATATCTTTCGCCTTACTTTGTTACAAA TGCTGAAAAAATGCTGGTGGAATTTGAAAAT 60
g-11_K_         GGATATCTTTCGCCTTACTTTGTTACAAA TGCTGAAAAAATGCTGGTGGAATTTGAAAAT 60
g-34            GGATATCTTTCGCCTTACTTTGTTACAAA TGCTGAAAAAATGCTGGTGGAATTTGAAAAT 60
g-21            GGATATCTTTCGCCTTACTTTGTTACAAA TGCTGAAAAAATGCTGGTGGAATTTGAAAAT 60
g-27            GGATATCTTTCGCCTTACTTTGTTACAAA TGCTGAAAAAATGCTGGTGGAATTTGAAAAT 60
g-22            GGATATCTTTCGCCTTACTTTGTTACAAA TGCTGAAAAAATGCTGGTGGAATTTGAAAAT 60
g-8_H_          GGATATCTTTCGCCTTACTTTGTTACAAA TGCTGAAAAAATGCTGGTGGAATTTGAAAAT 60
g-15_N_         GGATATCTTTCGCCTTACTTTGTTACAAG TGCTGAAAAAATGCTGGTGGAATTTGAAAAT 60
g-24            GGATATCTTTCGCCTTACTTTGTTACAAA TGCTGAAAAAATGCTGGTGGAATTTGAAAAT 60
g-10_J_         GGATATCTTTCGCCTTACTTTGTTACAAA TGCTGAAAAAATGCTGGTGGAATTTGAAAAT 60
g-35            GGATATCTTTCGCCTTACTTTGTTACAAA TGCTGAAAAAATGCTGGTGGAATTTGAAAAT 60
g-23            GGATATCTTTCGCCTTACTTTGTTACAAA TGCTGAAAAAATGCTGGTGGAATTTGAAAAT 60
g-16_O_         GGATATCTTTCGCCTTACTTTGTTACAAA TGCTGAAAAAATGCTGGTGGAATTTGAAAAT 60
g-9_I_          GGATATCTTTCGCCTTACTTTGTTACAAA TGCTGAAAAAATGCTGGTGGAATTTGAAAAT 60
g-33            GGATATCTTTCGCCTTACTTTGTTACAAA TGCTGAAAAAATGCTGGTGGAATTTGAAAAT 60
g-29            GGATATCTTTCGCCTTACTTTGTTACAAA TGCTGAAAAAATGCTGGTGGAATTTGAAAAT 60
g-4_D_          GGATATCTTTCGCCTTACTTTGTTACAAA TGCTGAAAAAATGCTGGTGGAATTTGAAAAT 60
g-5_E_          GGATATCTTTCGCCTTACTTTGTTACAAA TGCTGAAAAAATGCTGGTGGAATTTGAAAAT 60
g-12_K_         GGATATCTTTCGCCTTACTTTGTTACAAA TGCTGAAAAAATGCTGGTGGAATTTGAAAAT 60
g-31            GGATATCTTTCGCCTTACTTTGTTACAAA TGCTGAAAAAATGCTGGTGGAATTTGAAAAT 60
g-20            GGATATCTTTCGCCTTACTTTGTTACAAA TGCTGAAAAAATGCTGGTGGAATTTGAAAAT 60
g-6_F_          GGATATCTTTCGCCTTACTTTGTTACAAA TGCTGAAAAAATGCTGGTGGAATTTGAAAAT 60
g-7_G_          GGATATCTTTCGCCTTACTTTGTTACAAA TGCTGAAAAAATGCTGGTGGAATTTGAAAAT 60
g-30            GGATATCTTTCGCCTTACTTTGTTACAAA TGCTGAAAAAATGCTGGTGGAATTTGAAAAT 60
g-26            GGATATCTTTCGCCTTACTTTGTTACAAA TGCTGAAAAAATGCTGGTGGAATTTGAAAAT 60
Position        ****************************29*******************************


g-1_A_      CCATACATATTC CTTACTGAAAAGAAGATTAA TCTTGTACAAAGCATTC  TACCAATC  TTA 120
g-2_B_      CCATACATATTT CTTACTGAAAAGAAGATTAA TCTTGTACAAAGCATTC  TACCAATC  TTA 120
g-3_C_      CCATACATATTC CTTACTGAAAAGAAGATTAA TCTTGTACAAAGCATTC  TACCAATC  TTA 120
g-14_M_     CCATACATATTC CTTACTGAAAAGAAGATTAA TCTTGTACAAAGCATTC  TACCAATC  TTA 120
g-18_X_     CCATACATATTC CTTACTGAAAAGAAGATTAA TCTTGTACAAAGCATTC  TACCAATC  TTA 120
g-19_Y_     CCATACATATTC CTTACTGAAAAGAAGATTAA TCTTGTACAAAGCATTC  TACCAATC  TTA 120
g-25        CCATACATATTC CTTACTGAAAAGAAGATTAA TCTTGTACAAAGCATTC  TACCAATC  TTA 120
g-28        CCATACATATTC CTTACTGAAAAGAAGATTAA TCTTGTACAAAGCATTC  TACCAATC  TTA 120
g-17_P_     CCATACATATTC CTTACTGAAAAGAAGATTAA TCTTGTACAAAGCATTC  TACCAATC  TTA 120
g-32        CCATACATATTC CTTACTGAAAAGAAGATTAA TCTTGTACAAAGCATTC  TACCAATC  TTA 120
g-13_L_     CCATACATATTC CTTACTGAAAAGAAGATTAA TCTTGTACAAAGCATTC  TACCAATC  TTA 120
g-11_K_     CCATACATATTT CTTACTGAAAAGAAGATTAA TCTTGTACAAAGCATTC  TACCAATC  TTA 120
g-34        CCATACATATTC CTTACTGAAAAGAAGATTAA TCTTGTACAAAGCATTC  TACCAATC  TTA 120
g-21        CCATACATATTC CTTACTGAAAAGAAGATTAA TCTTGTACAAAGCATTC  TACCAATC  TTA 120
g-27        CCATACATATTC CTTACTGAAAAGAAGATTAG TCTTGTACAAAGCATTC  TACCAATC  TTA 120
g-22        CCATACATATTC CTTACTGAAAAGAAGATTAA TCTTGTACAAAGCATTC  TACCAATC  TTA 120
g-8_H_      CCATACATATTC CTTACTGAAAAGAAGATTAA TCTTGTACAAAGCATTC  TACCAATC  TTA 120
g-15_N_     CCATACATATTC CTTACTGAAAAGAAGATTAA TCTTGTACAAAGCATTC  TACCAATC  TTA 120
g-24        CCATACATATTC CTTACTGAAAAGAAGATTAA TCTTGTACAAAGCATTC  TACCAATC  TTA 120
g-10_J_     CCATACATATTC CTTACTGAAAAGAAGATTAA TCTTGTACAAAGCATTC  TACCAATC  TTA 120
g-35        CCATACATATTC CTTACTGAAAAGAAGATTAA TCTTGTACAAAGCATTC  TACCAATC  TTA 120
g-23        CCATACATATTC CTTACTGAAAAGAAGATTAA TCTTGTACAAAGCATTC  TACCAATC  TTA 120
g-16_O_     CCATACATATTC CTTACTGAAAAGAAGATTAA TCTTGTACAAAGCATTC  TACCAATC  TTA 120
g-9_I_      CCATACATATTC CTTACTGAAAAGAAGATTAA TCTTGTACAAAGCATTC  TACCAATC  TTA 120
g-33        CCATACATATTC CTTACTGAAAAGAAGATTAA TCTTGTACAAAGCATTC  TACCAATC  TTA 120
g-29        CCATACATATTC CTTACTGAAAAGAAGATTAA TCTTGTACAAAGCATTT  TACCAATM  TTA 120
g-4_D_      CCATACATATTC CTTACTGAAAAGAAGATTAA TCTTGTACAAAGCATTT  TACCAATA  TTA 120
```

```
g-5_E_      CCATACATATTC CTTACTGAAAAGAAGATTAA TCTTGTACAAAGCATTC  TACCAATA  TTA 120
g-12_K_     CCATACATATTC CTTACTGAAAAGAAGATTAA TCTTGTACAAAGCATTT  TACCAATA  TTA 120
g-31        CCATACATATTC CTTACTGAAAAGAAGATTAA TCTTGTACAAAGCATTT  TACCAATA  TTA 120
g-20        CCATACATATTC CTTACTGAAAAGAAGATTAA TCTTGTACAAAGCATTT  TACCAATA  TTA 120
g-6_F_      CCATACATATTC CTTACTGAAAAGAAGATTAA TCTTGTACAAAGCATTT  TACCAATA  TTA 120
g-7_G_      CCATACATATTC CTTACTGAAAAGAAGATTAA TCTTGTACAAAGCATTC  TACCAATA  TTA 120
g-30        CCATACATATTC CTTACTGAAAAGAAGATTAA TCTTGTACAAAGCATTT  TACCAATA  TTA 120
g-26        CCATACATATTC CTTACTGAAAAGAAGATTAA TCTTGTACAAAGCATTT  TACCAATA  TTA 120
Position    ***********72******************92***************109*******117***

g-1_A_  GAAAACGTTGCA   C  GG   T  CTGGA  AGA  CCATTGCTC  ATCATAGCTGAAGAC  GTTGAAGGTGAA 180
g-2_B_  GAAAACGTTGCA   C  GG   T  CTGGA  AGA  CCATTGCTC  ATCATAGCTGAAGAC  GTTGAAGGTGAA 180
g-3_C_  GAAAACGTTGCA   C  GG   T  CTGGA  AGA  CCATTGCTC  ATCATAGCTGAAGAC  GTTGAAGGTGAA 180
g-14_M_ GAAAACGTTGCA   C  GG   T  CTGGA  AGA  CCATTGCTC  ATCATAGCTGAAGAC  GTTGAAGGTGAA 180
g-18_X_ GAAAACGTTGCA   C  GG   T  CTGGA  AGA  CCATTGCTC  ATCATAGCTGAAGAC  GTTGAAGGTGAA 180
g-19_Y_ GAAAACGTTGCA   C  GG   T  CTGGA  AGA  CCATTGCTC  ATCATAGCTGAAGAC  GTTGAAGGTGAA 180
g-25    GAAAACGTTGCA   C  GG   T  CTGGA  AGA  CCATTGCTC  ATCATAGCTGAAGAC  GTTGAAGGTGAA 180
g-28    GAAAACGTTGCA   C  GG   T  CTGGA  AGA  CCATTGCTC  ATCATAGCTGAAGAC  GTTGAAGGTGAA 180
g-17_P_ GAAAACGTTGCA   C  GG   T  CTGGA  AGA  CCATTGCTC  ATCATAGCTGAAGAC  GTTGAAGGTGAA 180
g-32    GAAAACGTTGCA   C  GG   T  CTGGA  AGA  CCATTGCTC  ATCATAGCTGAAGAC  GTTGAAGGTGAA 180
g-13_L_ GAAAACGTTGCA   C  GG   T  CTGGA  AGA  CCATTGCTC  ATCATAGCTGAAGAC  GTTGAAGGTGAA 180
g-11_K_ GAAAACGTTGCA   C  GG   T  CTGGA  AGA  CCATTGCTC  ATCATAGCTGAAGAC  GTTGAAGGTGAA 180
g-34    GAAAACGTTGCA   C  GG   T  CTGGA  AGA  CCATTGCTC  ATCATAGCTGAAGAC  GTTGAAGGTGAA 180
g-21    GAAAACGTTGCA   C  GG   T  CTGGA  AGG  CCATTGCTC  ATCATAGCTGAAGAC  GTTGAAGGTGAA 180
g-27    GAAAACGTTGCA   C  GG   T  CTGGA  AGG  CCATTGCTC  ATCATAGCTGAAGAC  GTTGAAGGTGAA 180
g-22    GAAAACGTTGCA   C  GG   T  CTGGA  AGG  CCATTGCTC  ATCATAGCTGAAGAC  GTTGAAGGTGAA 180
g-8_H_  GAAAACGTTGCA   C  GG   T  CTGGA  AGA  CCATTGCTC  ATCATAGCTGAAGAC  GTTGAAGGTGAA 180
g-15_N_ GAAAACGTTGCA   C  GG   T  CTGGA  AGA  CCATTGCTC  ATCATAGCTGAAGAC  GTTGAAGGTGAA 180
g-24    GAAAACGTTGCA   C  GG   T  CTGGA  AGA  CCATTGCTC  ATCATAGCTGAAGAC  GTTGAAGGTGAA 180
g-10_J_ GAAAACGTTGCA   C  GG   T  CTGGA  AGA  CCATTGCTC  ATCATAGCTGAAGAC  GTTGAAGGTGAA 180
g-35    GAAAACGTTGCA   C  GG   T  CTGGA  AGA  CCATTGCTC  ATCATAGCTGAAGAC  GTTGAAGGTGAA 180
g-23    GAAAACGTTGCA   C  GG   T  CTGGA  AGA  CCATTGCTC  ATCATAGCTGAAGAC  GTTGAAGGTGAA 180
g-16_O_ GAAAACGTTGCA   C  GG   T  CTGGA  AGA  CCATTGCTC  ATCATAGCTGAAGAC  GTTGAAGGTGAA 180
g-9_I_  GAAAACGTTGCA   C  GG   T  CTGGA  AGA  CCATTGCTC  ATCATAGCTGAAGAC  GTTGAAGGTGAA 180
g-33    GAAAACGTTGCG   C  GG   T  CTGGA  AGA  CCATTGCTC  ATCATAGCTGAAGAC  GTTGAAGGTGAA 180
g-29    GAAAACGTTGCA   C  GG   T  CTGGA  AGA  CCATTGCTC  ATCATAGCTGAAGAC  GTTGAAGGTGAA 180
g-4_D_  GAAAACGTTGCG   A  GA   G  CTGGT  AGA  CCATTGCTT  ATCATAGCTGAAGAT  GTTGAAGGTGAA 180
g-5_E_  GAAAACGTTGCG   A  GA   G  CTGGT  AGA  CCATTGCTT  ATCATAGCTGAAGAT  GTTGAAGGTGAA 180
g-12_K_ GAAAACGTTGCG   A  GA   G  CTGGT  AGA  CCATTGCTT  ATCATAGCTGAAGAT  GTTGAAGGTGAA 180
g-31    GAAAACGTTGCG   A  GA   G  CTGGT  AGA  CCATTGCTT  ATCATAGCTGAAGAT  GTTGAAGGTGAA 180
g-20    GAAAACGTTGCG   A  GA   G  CTGGT  AGA  CCATTGCTT  ATCATAGCTGAAGAT  GTTGAAGGTGAA 180
g-6_F_  GAAAACGTTGCG   A  GA   G  CTGGT  AGA  CCATTGCTT  ATCATAGCTGAAGAT  GTTGAAGGTGAA 180
g-7_G_  GAAAACGTTGCG   A  GA   G  CTGGT  AGA  CCATTGCTT  ATCATAGCTGAAGAT  GTTGAAGGTGAA 180
g-30    GAAAACGTTGCG   A  GA   G  CTGGC  AGA  CCATTGCTC  ATCATAGCTGAAGAT  GTTGAAGGTGAA 180
g-26    GAAAACGTTGCG   A  GA   G  CTGGT  AGA  CCATTGCTC  ATCATAGCTGAAGAT  GTTGAAGGTGAA 180
Position***********132 133*135 136****141**144********153**************168************

g-1_A_      GCTCTG  AGCACG  CTTGTACTCAATAAGCTCCGTGGTGGC  CTTCAAGTTGCTGCTGTAAAG 240
g-2_B_      GCTCTG  AGCACG  CTTGTACTCAATAAGCTCCGTGGTGGC  CTTCAAGTTGCTGCTGTAAAG 240
g-3_C_      GCTCTT  AGCACG  CTTGTACTCAATAAGCTCCGTGGTGGC  CTTCAAGTTGCTGCTGTAAAG 240
g-14_M_     GCTCTG  AGCACG  CTTGTACTCAATAAGCTCCGTGGTGGC  CTTCAAGTTGCTGCTGTAAAG 240
g-18_X_     GCTCTG  AGCACG  CTTGTACTCAATAAGCTCCGTGGTGGC  CTTCAAGTTGCTGCTGTAAAG 240
g-19_Y_     GCTCTG  AGCACG  CTTGTACTCAATAAGCTCCGTGGTGGC  CTTCAAGTTGCTGCTGTAAAG 240
g-25        GCTCTG  AGCACG  CTTGTACTCAATAAGCTCCGTGGTGGC  CTTCAAGTTGCTGCTGTAAAG 240
g-28        GCTCTG  AGCACG  CTTGTACTCAATAAGCTCCGTGGTGGC  CTTCAAGTTGCTGCTGTAAAG 240
g-17_P_     GCTCTG  AGCACG  CTTGTACTCAATAAGCTCCGTGGTGGC  CTTCAAGTTGCTGCTGTAAAG 240
g-32        GCTCTG  AGCACA  CTTGTACTCAATAAGCTCCGTGGTGGC  CTTCAAGTTGCTGCTGTAAAG 240
g-13_L_     GCTCTG  AGCACG  CTTGTACTCAATAAGCTCCGTGGTGGC  CTTCAAGTTGCTGCTGTAAAG 240
g-11_K_     GCTCTG  AGCACG  CTTGTACTCAATAAGCTCCGTGGTGGC  CTTCAAGTTGCTGCTGTAAAG 240
g-34        GCTCTG  AGCACG  CTTGTACTCAATAAGCTCCGTGGTGGC  CTTCAAGTTGCTGCTGTAAAG 240
g-21        GCTCTT  AGCACG  CTTGTACTCAATAAGCTCCGTGGTGGC  CTTCAAGTTGCTGCTGTAAAG 240
g-27        GCTCTT  AGCACG  CTTGTACTCAATAAGCTCCGTGGTGGC  CTTCAAGTTGCTGCTGTAAAG 240
g-22        GCTCTT  AGCACG  CTTGTACTCAATAAGCTCCGTGGTGGC  CTTCAAGTTGCTGCTGTAAAG 240
g-8_H_      GCTCTT  AGCACG  CTTGTACTCAATAAGCTCCGTGGTGGC  CTTCAAGTTGCTGCTGTAAAG 240
g-15_N_     GCTCTT  AGCACG  CTTGTACTCAATAAGCTCCGTGGTGGC  CTTCAAGTTGCTGCTGTAAAG 240
g-24        GCTCTT  AGCACG  CTTGTACTCAATAAGCTCCGTGGTGGC  CTTCAAGTTGCTGCTGTAAAG 240
g-10_J_     GCTCTT  AGCACG  CTTGTACTCAATAAGCTCCGTGGTGGC  CTTCAAGTTGCTGCTGTAAAG 240
g-35        GCTCTG  AGCACG  CTTGTACTCAATAAGCTCCGTGGTGGC  CTTCAAGTTGCTGCTGTAAAG 240
g-23        GCTCTT  AGCACG  CTTGTACTCAATAAGCTCCGTGGTGGC  CTTCAAGTTGCTGCTGTAAAG 240
g-16_O_     GCTCTG  AGCACG  CTTGTACTCAATAAGCTCCGTGGTGGC  CTTCAAGTTGCTGCTGTAAAG 240
g-9_I_      GCTCTG  AGCACG  CTTGTACTCAATAAGCTCCGTGGTGGC  CTTCAAGTTGCTGCTGTAAAG 240
g-33        GCTCTG  AGCACG  CTTGTACTCAATAAGCTCCGTGGTGGC  CTTCAAGTTGCTGCTGTAAAG 240
g-29        GCTCTG  AGCACG  CTTGTACTCAATAAGCTCCGTGGTGGC  CTTCAAGTTGCTGCTGTAAAG 240
g-4_D_      GCTCTG  AGCACG  CTTGTACTCAATAAGCTCCGTGGTGGG  CTTCAAGTTGCTGCTGTAAAG 240
g-5_E_      GCTCTG  AGCACG  CTTGTACTCAATAAGCTCCGTGGTGGG  CTTCAAGTTGCTGCTGTAAAG 240
g-12_K_     GCTCTG  AGCACG  CTTGTACTCAATAAGCTCCGTGGTGGG  CTTCAAGTTGCTGCTGTAAAG 240
g-31        GCTCTG  AGCACG  CTTGTACTCAATAAGCTCCGTGGTGGG  CTTCAAGTTGCTGCTGTAAAG 240
g-20        GCTCTG  AGCACG  CTTGTACTCAATAAGCTCCGTGGTGGG  CTTCAAGTTGCTGCTGTAAAG 240
g-6_F_      GCTCTG  AGCACG  CTTGTACTCAATAAGCTCCGTGGTGGG  CTTCAAGTTGCTGCTGTAAAG 240
g-7_G_      GCTCTG  AGCACG  CTTGTACTCAATAAGCTCCGTGGTGGG  CTTCAAGTTGCTGCTGTAAAG 240
g-30        GCTCTG  AGCACG  CTTGTACTCAATAAGCTCCGTGGTGGG  CTTCAAGTTGCTGCTGTAAAG 240
g-26        GCTCTG  AGCACG  CTTGTACTCAATAAGCTCCGTGGTGGC  CTTCAAGTTGCTGCTGTAAAG 240
Position    *****186*****192**************************219**********************

g-1_A_  GCGCCTGGTTTC  GGTGAC  AGG  AGAAAAGACATGCTT  GGCGATATTGCT  GTAATAGTAGGC 300
g-2_B_  GCGCCTGGTTTC  GGTGAC  AGG  AGAAAAGACATGCTT  GGCGATATTGCT  GTAATAGTAGGC 300
g-3_C_  GCGCCTGGTTTC  GGTGAC  AGG  AGAAAAGACATGCTT  GGCGATATTGCT  GTAATAGTAGGC 300
g-14_M_ GCGCCTGGTTTC  GGTGAC  AGG  AGAAAAGACATGCTT  GGCGATATTGCT  GTAATAGTAGGC 300
g-18_X_ GCGCCTGGTTTC  GGTGAC  AGG  AGAAAAGACATGCTT  GGCGATATTGCT  GTAATAGTAGGC 300
g-19_Y_ GCGCCTGGTTTC  GGTGAC  AGG  AGAAAAGACATGCTT  GGCGATATTGCT  GTAATAGTAGGC 300
```

```
g-25    GCGCCTGGTTTC  GGTGAC  AGG  AGAAAAGACATGCTT  GGCGATATTGCT  GTAATAGTAGGC 300
g-28    GCGCCTGGTTTC  GGTGAC  AGG  AGAAAAGACATGCTT  GGCGATATTGCT  GTAATAGTAGGC 300
g-17_P_ GCGCCTGGTTTT  GGTGAC  AGG  AGAAAAGACATGCTT  GGCGATATTGCT  GTAATAGTAGGC 300
g-32    GCGCCTGGTTTT  GGTGAC  AGG  AGAAAAGACATGCTT  GGCGATATTGCT  GTAATAGTAGGC 300
g-13_L_ GCGCCTGGTTTT  GGTGAC  AGG  AGAAAAGACATGCTT  GGCGATATTGCT  GTAATAGTAGGC 300
g-11_K_ GCGCCTGGTTTT  GGTGAC  AGG  AGAAAAGACATGCTT  GGCGATATTGCT  GTAATAGTAGGC 300
g-34    GCGCCTGGTTTT  GGTGAC  AGG  AGAAAAGACATGCTT  GGCGATATTGCT  GTAATAGTAGGC 300
g-21    GCGCCTGGTTTT  GGTGAC  AGG  AGAAAAGACATGCTT  GGCGATATTGCT  GTAATAGTAGGC 300
g-27    GCGCCTGGTTTT  GGTGAC  AGG  AGAAAAGACATGCTT  GGCGATATTGCT  GTAATAGTAGGC 300
g-22    GCGCCTGGTTTT  GGTGAC  AGG  AGAAAAGACATGCTT  GGCGATATTGCT  GTAATAGTAGGC 300
g-8_H_  GCGCCTGGTTTT  GGTGAC  AGG  AGAAAAGACATGCTT  GGCGATATTGCT  GTAATAGTAGGC 300
g-15_N_ GCGCCTGGTTTT  GGTGAC  AGG  AGAAAAGACATGCTT  GGCGATATTGCT  GTAATAGTAGGC 300
g-24    GCGCCTGGTTTT  GGTGAC  AGG  AGAAAAGACATGCTT  GGCGATATTGCT  GTAATAGTAGGC 300
g-10_J_ GCGCCTGGTTTT  GGTGAC  AGG  AGAAAAGACATGCTT  GGCGATATTGCT  GTAATAGTAGGC 300
g-35    GCGCCTGGTTTC  GGTGAC  AGG  AGAAAAGACATGCTT  GGCGATATTGCT  GTAATAGTAGGC 300
g-23    GCGCCTGGTTTC  GGTGAC  AGG  AGAAAAGACATGCTT  GGCGATATTGCT  GTAATAGTAGGC 300
g-16_O_ GCGCCTGGTTTC  GGTGAC  AGG  AGAAAAGACATGCTT  GGCGATATTGCT  GTAATAGTAGGC 300
g-9_I_  GCGCCTGGTTTC  GGTGAC  AGA  AGAAAAGACATGCTT  GGCGATATTGCT  GTAATAGTAGGC 300
g-33    GCGCCTGGTTTC  GGTGAT  AGA  AGAAAAGACATGCTT  GGCGATATTGCT  GTAATAGTAGGC 300
g-29    GCGCCTGGTTTC  GGTGAC  AGG  AGAAAAGACATGCTT  GGCGATATTGCY  GTAATAGTAGGC 300
g-4_D_  GCGCCTGGTTTC  GGTGAC  AGA  AGAAAAGACATGCTA  GGCGATATTGCC  GTAATAGTAGGC 300
g-5_E_  GCGCCTGGTTTC  GGTGAC  AGA  AGAAAAGACATGCTA  GGCGATATTGCC  GTAATAGTAGGC 300
g-12_K_ GCGCCTGGTTTC  GGTGAC  AGA  AGAAAAGACATGCTA  GGCGATATTGCC  GTAATAGTAGGC 300
g-31    GCGCCTGGTTTC  GGTGAC  AGA  AGAAAAGACATGCTA  GGCGATATTGCC  GTAATAGTAGGC 300
g-20    GCGCCTGGTTTC  GGTGAC  AGA  AGAAAAGACATGCTA  GGCGATATTGCC  GTAATAGTAGGC 300
g-6_F_  GCGCCTGGTTTC  GGTGAC  AGA  AGAAAAGACATGCTA  GGCGATATTGCC  GTAATAGTAGGC 300
g-7_G_  GCGCCTGGTTTC  GGTGAC  AGA  AGAAAAGACATGCTA  GGCGATATTGCC  GTAATAGTAGGC 300
g-30    GCGCCTGGTTTC  GGTGAT  AGA  AGAAAAGACATGCTA  GGCGATATTGCC  GTAATAGTAGGC 300
g-26    GCGCCTGGTTTC  GGTGAC  AGA  AGAAAAGACATGCTA  GGCGATATTGCC  GTAATAGTAGGC 300
Position***********252*****258**261**************276***********288***********

g-1_A_          GCTAAGTATGTAGTAAATGACGAGCTTGCTGTTAAGATGGAAGAC  ATCGCTCTAAGCGAT 360
g-2_B_          GCTAAGTATGTAGTAAATGACGAGCTTGCTGTTAAGATGGAAGAC  ATCGCTCTAAGCGAT 360
g-3_C_          GCTAAGTATGTAGTAAATGACGAGCTTGCTGTTAAGATGGAAGAC  ATCGCTCTAAGCGAT 360
g-14_M_         GCTAAGTATGTAGTAAATGACGAGCTTGCTGTTAAGATGGAAGAC  ATCGCTCTAAGCGAT 360
g-18_X_         GCTAAGTATGTAGTAAATGACGAGCTTGCTGTTAAGATGGAAGAC  ATCGCTCTAAGCGAT 360
g-19_Y_         GCTAAGTATGTAGTAAATGACGAGCTTGCTGTTAAGATGGAAGAC  ATCGCTCTAAGCGAT 360
g-25            GCTAAGTATGTAGTAAATGACGAGCTTGCTGTTAAGATGGAAGAC  ATCGCTCTAAGCGAT 360
g-28            GCTAAGTATGTAGTAAATGACGAGCTTGCTGTTAAGATGGAAGAC  ATCGCTCTAAGCGAT 360
g-17_P_         GCTAAGTATGTAGTAAATGACGAGCTTGCTGTTAAGATGGAAGAT  ATCGCTCTAAGCGAT 360
g-32            GCTAAGTATGTAGTAAATGACGAGCTTGCTGTTAAGATGGAAGAT  ATCGCTCTAAGCGAT 360
g-13_L_         GCTAAGTATGTAGTAAATGACGAGCTTGCTGTTAAGATGGAAGAT  ATCGCTCTAAGCGAT 360
g-11_K_         GCTAAGTATGTAGTAAATGACGAGCTTGCTGTTAAGATGGAAGAT  ATCGCTCTAAGCGAT 360
g-34            GCTAAGTATGTAGTAAATGACGAGCTTGCTGTTAAGATGGAAGAT  ATCGCTCTAAGCGAT 360
g-21            GCTAAGTATGTAGTAAATGACGAGCTTGCTGTTAAGATGGAAGAT  ATCGCTCTAAGCGAT 360
g-27            GCTAAGTATGTAGTAAATGACGAGCTTGCTGTTAAGATGGAAGAT  ATCGCTCTAAGCGAT 360
g-22            GCTAAGTATGTAGTAAATGACGAGCTTGCTGTTAAGATGGAAGAT  ATCGCTCTAAGCGAT 360
g-8_H_          GCTAAGTATGTAGTAAATGACGAGCTTGCTGTTAAGATGGAAGAT  ATCGCTCTAAGCGAT 360
g-15_N_         GCTAAGTATGTAGTAAATGACGAGCTTGCTGTTAAGATGGAAGAT  ATCGCTCTAAGCGAT 360
g-24            GCTAAGTATGTAGTAAATGACGAGCTTGCTGTTAAGATGGAAGAT  ATCGCTCTAAGCGAT 360
g-10_J_         GCTAAGTATGTAGTAAATGACGAGCTTGCTGTTAAGATGGAAGAC  ATCGCTCTAAGCGAT 360
g-35            GCTAAGTATGTAGTAAATGACGAGCTTGCTGTTAAGATGGAAGAT  ATCGCTCTAAGCGAT 360
g-23            GCTAAGTATGTAGTAAATGACGAGCTTGCTGTTAAGATGGAAGAT  ATCGCTCTAAGCGAT 360
g-16_O_         GCTAAGTATGTAGTAAATGACGAGCTTGCTGTTAAGATGGAAGAT  ATCGCTCTAAGCGAT 360
g-9_I_          GCTAAGTATGTAGTAAATGACGAGCTTGCTGTTAAGATGGAAGAT  ATCGCTCTAAGCGAT 360
g-33            GCTAAGTATGTAGTAAATGACGAGCTTGCTGTTAAGATGGAAGAT  ATCGCTCTAAGCGAT 360
g-29            GCTAAGTATGTAGTAAATGACGAGCTTGCTGTTAAGATGGAAGAT  ATCGCTCTAAGCGAT 360
g-4_D_          GCTAAGTATGTAGTAAATGACGAGCTTGCTGTTAAGATGGAAGAT  ATCGCTCTAAGCGAT 360
g-5_E_          GCTAAGTATGTAGTAAATGACGAGCTTGCTGTTAAGATGGAAGAT  ATCGCTCTAAGCGAT 360
g-12_K_         GCTAAGTATGTAGTAAATGACGAGCTTGCTGTTAAGATGGAAGAT  ATCGCTCTAAGCGAT 360
g-31            GCTAAGTATGTAGTAAATGACGAGCTTGCTGTTAAGATGGAAGAT  ATCGCTCTAAGCGAT 360
g-20            GCTAAGTATGTAGTAAATGACGAGCTTGCTGTTAAGATGGAAGAT  ATCGCTCTAAGCGAT 360
g-6_F_          GCTAAGTATGTAGTAAATGACGAGCTTGCTGTTAAGATGGAAGAT  ATCGCTCTAAGCGAT 360
g-7_G_          GCTAAGTATGTAGTAAATGACGAGCTTGCTGTTAAGATGGAAGAT  ATCGCTCTAAGCGAT 360
g-30            GCTAAGTATGTAGTAAATGACGAGCTTGCTGTTAAGATGGAAGAT  ATCGCTCTAAGCGAT 360
g-26            GCTAAGTATGTAGTAAATGACGAGCTTGCTGTTAAGATGGAAGAT  ATCGCTCTAAGCGAT 360
Position        ******************************************345***************

g-1_A_  CTTGGTACTGCTAAGAGCGTA  CGC  ATCACAAAAGACGCAACTACTATC  ATAGGT  AGTG  TT 420
g-2_B_  CTTGGTACTGCTAAGAGCGTA  CGC  ATCACAAAAGACGCAACTACTATC  ATAGGT  AGTG  TT 420
g-3_C_  CTTGGTACTGCTAAGAGCGTA  CGC  ATCACAAAAGACGCAACTACTATC  ATAGGT  AGTG  TT 420
g-14_M_ CTTGGTACTGCTAAGAGCGTG  CGC  ATCACAAAAGACGCAACTACTATC  ATAGGC  AGTG  TT 420
g-18_X_ CTTGGTACTGCTAAGAGCGTA  CGC  ATCACAAAAGACGCAACTACTATC  ATAGGC  AGTG  TT 420
g-19_Y_ CTTGGTACTGCTAAGAGCGTA  CGC  ATCACAAAAGACGCAACTACTATC  ATAGGC  AGTG  TT 420
g-25    CTTGGTACTGCTAAGAGCGTA  CGC  ATCACAAAAGACGCAACTACTATC  ATAGGT  AGTG  TT 420
g-28    CTTGGTACTGCTAAGAGCGTA  CGC  ATCACAAAAGACGCAACTACTATC  ATAGGT  AGTT  TT 420
g-17_P_ CTTGGTACTGCTAAGAGCGTA  CGC  ATCACAAAAGACGCAACTACTATC  ATAGGT  AGTG  TT 420
g-32    CTTGGTACTGCTAAGAGCGTA  CGC  ATCACAAAAGACGCAACTACTATC  ATAGGT  AGTG  TT 420
g-13_L_ CTTGGTACTGCTAAGAGCGTA  CGC  ATCACAAAAGACGCAACTACTATC  ATAGGT  AGTG  TT 420
g-11_K_ CTTGGTACTGCTAAGAGCGTA  CGC  ATCACAAAAGACGCAACTACTATC  ATAGGT  AGTG  TT 420
g-34    CTTGGTACTGCTAAGAGCGTA  CGC  ATCACAAAAGACGCAACTACTATC  ATAGGT  AGTG  TT 420
g-21    CTTGGTACTGCTAAGAGCGTA  CGC  ATCACAAAAGACGCAACTACTATC  ATAGGT  AGTG  TT 420
g-27    CTTGGTACTGCTAAGAGCGTA  CGC  ATCACAAAAGACGCAACTACTATC  ATAGGT  AGTG  TT 420
g-22    CTTGGTACTGCTAAGAGCGTA  CGC  ATCACAAAAGACGCAACTACTATC  ATAGGT  AGTG  TT 420
g-8_H_  CTTGGTACTGCTAAGAGCGTA  CGC  ATCACAAAAGACGCAACTACTATC  ATAGGT  AGTG  TT 420
g-15_N_ CTTGGTACTGCTAAGAGCGTA  CGC  ATCACAAAAGACGCAACTACTATC  ATAGGT  AGTG  TT 420
g-24    CTTGGTACTGCTAAGAGCGTA  CGC  ATCACAAAAGACGCAACTACTATC  ATAGGT  AGTG  TT 420
g-10_J_ CTTGGTACTGCTAAGAGCGTA  CGC  ATCACAAAAGACGCAACTACTATC  ATAGGC  AGTG  TT 420
g-35    CTTGGTACTGCTAAGAGCGTA  CGC  ATCACAAAAGACGCAACTACTATC  ATAGGT  AGTG  TT 420
g-23    CTTGGTACTGCTAAGAGCGTA  CGC  ATCACAAAAGACGCAACTACTATC  ATAGGT  AGTG  TT 420
g-16_O_ CTTGGTACTGCTAAGAGCGTA  CGC  ATCACAAAAGACGCAACTACTATC  ATAGGT  AGTG  TT 420
```

```
g-9_I_  CTTGGTACTGCTAAGAGCGTA  CGC  ATCACAAAAGACGCAACTACTATC  ATAGGT  AGTG  TT 420
g-33    CTTGGTACTGCTAAGAGCGTA  CGC  ATCACAAAAGACGCAACTACTATC  ATAGGT  AGTG  TT 420
g-29    CTTGGTACTGCTAAGAGCGTG  CGC  ATCACAAAAGACGCAACTACTATC  ATAGGT  AGTG  TT 420
g-4_D_  CTTGGTACTGCTAAGAGCGTG  CGC  ATCACAAAAGACGCAACTACTATC  ATAGGT  AGTG  TT 420
g-5_E_  CTTGGTACTGCTAAGAGCGTG  CGC  ATCACAAAAGACGCAACTACTATC  ATAGGT  AGTG  TT 420
g-12_K_ CTTGGTACTGCTAAGAGCGTG  CGC  ATCACAAAAGACGCAACTACTATC  ATAGGT  AGTG  TT 420
g-31    CTTGGTACTGCTAAGAGCGTG  CGC  ATCACAAAAGACGCAACTACTATC  ATAGGT  AGTG  TT 420
g-20    CTTGGTACTGCTAAGAGCGTG  CGC  ATCACAAAAGACGCAACTACTATC  ATAGGT  AGTG  TT 420
g-6_F_  CTTGGTACTGCTAAGAGCGTG  CGC  ATCACAAAAGACGCAACTACTATT  ATAGGT  AGTG  TT 420
g-7_G_  CTTGGTACTGCTAAGAGCGTG  CGC  ATCACAAAAGACGCAACTACTATT  ATAGGT  AGTG  TT 420
g-30    CTTGGTACTGCTAAGAGCGTG  CGT  ATCACAAAAGACGCAACTACTATT  ATAGGT  AGTG  TT 420
g-26    CTTGGTACTGCTAAGAGCGTG  CGC  ATCACAAAAGACGCAACTACTATC  ATAGGT  AGTG  TT 420
Position********************381**384************************408*****414***418**

g-1_A_        GAT  AGCAGTTCTGAAAGCATAGCTAGCAGGACTAAT  CAAATCAAAGCTCAGATAGAAAAT 480
g-2_B_        GAT  AGCAGTTCTGAAAGCATAGCTAGCAGGACTAAT  CAAATCAAAGCTCAGATAGAAAAT 480
g-3_C_        GAT  AGCAGTTCTGAAAGCATAGCTAGCAGGACTAAT  CAAATCAAAGCTCAGATAGAAAAT 480
g-14_M_       GAT  AGCAGTTCTGAAAGCATAGCTAGCAGGACTAAT  CAAATCAAAGCTCAGATAGAAAAT 480
g-18_X_       GAT  AGCAGTTCTGAAAGCATAGCTAGCAGGACTAAT  CAAATCAAAGCTCAGATAGAAAAT 480
g-19_Y_       GAT  AGCAGTTCTGAAAGCATAGCTAGCAGGACTAAT  CAAATCAAAGCTCAGATAGAAAAT 480
g-25          GAT  AGCAGTTCTGAAAGCATAGCTAGCAGGACTAAT  CAAATCAAAGCTCAGATAGAAAAT 480
g-28          GAT  AGCAGTTCTGAAAGCATAGCTAGCAGGACTAAT  CAAATCAAAGCTCAGATAGAAAAT 480
g-17_P_       GAT  AGCAGTTCTGAAAGCATAGCTAGCAGGACTAAT  CAAATCAAAGCTCAGATAGAAAAT 480
g-32          GAT  AGCAGTTCTGAAAGCATAGCTAGCAGGACTAAT  CAAATCAAAGCTCAGATAGAAAAT 480
g-13_L_       GAT  AGCAGTTCTGAAAGCATAGCTAGCAGGACTAAT  CAAATCAAAGCTCAGATAGAAAAT 480
g-11_K_       GAT  AGCAGTTCTGAAAGCATAGCTAGCAGGACTAAT  CAAATCAAAGCTCAGATAGAAAAT 480
g-34          GAT  AGCAGTTCTGAAAGCATAGCTAGCAGGACTAAT  CAAATCAAAGCTCAGATAGAAAAT 480
g-21          GAT  AGCAGTTCTGAAAGCATAGCTAGCAGGACTAAT  CAAATCAAAGCTCAGATAGAAAAT 480
g-27          GAT  AGCAGTTCTGAAAGCATAGCTAGCAGGACTAAT  CAAATCAAAGCTCAGATAGAAAAT 480
g-22          GAT  AGCAGTTCTGAAAGCATAGCTAGCAGGACTAAT  CAAATCAAAGCTCAGATAGAAAAT 480
g-8_H_        GAT  AGCAGTTCTGAAAGCATAGCTAGCAGGACTAAT  CAAATCAAAGCTCAGATAGAAAAT 480
g-15_N_       GAT  AGCAGTTCTGAAAGCATAGCTAGCAGGACTAAT  CAAATCAAAGCTCAGATAGAAAAT 480
g-24          GAT  AGCAGTTCTGAAAGCATAGCTAGCAGGACTAAT  CAAATCAAAGCTCAGATAGAAAAT 480
g-10_J_       GAC  AGCAGTTCTGAAAGCATAGCTAGCAGGACTAAT  CAAATCAAAGCTCAGATAGAAAAT 480
g-35          GAT  AGCAGTTCTGAAAGCATAGCTAGCAGGACTAAT  CAAATCAAAGCTCAGATAGAAAAT 480
g-23          GAT  AGCAGTTCTGAAAGCATAGCTAGCAGGACTAAT  CAAATCAAAGCTCAGATAGAAAAT 480
g-16_O_       GAT  AGCAGTTCTGAAAGCATAGCTAGCAGGACTAAT  CAAATCAAAGCTCAGATAGAAAAT 480
g-9_I_        GAT  AGCAGTTCTGAAAGCATAGCTAGCAGGACTAAT  CAAATCAAAGCTCAGATAGAAAAT 480
g-33          GAT  AGCAGTTCTGAAAGCATAGCTAGCAGGACTAAT  CAAATCAAAGCTCAGATAGAAAAT 480
g-29          GAT  AGCAGTTCTGAAAGCATAGCTAGCAGGACTAAT  CAAATCAAAGCTCAGATAGAAAAT 480
g-4_D_        GAT  AGCAGTTCTGAAAGCATAGCTAGCAGGACTAAC  CAAATCAAAGCTCAGATAGAAAAT 480
g-5_E_        GAT  AGCAGTTCTGAAAGCATAGCTAGCAGGACTAAC  CAAATCAAAGCTCAGATAGAAAAT 480
g-12_K_       GAT  AGCAGTTCTGAAAGCATAGCTAGCAGGACTAAC  CAAATCAAAGCTCAGATAGAAAAT 480
g-31          GAT  AGCAGTTCTGAAAGCATAGCTAGCAGGACTAAT  CAAATCAAAGCTCAGATAGAAAAT 480
g-20          GAT  AGCAGTTCTGAAAGCATAGCTAGCAGGACTAAT  CAAATCAAAGCTCAGATAGAAAAT 480
g-6_F_        GAT  AGCAGTTCTGAAAGCATAGCTAGCAGGACTAAT  CAAATCAAAGCTCAGATAGAAAAT 480
g-7_G_        GAT  AGCAGTTCTGAAAGCATAGCTAGCAGGACTAAT  CAAATCAAAGCTCAGATAGAAAAT 480
g-30          GAT  AGCAGTTCTGAAAGCATAGCTAGCAGGACTAAT  CAAATCAAAGCTCAGATAGAAAAT 480
g-26          GAT  AGCAGTTCTGAAAGCATAGCTAGCAGGACTAAT  CAAATCAAAGCTCAGATAGAAAAT 480
Position      **423*********************************458************************

g-1_A_        TCTAGTTCTGATTATGACAAGGAAAAGCTTAGAGAA  CGTTTA  GCA  AAGCT 530
g-2_B_        TCTAGTTCTGATTATGACAAGGAAAAGCTTAGAGAA  CGTTTA  GCA  AAGCT 530
g-3_C_        TCTAGTTCTGATTATGACAAGGAAAAGCTTAGAGAA  CGTTTA  GCA  AAGCT 530
g-14_M_       TCTAGTTCTGATTATGACAAGGAAAAGCTTAGAGAA  CGTTTA  GCA  AAGCT 530
g-18_X_       TCTAGTTCTGATTATGACAAGGAAAAGCTTAGAGAA  CGTTTA  GCA  AAGCT 530
g-19_Y_       TCTAGTTCTGATTATGACAAGGAAAAGCTTAGAGAA  CGTTTA  GCG  AAGCT 530
g-25          TCTAGTTCTGATTATGACAAGGAAAAGCTTAGAGAA  CGTTTA  GCG  AAGCT 530
g-28          TCTAGTTCTGATTATGACAAGGAAAAGCTTAGAGAA  CGTTTA  GCG  AAGCT 530
g-17_P_       TCTAGTTCTGATTATGACAAGGAAAAGCTTAGAGAA  CGTTTG  GCG  AAGCT 530
g-32          TCTAGTTCTGATTATGACAAGGAAAAGCTTAGAGAA  CGTTTA  GCG  AAGCT 530
g-13_L_       TCTAGTTCTGATTATGACAAGGAAAAGCTTAGAGAA  CGTTTA  GCG  AAGCT 530
g-11_K_       TCTAGTTCTGATTATGACAAGGAAAAGCTTAGAGAA  CGTTTA  GCG  AAGCT 530
g-34          TCTAGTTCTGATTATGACAAGGAAAAGCTTAGAGAA  CGTTTA  GCA  AAGCT 530
g-21          TCTAGTTCTGATTATGACAAGGAAAAGCTTAGAGAA  CGTTTA  GCG  AAGCT 530
g-27          TCTAGTTCTGATTATGACAAGGAAAAGCTTAGAGAA  CGTTTA  GCG  AAGCT 530
g-22          TCTAGTTCTGATTATGACAAGGAAAAGCTTAGAGAA  CGTTTA  GCA  AAGCT 530
g-8_H_        TCTAGTTCTGATTATGACAAGGAAAAGCTTAGAGAA  CGTTTA  GCA  AAGCT 530
g-15_N_       TCTAGTTCTGATTATGACAAGGAAAAGCTTAGAGAA  CGTTTA  GCG  AAGCT 530
g-24          TCTAGTTCTGATTATGACAAGGAAAAGCTTAGAGAA  CGTTTA  GCG  AAGCT 530
g-10_J_       TCTAGTTCTGATTATGACAAGGAAAAGCTTAGAGAA  CGTTTA  GCG  AAGCT 530
g-35          TCTAGTTCTGATTATGACAAGGAAAAGCTTAGAGAA  CGTTTA  GCG  AAGCT 530
g-23          TCTAGTTCTGATTATGACAAGGAAAAGCTTAGAGAA  CGTTTA  GCA  AAGCT 530
g-16_O_       TCTAGTTCTGATTATGACAAGGAAAAGCTTAGAGAA  CGTTTA  GCA  AAGCT 530
g-9_I_        TCTAGTTCTGATTATGACAAGGAAAAGCTTAGAGAA  CGTTTA  GCA  AAGCT 530
g-33          TCTAGTTCTGATTATGACAAGGAAAAGCTTAGAGAA  CGTTTA  GCA  AAGCT 530
g-29          TCTAGTTCTGATTATGACAAGGAAAAGCTTAGAGAA  CGTTTA  GCG  AAGCT 530
g-4_D_        TCTAGTTCTGATTATGACAAGGAAAAGCTTAGAGAG  CGTTTA  GCG  AAGCT 530
g-5_E_        TCTAGTTCTGATTATGACAAGGAAAAGCTTAGAGAG  CGTTTA  GCG  AAGCT 530
g-12_K_       TCTAGTTCTGATTATGACAAGGAAAAGCTTAGAGAA  CGTTTA  GCG  AAGCT 530
g-31          TCTAGTTCTGATTATGACAAGGAAAAGCTTAGAGAG  CGTTTA  GCG  AAGCT 530
g-20          TCTAGTTCTGATTATGACAAGGAAAAGCTTAGAGAA  CGTTTA  GCG  AAGCT 530
g-6_F_        TCTAGTTCTGATTATGACAAGGAAAAGCTTAGAGAA  CGTTTA  GCG  AAGCT 530
g-7_G_        TCTAGTTCTGATTATGACAAGGAAAAGCTTAGAGAA  CGTTTA  GCG  AAGCT 530
g-30          TCTAGTTCTGATTATGACAAGGAAAAGCTTAGAGAA  CGTTTA  GCA  AAGCT 530
g-26          TCTAGTTCTGATTATGACAAGGAAAAGCTTAGAGAA  CGTTTA  GCG  AAGCT 530
Position      ***********************************516*****522**525*****
```

Table 36: Alignment of the amino acid sequence of the *groEL* gene – highlighted amino acid positions are differing within the variants

```
CLUSTAL 2.1 multiple sequence alignment


g-4_D_          GYLSPYFVTNAEKMETLVEFENPYIFLTEKKINLVQSILPILENVARAGRPLLIIAEDVE 60
g-5_E_          GYLSPYFVTNAEKMETLVEFENPYIFLTEKKINLVQSILPILENVARAGRPLLIIAEDVE 60
g-6_F_          GYLSPYFVTNAEKMETLVEFENPYIFLTEKKINLVQSILPILENVARAGRPLLIIAEDVE 60
g-7_G_          GYLSPYFVTNAEKMETLVEFENPYIFLTEKKINLVQSILPILENVARAGRPLLIIAEDVE 60
g-12_K_         GYLSPYFVTNAEKMETLVEFENPYIFLTEKKINLVQSILPILENVARAGRPLLIIAEDVE 60
g-20            GYLSPYFVTNAEKMETLVEFENPYIFLTEKKINLVQSILPILENVARAGRPLLIIAEDVE 60
g-26            GYLSPYFVTNAEKMETLVEFENPYIFLTEKKINLVQSILPILENVARAGRPLLIIAEDVE 60
g-30            GYLSPYFVTNAEKMETLVEFENPYIFLTEKKINLVQSILPILENVARAGRPLLIIAEDVE 60
g-31            GYLSPYFVTNAEKMETLVEFENPYIFLTEKKINLVQSILPILENVARAGRPLLIIAEDVE 60
g-15_N_         GYLSPYFVTSAEKMETLVEFENPYIFLTEKKINLVQSILPILENVARSGRPLLIIAEDVE 60
g-27            GYLSPYFVTNAEKMETLVEFENPYIFLTEKKISLVQSILPILENVARSGRPLLIIAEDVE 60
g-28            GYLSPYFVTNAEKMETLVEFENPYIFLTEKKINLVQSILPILENVARSGRPLLIIAEDVE 60
g-34            GYLSPYFVTNAEKMETLVEFENPYIFLTEKKINLVQSILPILENVARSGRPLLIIAEDVE 60
g-1_A_          GYLSPYFVTNAEKMETLVEFENPYIFLTEKKINLVQSILPILENVARSGRPLLIIAEDVE 60
g-25            GYLSPYFVTNAEKMETLVEFENPYIFLTEKKINLVQSILPILENVARSGRPLLIIAEDVE 60
g-2_B_          GYLSPYFVTNAEKMETLVEFENPYIFLTEKKINLVQSILPILENVARSGRPLLIIAEDVE 60
g-3_C_          GYLSPYFVTNAEKMETLVEFENPYIFLTEKKINLVQSILPILENVARSGRPLLIIAEDVE 60
g-8_H_          GYLSPYFVTNAEKMETLVEFENPYIFLTEKKINLVQSILPILENVARSGRPLLIIAEDVE 60
g-9_I_          GYLSPYFVTNAEKMETLVEFENPYIFLTEKKINLVQSILPILENVARSGRPLLIIAEDVE 60
g-10_J_         GYLSPYFVTNAEKMETLVEFENPYIFLTEKKINLVQSILPILENVARSGRPLLIIAEDVE 60
g-11_K_         GYLSPYFVTNAEKMETLVEFENPYIFLTEKKINLVQSILPILENVARSGRPLLIIAEDVE 60
g-13_L_         GYLSPYFVTNAEKMETLVEFENPYIFLTEKKINLVQSILPILENVARSGRPLLIIAEDVE 60
g-14_M_         GYLSPYFVTNAEKMETLVEFENPYIFLTEKKINLVQSILPILENVARSGRPLLIIAEDVE 60
g-16_O_         GYLSPYFVTNAEKMETLVEFENPYIFLTEKKINLVQSILPILENVARSGRPLLIIAEDVE 60
g-17_P_         GYLSPYFVTNAEKMETLVEFENPYIFLTEKKINLVQSILPILENVARSGRPLLIIAEDVE 60
g-18_X_         GYLSPYFVTNAEKMETLVEFENPYIFLTEKKINLVQSILPILENVARSGRPLLIIAEDVE 60
g-19_Y_         GYLSPYFVTNAEKMETLVEFENPYIFLTEKKINLVQSILPILENVARSGRPLLIIAEDVE 60
g-21            GYLSPYFVTNAEKMETLVEFENPYIFLTEKKINLVQSILPILENVARSGRPLLIIAEDVE 60
g-22            GYLSPYFVTNAEKMETLVEFENPYIFLTEKKINLVQSILPILENVARSGRPLLIIAEDVE 60
g-23            GYLSPYFVTNAEKMETLVEFENPYIFLTEKKINLVQSILPILENVARSGRPLLIIAEDVE 60
g-32            GYLSPYFVTNAEKMETLVEFENPYIFLTEKKINLVQSILPILENVARSGRPLLIIAEDVE 60
g-35            GYLSPYFVTNAEKMETLVEFENPYIFLTEKKINLVQSILPILENVARSGRPLLIIAEDVE 60
g-33            GYLSPYFVTNAEKMETLVEFENPYIFLTEKKINLVQSILPILENVARSGRPLLIIAEDVE 60
g-24            GYLSPYFVTNAEKMETLVEFENPYIFLTEKKINLVQSILPILENVARSGRPLLIIAEDVE 60
g-29            GYLSPYFVTNAEKMETLVEFENPYIFLTEKKINLVQSILPXLENVARSGRPLLIIAEDVE 60
                ********* ********************** ******* ****** ************

g-4_D_          GEALSTLVLNKLRGGLQVAAVKAPGFGDRRKDMETLGDIAVIVGAKYVVNDELAVKMETE 120
g-5_E_          GEALSTLVLNKLRGGLQVAAVKAPGFGDRRKDMETLGDIAVIVGAKYVVNDELAVKMETE 120
g-6_F_          GEALSTLVLNKLRGGLQVAAVKAPGFGDRRKDMETLGDIAVIVGAKYVVNDELAVKMETE 120
g-7_G_          GEALSTLVLNKLRGGLQVAAVKAPGFGDRRKDMETLGDIAVIVGAKYVVNDELAVKMETE 120
g-12_K_         GEALSTLVLNKLRGGLQVAAVKAPGFGDRRKDMETLGDIAVIVGAKYVVNDELAVKMETE 120
g-20            GEALSTLVLNKLRGGLQVAAVKAPGFGDRRKDMETLGDIAVIVGAKYVVNDELAVKMETE 120
g-26            GEALSTLVLNKLRGGLQVAAVKAPGFGDRRKDMETLGDIAVIVGAKYVVNDELAVKMETE 120
g-30            GEALSTLVLNKLRGGLQVAAVKAPGFGDRRKDMETLGDIAVIVGAKYVVNDELAVKMETE 120
g-31            GEALSTLVLNKLRGGLQVAAVKAPGFGDRRKDMETLGDIAVIVGAKYVVNDELAVKMETE 120
g-15_N_         GEALSTLVLNKLRGGLQVAAVKAPGFGDRRKDMETLGDIAVIVGAKYVVNDELAVKMETE 120
g-27            GEALSTLVLNKLRGGLQVAAVKAPGFGDRRKDMETLGDIAVIVGAKYVVNDELAVKMETE 120
g-28            GEALSTLVLNKLRGGLQVAAVKAPGFGDRRKDMETLGDIAVIVGAKYVVNDELAVKMETE 120
g-34            GEALSTLVLNKLRGGLQVAAVKAPGFGDRRKDMETLGDIAVIVGAKYVVNDELAVKMETE 120
g-1_A_          GEALSTLVLNKLRGGLQVAAVKAPGFGDRRKDMETLGDIAVIVGAKYVVNDELAVKMETE 120
g-25            GEALSTLVLNKLRGGLQVAAVKAPGFGDRRKDMETLGDIAVIVGAKYVVNDELAVKMETE 120
g-2_B_          GEALSTLVLNKLRGGLQVAAVKAPGFGDRRKDMETLGDIAVIVGAKYVVNDELAVKMETE 120
g-3_C_          GEALSTLVLNKLRGGLQVAAVKAPGFGDRRKDMETLGDIAVIVGAKYVVNDELAVKMETE 120
g-8_H_          GEALSTLVLNKLRGGLQVAAVKAPGFGDRRKDMETLGDIAVIVGAKYVVNDELAVKMETE 120
g-9_I_          GEALSTLVLNKLRGGLQVAAVKAPGFGDRRKDMETLGDIAVIVGAKYVVNDELAVKMETE 120
g-10_J_         GEALSTLVLNKLRGGLQVAAVKAPGFGDRRKDMETLGDIAVIVGAKYVVNDELAVKMETE 120
g-11_K_         GEALSTLVLNKLRGGLQVAAVKAPGFGDRRKDMETLGDIAVIVGAKYVVNDELAVKMETE 120
g-13_L_         GEALSTLVLNKLRGGLQVAAVKAPGFGDRRKDMETLGDIAVIVGAKYVVNDELAVKMETE 120
g-14_M_         GEALSTLVLNKLRGGLQVAAVKAPGFGDRRKDMETLGDIAVIVGAKYVVNDELAVKMETE 120
g-16_O_         GEALSTLVLNKLRGGLQVAAVKAPGFGDRRKDMETLGDIAVIVGAKYVVNDELAVKMETE 120
g-17_P_         GEALSTLVLNKLRGGLQVAAVKAPGFGDRRKDMETLGDIAVIVGAKYVVNDELAVKMETE 120
g-18_X_         GEALSTLVLNKLRGGLQVAAVKAPGFGDRRKDMETLGDIAVIVGAKYVVNDELAVKMETE 120
g-19_Y_         GEALSTLVLNKLRGGLQVAAVKAPGFGDRRKDMETLGDIAVIVGAKYVVNDELAVKMETE 120
g-21            GEALSTLVLNKLRGGLQVAAVKAPGFGDRRKDMETLGDIAVIVGAKYVVNDELAVKMETE 120
g-22            GEALSTLVLNKLRGGLQVAAVKAPGFGDRRKDMETLGDIAVIVGAKYVVNDELAVKMETE 120
g-23            GEALSTLVLNKLRGGLQVAAVKAPGFGDRRKDMETLGDIAVIVGAKYVVNDELAVKMETE 120
g-32            GEALSTLVLNKLRGGLQVAAVKAPGFGDRRKDMETLGDIAVIVGAKYVVNDELAVKMETE 120
g-35            GEALSTLVLNKLRGGLQVAAVKAPGFGDRRKDMETLGDIAVIVGAKYVVNDELAVKMETE 120
g-33            GEALSTLVLNKLRGGLQVAAVKAPGFGDRRKDMETLGDIAVIVGAKYVVNDELAVKMETE 120
g-24            GEALSTLVLNKLRGGLQVAAVKAPGFGDRRKDMETLGDIAVIVGAKYVVNDELAVKMETE 120
g-29            GEALSTLVLNKLRGGLQVAAVKAPGFGDRRKDMETLGDIXVIVGAKYVVNDELAVKMETE 120
```

```
                ************************************ ********************

g-4_D_          DIALSDLGTAKSVRITKDATTIIGSVDSSSESIASRTNQIKAQIENSSSDYDKEKLRERL 180
g-5_E_          DIALSDLGTAKSVRITKDATTIIGSVDSSSESIASRTNQIKAQIENSSSDYDKEKLRERL 180
g-6_F_          DIALSDLGTAKSVRITKDATTIIGSVDSSSESIASRTNQIKAQIENSSSDYDKEKLRERL 180
g-7_G_          DIALSDLGTAKSVRITKDATTIIGSVDSSSESIASRTNQIKAQIENSSSDYDKEKLRERL 180
g-12_K_         DIALSDLGTAKSVRITKDATTIIGSVDSSSESIASRTNQIKAQIENSSSDYDKEKLRERL 180
g-20            DIALSDLGTAKSVRITKDATTIIGSVDSSSESIASRTNQIKAQIENSSSDYDKEKLRERL 180
g-26            DIALSDLGTAKSVRITKDATTIIGSVDSSSESIASRTNQIKAQIENSSSDYDKEKLRERL 180
g-30            DIALSDLGTAKSVRITKDATTIIGSVDSSSESIASRTNQIKAQIENSSSDYDKEKLRERL 180
g-31            DIALSDLGTAKSVRITKDATTIIGSVDSSSESIASRTNQIKAQIENSSSDYDKEKLRERL 180
g-15_N_         DIALSDLGTAKSVRITKDATTIIGSVDSSSESIASRTNQIKAQIENSSSDYDKEKLRERL 180
g-27            DIALSDLGTAKSVRITKDATTIIGSVDSSSESIASRTNQIKAQIENSSSDYDKEKLRERL 180
g-28            DIALSDLGTAKSVRITKDATTIIGSFDSSSESIASRTNQIKAQIENSSSDYDKEKLRERL 180
g-34            DIALSDLGTAKSVRITKDATTIIGSVDSSSESIASRTNQIKAQIENSSSDYDKEKLRERL 180
g-1_A_          DIALSDLGTAKSVRITKDATTIIGSVDSSSESIASRTNQIKAQIENSSSDYDKEKLRERL 180
g-25            DIALSDLGTAKSVRITKDATTIIGSVDSSSESIASRTNQIKAQIENSSSDYDKEKLRERL 180
g-2_B_          DIALSDLGTAKSVRITKDATTIIGSVDSSSESIASRTNQIKAQIENSSSDYDKEKLRERL 180
g-3_C_          DIALSDLGTAKSVRITKDATTIIGSVDSSSESIASRTNQIKAQIENSSSDYDKEKLRERL 180
g-8_H_          DIALSDLGTAKSVRITKDATTIIGSVDSSSESIASRTNQIKAQIENSSSDYDKEKLRERL 180
g-9_I_          DIALSDLGTAKSVRITKDATTIIGSVDSSSESIASRTNQIKAQIENSSSDYDKEKLRERL 180
g-10_J_         DIALSDLGTAKSVRITKDATTIIGSVDSSSESIASRTNQIKAQIENSSSDYDKEKLRERL 180
g-11_K_         DIALSDLGTAKSVRITKDATTIIGSVDSSSESIASRTNQIKAQIENSSSDYDKEKLRERL 180
g-13_L_         DIALSDLGTAKSVRITKDATTIIGSVDSSSESIASRTNQIKAQIENSSSDYDKEKLRERL 180
g-14_M_         DIALSDLGTAKSVRITKDATTIIGSVDSSSESIASRTNQIKAQIENSSSDYDKEKLRERL 180
g-16_O_         DIALSDLGTAKSVRITKDATTIIGSVDSSSESIASRTNQIKAQIENSSSDYDKEKLRERL 180
g-17_P_         DIALSDLGTAKSVRITKDATTIIGSVDSSSESIASRTNQIKAQIENSSSDYDKEKLRERL 180
g-18_X_         DIALSDLGTAKSVRITKDATTIIGSVDSSSESIASRTNQIKAQIENSSSDYDKEKLRERL 180
g-19_Y_         DIALSDLGTAKSVRITKDATTIIGSVDSSSESIASRTNQIKAQIENSSSDYDKEKLRERL 180
g-21            DIALSDLGTAKSVRITKDATTIIGSVDSSSESIASRTNQIKAQIENSSSDYDKEKLRERL 180
g-22            DIALSDLGTAKSVRITKDATTIIGSVDSSSESIASRTNQIKAQIENSSSDYDKEKLRERL 180
g-23            DIALSDLGTAKSVRITKDATTIIGSVDSSSESIASRTNQIKAQIENSSSDYDKEKLRERL 180
g-32            DIALSDLGTAKSVRITKDATTIIGSVDSSSESIASRTNQIKAQIENSSSDYDKEKLRERL 180
g-35            DIALSDLGTAKSVRITKDATTIIGSVDSSSESIASRTNQIKAQIENSSSDYDKEKLRERL 180
g-33            DIALSDLGTAKSVRITKDATTIIGSVDSSSESIASRTNQIKAQIENSSSDYDKEKLRERL 180
g-24            DIALSDLGTAKSVRITKDATTIIGSVDSSSESIASRTNQIKAQIENSSSDYDKEKLRERL 180
g-29            DIALSDLGTAKSVRITKDATTIIGSVDSSSESIASRTNQIKAQIENSSSDYDKEKLRERL 180
                ************************* **********************************

g-4_D_          AK 182
g-5_E_          AK 182
g-6_F_          AK 182
g-7_G_          AK 182
g-12_K_         AK 182
g-20            AK 182
g-26            AK 182
g-30            AK 182
g-31            AK 182
g-15_N_         AK 182
g-27            AK 182
g-28            AK 182
g-34            AK 182
g-1_A_          AK 182
g-25            A- 181
g-2_B_          AK 182
g-3_C_          AK 182
g-8_H_          AK 182
g-9_I_          AK 182
g-10_J_         AK 182
g-11_K_         AK 182
g-13_L_         AK 182
g-14_M_         AK 182
g-16_O_         AK 182
g-17_P_         AK 182
g-18_X_         AK 182
g-19_Y_         AK 182
g-21            AK 182
g-22            AK 182
g-23            AK 182
g-32            AK 182
g-35            AK 182
g-33            AK 182
g-24            AK 182
g-29            AK 182
                *
```

Table 37: Assigned accession numbers of the new *groEL* sequences

Accession number	*groEL* variant	Sample	Animal species
KU712086	g-2(B)	971	dog
KU712087	g-1(A)	69	dog
KU712088	g-1(A)	413	dog
KU712089	g-1(A)	438	dog
KU712090	g-2(B)	6473	dog
KU712091	g-1(A)	14333	dog
KU712092	g-1(A)	155480	dog
KU712093	g-24	BwDa3	fallow deer
KU712095	g-8(H)	M11	fallow deer
KU712096	g-8(H)	M19	fallow deer
KU712097	g-29	M20	fallow deer
KU712098	g-22	Mf7	mouflon
KU712099	g-21	Mf8	mouflon
KU712100	g-21	A31	mouflon
KU712101	g-13(L)	K26	mouflon
KU712102	g-8(H)	K37	mouflon
KU712103	g-13(L)	l38	mouflon
KU712104	g-34	N04	mouflon
KU712105	g-23	O32	mouflon
KU712106	g-13(L)	Ro23	red deer
KU712107	g-32	BwRo2	red deer
KU712108	g-33	BwRo4	red deer
KU712109	g-6(F)	BwReh4	roe deer
KU712110	g-7(G)	BwReh8	roe deer
KU712111	g-24	BwReh9	roe deer
KU712112	g-20	BwReh10	roe deer
KU712113	g-7(G)	BwReh12	roe deer
KU712114	g-30	BwReh13	roe deer
KU712115	g-5(E)	BwReh14	roe deer
KU712116	g-7(G)	BwReh16	roe deer
KU712117	g-31	BwReh24	roe deer
KU712118	g-6(F)	BwReh32	roe deer
KU712119	g-7(G)	BwReh33	roe deer
KU712120	g-7(G)	A04	roe deer
KU712121	g-4(D)	D31	roe deer
KU712122	g-7(G)	L12	roe deer

KU712123	g-26	O08	roe deer
KU712124	g-20	P10	roe deer
KU712125	g-34	BwSi3	sika deer
KU712126	g-35	BwSi5	sika deer
KU712127	g-28	BwSi11	sika deer
KU712128	g-13(L)	BwSi12	sika deer
KU712129	g-24	BwSi16	sika deer
KU712130	g-1(A)	S133	red fox
KU712131	g-1(A)	S146	red fox
KU712132	g-2(B)	S226	red fox
KU712133	g-1(A)	262	hedgehog

1.1.3. *msp4* gene

Table 38: Distribution of the *msp4* variants in the different examined animal species

	Dog	Horse	Cat	Hedgehog	Goat	Cattle	Roe deer	Red deer	Sika deer	Fallow deer	Mouflon	Chamois	Red fox	Ibex
m4-1(A)					2									
m4-2(B/C)	19	12	1		2		1							
m4-3(D)	1	1												
m4-4(J)								1						
m4-5(I)								1			4	2		
m4-6(E)								1						
m4-7(F)								2						
m4-8(H)								1			1			
m4-9(G)								1						
m4-10(K)												2		
m4-12(M)							2							
m4-13(N)						1	14		1					2
m4-14(O)						9								
m4-15(P)						1								
m4-16(Q)						1								
m4-17(R)						1								
m4-18(S)		1												
m4-19	1													
m4-20	9			18									3	
m4-21									2		1			
m4-22											1			
m4-23							2	1						

m4-24								1						
m4-25								1						
m4-26								1						
m4-27								1						
m4-28								1						
m4-29								1						
m4-30								1						
m4-31									1					
m4-32									1					
m4-33									1					
m4-34									1					
m4-35							1							
m4-36										2				
m4-37								1						
m4-38								1						
m4-39							1							
m4-40										1				
m4-41							1							
m4-42							1							
m4-43								1						
m4-44									1					
m4-45							1		1					
m4-46									1					
m4-47									1					
m4-48							1							
m4-49						12								
m4-50						4								
m4-51						5								
Total	**30**	**14**	**1**	**18**	**4**	**13**	**25**	**18**	**11**	**3**	**7**	**4**	**3**	**2**

Table 39: Alignment of the *msp4* nucleotide sequences – highlighted nucleotide positions are differing within the variants

```
CLUSTAL 2.1 multiple sequence alignment

m4-7_F_             GGAAGTCGAAGTTGGATAT AAG AAGTTTGAAACGCTT GCTGAG AGT GACTA CAAA CATG  T 60
m4-35               GGAAGTCGAAGTTGGATAT AAG AAGTTTGAAACGCTT GCTGAG AGT GACTA CAAA CATG  T 60
m4-5_I_             GGAAGTTGAAGTTGGATAT AAG AAGTTTGAAACGCTT GCTGAG AGT GACTA CAAA CATG  T 60
m4-24               GGAAGTCGAAGTTGGATAT AAG AAGTTTGAAACGCTT GCTGAG AGT GACTA CAAA CATG  T 60
m4-26               GGAAGTCGAAGTTGGATAT AAG AAGTTTGAAACGCTT GCTGAG AGT GACTA CAAA CATG  T 60
m4-1_A_             GGAAGTGGAAGTTGGATAT AAG AAGTTTGAAACGCTT GCTGAG AGT GACTA CAAG CATA  T 60
m4-41               GGAAGTGGAAGTTGGATAT AAG AAGTTTGAAACGCTT GCTGAG AGT GACTA CAAG CATA  T 60
m4-13_N_            GGAAGTGGAAGTTGGATAT AAG AAGTTTGAAACGCTT GCTGAG AGT GACTA CAAG CATA  T 60
m4-2_B/C_           GGAAGTGGAAGTTGGATAT AAG AAGTTTGAAACGCTT GCTGAG AGT GACTA CAAG CATA  T 60
m4-12_M_            GGAAGTGGAAGTTGGATAT AAG AAGTTTGAAACGCTT GCTGAG AGT GACTA CAAG CATA  T 60
m4-48               GGAAGTGGAAGTTGGATAT AAG AAGTTTGAAACGCTT GCTGAG AGT GACTA CAAG CATA  T 60
m4-40               GGAAGTCGAAGTTGGATAT AAG AAGTTTGAAACGCTT GCTGAG AGT GACTA CAAG CATA  T 60
m4-19               GGAAGTCGAAGTTGGATAT AAG AAGTTTGAAACGCTT GCTGAG AGT GACTA CAAG CATA  T 60
m4-3_D_             GGAAGTCGAAGTTGGATAT AAG AAGTTTGAAACGCTT GCTGAG AGT GACTA CAAG CATA  T 60
m4-51               GGAAGTCGAAGTTGGATAT AAG AAGTTTGAAACGCTT GCTGAG AGT GACTA CAAG CATA  T 60
m4-22               GGAAGTCGAAGTTGGATAT AAG AAGTTTGAAACGCTT GCTGAG AGT GACTA CAAG CATA  T 60
m4-32               GGAAGTCGAAGTTGGATAT AAG AAGTTTGAAACGCTT GCTGAG AGT GACTA CAAG CATA  T 60
m4-8_H_             GGAAGTCGAAGTTGGATAT AAG AAGTTTGAAACGCTT GCTGAG AGT GACTA CAAG CATA  T 60
m4-36               GGAAGTCGAAGTTGGATAT AAG AAGTTTGAAACGCTT GCTGAG AGT GACTA CAAG CATA  T 60
m4-14_O_            GGAAGTCGAAGTTGGATAT AAG AAGTTTGAAACGCTT GCTGAG AGT GACTA CAAG CATA  T 60
```

```
m4                GGAAGTCGAAGTTGGATAT AAG AAGTTTGAAACGCTT GCTGAG AGT GACTA CAAG CATA  T 60
m4-17_R_          GGAAGTCGAAGTTGGATAT AAG AAGTTTGAAACGCTT GCTGAG AGT GACTA CAAG CATG  T 60
m4-27             GGAAGTTGAAGTTGGATAT AAG AAGTTTGAAACGCTT GCTGAG AGT GACTA CAAG CATG  T 60
m4-30             GGAAGTCGAAGTTGGATAT AAG AAGTTTGAAACGCTT GCTGAA AGC GACTA CAAG CATG  T 60
m4-47             GGAAGTCGAAGTTGGATAT AAG AAGTTTGAAACGCTT GCTGAA AGC GACTA CAAG CATG  T 60
m4-9_G_           GGAAGTCGAAGTTGGATAT AAG AAGTTTGAAACGCTT GCTGAA AGT GACTA CAAG CATG  T 60
m4-11_L_          GGAAGTCGAAGTTGGATAT AAG AAGTTTGAAACGCTT GCTGAA AGT GACTA CAAG CATG  T 60
m4-31             GGAAGTCGAAGTTGGATAT AAG AAGTTTGAAACGCTT GCTGAA AGT GACTA CAAG CATG  T 60
m4-46             GGAAGTCGAAGTTGGATAT AAG AAGTTTGAAACGCTT GCTGAA AGC GACTA CAAG CATG  T 60
m4-34             GGAAGTCGAAGTTGGATAT AAG AAGTTTGAAACGCTT GCTGAA AGC GACTA CAAG CATG  T 60
m4-37             GGAAGTCGAAGTTGGATAT AAG AAGTTTGAAACGCTT GCTGAA AGC GACTA CAAG CATG  T 60
m4-25             GGAAGTCGAAGTTGGATAT AAG AAGTTTGAAACGCTT GCTGAA AGT GACTA CAAG CATG  T 60
m4-18_S_          GGAAGTCGAAGTTGGATAT AAG AAGTTTGAAACGCTT GCTGAG AGT GACTA CAAA CATG  T 60
m4-10_K_          GGAAGTTGAAGTTGGATAT AAA AAGTTTGAAACGCTC GCTGAA AGT GACTA CAAG CATG  T 60
m4-44             GGAAGTTGAAGTTGGATAT AAA AAGTTTGAAACGCTC GCTGAA AGT GACTA CAAG CATG  T 60
m4-20             GGAAGTTGAAGTTGGATAT AAA AAGTTTGAAACGCTC GCTGAA AGT GACTA CAAG CATG  T 60
m4-33             GGAAGTTGAAGTTGGATAT AAA AAGTTTGAAACGCTC GCTGAA AGT GACTA CAAG CATG  T 60
m4-50             GGAAGTTGAAGTTGGATAT AAA AAGTTTGAAACGCTC GCTGAA AGT GACTA CAAG CATG  T 60
m4-15_P_          GGAAGTTGAAGTTGGATAT AAA AAGTTTGAAACGCTC GCTGAA AGT GACTA CAAG CATG  T 60
m4-49             GGAAGTTGAAGTTGGATAT AAA AAGTTTGAAACGCTC GCTGAA AGT GACTA CAAG CATG  T 60
m4-16_Q_          GGAAGTTGAAGTTGGATAT AAA AAGTTTGAAACGCTC GCTGAA AGT GACTA CAAG CATG  T 60
m4-38             GGAAGTTGAAGTTGGATAT AAA AAGTTTGAAACGCTC GCTGAA AGT GACTA CAAA CATG  T 60
m4-29             GGAAGTCGAAGTTGGATAT AAG AAGTTTGAAACGCTT GCTGAA AGT GACTA CAAA CATG  T 60
m4-4_J_           GGAAGTTGAAGTTGGATAT AAG AAGTTTGAAACGCTT GCTGAG AGT GACTA CAAG CATG  T 60
m4-6_E_           GGAAGTCGAAGTTGGATAT AAG AAGTTTGAAACGCTC GCTGAA AGT GACTA CAAA CATG  T 60
m4-21             GGAAGTCGAAGTTGGATAT AAG AAGTTTGAAACGCTT GCTGAA AGT GACTA CAAA CATG  T 60
m4-42             GGAAGTCGAAGTTGGATAT AAG AAGTTTGAAACGCTT GCTGAA AGT GACTA CAAA CATG  T 60
m4-43             GGAAGTCGAAGTTGGATAT AAG AAGTTTGAAACGCTT GCTGAG AGT GACTA CAAA CATG  T 60
m4-23             GGAAGTTGAAGTTGGATAC AAA AAGTTTGAAACGCTA GCTGAC AGT GACTT CAAA CATG  C 60
m4-45             GGAAGTTGAAGTTGGATAC AAA AAGTTTGAAACGCTA GCTGAC AGT GACTT CAAA CATG  C 60
m4-39             GGAAGTTGAAGTTGGATAC AAA AAGTTTGAAACGCTA GCTGAC AGT GACTT CAAA CATG  C 60
Position          ******7**********19**22**************37*****43**46****51***55***59 60

m4-7_F_       AGAG TCG CATAATTTTGTTGCT GTG GGC CGTGAT  G  C GA CATTAACTCC   G  GACAACTTCTT 120
m4-35         AGAG TCG CATAATTTTGTTGCT GTG GGC CGTGAT  G  C GA CATTAACTCT   G  GACAACTTCTT 120
m4-5_I_       AGAG TCG CATAATTTTGTTGCT GTG GGC CGTGAT  G  C GA CATTAACTCC   G  GACAACTTCTT 120
m4-24         AGAG TCG CATAATTTTGTTGCT GTG GGC CGTGAC  G  C GA CATTAACTCC   G  GACAACTTCTT 120
m4-26         AGAG TCG CATAATTTTGTTGCT GTG GGC CGTGAT  G  C GA CATTAACTCC   G  GACAACTTCTT 120
m4-1_A_       AGAA TCA CATAATTTTGTTGCT GTG GGT CGTGAT  G  C GA CATTAACTCC   G  GACAACTTCTT 120
m4-41         AGAA TCA CATAATTTTGTTGCT GTG GGT CGTGAT  G  C GA CATTAACTCC   G  GACAACTTCTT 120
m4-13_N_      AGAA TCA CATAATTTTGTTGCT GTG GGT CGTGAT  G  C GA CATTAACTCC   G  GACAACTTCTT 120
m4-2_B/C_     AGAA TCA CATAATTTTGTTGCT GTG GGT CGTGAT  G  C GA CATTAACTCC   G  GACAACTTCTT 120
m4-12_M_      AGAA TCG CATAATTTTGTTGCT GTG GGT CGTGAT  G  C GA CATTAACTCC   G  GACAACTTCTT 120
m4-48         AGAA TCG CATAATTTTGTTGCT GTG GGT CGTGAT  A  C GA CATTAACTCC   G  GACAACTTCTT 120
m4-40         AGAA TCA CATAATTTTGTTGCT GTG GGT CGTGAT  G  C GA CATTAACTCC   G  GACAACTTCTT 120
m4-19         AGAA TCA CATAATTTTGTTGCT GTG GGT CGTGAC  G  C GA CATTAACTCC   G  GACAACTTCTT 120
m4-3_D_       AGAA TCA CATAATTTTGTTGCT GTG GGT CGTGAT  G  C GA CATTAACTCC   G  GACAACTTCTT 120
m4-51         AGAA TCA CATAATTTTGTTGCT GTG GGT CGTGAT  G  C GA CATTAACTCC   G  GACAACTTCTT 120
m4-22         AGAA TCA CATAATTTTGTTGCT GTG GGT CGTGAT  G  C GA CATTAACTCC   G  GACAACTTCTT 120
m4-32         AGAA TCA CATAATTTTGTTGCT GTG GGT CGTGAT  G  C GA CATTAACTCC   G  GACAACTTCTT 120
m4-8_H_       AGAA TCA CATAATTTTGTTGCT GTG GGT CGTGAT  G  C GA CATTAACTCC   G  GACAACTTCTT 120
m4-36         AGAA TCA CATAATTTTGTTGCT GTG GGT CGTGAT  G  C GA CATTAACTCC   G  GACAACTTCTT 120
m4-14_O_      AGAA TCA CATAATTTTGTTGCT GTG GGT CGTGAT  G  C GA CATTAACTCC   G  GACAACTTCTT 120
m4            AGAA TCA CATAATTTTGTTGCT GTG GGC CGTGAT  G  C GA CATTAACTCC   G  GACAACTTCTT 120
m4-17_R_      AGAA TCA CATAATTTTGTTGCC GTG GGT CGTGAC  G  C GA CATTAACTCC   G  GACAACTTCTT 120
m4-27         AGAA TCA CATAATTTTGTTGCT GTG GGT CGTGAC  G  C GA CATTAACTCC   A  GACAACTTCTT 120
m4-30         AGAA TCA CATAATTTTGTTGCT GTG GGT CGTGAC  G  C GA CATTAACTCC   A  GACAACTTCTT 120
m4-47         AGAA TCA CATAATTTTGTTGCT GTG GGT CGTGAC  G  C GA CATTAACTCC   A  GACAACTTCTT 120
m4-9_G_       AGAA TCA CATAATTTTGTTGCT GTG GGT CGTGAT  G  C GA CATTAACTCC   G  GACAACTTCTT 120
m4-11_L_      AGAA TCA CATAATTTTGTTGCT GTG GGT CGTGAT  G  C GA CATTAACTCC   G  GACAACTTCTT 120
m4-31         AGAA TCA CATAATTTTGTTGCT GTG GGT CGTGAC  G  C GA CATTAACTCC   G  GACAACTTCTT 120
m4-46         AGAA TCA CATAATTTTGTTGCT GTG GGT CGTGAC  G  C GA CATTAACTCC   G  GACAACTTCTT 120
m4-34         AGAA TCA CATAATTTTGTTGCT GTG GGT CGTGAC  G  C GA CATTAACTCC   A  GACAACTTCTT 120
m4-37         AGAA TCA CATAATTTTGTTGCT GTG GGT CGTGAC  G  C GA CATTAACTCC   G  GACAACTTCTT 120
m4-25         AGAA TCA CATAATTTTGTTGCT GTG GGT CGTGAT  G  C GA CATTAACTCC   G  GACAACTTCTT 120
m4-18_S_      AGAA TCA CATAATTTTGTTGCT GTG GGT CGTGAC  G  C GA CATTAACTCC   G  GACAACTTCTT 120
m4-10_K_      AGAA TCA CATAATTTTGTTGCT GTG GGT CGTGAC  G  C GA CATTAACTCC   G  GACAACTTCTT 120
m4-44         AGAA TCA CATAATTTTGTTGCT GTG GGT CGTGAC  G  C GA CATTAACTCC   G  GACAACTTCTT 120
m4-20         AGAA TCA CATAATTTTGTTGCT GTG GGT CGTGAC  G  C GA CATTAACTCC   G  GACAACTTCTT 120
m4-33         AGAA TCA CATAATTTTGTTGCT GTG GGT CGTGAC  G  C GA CATTAACTCC   G  GACAACTTCTT 120
m4-50         AGAA TCA CATAATTTTGTTGCC GTG GGT CGTGAC  G  C GA CATTAACTCC   G  GACAACTTCTT 120
m4-15_P_      AGAA TCA CATAATTTTGTTGCC GTG GGT CGTGAC  G  C GA CATTAACTCC   G  GACAACTTCTT 120
m4-49         AGAA TCA CATAATTTTGTTGCT GTG GGT CGTGAC  G  C GA CATTAACTCC   G  GACAACTTCTT 120
m4-16_Q_      AGAA TCA CATAATTTTGTTGCT GTG GGT CGTGAT  G  C GA CATTAACTCC   G  GACAACTTCTT 120
m4-38         AGAA TCA CATAATTTTGTTGCT GTG GGC CGTGAC  G  C GA CATTAACTCC   G  GACAACTTCTT 120
m4-29         AGAR TCG CATAATTTTGTTGCT GTG GGT CGTGAT  G  C GA CATTAACTCC   G  GACAACTTCTT 120
m4-4_J_       AGAG TCG CATAATTTTGTTGCT GTG GGC CGTGAT  G  C GA CATTAACTCC   G  GACAACTTCTT 120
m4-6_E_       AGAG TCG CATAATTTTGTTGCT GTG GGT CGTGAC  G  C GA CATTAACTCC   G  GACAACTTCTT 120
m4-21         AGAG TCG CATAATTTTGTTGCT GTG GGC CGTGAT  G  C GA CATTAACTCC   G  GACAACTTCTT 120
m4-42         AGAG TCG CATAATTTTGTTGCT GTG GGC CGTGAT  G  C GA CATTAACTCC   G  GACAACTTCTT 120
m4-43         AGAG TCG CATAATTTTGTTGCT GTG GGC CGTGAT  G  C GA CATTAACTCC   G  GACAACTTCTT 120
m4-23         AGAG TCG CATAATTTTGTTGCT GTA GGC CGTGAT  A  G GG CATTAACTCC   A  GACAACTTCTT 120
m4-45         AGAG TCG CATAATTTTGTTGCT GTA GGC CGTGAT  A  T GG CATTAACTCC   A  GACAACTTCTT 120
m4-39         AGAG TCG CATAATTTTGTTGCT GTA GGC CGTGAT  A  G GG CATTAACTCC   A  GACAACTTCTT 120
Position      ***64**67**************82**85**88******94 95 96*98**********108 109***********

m4-7_F_        TGTA  ATGAAAATAGAC   A  GC  GTC  AAAGAT  ATATCTGTA  ATGCTTAACGCTTGTTACGACGT 180
m4-35          TGTA  ATGAAAATAGAC   A  GC  GTC  AAAGAT  ATATCTGTA  ATGCTTAACGCTTGTTACGACGT 180
m4-5_I_        TGTA  ATGAAAATAGAC   A  GC  GTC  AAAGAT  ATATCTGTA  ATGCTTAACGCTTGTTACGACGT 180
m4-24          TGTA  ATGAAAATAGAC   A  GC  GTC  AAAGAT  ATATCTGTA  ATGCTTAACGCTTGTTACGACGT 180
```

```
m4-26          TGTA  ATGAAAATAGAC   A  GT  GTC  AAAGAT  ATATCTGTA  ATGCTTAACGCTTGTTACGACGT 180
m4-1_A_        TGTA  ATGAAAATAGAA   A  GC  GTC  AAAGAC  ATATCTGTA  ATGCTTAACGCTTGTTACGACGT 180
m4-41          TGTA  ATGAAAATAGAC   A  GC  GTC  AAAGAC  ATATCTGTA  ATGCTTAACGCTTGTTACGACGT 180
m4-13_N_       TGTA  ATGAAAATAGAC   A  GC  GTC  AAAGAC  ATATCTGTA  ATGCTTAACGCTTGTTACGACGT 180
m4-2_B/C_      TGTA  ATGAAAATAGAC   A  GC  GTC  AAAGAC  ATATCTGTA  ATGCTTAACGCTTGTTACGACGT 180
m4-12_M_       TGTA  ATGAAAATAGAC   A  GC  GTC  AAAGAC  ATATCTGTA  ATGCTTAACGCTTGTTACGACGT 180
m4-48          TGTA  ATGAAAATAGAC   A  GC  GTC  AAAGAC  ATATCTGTA  ATGCTTAACGCTTGTTACGACGT 180
m4-40          TGTA  ATGAAAATAGAC   A  GC  GTC  AAAGAC  ATATCTGTA  ATGCTTAACGCTTGTTACGACGT 180
m4-19          TGTA  ATGAAAATAGAC   A  GC  GTT  AAAGAC  ATATCTGTA  ATGCTTAACGCTTGTTACGACGT 180
m4-3_D_        TGTA  ATGAAAATAGAC   A  GC  GTC  AAAGAT  ATATCTGTA  ATGCTTAACGCTTGTTACGACGT 180
m4-51          TGTA  ATGAAAATAGAC   A  GC  GTC  AAAGAT  ATATCTGTA  ATGCTTAACGCTTGTTACGACGT 180
m4-22          TGTA  ATGAAAATAGAC   A  GC  GTC  AAAGAT  ATATCTGTA  ATGCTTAACGCTTGTTACGACGT 180
m4-32          TGTA  ATGAAAATAGAC   A  GC  GTC  AAAGAT  ATATCTGTA  ATGCTTAACGCTTGTTACGACGT 180
m4-8_H_        TGTA  ATGAAAATAGAC   A  GC  GTC  AAAGAT  ATATCTGTA  ATGCTTAACGCTTGTTACGACGT 180
m4-36          TGTA  ATGAAAATAGAC   A  GC  GTC  AAAGAT  ATATCTGTA  ATGCTTAACGCTTGTTACGACGT 180
m4-14_O_       TGTA  ATGAAAATAGAC   A  GC  GTC  AAAGAT  ATATCTGTA  ATGCTTAACGCTTGTTACGACGT 180
m4             TGTA  ATGAAAATAGAC   A  GC  GTC  AAAGAC  ATATCTGTA  ATGCTTAACGCTTGTTACGACGT 180
m4-17_R_       TGTA  ATGAAAATAGAC   A  GC  GTC  AAAGAT  ATATCTGTA  ATGCTTAACGCTTGTTACGACGT 180
m4-27          TGTA  ATGAAAATAGAC   A  GC  GTC  AAAGAT  ATATCTGTA  ATGCTTAACGCTTGTTACGACGT 180
m4-30          TGTG  ATGAAAATAGAC   A  GC  GTC  AAAGAT  ATATCTGTA  ATGCTTAACGCTTGTTACGACGT 180
m4-47          TGTG  ATGAAAATAGAC   A  GC  GTC  AAAGAT  ATATCTGTA  ATGCTTAACGCTTGTTACGACGT 180
m4-9_G_        TGTA  ATGAAAATAGAC   A  GC  GTC  AAAGAC  ATATCTGTA  ATGCTTAACGCTTGTTACGACGT 180
m4-11_L_       TGTA  ATGAAAATAGAC   A  GC  GTC  AAAGAT  ATATCTGTA  ATGCTTAACGCTTGTTACGACGT 180
m4-31          TGTA  ATGAAAATAGAC   A  GC  GTC  AAAGAT  ATATCTGTA  ATGCTTAACGCTTGTTACGACGT 180
m4-46          TGTA  ATGAAAATAGAC   A  GC  GTC  AAAGAT  ATATCTGTA  ATGCTTAACGCTTGTTACGACGT 180
m4-34          TGTA  ATGAAAATAGAC   A  GC  GTC  AAAGAT  ATATCTGTA  ATGCTTAACGCTTGTTACGACGT 180
m4-37          TGTA  ATGAAAATAGAC   A  GC  GTC  AAAGAT  ATATCTGTA  ATGCTTAACGCTTGTTACGACGT 180
m4-25          TGTA  ATGAAAATAGAC   A  GC  GTC  AAAGAT  ATATCTGTA  ATGCTTAACGCTTGTTACGACGT 180
m4-18_S_       TGTA  ATGAAAATAGAC   A  GT  GTC  AAAGAT  ATATCTGTA  ATGCTTAACGCTTGTTACGACGT 180
m4-10_K_       TGTA  ATGAAAATAGAC   A  GT  GTC  AAAGAT  ATATCTGTA  ATGCTTAACGCTTGTTACGACGT 180
m4-44          TGTA  ATGAAAATAGAC   A  GT  GTC  AAAGAT  ATATCTGTA  ATGCTTAACGCTTGTTACGACGT 180
m4-20          TGTA  ATGAAAATAGAC   A  GC  GTT  AAAGAC  ATATCTGTA  ATGCTTAACGCTTGTTACGACGT 180
m4-33          TGTA  ATGAAAATAGAC   A  GC  GTT  AAAGAC  ATATCTGTA  ATGCTTAACGCTTGTTACGACGT 180
m4-50          TGTA  ATGAAAATAGAC   A  GC  GTT  AAAGAC  ATATCTGTA  ATGCTTAACGCTTGTTACGACGT 180
m4-15_P_       TGTA  ATGAAAATAGAC   A  GC  GTC  AAAGAT  ATATCTGTA  ATGCTTAACGCTTGTTACGACGT 180
m4-49          TGTA  ATGAAAATAGAC   A  GC  GTC  AAAGAT  ATATCTGTA  ATGCTTAACGCTTGTTACGACGT 180
m4-16_Q_       TGTA  ATGAAAATAGAC   A  GC  GTC  AAAGAT  ATATCTGTA  ATGCTTAACGCTTGTTACGACGT 180
m4-38          TGTA  ATGAAAATAGAC   A  GC  GTC  AAAGAT  ATATCTGTA  ATGCTTAACGCTTGTTACGACGT 180
m4-29          TGTA  ATGAAAATAGAC   A  GC  GTC  AAAGAT  ATATCTGTA  ATGCTTAACGCTTGTTACGACGT 180
m4-4_J_        TGTA  ATGAAAATAGAC   A  GC  GTC  AAAGAT  ATATCTGTA  ATGCTTAACGCTTGTTACGACGT 180
m4-6_E_        TGTA  ATGAAAATAGAC   A  GC  GTC  AAAGAC  ATATCTGTA  ATGCTTAACGCTTGTTACGACGT 180
m4-21          TGTA  ATGAAAATAGAC   A  GC  GTC  AAAGAT  ATATCTGTA  ATGCTTAACGCTTGTTACGACGT 180
m4-42          TGTA  ATGAAAATAGAC   A  GC  GTC  AAAGAC  ATATCTGTA  ATGCTTAACGCTTGTTACGACGT 180
m4-43          TGTA  ATGAAAATAGAC   A  GC  GTC  AAAGAC  ATATCTGTA  ATGCTTAACGCTTGTTACGACGT 180
m4-23          TGTT  ATGAAAATAGAC   G  GT  GTT  AAAGAT  ATATCTGTG  ATGCTTAACGCTTGTTACGACGT 180
m4-45          TGTT  ATGAAAATAGAC   G  GT  GTT  AAAGAT  ATATCTGTG  ATGCTTAACGCTTGTTACGACGT 180
m4-39          TGTT  ATGAAAATAGAC   G  GT  GTT  AAAGAT  ATATCTGTG  ATGCTTAACGCTTGTTACGACGT 180
Position       ***124***********136 137*139**142*****148********157**********************

m4-7_F_ T  ATGCATACTGAC  TTGCCT  GTA  TCTCCT  TAC  ATGTGC  GCT  GGGT  TAGGG  GCGAGT  TTTAT 240
m4-35   C  ATGCATACTGAC  TTGCCT  GTA  TCTCCT  TAC  ATGTGC  GCT  GGGT  TAGGG  GCGAGT  TTTAT 240
m4-5_I_ T  ATGCATACTGAC  TTGCCT  GTA  TCTCCT  TAC  ATGTGT  GCT  GGGT  TAGGG  GCGAGT  TTTAT 240
m4-24   T  ATGCATACTGAC  TTGCCT  GTA  TCTCCT  TAC  ATGTGT  GCT  GGGT  TAGGG  GCGAGT  TTTAT 240
m4-26   T  ATGCATACTGAC  TTGCCT  GTA  TCTCCT  TAC  ATGTGT  GCT  GGGT  TAGGG  GCGAGC  TTTAT 240
m4-1_A_ T  ATGCATACTGAC  TTGCCT  GTA  TCTCCT  TAC  ATGTGT  GCT  GGGT  TAGGG  GCGAGT  TTTAT 240
m4-41   T  ATGCATACTGAC  TTGCCT  GTA  TCTCCT  TAC  ATGTGT  GCT  GGGT  TAGGG  GCGAGT  TTTAT 240
m4-13_N_T  ATGCATACTGAC  TTGCCT  GTA  TCTCCT  TAC  ATGTGT  GCT  GGGT  TAGGG  GCGAGT  TTTAT 240
m4-2_B/CT  ATGCATACTGAC  TTGCCT  GTA  TCTCCT  TAC  ATGTGT  GCT  GGGC  TAGGG  GCGAGT  TTTAT 240
m4-12_M_T  ATGCATACTGAC  TTGCCT  GTA  TCTCCT  TAC  ATGTGT  GCT  GGGT  TAGGG  GCGAGT  TTTAT 240
m4-48   T  ATGCATACTGAC  TTGCCT  GTA  TCTCCT  TAC  ATGTGT  GCT  GGGT  TAGGG  GCGAGT  TTTAT 240
m4-40   T  ATGCATACTGAC  TTGCCT  GTA  TCTCCT  TAC  ATGTGT  GCT  GGGT  TAGGG  GCGAGT  TTTAT 240
m4-19   T  ATGCATACTGAC  TTGCCT  GTA  TCTCCT  TAC  ATGTGT  GCT  GGGT  TAGGG  GCGAGT  TTTAT 240
m4-3_D_ T  ATGCATACTGAC  TTGCCT  GTA  TCTCCT  TAC  ATGTGT  GCT  GGGT  TAGGG  GCGAGT  TTTAT 240
m4-51   T  ATGCATACTGAC  TTGCCT  GTA  TCTCCT  TAC  ATGTGT  GCT  GGGT  TAGGG  GCGAGC  TTTAT 240
m4-22   T  ATGCATACTGAC  TTGCCT  GTG  TCTCCT  TAC  ATGTGT  GCT  GGGT  TAGGG  GCGAGT  TTTAT 240
m4-32   T  ATGCATACTGAC  TTGCCT  GTG  TCTCCT  TAC  ATGTGT  GCT  GGGT  TAGGG  GCGAGT  TTTAT 240
m4-8_H_ T  ATGCATACTGAC  TTGCCT  GTA  TCTCCT  TAC  ATGTGT  GCT  GGGT  TAGGG  GCGAGT  TTTAT 240
m4-36   T  ATGCATACTGAC  TTGCCT  GTA  TCTCCT  TAC  ATGTGT  GCT  GGGT  TAGGG  GCGAGT  TTTAT 240
m4-14_O_T  ATGCATACTGAC  TTGCCT  GTA  TCTCCT  TAC  ATGTGT  GCT  GGGT  TAGGG  GCGAGC  TTTAT 240
m4      T  ATGCATACTGAT  TTGCCT  GTA  TCTCCT  TAC  ATGTGT  GCT  GGGT  TAGGG  GCGAGT  TTTAT 240
m4-17_R_T  ATGCATACTGAC  TTGCCT  GTA  TCTCCT  TAC  ATGTGT  GCT  GGGT  TAGGG  GCGAGC  TTTAT 240
m4-27   T  ATGCATACTGAC  TTGCCT  GTA  TCTCCT  TAC  ATGTGT  GCT  GGGT  TAGGG  GCGAGC  TTTAT 240
m4-30   T  ATGCATACTGAC  TTGCCT  GTA  TCTCCT  TAC  ATGTGT  GCT  GGGT  TAGGG  GCGAGT  TTTAT 240
m4-47   T  ATGCATACTGAC  TTGCCT  GTA  TCTCCT  TAC  ATGTGT  GCT  GGGT  TAGGG  GCGAGT  TTTAT 240
m4-9_G_ T  ATGCATACTGAC  TTGCCG  GTA  TCTCCT  TAC  ATGTGT  GCT  GGGT  TAGGG  GCGAGC  TTTAT 240
m4-11_L_T  ATGCATACTGAC  TTGCCT  GTA  TCTCCT  TAC  ATGTGT  GCT  GGGT  TAGGG  GCGAGT  TTTAT 240
m4-31   T  ATGCATACTGAC  TTGCCT  GTA  TCTCCT  TAC  ATGTGT  GCT  GGGT  TAGGG  GCGAGT  TTTAT 240
m4-46   T  ATGCATACTGAC  TTGCCT  GTA  TCTCCT  TAC  ATGTGT  GCT  GGGT  TAGGG  GCGAGT  TTTAT 240
m4-34   T  ATGCATACTGAC  TTGCCT  GTA  TCTCCT  TAC  ATGTGT  GCT  GGGT  TAGGG  GCGAGT  TTTAT 240
m4-37   T  ATGCATACTGAC  TTGCCT  GTA  TCTCCT  TAC  ATGTGT  GCT  GGGT  TAGGG  GCGAGT  TTTAT 240
m4-25   T  ATGCATACTGAC  TTGCCT  GTA  TCTCCT  TAC  ATGTGT  GCT  GGGT  TAGGG  GCGAGT  TTTAT 240
m4-18_S_T  ATGCATACTGAC  TTGCCT  GTA  TCTCCT  TAC  ATGTGT  GCC  GGGT  TAGGG  GCGAGT  TTTAT 240
m4-10_K_T  ATGCATACTGAC  TTGCCT  GTA  TCTCCT  TAC  ATGTGT  GCT  GGGT  TAGGG  GCGAGC  TTTAT 240
m4-44   T  ATGCATACTGAC  TTGCCT  GTA  TCTCCT  TAC  ATGTGT  GCT  GGGT  TAGGG  GCGAGC  TTTAT 240
m4-20   T  ATGCATACTGAC  TTGCCT  GTA  TCTCCT  TAC  ATGTGT  GCT  GGGT  TAGGG  GCGAGT  TTTAT 240
m4-33   T  ATGCATACTGAC  TTGCCT  GTA  TCTCCT  TAC  ATGTGT  GCT  GGGT  TAGGG  GCGAGT  TTTAT 240
m4-50   T  ATGCATACTGAC  TTGCCT  GTA  TCTCCT  TAC  ATGTGT  GCT  GGGT  TAGGG  GCGAGC  TTTAT 240
m4-15_P_T  ATGCATACTGAC  TTGCCT  GTA  TCTCCT  TAC  ATGTGT  GCT  GGGT  TAGGG  GCGAGC  TTTAT 240
m4-49   T  ATGCATACTGAC  TTGCCT  GTA  TCTCCT  TAC  ATGTGT  GCT  GGGT  TAGGG  GCGAGC  TTTAT 240
m4-16_Q_T  ATGCATACTGAC  TTGCCT  GTA  TCTCCT  TAC  ATGTGT  GCT  GGGT  TAGGG  GCGAGC  TTTAT 240
m4-38   T  ATGCATACTGAC  TTGCCT  GTA  TCTCCT  TAC  ATGTGT  GCT  GGGT  TAGGG  GCGAGT  TTTAT 240
```

```
m4-29    T  ATGCATACTGAC  TTGCCT  GTA  TCTCCT  TAC  ATGTGT  GCT  GGGT  TAGGG  GCGAGY  TTTAT 240
m4-4_J_  T  ATGCATACTGAC  TTGCCT  GTA  TCTCCT  TAC  ATGTGT  GCT  GGGT  TAGGG  GCGAGT  TTTAT 240
m4-6_E_  T  ATGCATACTGAC  TTGCCT  GTC  TCTCCT  TAC  ATGTGT  GCT  GGGT  TAGGG  GCGAGC  TTTAT 240
m4-21    T  ATGCATACTGAC  TTGCCT  GTA  TCTCCT  TAC  ATGTGC  GCT  GGGT  TAGGG  GCGAGC  TTTAT 240
m4-42    T  ATGCATACTGAT  TTGCCT  GTA  TCTCCT  TAC  ATGTGT  GCT  GGGT  TAGGG  GCGAGT  TTTAT 240
m4-43    T  ATGCATACTGAC  TTGCCT  GTA  TCTCCT  TAC  ATGTGT  GCT  GGGT  TAGGG  GCGAGT  TTTAT 240
m4-23    T  ATGCATACTGAT  TTGCCT  GTA  TCTCCC  TAT  ATGTGT  GCC  GGGC  TAGGA  GCGAGT  TTTAT 240
m4-45    T  ATGCATACTGAT  TTGCCT  GTA  TCTCCC  TAT  ATGTGT  GCC  GGGC  TAGGA  GCGAGT  TTTAT 240
m4-39    T  ATGCATACTGAT  TTGCCT  GTA  TCTCCC  TAT  ATGTGT  GCC  GGGC  TAGGG  GCGAGT  TTTAT 240
Position181**********193*****199**202*****208**211*****217**220***224****229*****235*****

m4-7_F_   TA  AT  ATC  GCC   G  ACCACGTTACAAGC  AAGTTGGCT  TATAGAGGAAAG  GTAGGC  GTA  AGTTA 300
m4-35     TA  AT  ATC  GCC   G  ACCACGTTACAAGC  AAGTTGGCT  TATAGAGGAAAG  GTAGGC  GTA  AGTTA 300
m4-5_I_   TA  AT  ATC  GCC   G  ACCACGTTACAAGC  AAGTTGGCT  TATAGAGGAAAG  GTAGGC  GTA  AGTTA 300
m4-24     TA  AT  ATC  GCC   G  ACCACGTTACAAGC  AAGTTGGCT  TATAGAGGAAAG  GTAGGC  GTA  AGTTA 300
m4-26     TA  AT  ATC  GCA   G  ACCACGTTACAAGC  AAGTTGGCT  TATAGAGGAAAG  GTAGGC  GTA  AGTTA 300
m4-1_A_   TA  AC  ATC  GCC   G  ACCACGTTACAAGC  AAGTTGGCT  TATAGAGGAAAG  GTAGGC  GTA  AGTTA 300
m4-41     TA  AC  ATC  GCC   G  ACCACGTTACAAGC  AAGTTGGCT  TATAGAGGAAAA  GTAGGC  GTA  AGTTA 300
m4-13_N_  TA  AC  ATC  GCC   G  ACCACGTTACAAGC  AAGTTGGCT  TATAGAGGAAAG  GTAGGC  GTA  AGTTA 300
m4-2_B/C_ TA  AC  ATC  GCC   G  ACCACGTTACAAGC  AAGTTGGCT  TATAGAGGAAAG  GTAGGC  GTA  AGTTA 300
m4-12_M_  TA  AC  ATC  GCC   G  ACCACGTTACAAGC  AAGTTGGCT  TATAGAGGAAAG  GTAGGC  GTA  AGTTA 300
m4-48     TA  AC  ATC  GCC   G  ACCACGTTACAAGC  AAGTTGGCT  TATAGAGGAAAG  GTAGGC  GTA  AGTTA 300
m4-40     TA  AC  ATC  GCC   G  ACCACGTTACAAGC  AAGTTGGCT  TATAGAGGAAAG  GTAGGC  GTA  AGTTA 300
m4-19     TA  AT  ATC  GCA   G  ACCACGTTACAAGC  AAGTTGGCT  TATAGAGGAAAG  GTAGGC  GTA  AGTTA 300
m4-3_D_   TA  AT  ATC  GCC   G  ACCACGTTACAAGC  AAGTTGGCT  TATAGAGGAAAG  GTAGGC  GTA  AGTTA 300
m4-51     TA  AT  ATC  GCC   G  ACCACGTTACAAGC  AAGTTGGCT  TATAGAGGAAAG  GTAGGC  GTA  AGTTA 300
m4-22     TA  AT  ATC  GCC   G  ACCACGTTACAAGC  AAGTTGGCT  TATAGAGGAAAG  GTAGGC  GTA  AGTTA 300
m4-32     TA  AT  ATC  GCC   G  ACCACGTTACAAGT  AAGTTGGCT  TATAGAGGAAAG  GTAGGC  GTA  AGTTA 300
m4-8_H_   TA  AT  ATC  GCC   G  ACCACGTTACAAGT  AAGTTGGCT  TATAGAGGAAAG  GTAGGC  GTA  AGTTA 300
m4-36     TA  AT  ATC  GCC   G  ACCACGTTACAAGC  AAGTTGGCT  TATAGAGGAAAG  GTAGGC  GTA  AGTTA 300
m4-14_O_  TA  AT  ATC  GCC   G  ACCACGTTACAAGC  AAGTTGGCT  TATAGAGGAAAG  GTAGGC  GTA  AGTTA 300
m4        TA  AC  ATC  GCC   G  ACCACGTTACAAGC  AAGTTGGCT  TATAGAGGAAAG  GTAGGC  GTA  AGTTA 300
m4-17_R_  TA  AT  ATC  GCC   G  ACCACGTTACAAGC  AAGTTGGCT  TATAGAGGAAAG  GTAGGC  GTA  AGTTA 300
m4-27     TA  AT  ATC  GCC   G  ACCACGTTACAAGC  AAGTTGGCT  TATAGAGGAAAG  GTAGGC  GTA  AGTTA 300
m4-30     TA  AC  ATC  GCC   G  ACCACGTTACAAGC  AAGTTGGCT  TATAGAGGAAAG  GTAGGC  GTA  AGTTA 300
m4-47     TA  AT  ATC  GCC   G  ACCACGTTACAAGC  AAGTTGGCT  TATAGAGGAAAG  GTAGGC  GTA  AGTTA 300
m4-9_G_   TA  AT  ATC  GCC   G  ACCACGTTACAAGC  AAGTTGGCT  TATAGAGGAAAG  GTAGGC  GTA  AGTTA 300
m4-11_L_  TA  AT  ATC  GCC   G  ACCACGTTACAAGC  AAGTTGGCT  TATAGAGGAAAG  GTAGGC  GTA  AGTTA 300
m4-31     TA  AT  ATC  GCC   G  ACCACGTTACAAGT  AAGTTGGCT  TATAGAGGAAAG  GTAGGC  GTG  AGTTA 300
m4-46     TA  AT  ATC  GCC   G  ACCACGTTACAAGT  AAGTTGGCT  TATAGAGGAAAG  GTAGGC  GTG  AGTTA 300
m4-34     TA  AT  ATC  GCC   G  ACCACGTTACAAGT  AAGTTGGCT  TATAGAGGAAAG  GTAGGC  GTA  AGTTA 300
m4-37     TA  AT  ATC  GCC   G  ACCACGTTACAAGT  AAGTTGGCT  TATAGAGGAAAG  GTAGGC  GTA  AGTTA 300
m4-25     TA  AT  ATC  GCC   G  ACCACGTTACAAGT  AAGTTGGCT  TATAGAGGAAAG  GTAGGC  GTA  AGTTA 300
m4-18_S_  TA  AT  ATC  GCC   G  ACCACGTTACAAGC  AAGTTGGCT  TATAGAGGAAAG  GTAGGC  GTA  AGTTA 300
m4-10_K_  TA  AT  ATC  GCA   G  ACCACGTTACAAGC  AAGTTGGCT  TATAGAGGAAAG  GTAGGC  GTA  AGTTA 300
m4-44     TA  AT  ATC  GCA   G  ACCACGTTACAAGC  AAGTTGGCT  TATAGAGGAAAG  GTAGGC  GTA  AGTTA 300
m4-20     TA  AT  ATC  GCA   G  ACCACGTTACAAGC  AAGTTGGCT  TATAGAGGAAAG  GTAGGC  GTA  AGTTA 300
m4-33     TA  AT  ATC  GCA   G  ACCACGTTACAAGC  AAGTTGGCT  TATAGAGGAAAG  GTAGGC  GTA  AGTTA 300
m4-50     TA  AT  ATC  GCC   G  ACCACGTTACAAGC  AAGTTGGCT  TATAGAGGAAAG  GTAGGC  GTA  AGTTA 300
m4-15_P_  TA  AT  ATC  GCC   G  ACCACGTTACAAGC  AAGTTGGCT  TATAGAGGAAAG  GTAGGC  GTA  AGTTA 300
m4-49     TA  AT  ATC  GCC   G  ACCACGTTACAAGC  AAGTTGGCT  TATAGAGGAAAG  GTAGGC  GTA  AGTTA 300
m4-16_Q_  TA  AT  ATC  GCC   G  ACCACGTTACAAGC  AAGTTGGCT  TATAGAGGAAAG  GTAGGC  GTA  AGTTA 300
m4-38     TA  AT  ATC  GCC   G  ACCACGTTACAAGC  AAGTTGGCT  TATAGAGGAAAG  GTAGGC  GTA  AGTTA 300
m4-29     TA  AT  ATC  GCC   G  ACCACGTTACAAGC  AAGTTGGCT  TATAGAGGAAAG  GTAGGC  GTA  AGTTA 300
m4-4_J_   TA  AT  ATC  GCC   G  ACCACGTTACAAGC  AAGTTGGCT  TATAGAGGAAAG  GTAGGC  GTA  AGTTA 300
m4-6_E_   TA  AT  ATC  GCC   G  ACCACGTTACAAGC  AAGTTGGCT  TATAGAGGAAAG  GTAGGC  GTA  AGTTA 300
m4-21     TA  AT  ATC  GCC   G  ACCACGTTACAAGC  AAGTTGGCT  TATAGAGGAAAG  GTAGGC  GTA  AGTTA 300
m4-42     TA  AT  ATC  GCA   G  ACCACGTTACAAGT  AAGTTGGCT  TATAGAGGAAAG  GTAGGC  GTA  AGTTA 300
m4-43     TA  AT  ATC  GCC   G  ACCACGTTACAAGC  AAGTTGGCT  TATAGAGGAAAG  GTAGGC  GTA  AGTTA 300
m4-23     TG  AC  ATA  GCC   A  ACCACGTTACAAGC  AAGTTGGCA  TATAGAGGAAAG  GTAGGT  GTA  AGTTA 300
m4-45     TG  AC  ATA  GCC   A  ACCACGTTACAAGC  AAGTTGGCA  TATAGAGGAAAG  GTAGGT  GTA  AGTTA 300
m4-39     TG  AC  ATA  GCC   A  ACCACGTTACAAGC  AAGTTGGCA  TATAGAGGAAAG  GTAGGT  GTA  AGTTA 300
Position *242*244**247**250 251*************265********274***********286*****292**295*****

m4-7_F_   CAA   G  CTT  ACT  CCTGAA  ATA  TCTTTAATAGCT  GGAGGTTTT 340
m4-35     CAA   G  CTT  ACT  CCTGAA  ATA  TCTTTAATAGCT  GGAGGTTTT 340
m4-5_I_   CAA   G  CTT  ACT  CCTGAA  ATA  TCTTTAATAGCT  GGAGGTTTT 340
m4-24     CAA   G  CTT  ACT  CCTGAA  ATA  TCTTTAATAGCT  GGAGGTTTT 340
m4-26     CAG   G  CTT  ACT  CCTGAA  ATA  TCTTTAATAGCT  GGAGGTTTT 340
m4-1_A_   CAA   A  CTT  ACC  CCTGAA  ATA  TCTTTAATAGCT  GGAGGTTTT 340
m4-41     CAA   A  CTT  ACC  CCTGAA  ATA  TCTTTAATAGCT  GGAGGTTTT 340
m4-13_N_  CAA   A  CTT  ACC  CCTGAA  ATA  TCTTTAATAGCT  GGAGGTTTT 340
m4-2_B/C_ CAA   A  CTT  ACC  CCTGAA  ATA  TCTTTAATAGCT  GGAGGTTTT 340
m4-12_M_  CAA   A  CTT  ACC  CCTGAA  ATA  TCTTTAATAGCT  GGAGGTTTT 340
m4-48     CAA   A  CTT  ACC  CCTGAA  ATA  TCTTTAATAGCT  GGAGGTTTT 340
m4-40     CAA   A  CTT  ACC  CCTGAA  ATA  TCTTTAATAGCT  GGAGGTTTT 340
m4-19     CAA   A  CTT  ACC  CCTGAA  ATA  TCTTTAATAGCT  GGAGGTTTT 340
m4-3_D_   CAA   A  CTT  ACC  CCTGAA  ATA  TCTTTAATAGCT  GGAGGTTTT 340
m4-51     CAA   A  CTT  ACC  CCTGAA  ATA  TCTTTAATAGCT  GGAGGTTTT 340
m4-22     CAG   G  CTT  ACT  CCTGAA  ATA  TCTTTAATAGCT  GGAGGTTTT 340
m4-32     CAA   G  CTT  ACT  CCTGAA  ATA  TCTTTAATAGCT  GGAGGTTTT 340
m4-8_H_   CAA   G  CTT  ACT  CCTGAA  ATA  TCTTTAATAGCT  GGAGGTTTT 340
m4-36     CAA   G  CTT  ACT  CCTGAA  ATA  TCTTTAATAGCT  GGAGGTTTT 340
m4-14_O_  CAA   G  CTT  ACT  CCTGAA  ATA  TCTTTAATAGCT  GGAGGTTTT 340
m4        CAA   G  CTT  ACT  CCTGAA  ATA  TCTTTAATAGCT  GGAGGTTTT 340
m4-17_R_  CAA   A  CTT  ACC  CCTGAA  ATA  TCTTTAATAGCT  GGAGGTTTT 340
m4-27     CAA   A  CTT  ACC  CCTGAA  ATA  TCTTTAATAGCT  GGAGGTTTT 340
m4-30     CAA   A  CTT  ACC  CCTGAA  ATA  TCTTTAATAGCT  GGAGGTTTT 340
m4-47     CAA   A  CTT  ACC  CCTGAA  ATA  TCTTTAATAGCT  GGAGGTTTT 340
m4-9_G_   CAA   G  CTT  ACT  CCTGAA  ATA  TCTTTAATAGCT  GGAGGTTTT 340
m4-11_L_  CAA   G  CTT  ACT  CCTGAA  ATA  TCTTTAATAGCT  GGAGGTTTT 340
```

```
m4-31     CAA   G  CTT  ACT  CCTGAA  ATA  TCTTTAATAGCT  GGAGGTTTT 340
m4-46     CAA   G  CTT  ACT  CCTGAA  ATA  TCTTTAATAGCT  GGAGGTTTT 340
m4-34     CAA   G  CTT  ACT  CCTGAA  ATA  TCTTTAATAGCT  GGAGGTTTT 340
m4-37     CAA   G  CTT  ACT  CCTGAA  ATA  TCTTTAATAGCT  GGAGGTTTT 340
m4-25     CAA   G  CTT  ACC  CCTGAA  ATA  TCTTTAATAGCT  GGAGGTTTT 340
m4-18_S_  CAA   G  CTT  ACT  CCTGAA  ATA  TCTTTAATAGCT  GGAGGTTTT 340
m4-10_K_  CAA   G  CTT  ACT  CCTGAA  ATA  TCTTTAATAGCT  GGAGGTTTT 340
m4-44     CAG   G  CTT  ACT  CCTGAA  ATA  TCTTTAATAGCT  GGAGGTTTT 340
m4-20     CAA   A  CTT  ACC  CCTGAA  ATA  TCTTTAATAGCT  GGAGGTTTT 340
m4-33     CAA   A  CTT  ACC  CCTGAG  ATA  TCTTTAATAGCT  GGAGGTTTT 340
m4-50     CAA   G  CTT  ACT  CCTGAA  ATA  TCTTTAATAGCT  GGAGGTTTT 340
m4-15_P_  CAA   A  CTT  ACC  CCTGAA  ATA  TCTTTAATAGCT  GGAGGTTTT 340
m4-49     CAA   A  CTT  ACC  CCTGAA  ATA  TCTTTAATAGCT  GGAGGTTTT 340
m4-16_Q_  CAA   A  CTT  ACC  CCTGAA  ATA  TCTTTAATAGCT  GGAGGTTTT 340
m4-38     CAA   G  CTT  ACT  CCTGAA  ATA  TCTTTAATAGCT  GGAGGTTTT 340
m4-29     CAA   A  CTT  ACT  CCTGAA  ATA  TCTTTAATAGCT  GGAGGTTTT 340
m4-4_J_   CAA   A  CTT  ACC  CCTGAG  ATA  TCTTTAATAGCT  GGAGGTTTT 340
m4-6_E_   CAA   A  CTT  ACC  CCTGAA  ATA  TCTTTAATAGCT  GGAGGTTTT 340
m4-21     CAA   A  CTT  ACC  CCTGAA  ATA  TCTTTAATAGCT  GGAGGTTTT 340
m4-42     CAA   G  CTT  ACC  CCTGAA  ATA  TCTTTAATAGCT  GGAGGTTTT 340
m4-43     CAA   G  CTT  ACC  CCTGAA  ATA  TCTTTAATAGCT  GGAGGTTTT 340
m4-23     CAA   A  CTC  ACT  CCTGAA  ATC  TCTTTAATAGCC  GGAGGTTTT 340
m4-45     CAA   A  CTC  ACT  CCTGAA  ATC  TCTTTAATAGCC  GGAGGTTTT 340
m4-39     CAA   A  CTC  ACT  CCTGAA  ATC  TCTTTAATAGCC  GGAGGTTTT 340
Position **303 304**307**310*****316**319***********331*********
```

Table 40: Alignment of the translated amino acid sequence of the *msp4* gene – highlighted amino acid positions are differing within the variants

```
CLUSTAL 2.1 multiple sequence alignment
```

```
m4-1_A_          EVEVGYKKFETLAESDYKHIESHNFVAVGRD--AILTPDNFFVMETKIESVKDISVMET 57
m4-51            EVEVGYKKFETLAESDYKHIESHNFVAVGRD--AILTPDNFFVMETKIDSVKDISVMET 57
m4-2_B/C_        EVEVGYKKFETLAESDYKHIESHNFVAVGRD--AILTPDNFFVMETKIDSVKDISVMET 57
m4-19            EVEVGYKKFETLAESDYKHIESHNFVAVGRD--AILTPDNFFVMETKIDSVKDISVMET 57
m4-3_D_          EVEVGYKKFETLAESDYKHIESHNFVAVGRD--AILTPDNFFVMETKIDSVKDISVMET 57
m4-28            EVEVGYKKFETLAESDYKHIESHNFVAVGRD--AILTPDNFFVMETKIDSVKDISVMET 57
m4-8_H_          EVEVGYKKFETLAESDYKHIESHNFVAVGRD--AILTPDNFFVMETKIDSVKDISVMET 57
m4-32            EVEVGYKKFETLAESDYKHIESHNFVAVGRD--AILTPDNFFVMETKIDSVKDISVMET 57
m4-12_M_         EVEVGYKKFETLAESDYKHIESHNFVAVGRD--AILTPDNFFVMETKIDSVKDISVMET 57
m4-36            EVEVGYKKFETLAESDYKHIESHNFVAVGRD--AILTPDNFFVMETKIDSVKDISVMET 57
m4-13_N_         EVEVGYKKFETLAESDYKHIESHNFVAVGRD--AILTPDNFFVMETKIDSVKDISVMET 57
m4-40            EVEVGYKKFETLAESDYKHIESHNFVAVGRD--AILTPDNFFVMETKIDSVKDISVMET 57
m4-14_O_         EVEVGYKKFETLAESDYKHIESHNFVAVGRD--AILTPDNFFVMETKIDSVKDISVMET 57
m4-41            EVEVGYKKFETLAESDYKHIESHNFVAVGRD--AILTPDNFFVMETKIDSVKDISVMET 57
m4-15_P_         EVEVGYKKFETLAESDYKHVESHNFVAVGRD--AILTPDNFFVMETKIDSVKDISVMET 57
m4-30            EVEVGYKKFETLAESDYKHVESHNFVAVGRD--AILTPDNFFVMETKIDSVKDISVMET 57
m4-50            EVEVGYKKFETLAESDYKHVESHNFVAVGRD--AILTPDNFFVMETKIDSVKDISVMET 57
m4-4_J_          EVEVGYKKFETLAESDYKHVESHNFVAVGRD--AILTPDNFFVMETKIDSVKDISVMET 57
m4-29            EVEVGYKKFETLAESDYKHVXSHNFVAVGRD--AILTPDNFFVMETKIDSVKDISVMET 57
m4-37            EVEVGYKKFETLAESDYKHVESHNFVAVGRD--AILTPDNFFVMETKIDSVKDISVMET 57
m4-17_R_         EVEVGYKKFETLAESDYKHVESHNFVAVGRD--AILTPDNFFVMETKIDSVKDISVMET 57
m4-33            EVEVGYKKFETLAESDYKHVESHNFVAVGRD--AILTPDNFFVMETKIDSVKDISVMET 57
m4-6_E_          EVEVGYKKFETLAESDYKHVESHNFVAVGRD--AILTPDNFFVMETKIDSVKDISVMET 57
m4-20            EVEVGYKKFETLAESDYKHVESHNFVAVGRD--AILTPDNFFVMETKIDSVKDISVMET 57
m4-42            EVEVGYKKFETLAESDYKHVESHNFVAVGRD--AILTPDNFFVMETKIDSVKDISVMET 57
m4-10_K_         EVEVGYKKFETLAESDYKHVESHNFVAVGRD--AILTPDNFFVMETKIDSVKDISVMET 57
m4-25            EVEVGYKKFETLAESDYKHVESHNFVAVGRD--AILTPDNFFVMETKIDSVKDISVMET 57
m4-47            EVEVGYKKFETLAESDYKHVESHNFVAVGRD--AILTPDNFFVMETKIDSVKDISVMET 57
m4-5_I_          EVEVGYKKFETLAESDYKHVESHNFVAVGRD--AILTPDNFFVMETKIDSVKDISVMET 57
m4-35            EVEVGYKKFETLAESDYKHVESHNFVAVGRD--AILTLDNFFVMETKIDSVKDISVMET 57
m4-38            EVEVGYKKFETLAESDYKHVESHNFVAVGRD--AILTPDNFFVMETKIDSVKDISVMET 57
m4-18_S_         EVEVGYKKFETLAESDYKHVESHNFVAVGRD--AILTPDNFFVMETKIDSVKDISVMET 57
m4-34            EVEVGYKKFETLAESDYKHVESHNFVAVGRD--AILTPDNFFVMETKIDSVKDISVMET 57
m4-7_F_          EVEVGYKKFETLAESDYKHVESHNFVAVGRD--AILTPDNFFVMETKIDSVKDISVMET 57
m4-21            EVEVGYKKFETLAESDYKHVESHNFVAVGRD--AILTPDNFFVMETKIDSVKDISVMET 57
m4-43            EVEVGYKKFETLAESDYKHVESHNFVAVGRD--AILTPDNFFVMETKIDSVKDISVMET 57
m4-11_L_         EVEVGYKKFETLAESDYKHVESHNFVAVGRD--AILTPDNFFVMETKIDSVKDISVMET 57
m4-27            EVEVGYKKFETLAESDYKHVESHNFVAVGRD--AILTPDNFFVMETKIDSVKDISVMET 57
m4-49            EVEVGYKKFETLAESDYKHVESHNFVAVGRD--AILTPDNFFVMETKIDSVKDISVMET 57
m4-9_G_          EVEVGYKKFETLAESDYKHVESHNFVAVGRD--AILTPDNFFVMETKIDSVKDISVMET 57
m4-24            EVEVGYKKFETLAESDYKHVESHNFVAVGRD--AILTPDNFFVMETKIDSVKDISVMET 57
m4-46            EVEVGYKKFETLAESDYKHVESHNFVAVGRD--AILTPDNFFVMETKIDSVKDISVMET 57
m4-16_Q_         EVEVGYKKFETLAESDYKHVESHNFVAVGRD--AILTPDNFFVMETKIDSVKDISVMET 57
m4-31            EVEVGYKKFETLAESDYKHVESHNFVAVGRD--AILTPDNFFVMETKIDSVKDISVMET 57
m4-26            EVEVGYKKFETLAESDYKHVESHNFVAVGRD--AILTPDNFFVMETKIDSVKDISVMET 57
m4-44            EVEVGYKKFETLAESDYKHVESHNFVAVGRD--AILTPDNFFVMETKIDSVKDISVMET 57
m4-22            EVEVGYKKFETLAESDYKHIESHNFVAVGRD--AILTPDNFFVMETKIDSVKDISVMET 57
```

```
m4-48           EVEVGYKKFETLAESDYKHIESHNFVAVGRD--TTLTPDNFFVMETKIDSVKDISVMET 57
m4-23           EVEVGYKKFETLADSDFKHAESHNFVAVGRD--RALTPDNFFVMETKIDGVKDISVMET 57
m4-39           EVEVGYKKFETLADSDFKHAESHNFVAVGRD--RALTPDNFFVMETKIDGVKDISVMET 57
m4-45           EVEVGYKKFETLADSDFKHAESHNFVAVGRDMetALTPDNFFVMETKIDGVKDISVMET 57
                 ************:**:**  **********   :** **********:.*********

m4-1_A_         LNACYDVMETHTDLPVSPYMETCAGLGASFINIADHVTSKLAYRGKVGVSYKLTPEISLI 117
m4-51           LNACYDVMETHTDLPVSPYMETCAGLGASFINIADHVTSKLAYRGKVGVSYKLTPEISLI 117
m4-2_B/C_       LNACYDVMETHTDLPVSPYMETCAGLGASFINIADHVTSKLAYRGKVGVSYKLTPEISLI 117
m4-19           LNACYDVMETHTDLPVSPYMETCAGLGASFINIADHVTSKLAYRGKVGVSYKLTPEISLI 117
m4-3_D_         LNACYDVMETHTDLPVSPYMETCAGLGASFINIADHVTSKLAYRGKVGVSYKLTPEISLI 117
m4-28           LNACYDVMETHTDLPVSPYMETCAGLGASFINIADHVTSKLAYRGKVGVSYKLTPEISLI 117
m4-8_H_         LNACYDVMETHTDLPVSPYMETCAGLGASFINIADHVTSKLAYRGKVGVSYKLTPEISLI 117
m4-32           LNACYDVMETHTDLPVSPYMETCAGLGASFINIADHVTSKLAYRGKVGVSYKLTPEISLI 117
m4-12_M_        LNACYDVMETHTDLPVSPYMETCAGLGASFINIADHVTSKLAYRGKVGVSYKLTPEISLI 117
m4-36           LNACYDVMETHTDLPVSPYMETCAGLGASFINIADHVTSKLAYRGKVGVSYKLTPEISLI 117
m4-13_N_        LNACYDVMETHTDLPVSPYMETCAGLGASFINIADHVTSKLAYRGKVGVSYKLTPEISLI 117
m4-40           LNACYDVMETHTDLPVSPYMETCAGLGASFINIADHVTSKLAYRGKVGVSYKLTPEISLI 117
m4-14_O_        LNACYDVMETHTDLPVSPYMETCAGLGASFINIADHVTSKLAYRGKVGVSYKLTPEISLI 117
m4-41           LNACYDVMETHTDLPVSPYMETCAGLGASFINIADHVTSKLAYRGKVGVSYKLTPEISLI 117
m4-15_P_        LNACYDVMETHTDLPVSPYMETCAGLGASFINIADHVTSKLAYRGKVGVSYKLTPEISLI 117
m4-30           LNACYDVMETHTDLPVSPYMETCAGLGASFINIADHVTSKLAYRGKVGVSYKLTPEISLI 117
m4-50           LNACYDVMETHTDLPVSPYMETCAGLGASFINIADHVTSKLAYRGKVGVSYKLTPEISLI 117
m4-4_J_         LNACYDVMETHTDLPVSPYMETCAGLGASFINIADHVTSKLAYRGKVGVSYKLTPEISLI 117
m4-29           LNACYDVMETHTDLPVSPYMETCAGLGAXFINIADHVTSKLAYRGKVGVSYKLTPEISLI 117
m4-37           LNACYDVMETHTDLPVSPYMETCAGLGASFINIADHVTSKLAYRGKVGVSYKLTPEISLI 117
m4-17_R_        LNACYDVMETHTDLPVSPYMETCAGLGASFINIADHVTSKLAYRGKVGVSYKLTPEISLI 117
m4-33           LNACYDVMETHTDLPVSPYMETCAGLGASFINIADHVTSKLAYRGKVGVSYKLTPEISLI 117
m4-6_E_         LNACYDVMETHTDLPVSPYMETCAGLGASFINIADHVTSKLAYRGKVGVSYKLTPEISLI 117
m4-20           LNACYDVMETHTDLPVSPYMETCAGLGASFINIADHVTSKLAYRGKVGVSYKLTPEISLI 117
m4-42           LNACYDVMETHTDLPVSPYMETCAGLGASFINIADHVTSKLAYRGKVGVSYKLTPEISLI 117
m4-10_K_        LNACYDVMETHTDLPVSPYMETCAGLGASFINIADHVTSKLAYRGKVGVSYKLTPEISLI 117
m4-25           LNACYDVMETHTDLPVSPYMETCAGLGASFINIADHVTSKLAYRGKVGVSYKLTPEISLI 117
m4-47           LNACYDVMETHTDLPVSPYMETCAGLGASFINIADHVTSKLAYRGKVGVSYKLTPEISLI 117
m4-5_I_         LNACYDVMETHTDLPVSPYMETCAGLGASFINIADHVTSKLAYRGKVGVSYKLTPEISLI 117
m4-35           LNACYDVMETHTDLPVSPYMETCAGLGASFINIADHVTSKLAYRGKVGVSYKLTPEISLI 117
m4-38           LNACYDVMETHTDLPVSPYMETCAGLGASFINIADHVTSKLAYRGKVGVSYKLTPEISLI 117
m4-18_S_        LNACYDVMETHTDLPVSPYMETCAGLGASFINIADHVTSKLAYRGKVGVSYKLTPEISLI 117
m4-34           LNACYDVMETHTDLPVSPYMETCAGLGASFINIADHVTSKLAYRGKVGVSYKLTPEISLI 117
m4-7_F_         LNACYDVMETHTDLPVSPYMETCAGLGASFINIADHVTSKLAYRGKVGVSYKLTPEISLI 117
m4-21           LNACYDVMETHTDLPVSPYMETCAGLGASFINIADHVTSKLAYRGKVGVSYKLTPEISLI 117
m4-43           LNACYDVMETHTDLPVSPYMETCAGLGASFINIADHVTSKLAYRGKVGVSYKLTPEISLI 117
m4-11_L_        LNACYDVMETHTDLPVSPYMETCAGLGASFINIADHVTSKLAYRGKVGVSYKLTPEISLI 117
m4-27           LNACYDVMETHTDLPVSPYMETCAGLGASFINIADHVTSKLAYRGKVGVSYKLTPEISLI 117
m4-49           LNACYDVMETHTDLPVSPYMETCAGLGASFINIADHVTSKLAYRGKVGVSYKLTPEISLI 117
m4-9_G_         LNACYDVMETHTDLPVSPYMETCAGLGASFINIADHVTSKLAYRGKVGVSYKLTPEISLI 117
m4-24           LNACYDVMETHTDLPVSPYMETCAGLGASFINIADHVTSKLAYRGKVGVSYKLTPEISLI 117
m4-46           LNACYDVMETHTDLPVSPYMETCAGLGASFINIADHVTSKLAYRGKVGVSYKLTPEISLI 117
m4-16_Q_        LNACYDVMETHTDLPVSPYMETCAGLGASFINIADHVTSKLAYRGKVGVSYKLTPEISLI 117
m4-31           LNACYDVMETHTDLPVSPYMETCAGLGASFINIADHVTSKLAYRGKVGVSYKLTPEISLI 117
m4-26           LNACYDVMETHTDLPVSPYMETCAGLGASFINIADHVTSKLAYRGKVGVSYRLTPEISLI 117
m4-44           LNACYDVMETHTDLPVSPYMETCAGLGASFINIADHVTSKLAYRGKVGVSYRLTPEISLI 117
m4-22           LNACYDVMETHTDLPVSPYMETCAGLGASFINIADHVTSKLAYRGKVGVSYRLTPEISLI 117
m4-48           LNACYDVMETHTDLPVSPYMETCAGLGASFINIADHVTSKLAYRGKVGVSYKLTPEISLI 117
m4-23           LNACYDVMETHTDLPVSPYMETCAGLGASFIDIANHVTSKLAYRGKVGVSYKLTPEISLI 117
m4-39           LNACYDVMETHTDLPVSPYMETCAGLGASFIDIANHVTSKLAYRGKVGVSYKLTPEISLI 117
m4-45           LNACYDVMETHTDLPVSPYMETCAGLGASFIDIANHVTSKLAYRGKVGVSYKLTPEISLI 117
                ***************************** **:**:****************:********

m4-1_A_         AGGF 121
m4-51           AGGF 121
m4-2_B/C_       AGGF 121
m4-19           AGGF 121
m4-3_D_         AGGF 121
m4-28           AGGF 121
m4-8_H_         AGGF 121
m4-32           AGGF 121
m4-12_M_        AGGF 121
m4-36           AGGF 121
m4-13_N_        AGGF 121
m4-40           AGGF 121
m4-14_O_        AGGF 121
m4-41           AGGF 121
m4-15_P_        AGGF 121
m4-30           AGGF 121
m4-50           AGGF 121
m4-4_J_         AGGF 121
m4-29           AGGF 121
m4-37           AGGF 121
```

```
m4-17_R_        AGGF 121
m4-33           AGGF 121
m4-6_E_         AGGF 121
m4-20           AGGF 121
m4-42           AGGF 121
m4-10_K_        AGGF 121
m4-25           AGGF 121
m4-47           AGGF 121
m4-5_I_         AGGF 121
m4-35           AGGF 121
m4-38           AGGF 121
m4-18_S_        AGGF 121
m4-34           AGGF 121
m4-7_F_         AGGF 121
m4-21           AGGF 121
m4-43           AGGF 121
m4-11_L_        AGGF 121
m4-27           AGGF 121
m4-49           AGGF 121
m4-9_G_         AGGF 121
m4-24           AGGF 121
m4-46           AGGF 121
m4-16_Q_        AGGF 121
m4-31           AGGF 121
m4-26           AGGF 121
m4-44           AGGF 121
m4-22           AGGF 121
m4-48           AGGF 121
m4-23           AGGF 121
m4-39           AGGF 121
m4-45           AGGF 121
                ****
```

Table 41: Assigned accession numbers of the new *msp4* sequences

Accession number	*Msp4* variant	Sample	Animal species
KU712134	m4-2(B/C)	971	dog
KU712135	m4-20	69	dog
KU712136	m4-20	438	dog
KU712137	m4-20	7444	dog
KU712138	m4-20	14333	dog
KU712139	m4-20	S136009	dog
KU712140	m4-20	153681	dog
KU712141	m4-20	14260	dog
KU712142	m4-20	14524	dog
KU712143	m4-20	N	dog
KU712144	m4-36	BwDa8	fallow deer
KU712145	m4-36	M11	fallow deer
KU712146	m4-36	M19	fallow deer
KU712147	m4-40	M20	fallow deer
KU712148	m4-21	A31	mouflon
KU712149	m4-8(H)	E09	mouflon
KU712150	m4-5(I)	H24	mouflon
KU712151	m4-5(I)	L35	mouflon
KU712152	m4-37	Ro22	red deer
KU712153	m4-38	Ro23	red deer
KU712154	m4-43	BwRo1.1	red deer
KU712155	m4-7(F)	BwRo2	red deer

KU712156	m4-24	BwRo3	red deer
KU712157	m4-23	BwRo4	red deer
KU712158	m4-25	BwRo5	red deer
KU712159	m4-26	BwRo6	red deer
KU712160	m4-27	BwRo9	red deer
KU712161	m4-28	BwRo14	red deer
KU712162	m4-29	BwRo18	red deer
KU712163	m4-23	BWReh6	roe deer
KU712164	m4-2(B/C)	BwReh4	roe deer
KU712165	m4-13(N)	BwReh8	roe deer
KU712166	m4-42	BwReh9	roe deer
KU712167	m4-13(N)	BwReh10	roe deer
KU712168	m4-13(N)	BwReh12	roe deer
KU712169	m4-13(N)	BwReh13	roe deer
KU712170	m4-13(N)	BwReh14	roe deer
KU712171	m4-39	BwReh16	roe deer
KU712172	m4-13(N)	BwReh24	roe deer
KU712173	m4-23	BwReh31	roe deer
KU712174	m4-35	D31	roe deer
KU712175	m4-41	E25	roe deer
KU712176	m4-13(N)	G14	roe deer
KU712177	m4-48	L12	roe deer
KU712178	m4-31	BwSi3	sika deer
KU712179	m4-21	BwSi5	sika deer
KU712180	m4-44	BwSi7	sika deer
KU712181	m4-13(N)	BwSi9	sika deer
KU712182	m4-45	BwSi11	sika deer
KU712183	m4-32	BwSi12	sika deer
KU712184	m4-46	BwSi13	sika deer
KU712185	m4-33	BwSi14	sika deer
KU712186	m4-47	BwSi17	sika deer
KU712187	m4-34	BwSi19	sika deer
KU712188	m4-21	BwSi23	sika deer
KU712189	m4-20	S146	red fox
KU712190	m4-20	S220	red fox
KU712191	m4-20	S226	red fox
KU712192	m4-20	14	hedgehog
KU712193	m4-20	21	hedgehog
KU712194	m4-20	23	hedgehog
KU712195	m4-20	80	hedgehog
KU712196	m4-20	89	hedgehog
KU712197	m4-20	109	hedgehog
KU712198	m4-20	127	hedgehog
KU712199	m4-20	142	hedgehog

KU712200	m4-20	159	hedgehog
KU712201	m4-20	161	hedgehog
KU712202	m4-20	162	hedgehog
KU712203	m4-20	164	hedgehog
KU712204	m4-20	205	hedgehog
KU712205	m4-20	206	hedgehog
KU712206	m4-20	212	hedgehog
KU712207	m4-20	232	hedgehog

1.1.4. *msp2* gene

Table 42: Distribution of the *msp2* variants in the different examined animal species

	Dog	Horse	Cat	Hedgehog	Cattle	Goat	Fox	Roe deerr	Red deer	Fallow deer
m2-2(a)	11	4				1				
m2-3(b)					9					
m2-4(d)		1								
m2-5(e)	2	7								
m2-6(f)	4	2	1	1			1			
m2-11(n)									1	
m2-12(p)									1	
m2-13	2									
m2-14	1									
m2-15				1						
m2-16				1						
m2-17								1		
m2-18									1	
m2-19								1		
m2-20								1		
m2-21								1		
m2-22								1		
m2-23								1		
m2-24										1
m2-25				1						
m2-26					9					
m2-27					2					
Total	**20**	**14**	**1**	**4**	**20**	**1**	**1**	**6**	**3**	**1**

Table 43: Alignment of the *msp2* nucleotide sequences – highlighted nucleotide positions are differing within the variants

```
CLUSTAL O(1.2.2) multiple sequence alignment

m2-19          ----CCGCTGCCCCCTCCATGCCCAAAGG-CCTAATCATAAATSMAGCACAT-----GTG
m2-20          ------------------------------------------TCCGGCACAT-----GTG
m2-25          ------------------------------------------------------------
m2-22          -----------------------------------------ATCCGGCACAT-----GTG
m2-18          ------------------------------------------------------------
m2-24          ------------------------------------------------------------
m2-23          ------------------------------------------------------------
m2-26          TTGGGATATGGACTATCCGTGAGCCAGGTACATAATTTTAAAATCAATGATGGCGGAGAG
m2-27          TTGGGATATGGACTATCCGTGAGCCAGGTACATAATTTTAAAATCAATGATGGCGGAGAG
m2-11(n)       ------------------------------------------------------------
m2-7(h)        TTGGGATATGGACTATCCGTGAGCCAGGTACATAATTTTAAAATCAATGATGGCGGAGAG
m2-9(j)        TTGGGATATGGACTATCCGTGAGCCAGGTACATAATTTTAAAATCAATGATGGCGGAGAG
m2-21          TTGGGATATGGACTATCCGTGAGCCAGGTACATAATTTTAAAATCAATGATGGCGGAGAG
m2-1           TTGGGATATGGACTATCCGTGAGCCAGGTACATAATTTTAAAATCAATGATGGCGGAGAG
m2-10(k)       TTGGGATATGGACTATCCGTGAGCCAGGTACATAATTTTAAAATCAATGATGGCGGAGAG
m2-3(b)        TTGGGATATGGACTATCCGTGAGCCAGGTACATAATTTTAAAATCAATGATGGCGGAGAG
m2-17          ----------GACTATCCGTGAGCCAGGTACATAATTTTAAAATCAATGATGGCGGAGAG
m2-12(p)       TTGGGATATGGACTATCCGTGAGCCAGGTACATAATTTTAAAATCAATGATGGCGGAGAG
m2-15          ------------------------------------------------------------
m2-4(d)        TTGGGGTATGGACTATCCGTGAGCCAGGTACATAATTTTAAAATCAATGATGACGGAGAG
m2-16          ------------------------------------------------------------
m2-8(i)        TTGGGGTATGGACTATCCGTGAGCCAGGTACATAATTTTAAAATCAATGATGGCGGAGAG
m2-2(a)        TTGGGGTATGGACTATCCGTGAGCCAGGTACATAATTTTAAAATCAATGATGGCGGAGAG
m2-14          TTGGGGTATGGACTATCCGTGAGCCAGGTACATAATTTTAAAATCAATGATGGCGGAGAG
m2-6(f)        TTGGGGTATGGACTATCCGTGAGCCAGGTACATAATTTTAAAATCAATGATGGCGGAGAG
m2-5(e)        TTGGGGTATGGACTATCCGTGAGCCAGGTACATAATTTTAAAATCAATGATGGCGGAGAG
m2-13          TTGGGGTATGGACTATCCGTGAGCCAGGTACATAATTTTAAAATCAATGATGGCGGAGAG


m2-19          TAAGGGGAAATGTTGAGTTTAACCAACTCCATGCCAGGAAAGTCGTAACAC-----AAAT
m2-20          TAAGGTGAAATGTTGAGTTTAACCAACTTCATGCCAGGAAAGTCGTAACAC-----AAAT
m2-25          -------------------AACCAACTCCAGGCCAGGGAAAGTCATAGCAC-----AAAT
m2-22          TAAGGAGAAATGTTGAGTTTAACCAACTCCAGGCCAGGAAAGTCGTAGCAC-----AAAT
m2-18          -AAGGAGAAATGTGGAGTTTAACCAACTCCAGGCCAGGAAAGTCGTAGCAC-----AAAT
m2-24          ------------------------------------------------------------
m2-23          ------------------------------------------------------------
m2-26          ACGCGTTGGCTGTTTCCCTTTCATGAGCGAGTACACAGAGAGGAATTGCACTCTGCAAAT
m2-27          ACGCGTTGGCTGTTTCCCTTTCATGAGCGAGTACACAGAGAGGAATTGCACTCTGCAAAT
m2-11(n)       ------------------------------------------------------------
m2-7(h)        ACGCGTTGGCTGTTTCCCTTTCATGAGCGAGTGCACAGGGAGGAATTGCACTCTGCAAAT
m2-9(j)        ACGCGTTGGCTGTTTCCCTTTCATGAGCGAGTGCACAGAGAGGAATTGCACTCTGCAAAT
m2-21          ACGCGTTGGCTGTTTCCCTTTCATGAGCGAGTGCACAGAGAGGAATTGCACTCTGCAAAT
m2-1           ACGCGTTGGCTGTTTCCCTTTCATGAGCGAGTGCACAGAGAGGAATTGCACTCTGCAAAT
m2-10(k)       ACGCGTTGGCTGTTTCCCTTTCATGAGCGAGTGCACAGAGAGGAATTGCACTCTGCAAAT
m2-3(b)        ACGCGTTGGCTGTTTCCCTTTCATGAGCGAGTACACAGAGAGGAATTGCACTCTGCAAAT
m2-17          ACGCGTTGGCTGTTTCCCTTTCATGAACGGGTGCACAGAGAGAAATTACACTCTGCAAAT
m2-12(p)       ACGCGTTGGCTGTTTCCCTTTCATGAGCGAGTGCACAGAGAGGAATTGCACTCTGCAAAT
m2-15          ------------------------------------------AAATTACACTCTGCAAAT
m2-4(d)        ACGCGTTGGCTGTTTCCCTTTCATGAACGGGTGCACAGAGAGAAATTACACTCTGCAAAT
m2-16          ------------------------------------------AAATTACACTCTGCAAAT
m2-8(i)        ACGCGTTGGCTGTTTCCCTTTCATGAACGGGTGCACAGAGAGAAATTACACTCTGCAAAT
m2-2(a)        ACGCGTTGGCTGTTTCCCTTTCATGAACGGGTGCACAGAGAGAAATTACACTCTGCAAAT
m2-14          ACGCGTTGGCTGTTTCCCTTTCATGAACGGGTGCACAGAGAGAAATTACACTCTGCAAAT
m2-6(f)        ACGCGTTGGCTGTTTCCCTTTCATGAACGGGTGCACAGAGAGAAATTACACTCTGCAAAT
m2-5(e)        ACGCGTTGGCTGTTTCCCTTTCATGAACGGGTGCACAGAGAGAAATTACACTCTGCAAAT
m2-13          ACGCGTTGGCTGTTTCCCTTTCATGAACGGGTGCACAGAGAGAAATTACACTCTGCAAAT


m2-19          --------------------------------TCRACGTAGCAGTGGTAT---TCCTTAT
m2-20          --------------------------------TCAACGTAGCAGTGGTAT---TCCTTAT
m2-25          --------------------------------TCAACGTAGCAGTGGTAT---TCCTTAT
m2-22          --------------------------------TCAACGTAGCAGTGGTAT---TCCTTAT
m2-18          --------------------------------TCAACGTAGCAGTGGTAT---TCCTTAT
m2-24          ------------------------------------------------------TAT
m2-23          ------------------------------------------------------------
m2-26          TTCTATTGGGGTCCAGAGGTTGCTTCGGAGATAAGATTTAAGAGAGGTAACACTACTTTC
m2-27          TTCTATTGGGGTCCAGAGGTTGCATCGGAGATAAGATTTCAGAGAGGTAACACTACTTTC
m2-11(n)       ------------------------------------------------------------
m2-7(h)        TTCTATTGGGGTCCAGAGGTTGCTTCGGAGATAAGGTTTCGGAGAAGTAATGTTGCTTTC
m2-9(j)        TTCTATTGGGGTCCAGAGGTTGCTTCGGAGATAAGGTTTCGGAGAAGTAATGTTGCTTTC
m2-21          TTCTATTGGGGTCCAGAGGTTGCTTCGGAGATAAGGTTTCGGAGAAGTAATGTTGCTTTC
m2-1           TTCTATTGGGGTCCAGAGGTTGCTTCGGAGATAAGGTTTCGGAGAAGTAATGTTGCTTTC
m2-10(k)       TTCTATTGGGGTCCAGAGGTTGCTTCGGAGATAAGGTTTCGGAGAAGTAATGTTGCTTTC
m2-3(b)        TTCTATTGGGGTCCAGAGGTTGCTTCGGAGATAAGATTTCAGAGAGGTAACACTACTTTC
```

```
m2-17          TTCTATTGGGGTCCAGAGGTTGCTTCGGAGATAAGATTTCAGAGAGGTAACACTACTTTC
m2-12(p)       TTCTATTGGGGTCCAGAGGTTGCTTCGGAGATAAGATTTCAGAGAGGTAACTCTACTTTC
m2-15          TTCTATTGGGGTCCAGAGGTTGCTTCGGAGATAAGATTTCAGAGAGGAAACACTACTTTC
m2-4(d)        TTCTATTGGGGTCCAGAGGTTGCTTCGGAGATAAGATTTCAGAGAGGTAACGCTACTTTC
m2-16          TTCTATTGGGGTCCAGAGGTTGCTTCGGAGATAAGATTTCAGAGAGGTAACACTACTTTC
m2-8(i)        TTCTATTGGGGTCCAAAGGTTGCTTCGGAGATAAGATTTCAGAGAGGTAACACTACTTTC
m2-2(a)        TTCTATTGGGGTCCAAAGGTTGCTTCGGAGATAAGATTTCAGAGAGGTAACACTACTTTC
m2-14          TTCTATTGGGGTCCAGAGGTTGCTTCGGAGATAAGATTTCAGAGAGGTAACACTACTTTC
m2-6(f)        TTCTATTGGGGTCCAGAGGTTGCTTCGGAGATAAGATTTCAGAGAGGTAACACTACTTTC
m2-5(e)        TTCTATTGGGGTCCAGAGGTTGCTTCGGAGATAAGATTTCAGAGAGGTAACACTACTTTC
m2-13          TTCTATTGGGGTCCAGAGGTTGCTTCGGAGATAAGATTTCAGAGAGGTAACACTACTTTC


m2-19          AGCCATAATCTCTAGCACTTCATCAGCC--TCAACAGTCAATGCAAGAGCTTTTGTGAAA
m2-20          AGCCATAATCTCTAGCACTTCAGCAGCC--TCAACAGTCAAAGCAAGAGCTTTTGTGAAA
m2-25          AGCCATAATCTCTAGCACTTCAGCAGCC--TCAACAGACAAAGCAAGAGCTTTTGTGAAA
m2-22          AGCCATAATCTCTAGCACTTCAGCAGCC--TCAACAGTCAAAGCAAGAGCTTTTGTGAAA
m2-18          AGCCATAATCTCTAGCACTTCAGCAGCC--TCAACAGTCAAAGCAAGAGCTTTTGTGAAA
m2-24          AGCCATAATCTCTAGCACTTCAGCAGCC--TCAACAGTCAAAGCAAGAGCTTTTGTGAAA
m2-23          ------------------TTTATTCTCTGCATTACGGTTAGAAATTGAACTTACGCACGA
m2-26          GGGGGA-AGTGCTGGATATTTATTCTCTGCTGTACGGTTAGAAATTGATCTTACACACGA
m2-27          GGGGGG-AGTGCTGGATATTTATTCTCTGCTGTACGGTTAGAAATTGATCTTACACACGA
m2-11(n)       ------------------------------------------------------------
m2-7(h)        GGTGGT-AGTGCTGGATATTTATTCTCTGCATTACGGTTAGAAATTGAACTTACGCACGA
m2-9(j)        GGTGGT-AGTGCTGGATATTTATTCTCTGCATTACGGTTGGAAATTGAACTTACGCACGA
m2-21          GGTGGT-AGTGCTGGATATTTATTCTCTGCATTACGGTTGGAAATTGAACTTACGCACGA
m2-1           GGTGGT-AGTGCTGGATATTTATTCTCTGCATTACGGTTGGAAATTGAACTTACGCACGA
m2-10(k)       GGTGGT-AGTGCTGGATATTTATTCTCTGCATTACGGTTAGAAATTGAACTTACGCACGA
m2-3(b)        GGGGGG-AGTGCTGGATATTTATTCTCTGCTGTACGGTTAGAAATTGATCTTACACACGA
m2-17          GGGGGG-AGTGCTGGATATTTATTCTCTGCTGTACGGTTAGAAATTGATCTTACACACGA
m2-12(p)       GGGGGG-AGTGCTGGATATTTATTCTCTGCTGTACGGTTAGAAATTGATCTTACACACGA
m2-15          GGGGGG-AGTGCTGGATATTTATTCTCTGCTGTACGGTTAGAAATTGATCTTACACACGA
m2-4(d)        GGAGGG-AGTGCTGGATATTTATTCTCTGCTGTACGGTTAGAAATTGATCTTACACACGA
m2-16          GGGGGG-AGTGCTGGATATTTATTCTCTGCTGTACGGTTAGAAATTGATCTTACACACGA
m2-8(i)        GGGGGG-AGTGCTGGATATTTATTCTCTGCTGTACGGTTAGAAATTGATCTTACACACGA
m2-2(a)        GGGGGG-AGTGCTGGATATTTATTCTCTGCTGTACGGTTAGAAATTGATCTTACACACGA
m2-14          GGGGGG-AGTGCTGGATATTTATTCTCTGCTGTACGGTTAGAAATTGATCTTACACACGA
m2-6(f)        GGGGGG-AGTGCTGGATATTTATTCTCTGCTGTACGGTTAGAAATTGATCTTACACACGA
m2-5(e)        GGGGGG-AGTGCTGGATATTTATTCTCTGCTGTACGGTTAGAAATTGATCTTACACACGA
m2-13          GGGGGG-AGTGCTGGATATTTATTCTCTGCTGTACGGTTAGAAATTGATCTTACACACGA


m2-19          AGAACGCTTTGCTCAACCATCAGATCTA------GTAAATTAAAAAGATTATCGTCCGAG
m2-20          AGAACGCTTTGCTCAACCATCAGATCTA------GTAAATTAAAAAGATTATCGTCCGAG
m2-25          AGAACGCTTTGCTCAACCATCAGATCTA------GTAAGTTAAAAAGATTATCGTCCGTG
m2-22          AGAACGCTTTGCTCAACCATCAGATCTA------GTAAGTTAAAAAGATTATCGTCCGTG
m2-18          AGAACGCTTTGCTCAACCATCAGATCTA------GTAAGTTAAAAAGATTATCGTCCGTG
m2-24          AGAACGCTTTGCTCAACCATCAGATCTA------GTAAGTTAAAAAGATTATCGTCCGTG
m2-23          AAGATCTGAAATTCTAAAATCTGGGTTACAAAAGGGACGTAAAGGTGGTGGTATGCCCTT
m2-26          AAGATCCGAAATTCTAAAATCTGGATTACAAAAGGGCCGTAAAGGTGGTGGTATGCCCTT
m2-27          AAGATCCGAAATTCTAAAATCTGGATTACAAAAGGGCCGTAAAGGTGGTGGTATGCCCTT
m2-11(n)       ----------------------TGGTTACATAAGGGCAGCAAAGGTGGTGGTATGCCCTT
m2-7(h)        AAGATCTGAAATTCTAAAATCTGGGTTACAAAAGGGACGTAAAGGTGGTGGTATGCCCTT
m2-9(j)        AAGATCTGAAATTCTAAAATCTGGGTTACAAAAGGGACGTAAAGGTGGTGGTATGCCCTT
m2-21          AAGATCTGAAATTCTAAAATCTGGGTTACAAAAGGGACGTAAAGGTGGTGGTATGCCCTT
m2-1           AAGATCTGAAATTCTAAAATCTGGGTTACAAAAGGGACGTAAAGGTGGTGGTATGCCCTT
m2-10(k)       AAGATCTGAAATTCTAAAATCTGGGTTACAAAAGGGACGTAAAGGTGGTGGTATGCCCTT
m2-3(b)        AAGATCCGAAATTCTAAAATCTGGGTTACAAAAGGACCGTAAAGGTGGTGGTATGCCCTT
m2-17          AAGATCTGAAATTCTAAAATCTGGGTTACAAAAGGGACGTAAAGGTGGTGGTATGCCCTT
m2-12(p)       AAGATCTGAAATTCTAAAATCTGGGTTACAAAAGGGACGTAAAGGTGGTGGTATGCCCTT
m2-15          AAGATCTGAAATTCTAAAATCTGGGTTACAAAAGGGACGTAAAGGTGGTGGTATGCCCTT
m2-4(d)        AAGATCTGAAATTCTAAAATCTGGGTTACAAGAGGGACGTAAAGGTGGGGGTATGCCCTT
m2-16          AAGATCTGAAATTCTAAAATCTGGGTTACAAAAGGGACGTAAAGGTGGTGGTATGCCCTT
m2-8(i)        AAGATCTGAAATTCTAAAATCTGGGTTACAAAAGGGACGTAAAGGTGGTGGTATGCCCTT
m2-2(a)        AAGATCTGAAATTCTAAAATCTGGGTTACAAAAGGGACGTAAAGGTGGTGGTATGCCCTT
m2-14          AAGATCTGAAATTCTAAAATCTGGGTTACAAAAGGGACGTAAAGGTGGTGGTATGCCCTT
m2-6(f)        AAGATCTGAAATTCTAAAATCTGGGTTACAAAAGGGACGTAAAGGTGGTGGTATGCCCTT
m2-5(e)        AAGATCTGAAATTCTAAAATCTGGGTTACAAAAGGGACGTAAAGGTGGTGGTATGCCCTT
m2-13          AAGATCTGAAATTCTAAAATCTGGGTTACAAAAGGGACGTAAAGGTGGTGGTATGCCCTT
                                        **      *      **   *    *   **


m2-19          TCTTTCGGCAAAACCACAACAACTTCTGTATTTTTATTAACTCTCCTGCGTTTTTCTTCC
m2-20          TCTTTCGGCAAAACCACAACAACTTCCGTATTTTTATTAACTCTCCTGCGTTTTTCTTCC
m2-25          TCCTTCGGCAAAACCACAACAACTTCCGTATTTTTATTAACTCTCCTGCGTTTTTCTTCC
m2-22          TCCTTCGGCAAAACCACAACAACTTCCGTATTTTTATTAACTCTCCTGCGTTTTTCTTCC
m2-18          TCCTTCGGCAAAACCACAACAACTTCCGTATTTTTATTAACTCTCCTGCGTTTTTCTTCC
m2-24          TCCTTCGGCAAAACCACAACAACTTCCGTATTTTTATTAACTCTCCTGCGTTTTTCTTCC
m2-23          TGTATTAGGAAAACACGAGGCAATGGTATCTTTACAGAAG---GGCTATGATCGTATTGC
m2-26          TGTATTAGGAAAACACAAAGCGATGGTGTCTTTACAGAGA---GGCTATGATCGTATTGA
```

```
m2-27         TGTATTAGGAAAACACAAAGCGATGGTGTCTTTACAGAGA---GGCTATGATCGTATTGA
m2-11(n)      TGTATTAGGAAAACACGAAGCGATGGTGTCCTTACAGAAG---GGCTATGATCGTATTGC
m2-7(h)       TGTATTAGGAAAACACGAGGCAATGGTATCTTTACAGAAG---GGCTATGATCGTATTGC
m2-9(j)       TGTATTAGGAAAACACGAGGCAATGGTATCTTTACAGAAG---GGCTATGATCGTATTGC
m2-21         TGTATTAGGAAAACACGAGGCAATGGTATCTTTACAGAAG---GGCTATGATCGTATTGC
m2-1          TGTATTAGGAAAACACGAGGCAATGGTATCTTTACAGAAG---GGCTATGATCGTATTGC
m2-10(k)      TGTATTAGGAAAACACGAGGCAATGGTATCTTTACAGAAG---GGCTATGATCGTATTGC
m2-3(b)       TGTATTAGGAAAACACGAAGCGATGGTGTCCTTACAGAAG---GGCTATGATCGTATTGC
m2-17         TGTATTAGGAAAACACGAAGCGATGGTGTCCTTACAGAAG---GGCTATGATCGTATTGC
m2-12(p)      TGTATTAGGAAAACACGAAGCGATGGTGTCCTTACAGAAG---GGCTATGATCGTATTGC
m2-15         TGTATTAGGAAAACACGAAGCGATGGTGTCCTTACAGAAG---GGCTATGATCGTATTGC
m2-4(d)       TGTATTAGGAAAACACGAAGCGATGGTGTCCTTACAGAAG---GGCTATGATCGTATTGC
m2-16         TGTATTAGGAAAACACGAAGCGATGGTGTCCTTACAGAAG---GGCTATGATCGTATTGC
m2-8(i)       TGTATTAGGAAAACACGAAGCGATGGTGTCCTTACAGAAG---GGCTATGATCGTATTGC
m2-2(a)       TGTATTAGGAAAACACGAAGCGATGGTGTCCTTACAGAAG---GGCTATGATCGTATTGC
m2-14         TGTATTAGGAAAACACGAAGCGATGGTGTCCTTACAGAAG---GGCTATGATCGTATTGC
m2-6(f)       TGTATTAGGAAAACACGAAGCGATGGTGTCCTTACAGAAG---GGCTATGATCGTATTGC
m2-5(e)       TGTATTAGGAAAACACGAAGCGATGGTGTCCTTACAGAAG---GGCTATGATCGTATTGC
m2-13         TGTATTAGGAAAACACGAAGCGATGGTGTCCTTACAGAAG---GGCTATGATCGTATTGC
              *   *  * *****   *     *    *  **  *         **  * *  * **

m2-19         TTCATTTCAGCTAGTCTTCTCAGTTGATCATATCCTAGTTCTGCCACCATGTATTTTTCT
m2-20         TTCATTTCAGCTAGTCTTCTCAGTTGATCATATCCTAGTTCTGCCACCATGTATTTTTCT
m2-25         TTCATTTCTGCTAGTCTTCTCAGTTGATCATATCCTAGTCCTGCCACCATGTATTTTTCT
m2-22         TTCATTTCTGCTAGTCTTCTCAGTTGATCATATCCTAGTCCTGCCACCATGTATTTTTCT
m2-18         TTCATTTCTGCTAGTCTTCTCAGTTGATCATATCCTAGTCCTGCCACCATGTATTTTTCT
m2-24         TTCATTTCTGCTAGTCTTCTCAGTTGATCATATCCTAGTCCTGCCACCATGTATTTTTCT
m2-23         TTTAATC---GGGAGACTTTCGCGTGAAGATGTTATTGCTATAGAAAAATACATGGTGTC
m2-26         TTTAATC---GGAAGTCTTTCGCGTGAAAATGTTATCGCTATAGAAAAATACATGGTGTC
m2-27         TTTAATC---GGAAGTCTTTCGCGTGAAAATGTTATCGCTATAGAAAAATACATGGTGTC
m2-11(n)      TTTAATC---GGGAGACTTTCGCGTGAAGATGTTTGTTTAATAAAAAATACATGGTGGC
m2-7(h)       TTTAATC---GGGAGACTTTCGCGTGAAGATGTTATTGCTATAGAAAAATACATGGTGGC
m2-9(j)       TTTAATC---GGGAGACTTTCGCGTGAAGATGTTATTGCTATAGAAAAATACATGGTGGC
m2-21         TTTAATC---GGGAGACTTTCGCGTGAAGATGTTATTGCTATAGAAAAATACATGGTGGC
m2-1          TTTAATC---GGGAGACTTTCGCGTGAAGATGTTATTGCTATAGAAAAATACATGGTGTC
m2-10(k)      TTTAATC---GGGAGACTTTCGCGTGAAGATGTTATTGCTATAGAAAAATACATGGTGTC
m2-3(b)       TTTAATC---GGGAGACTTTCGCGTGAAGATGTTATTGCTATAGAAAAATACATGGTGGC
m2-17         TTTAATC---GGGAGACTTTCGCGTGAAGATGTTATTGCTATAGAAAAATACATGGTGGC
m2-12(p)      TTTAATC---GGGAGACTTTCGCGTGAAGATGTTATTGCTATAGAAAAATACATGGTGGC
m2-15         TTTAATC---GGGAGACTTTCGCGTGAAGATGTTATTGCTATAGAAAAATACATGGTGGC
m2-4(d)       TTTAATC---GGGAGACTTTCGCGTGAAGATGTTATTGCTATAGAAAAATACATGGTGGC
m2-16         TTTAATC---GGGAGACTTTCGCGTGAAGATGTTATTGCTATAGAAAAATACATGGTGGC
m2-8(i)       TTTAATC---GGGAGACTTTCGCGTGAAGATGTTATTGCTATAGAAAAATACATGGTGGC
m2-2(a)       TTTAATC---GGGAGACTTTCGCGTGAAGATGTTATTGCTATAGAAAAATACATGGTGGC
m2-14         TTTAATC---GGGAGACTTTCGCGTGAAGATGTTATTGCTATAGAAAAATACATGGTGGC
m2-6(f)       TTTAATC---GGGAGACTTTCGCGTGAAGATGTTATTGCTATAGAAAAATACATGGTGGC
m2-5(e)       TTTAATC---GGGAGACTTTCGCGTGAAGATGTTATTGCTATAGAAAAATACATGGTGGC
m2-13         TTTAATC---GGGAGACTTTCGCGTGAAGATGTTATTGCTATAGAAAAATACATGGTGGC
              ** * *           * **   ***  ** *           *  **  **  *

m2-19         ATAGCAATAACATCTTCACGCGA---AAGTCTCCCGATTAAAGCAATACGATCATAGCCC
m2-20         ATAGCAATAACATCTTCACGCGA---AAGTCTCCCGATTAAAGCAATACGATCATAGCCC
m2-25         ATAGCAATAACATCTTCACGCGA---AAGTCTCCCGATTAAAGCAATACGATCATAGCCC
m2-22         ATAGCAATAACATCTTCACGCGA---AAGTCTCCCGATTAAAGCAATACGATCATAGCCC
m2-18         ATAGCAATAACATCTTCACGCGA---AAGTCTCCCGATTAAAGCAATACGATCATAGCCC
m2-24         ATAGCAATAACATCTTCACGCGA---AAGTCTCCCGATTAAAGCAATACGATCATAGCCC
m2-23         AGAACTAGGATATGATCAACTGAGAAGACTAGCTGAAATGAAGGAAGAAAAACGCAGGAG
m2-26         AGAACTAGGATATGATCAACTGAGAAGACTTTCGGTAATGCAGGAAGAAGAACTCAGGAG
m2-27         AGAACTAGGATATGATCAACTGAGAAGACTTTCGGTAATGCAGGAAGAAGAACTCAGGAG
m2-11(n)      AGGACTAGTATATGATCAACTGAGAAGACTAGCAGAAATGAAGGAAGAAAAACGCAGGAG
m2-7(h)       AGAACTAGGATATGATCAACTGAGAAGACTAGCTGAAATGAAGGAAGAAAAACGCAGGAG
m2-9(j)       AGAACTAGGATATGATCAACTGAGAAGACTAGCTGAAATGAAGGAAGAAAAACGCAGGAG
m2-21         AGAACTAGGATATGATCAACTGAGAAGACTAGCTGAAATGAAGGAAGAAAAACGCAGGAG
m2-1          AGAACTAGGATATGATCAACTGAGAAGACTAGCTGAAATGAAGGAAGAAAAACGCAGGAG
m2-10(k)      AGAACTAGGATATGATCAACTGAGAAGACTAGCTGAAATGAAGGAAGAAAAACGCAGGAG
m2-3(b)       AGGACTAGGATATGATCAACTGAGAAGACTAGCAGAAATGAAGGAAGAAAAACGCAGGAG
m2-17         AGGACTAGGATATGATCAACTGAGAAGACTAGCAGAAATGAAGGAAGAAAAACGCAGGAG
m2-12(p)      AGGACTAGGATATGATCAACTGAGAAGACTAGCAGAAATGAAGGAAGAAAAACGCAGGAG
m2-15         AGGACTAGGATATGATCAACTGAGAAGACTAGCAGAAATGAAGGAAGAAAAACGCAGGAG
m2-4(d)       AGGACTAGGATATGATCAACTGAGAAGACTAGCAGAAATGAAGGAAGAAAAACGCAGGAG
m2-16         AGGACTAGGATATGATCAACTGAGAAGACTAGCAGAAATGAAGGAAGAAAAACGCAGGAG
m2-8(i)       AGGACTAGGATATGATCAACTGAGAAGACTAGCAGAAATGAAGGAAGAAAAACGCAGGAG
m2-2(a)       AGGACTAGGATATGATCAACTGAGAAGACTAGCAGAAATGAAGGAAGAAAAACGCAGGAG
m2-14         AGGACTAGGATATGATCAACTGAGAAGACTAGCAGAAATGAAGGAAGAAAAACGCAGGAG
m2-6(f)       AGGACTAGGATATGATCAACTGAGAAGACTAGCAGAAATGAAGGAAGAAAAACGCAGGAG
m2-5(e)       AGGACTAGGATATGATCAACTGAGAAGACTAGCAGAAATGAAGGAAGAAAAACGCAGGAG
m2-13         AGGACTAGGATATGATCAACTGAGAAGACTAGCAGAAATGAAGGAAGAAAAACGCAGGAG
              *   * *  * **  ***   **    * *  *   * *  ** ** *  * *  **
```

```
m2-19       TT---CTGTAAAGATACCATTGCCTCGTGTTTTCCTAATACAAAGGGCATACCACCACCT
m2-20       TT---CTGTAAAGATACCATTGCCTCGTGTTTTCCTAATACAAAGGGCATACCACCACCT
m2-25       TT---CTGTAAGGACACCATCGCTTCGTGTTTTCCTAATACAAAGGGCATACCACCACCT
m2-22       TT---CTGTAAGGACACCATCGCTTCGTGTTTTCCTAATACAAAGGGCATACCACCCCCT
m2-18       TT---CTGTAAGGACACCATCGCTTCGTGTTTTCCTAATACAAAGGGCATACCACCACCT
m2-24       TT---CTGTAAGGACACCATCGCTT-----------------------------------
m2-23       AGTTAATAAAAATACGGAAGTTGTTGTGGTTTTGCCGAAAGACTCGGACGATAATCTTTT
m2-26       AGTTAATAAAACTAAGAAAGTTGTTGGGGCTTTTCCCGAAAATAGGAGCGATAATTTTTT
m2-27       AGTTAATAAAACTAAGAAAGTTGTTGGGGCTTTTCCCGAAAATAGGAGCGATAATTTTTT
m2-11(n)    AGTTAATAAAAATACGGAAGTTGTTGTGGTTTTGCCGAAGGACTCGGACGATAATCTTTT
m2-7(h)     AGTTAATAAAAATACAGAAGTTGTTGTGGTTTTGCCGAAAGACTCGGACGATAATCTTTT
m2-9(j)     AGTTAATAAAAATACGGAAGTTGTTGTGGTTTTGCCGAAAGACTCGGACGATAATCTTTT
m2-21       AGTTAATAAAAATACGGAAGTTGTTGTGGTTTTGCCGAAAGACTCGGACGATAATCTTTT
m2-1        AGTTAATAAAAATACGGAAGTTGTTGTGGTTTTGCCGAAAGACTCGGACGATAATCTTTT
m2-10(k)    AGTTAATAAAAATACGGAAGTTGTTGTGGTTTTGCCGAAAGACTCGGACGATAATCTTTT
m2-3(b)     AGTTAATAAAAATACGGAAGTTGTTGTGGTTTTGCCGAAGGACTCGGACGATAATCTTTT
m2-17       AGTTAATAAAAATACGGAAGTTGTTGTGGTTTTGCCGAAGGACACGGACGATAATCTTTT
m2-12(p)    AGTTAATAAAAATACGGAAGTTGTTGTGGTTTTGCCGAAGGACACGGACGATAATCTTTT
m2-15       AGTTAATAAAAATACGGAAGTTGTTGTGGTTTTGCCGAAGGACACGGACGATAATCTTTT
m2-4(d)     AGTTAATAAAAATACGGAAGTTGTTGTGGTTTTGCCGAAGGACACGGACGATAATCTTTT
m2-16       AGTTAATAAAAATACGGAAGTTGTTGTGGTTTTGCCGAAGGACACGGACGATAATCTTTT
m2-8(i)     AGTTAATAAAAATACGGAAGTTGTTGTGGTTTTGCCGAAGGACACGGACGATAATCTTTT
m2-2(a)     AGTTAATAAAAATACGGAAGTTGTTGTGGTTTTGCCGAAGGACACGGACGATAATCTTTT
m2-14       AGTTAATAAAAATACGGAAGTTGTTGTGGTTTTGCCGAAGGACACGGACGATAATCTTTT
m2-6(f)     AGTTAATAAAAATACGGAAGTTGTTGTGGTTTTGCCGAAGGACACGGACGATAATCTTTT
m2-5(e)     AGTTAATAAAAATACGGAAGTTGTTGTGGTTTTGCCGAAGGACACGGACGATAATCTTTT
m2-13       AGTTAATAAAAATACGGAAGTTGTTGTGGTTTTGCCGAAGGACACGGACGATAATCTTTT
                  *  **  *    *     *

m2-19       TTACGGCTCTTTTGTAACCCAGATTTTAGAAT----------------TTCAGATCTTTC
m2-20       TTACGTCCCTTTTGTAACCCAGATTTTAGAAT----------------TTCAGATCTTTC
m2-25       TTACGTCCCTTTTGTAACCCAGATTTTAGAAT----------------TTCAGATCTTTC
m2-22       TTACGTCCCTTTTGTAACCCAGATTTTAGAAT----------------TTCGGATCTTTC
m2-18       TTACGGCCCTTTTGTAACCCAGATTTTAGAAT----------------TTCAGATCTTTC
m2-24       ------------------------------------------------------------
m2-23       ------TAATTTACTAGATCTGATGGTTGAGCAAAGCGTTCTTTTCACAAAAGCTCTTGC
m2-26       ------TAATTTACTGGATCTGATGATTGTGCAAAGCGTTCTTTTCACAAAAGCTCTTGC
m2-27       ------TAATTTACTGGATCTGATGATTGTGCAAAGCGTTCTTTTCACAAAAGCTCTTGC
m2-11(n)    ------TAACTTACTAGATCTGATGGTTGAGCAAAGCGTTCTTTTCACAAAAGCTCTTGC
m2-7(h)     ------TAATTTACTAGATCTGATGGTTGAGCAAAGCGTTCTTTTCACAAAAGCTCTTGC
m2-9(j)     ------TAATTTACTAGATCTGATGGTTGAGCAAAGCGTTCTTTTCACAAAAGCTCTTGC
m2-21       ------TAATTTACTAGATCTGATGGTTGAGCAAAGCGTTCTTTTCACAAAAGCTCTTGC
m2-1        ------TAATTTACTAGATCTGATGGTTGAGCAAAGCGTTCTTTTCACAAAAGCTCTTGC
m2-10(k)    ------TAATTTACTAGATCTGATGGTTGAGCAAAGCGTTCT------------------
m2-3(b)     ------TAACTTACTAGATCTGATGGTTGAGCAAAGCGTTCTTTTCACAAAAGCTCTTGC
m2-17       ------TAACTTACTAGATCTGATGGTTGAGCAAAGCGTTCTTTTCACAAAAGCTCTTGC
m2-12(p)    ------TAACTTACTAGATCTGATGGTTGAGCAAAGCGTTCTTTTCACAAAAGCTCTTGC
m2-15       ------TAACTTACTAGATCTGATGGTTGAGCAAAGCGTTCTTTTCACAAAAGCTCTTGC
m2-4(d)     ------TAACTTACTAGATCTGATGGTTGAGCAAAGCGTTCTTTTCACAAAAGCTCTTGC
m2-16       ------TAACTTACTAGATCTGATGGTTGAGCAAAGCGTTCTTTTCACAAAAGCTCTTGC
m2-8(i)     ------TAACTTACTAGATCTGATGGTTGAGCAAAGCGTTCTTTTCACAAAAGCTCTTGC
m2-2(a)     ------TAACTTACTAGATCTGATGGTTGAGCAAAGCGTTCTTTTCACAAAAGCTCTTGC
m2-14       ------TAACTTACTAGATCTGATGGTTGAGCAAAGCGTTCTTTTCACAAAAGCTCTTGC
m2-6(f)     ------TAACTTACTAGATCTGATGGTTGAGCAAAGCGTTCTTTTCACAAAAGCTCTTGC
m2-5(e)     ------TAACTTACTAGATCTGATGGTTGAGCAAAGCGTTCTTTTCACAAAAGCTCTTGC
m2-13       ------TAACTTACTAGATCTGATGGTTGAGCAAAGCGTTCTTTTCACAAAAGCTCTTGC


m2-19       GTGCGTAAGTTCAATTTCCAACCGTAATGCAGAGAATAAATATCCAGCACTACCACCGAA
m2-20       GTGCGTAAGTTCAATTTCCAACCGTAATGCAGAGAATAAATATCCAGCACTACCACCGAA
m2-25       GTGTGTAAGATCAATTTCTAACCGTACAGCAGAGAATAAATATCCAGCACTCCCCCCGAA
m2-22       GTGTGTAAGATCAATTTCTAACCGTACAGCAGAGAATAAATATCCAGCACTCCCCCCGAA
m2-18       GTGTGTAAGATCAATTTCTAACCGTACAGCAGAGAATAAATATCCAGCACTCCCCCCGAA
m2-24       ------------------------------------------------------------
m2-23       -ATTGACTGTTGAGGCTGCTGAAGTGCTAGAGATTATGGCTATAAGGAATACCACTGCTA
m2-26       -ATTGACTGTTGAGGGTGCTGAAGTGATAGAGATTATGGCTATAAGGAATACCACTGCTA
m2-27       -ATTGACTGTTGAGGGTGCTGAAGTGATAGAGATTATGGCTATAAGGAATACCACTGCTA
m2-11(n)    -ATTGACTGTTGAGGGTGC------------------------------------------
m2-7(h)     -ATTGACTGTTGAGGCTGATGAAGTGCTAGAGATTATGGCTATAAGGAATACCACTGCTA
m2-9(j)     -TTTGACTGTTGAGGCTGCTGAAGTGCTAGAGATTATGGCTATAAGGAATACCACTGCTA
m2-21       -ATTGACTGTTGAGGCTGCTGAAGTGCTAGAGATTATGGCTATAAGGAATACCACTGCTA
m2-1        -ATTGACTGTTGAGGCTGCTGAAGTGCTAGAGATTATGGCTATAAGGAATACCACTGCTA
m2-10(k)    ------------------------------------------------------------
m2-3(b)     -TTTGACTGTTGAGGGTGCTGAAGTGATAGAGATTATGGCTATAAGGAATACCACTGCTA
m2-17       -ATTGACTGTTGAGGGTGCTGAAGTGATAGAGATTATGGCTATAAGGAATACCACTGCTA
m2-12(p)    -ATTGACGGTTGAGGCTGCTGAAGTGCTAGAGATTATGGCTATAAGGAATACCACTGCTA
m2-15       -TTTGTCTGTTGAGGCTGCTGAAGTGCTAGAGATTATGGCTATAAGGAATACCACTGATA
m2-4(d)     -TTTGTCTGTTGAGGCTGCTGAAGTGCTAGAGATTATGGCTATAAGGAATACCACTGCTA
m2-16       -TTTGTCTGTTGAGGCTGCTGAAGTGCTAGAGATTATGGCTATAAGGAATACCACTGCTA
```

```
m2-8(i)       -TTTGTCTGTTGAGGCTGCTGAAGTGCTAGAGATTATGGCTATAAGGAATACCACTGCTA
m2-2(a)       -TTTGTCTGTTGAGGCTGCTGAAGTGCTAGAGATTATGGCTATAAGGAATACCACTGCTA
m2-14         -TTTGTCTGTTGAGGCTGCTGAAGTGCTAGAGATTATGGCTATAAGGAATACCACTGCTA
m2-6(f)       -TTTGTCTGTTGAGGCTGCTGAAGTGCTAGAGATTATGGCTATAAGGAATACCACTGCTA
m2-5(e)       -TTTGTCTGTTGAGGCTGCTGAAGTGCTAGAGATTATGGCTATAAGGAATACCACTGCTA
m2-13         -TTTGTCTGTTGAGGCTGCTGAAGTGCT--------------------------------

m2-19         AGCA----------------------------------------------------------
m2-20         AGCAA------CATT-----------------------------------------------
m2-25         AGTAG------TGTTACCTCTCT---GAAATCTTATCTCCGAAGCAACCTCTGGACCCCA
m2-22         AGTAG------TGTT-----------------------------------------------
m2-18         AGTAG------AGTTACCTCTTT---GAAATCTTATCTCCGAAGCAACCTCTGGACCCCA
m2-24         ----------------------------------------------------------------
m2-23         CGTTGAATTTGTGTTACGACTTTCCTGGCATGAAGTTCCAGAAAGAAAACATTGCCCCC-
m2-26         CGTTGAATTTGTGCTACGACTTTCCTAGCATGGAGTTGGTTAAACTCAACATTTCACCT-
m2-27         CGTTGAATTTGTGCTACGACTTTCCTAGCATGGAGTTGGTTAAACTCAACATTTCACCT-
m2-11(n)      ----------------------------------------------------------------
m2-7(h)       CGTTGAATTTGTGTTACGACTTTCCTGGCATGGAGTTGGTTAAACTCAACATTTCACCT-
m2-9(j)       CGTTGAATTTGTGTTACGACTTTCCTGGCATGAAGTT------------------------
m2-21         CGTTGAATTTGTGTTACGACTTTCCTGGCATGAAGTTGGTTAAACTCAACATTTCACCT-
m2-1          CGTTGAATTTGTGTTACGACTTTCCTGGCCTGAAGTTGGTTAAACTCAACATTTCACCT-
m2-10(k)      ----------------------------------------------------------------
m2-3(b)       CGTTGAATTTGTGCTACGACTTTCCTGGCCTGGAGTTGGTTAAACTCAACATTTCACCT-
m2-17         CGTTGAATTTGTGCTACGACTTTCCTAGCATGGAGTTGGTTAAACTCAACATTTCACCT-
m2-12(p)      CGTTGAATTTGTGCTACGACTTTCCTGGCCTGGAGTTGGTTAAACTCAACATTTCACCT-
m2-15         CGTTGAATTTGTGCTATGACTTTCCTGGCCTGGAGTTGGTTAAACTCAACATTTCTCCT-
m2-4(d)       CGTTGAATTTGTGCTATGACTTTCCTGGCCTGGAGTTGGTTAAACTCAACATTTCTCCT-
m2-16         CGTAGAATTTGTGCTATGACTTTCCTGGCCTGGAGTTGGTTAAACTCAACATTTCTCCT-
m2-8(i)       CGTTGAATTTGTGCTATGACTTTCCTGGCCTGGAGTTGGTTAAACTCAACATTTCTCCT-
m2-2(a)       CGTTGAATTTGTGCTATGACTTTCCTGGCCTGGAGTTGGTTAAACTCAACATTTCTCCT-
m2-14         CGTTGAATTTGTGCTATGACTTTCCTGGCCTGGAGTTGGTTAAACTCAACATTTCTCCT-
m2-6(f)       CGTTGAATTTGTGCTATGACTTTCCTGGCCTGGAGTTGGTTAAACTCAACATTTCTCCT-
m2-5(e)       CGTTGAATTTGTGCTATGACTTTCCTGGCCTGGAGTTGGTTAAACTCAACATTTCTCCT-
m2-13         ----------------------------------------------------------------

m2-19         ----------------------------------------------------------------
m2-20         ----------------------------------------------------------------
m2-25         ATAGAAATTTGCAGAG---TGTAATTTCTCTCTGTGCACCCGTTCATGAAAGGGAA---A
m2-22         ----------------------------------------------------------------
m2-18         ATAGAAATTTG---------------------------------------------------
m2-24         ----------------------------------------------------------------
m2-23         -CACGCGACCGGCGCCTTTATGATTAGGCCTTTGGGCATGGAGGGGGCAGCGGGTATAAC
m2-26         -TACACATGCGCCGGAATAGGTGGAAGCGTTATAGGTATTACAAAAGGACACGCCA---A
m2-27         -TACACATGCGCCGGAATAGGTGGAAGCGTTATAGGTATTACAAAAGGGCACGCCA---A
m2-11(n)      ----------------------------------------------------------------
m2-7(h)       -TACACATGTGCCGGAATAGGCGGAAGCGTTATAGGTATTACTAAAGGACATGCCA---A
m2-9(j)       ----------------------------------------------------------------
m2-21         -TACACATGTGCCGGAATAGGCGGAAGCGTTATAGGTATTACTAAAGGACATGCCA---A
m2-1          -TACACATGTGCCGGAATAGGCGGAAGCGTTATAGGTATTACTAAAGGACATGCCA---A
m2-10(k)      ----------------------------------------------------------------
m2-3(b)       -TACACATGTGCCGGAATAGGTGGAAGCGTTATAGGTATTACAAAAGGACATGCAA---A
m2-17         -TACACATGTGCCGGAATAGGTGGAAGCGTTATAGGTATTACAAAAGGACACGCCA---A
m2-12(p)      -TACACATGTGCCGGAATAGGTGGAAGCGTTATAGGTATTACAAAAGGACATGCAA---A
m2-15         -TACACATGTGCCGGAATAGGTGGAAGCGTTATAGGTATTACAAAAGGACATGCAA---A
m2-4(d)       -TACACATGTGCCGGAATAGGTGGAAGCGTTATAGGTATTACAAAAGGACATGCAA---A
m2-16         -TACACATGTGCCGGAATAGGTGGAAGCGTTATAGGTATTACAAAAGGACATGCAA---A
m2-8(i)       -TACACATGTGCCGGAATAGGTGGAAGCGTTATAGGTATTACAAAAGGACATGCAA---A
m2-2(a)       -TACACATGTGCCGGAATAGGTGGAAGCGTTATAGGTATTACAAAAGGACATGCAA---A
m2-14         -TACACATGTGCCGGAATAGGTGGAAGCGTTATAGGTATTACAAAAGGACATGCAA---A
m2-6(f)       -TACACATGTGCCGGAATAGGTGGAAGCGTTATAGGTATTACAAAAGGACATGCAA---A
m2-5(e)       -TACACATGTGCCGGAATAGGTGGAAGCGTTATAGGTATTACAAAAGGACATGCAA---A
m2-13         ----------------------------------------------------------------

m2-19         ----------------------------------------------------------------
m2-20         ----------------------------------------------------------------
m2-25         CAGCCAACGCGTCTCTCCGCCATCATTGATTTTAAAATTATGTACCTGGCTCACGGATAG
m2-22         ----------------------------------------------------------------
m2-18         ----------------------------------------------------------------
m2-24         ----------------------------------------------------------------
m2-23         TTTCCGTGCCTTCTACGCAATTATG------------------------------CTGT
m2-26         TTTACAACTTTCATACAAGCTCAAACTTGGTTTAAATTACCGGTTCCGCTCAAATGCTGT
m2-27         TTTACAACTTTCATACAAGCTCAAACTTGGTTTAAATTACCGTTTCCGCTCAAATGCTGT
m2-11(n)      ----------------------------------------------------------------
m2-7(h)       TTTTCAACTTTCATACAAGCTCAAACTTGGTTTAAATTACCATTTTCGCTCGAATGCTGT
m2-9(j)       ----------------------------------------------------------------
m2-21         TTTTCAACTTTCATACAAGCTCAAACTTGGTTTAAATTACCATTTTCGCTCGAATGCTGT
```

```
m2-1          TTTTCAACTTTCATACAAGCTCAAACTTGGTTTAAATTATCATTTTCGCTCGAATGCTGT
m2-10(k)      ------------------------------------------------------------
m2-3(b)       TTTACAACTTTCTTACAAGCTCAAACTTGGTTTAAATTACCGTTTCCGCTCAAATGCTGT
m2-17         TTTACAACTTTCATACAAGCTCAAACTTGGTTTAAATTACCGTTTCCGCTCAAATGCTGT
m2-12(p)      TTTACAACTTTCATACAAGCTCAAACTTGGTTTAAATTACCGTTTCCGCTCGAATGCTGT
m2-15         TTTACAACTTTCATACAAGCTCAAACTTGGTTTAAATTRCGGTTTCCGC-----------
m2-4(d)       TTTACAACTTTCATACAAGCTCAAACTTGGTTTAAATTACCGTTTCCGCTCGAATGCTGT
m2-16         TTTACAACTTTCATACAAGCTCAAACTTGGTTTAAATTACCGTTTCCGCTCGAATGCTGT
m2-8(i)       TTTACAACTTTCATACAAGCTCAAACTTGGTTTAAATTACCGTTTCCGCTCAAATGCTGT
m2-2(a)       TTTACAACTTTCATACAAGCTCAAACTTGGTTTAAATTACCGTTTCCGCTCAAATGCTGT
m2-14         TTTCCAACTTTCATACAAGCTCAAACTTGGTTTAAATTACCGTTTCCGCTCGAATGCTGT
m2-6(f)       TTTACAACTTTCATACAAGCTCAAACTTGGTTTAAATTACCGTTTCCGCTCGAATGCTGT
m2-5(e)       TTTACAACTTTCATACAAGCTCAAACTTGGTTTAAATTACCGTTTCCGCTCAAATGCTGT
m2-13         ------------------------------------------------------------


m2-19         ------------------------------------------------------------
m2-20         ------------------------------------------------------------
m2-25         TCCATACCCCAAGCCCACAAAAAAACCTGAC---TTGTC--------ATCATCCGCACGT
m2-22         ------------------------------------------------------------
m2-18         ------------------------------------------------------------
m2-24         ------------------------------------------------------------
m2-23         G-----------------------------------------------------------
m2-26         GGCATATATAGGTACTTCATATCAGAAAGTTTTAGGGTCGGAGTACTATAATGTGCCATT
m2-27         GGCATATATAGGGACTTCATATCAGAAAGTTTTAGGGTCGGAGTACTATAACGTGCCACT
m2-11(n)      ------------------------------------------------------------
m2-7(h)       AGCATATATAGGTACTGCATATCAGAAAGTTTTAGGGTCGGAATACTATAATGTGCCCCT
m2-9(j)       ------------------------------------------------------------
m2-21         AGCATATATAGGTACTGCATATCAGAAAGTTTTAGGGTCGGAATACTATAATGTGCCCCT
m2-1          AGCATATATAGGTACTGCATATCAGAAAGTTTTAGGGTCGGAATACTATAATGTGCCCCT
m2-10(k)      ------------------------------------------------------------
m2-3(b)       AGCATATATAGGGACTTCATATCAGAAAGTTTTAGGGTCAGAATACTATAATGTGCCCCT
m2-17         GGCATATATAGGGACTTCATATCAGAAAGTTTTAGGGTCGGAGTACTATAACGTG-----
m2-12(p)      GGCATATATAGGTACTGCATATCAGAAAGTTTTAGGGTCAGAATACTATAATGTGCCCCT
m2-15         ------------------------------------------------------------
m2-4(d)       GGCATATATAGGTACTGCATATCAGAAAGTTTTAGGGTCAGAATACTATAATGTGCCCCT
m2-16         GGCATATATAGGTACTGCATATCAGAAAGTTTTAGGGTCAGAATACTATAA---------
m2-8(i)       AGCATATATAGGGACTTCATATCAGAAAGTTTTAGGGTCGGAATACTATAATGTGCCCCT
m2-2(a)       AGCATATATAGGGACTTCATATCAGAAAGTTTTAGGGTCGGAATACTATAATGTGCCCCT
m2-14         GGCATATATAGGTACTGCATATCAGAAAGTTTTAGGGTCAGAATACTATAATGTGCCCCT
m2-6(f)       GGCATATATAGGTACTGCATATCAGAAAGTTTTAGGGTCAGAATACTATAATGTGCCCCT
m2-5(e)       AGCATATATAGGGACTTCATATCAGAAAGTTTTAGGGTCGGAATACTATAATGTGCCCCT
m2-13         ------------------------------------------------------------


m2-19         ------------
m2-20         ------------
m2-25         GCAGATGTTTCC
m2-22         ------------
m2-18         ------------
m2-24         ------------
m2-23         ------------
m2-26         AAAGCGGTTGGT
m2-27         AAAGCGGTTGGT
m2-11(n)      ------------
m2-7(h)       AAAGCGGTTAGT
m2-9(j)       ------------
m2-21         AAAGCGGTTGGT
m2-1          AAAGCGGTTGGT
m2-10(k)      ------------
m2-3(b)       AAAGCGGTTAGT
m2-17         ------------
m2-12(p)      AAAGCGGTTGGT
m2-15         ------------
m2-4(d)       AAAGCGGTTGGT
m2-16         ------------
m2-8(i)       AAAGCGGTGGGT
m2-2(a)       AAAGCGGTTGGT
m2-14         AAAGCGCTTGGT
m2-6(f)       AAAGCGGTTGGT
m2-5(e)       AAAGCGGTTGGT
m2-13         ------------
```

Table 44: Alignment of the amino acids of the *msp2* gene variants – highlighted amino acid positions are differing within the variants

```
CLUSTAL O(1.2.2) multiple sequence alignment

m2-26          LGYGLSVSQVHNFKINDGGETRWLFPFHERVHREELHSANFYWGPEVASEIRFKRGNTTF
m2-27          LGYGLSVSQVHNFKINDGGETRWLFPFHERVHREELHSANFYWGPEVASEIRFQRGNTTF
m2-11(n)       ------------------------------------------------------------
m2-20          ------------------------------------------------------NVAF
m2-9(j)        LGYGLSVSQVHNFKINDGGETRWLFPFHERVHREELHSANFYWGPEVASEIRFRRSNVAF
m2-21          LGYGLSVSQVHNFKINDGGETRWLFPFHERVHREELHSANFYWGPEVASEIRFRRSNVAF
m2-7(h)        LGYGLSVSQVHNFKINDGGETRWLFPFHERVHREELHSANFYWGPEVASEIRFRRSNVAF
m2-1           LGYGLSVSQVHNFKINDGGETRWLFPFHERVHREELHSANFYWGPEVASEIRFRRSNVAF
m2-10(k)       LGYGLSVSQVHNFKINDGGETRWLFPFHERVHREELHSANFYWGPEVASEIRFRRSNVAF
m2-3(b)        LGYGLSVSQVHNFKINDGGETRWLFPFHERVHREELHSANFYWGPEVASEIRFQRGNTTF
m2-8(i)        LGYGLSVSQVHNFKINDGGETRWLFPFHERVHREKLHSANFYWGPKVASEIRFQRGNTTF
m2-2(a)        LGYGLSVSQVHNFKINDGGETRWLFPFHERVHREKLHSANFYWGPKVASEIRFQRGNTTF
m2-25          LGYGLSVSQVHNFKINDGGETRWLFPFHERVHREKLHSANFYWGPEVASEIRFQRGNTTF
m2-14          LGYGLSVSQVHNFKINDGGETRWLFPFHERVHREKLHSANFYWGPEVASEIRFQRGNTTF
m2-6(f)        LGYGLSVSQVHNFKINDGGETRWLFPFHERVHREKLHSANFYWGPEVASEIRFQRGNTTF
m2-5(e)        LGYGLSVSQVHNFKINDGGETRWLFPFHERVHREKLHSANFYWGPEVASEIRFQRGNTTF
m2-13          LGYGLSVSQVHNFKINDGGETRWLFPFHERVHREKLHSANFYWGPEVASEIRFQRGNTTF
m2-23          ------------------------------------------------------------
m2-19          ----------------------------------------------------------AF
m2-17          ---XLSVSQVHNFKINDGGETRWLFPFHERVHREKLHSANFYWGPEVASEIRFQRGNTTF
m2-16          ----------------------------------KLHSANFYWGPEVASEIRFQRGNTTF
m2-15          ----------------------------------KLHSANFYWGPEVASEIRFQRGNTTF
m2-4(d)        LGYGLSVSQVHNFKINDDGETRWLFPFHERVHREKLHSANFYWGPEVASEIRFQRGNATF
m2-22          ----------------------------------------------------------NTTF
m2-18          ------------------------------------NFYWGPEVASEIRFQRGNSTF
m2-12(p)       LGYGLSVSQVHNFKINDGGETRWLFPFHERVHREHLHSANFYWGPEVASEIRFQRGNSTF
m2-24          ------------------------------------------------------------


m2-26          GGSAGYLFSAVRLEIDLTHERSEILKSGLQKGRKGGGMetPFVLGKHKAMetVSLQRGYD
m2-27          GGSAGYLFSAVRLEIDLTHERSEILKSGLQKGRKGGGMetPFVLGKHKAMetVSLQRGYD
m2-11(n)       ---------------------------WLHKGSKGGGMetPFVLGKHEAMetVSLQKGYD
m2-20          GGSAGYLFSALRLEIELTHERSEILKSGLQKGRKGGGMetPFVLGKHEAMetVSLQKGYD
m2-9(j)        GGSAGYLFSALRLEIELTHERSEILKSGLQKGRKGGGMetPFVLGKHEAMetVSLQKGYD
m2-21          GGSAGYLFSALRLEIELTHERSEILKSGLQKGRKGGGMetPFVLGKHEAMetVSLQKGYD
m2-7(h)        GGSAGYLFSALRLEIELTHERSEILKSGLQKGRKGGGMetPFVLGKHEAMetVSLQKGYD
m2-1           GGSAGYLFSALRLEIELTHERSEILKSGLQKGRKGGGMetPFVLGKHEAMetVSLQKGYD
m2-10(k)       GGSAGYLFSALRLEIELTHERSEILKSGLQKGRKGGGMetPFVLGKHEAMetVSLQKGYD
m2-3(b)        GGSAGYLFSAVRLEIDLTHERSEILKSGLQKDRKGGGMetPFVLGKHEAMetVSLQKGYD
m2-8(i)        GGSAGYLFSAVRLEIDLTHERSEILKSGLQKGRKGGGMetPFVLGKHEAMetVSLQKGYD
m2-2(a)        GGSAGYLFSAVRLEIDLTHERSEILKSGLQKGRKGGGMetPFVLGKHEAMetVSLQKGYD
m2-25          GGSAGYLFSAVRLEIDLTHERSEILKSGLQKGRKGGGMetPFVLGKHEAMetVSLQKGYD
m2-14          GGSAGYLFSAVRLEIDLTHERSEILKSGLQKGRKGGGMetPFVLGKHEAMetVSLQKGYD
m2-6(f)        GGSAGYLFSAVRLEIDLTHERSEILKSGLQKGRKGGGMetPFVLGKHEAMetVSLQKGYD
m2-5(e)        GGSAGYLFSAVRLEIDLTHERSEILKSGLQKGRKGGGMetPFVLGKHEAMetVSLQKGYD
m2-13          GGSAGYLFSAVRLEIDLTHERSEILKSGLQKGRKGGGMetPFVLGKHEAMetVSLQKGYD
m2-23          ------LFSALRLEIELTHERSEILKSGLQKGRKGGGMetPFVLGKHEAMetVSLQKGYD
m2-19          GGSAGYLFSALRLEIELTHERSEILKSGLQKSRKGGGMetPFVLGKHEAMetVSLQKGYD
m2-17          GGSAGYLFSAVRLEIDLTHERSEILKSGLQKGRKGGGMetPFVLGKHEAMetVSLQKGYD
m2-16          GGSAGYLFSAVRLEIDLTHERSEILKSGLQKGRKGGGMetPFVLGKHEAMetVSLQKGYD
m2-15          GGSAGYLFSAVRLEIDLTHERSEILKSGLQKGRKGGGMetPFVLGKHEAMetVSLQKGYD
m2-4(d)        GGSAGYLFSAVRLEIDLTHERSEILKSGLQEGRKGGGMetPFVLGKHEAMetVSLQKGYD
m2-22          GGSAGYLFSAVRLEIDLTHERSEILKSGLQKGRKGGGMetPFVLGKHEAMetVSLQKGYD
m2-18          GGSAGYLFSAVRLEIDLTHERSEILKSGLQKGRKGGGMetPFVLGKHEAMetVSLQKGYD
m2-12(p)       GGSAGYLFSAVRLEIDLTHERSEILKSGLQKGRKGGGMetPFVLGKHEAMetVSLQKGYD
m2-24          ------------------------------------------------AMetVSLQKGYD
                                                                ********:***

m2-26          RIDLIGSLSRENVIAIEKYMetVSELGYDQLRRLSVMetQEEELRRVNKTKKVVGAFPEN
m2-27          RIDLIGSLSRENVIAIEKYMetVSELGYDQLRRLSVMetQEEELRRVNKTKKVVGAFPEN
m2-11(n)       RIALIGRLSREDVCFNKKYMetVAGLVYDQLRRLAEMetKEEKRRRVNKNTEVVVVLPKD
m2-20          RIALIGRLSREDVIAIEKYMetVAELGYDQLRRLAEMetKEEKRRRVNKNTEVVVVLPKD
m2-9(j)        RIALIGRLSREDVIAIEKYMetVAELGYDQLRRLAEMetKEEKRRRVNKNTEVVVVLPKD
m2-21          RIALIGRLSREDVIAIEKYMetVAELGYDQLRRLAEMetKEEKRRRVNKNTEVVVVLPKD
m2-7(h)        RIALIGRLSREDVIAIEKYMetVAELGYDQLRRLAEMetKEEKRRRVNKNTEVVVVLPKD
m2-1           RIALIGRLSREDVIAIEKYMetVSELGYDQLRRLAEMetKEEKRRRVNKNTEVVVVLPKD
m2-10(k)       RIALIGRLSREDVIAIEKYMetVSELGYDQLRRLAEMetKEEKRRRVNKNTEVVVVLPKD
m2-3(b)        RIALIGRLSREDVIAIEKYMetVAGLGYDQLRRLAEMetKEEKRRRVNKNTEVVVVLPKD
m2-8(i)        RIALIGRLSREDVIAIEKYMetVAGLGYDQLRRLAEMetKEEKRRRVNKNTEVVVVLPKD
m2-2(a)        RIALIGRLSREDVIAIEKYMetVAGLGYDQLRRLAEMetKEEKRRRVNKNTEVVVVLPKD
m2-25          RIALIGRLSREDVIAIEKYMetVAGLGYDQLRRLAEMetKEEKRRRVNKNTEVVVVLPKD
m2-14          RIALIGRLSREDVIAIEKYMetVAGLGYDQLRRLAEMetKEEKRRRVNKNTEVVVVLPKD
m2-6(f)        RIALIGRLSREDVIAIEKYMetVAGLGYDQLRRLAEMetKEEKRRRVNKNTEVVVVLPKD
m2-5(e)        RIALIGRLSREDVIAIEKYMetVAGLGYDQLRRLAEMetKEEKRRRVNKNTEVVVVLPKD
```

```
m2-13          RIALIGRLSREDVIAIEKYMetVAGLGYDQLRRLAEMetKEEKRRRVNKNTEVVVVLPKD
m2-23          RIALIGRLSREDVIAIEKYMetVSELGYDQLRRLAEMetKEEKRRRVNKNTEVVVVLPKD
m2-19          RIALIGRLSREDVIAIEKYMetVAELGYDQLRRLAEMetKEEKRRRVNKNTEVVVVLPKD
m2-17          RIALIGRLSREDVIAIEKYMetVAGLGYDQLRRLAEMetKEEKRRRVNKNTEVVVVLPKD
m2-16          RIALIGRLSREDVIAIEKYMetVAGLGYDQLRRLAEMetKEEKRRRVNKNTEVVVVLPKD
m2-15          RIALIGRLSREDVIAIEKYMetVAGLGYDQLRRLAEMetKEEKRRRVNKNTEVVVVLPKD
m2-4(d)        RIALIGRLSREDVIAIEKYMetVAGLGYDQLRRLAEMetKEEKRRRVNKNTEVVVVLPKD
m2-22          RIALIGRLSREDVIAIEKYMetVAGLGYDQLRRLAEMetKEEKRRRVNKNTEVVVVLPKD
m2-18          RIALIGRLSREDVIAIEKYMetVAGLGYDQLRRLAEMetKEEKRRRVNKNTEVVVVLPKD
m2-12(p)       RIALIGRLSREDVIAIEKYMetVAGLGYDQLRRLAEMetKEEKRRRVNKNTEVVVVLPKD
m2-24          RIALIGRLSREDVIAIEKYMetVAGLGYDQLRRLAEMetKEEKRRRVNKNTEVVVVLPKD
               ** *** ****:*   :******: * *******: ***:**: *****..:** .:*::

m2-26          RSDNFFNLLDLMetIVQSVLFTKALALTVEGAEVIEIMetAIRNTTAT---LNLCYDFPS
m2-27          RSDNFFNLLDLMetIVQSVLFTKALALTVEGAEVIEIMetAIRNTTAT---LNLCYDFPS
m2-11(n)       SDDNLFNLLDLMetVEQSVLFTKALALTVEG----------------------------
m2-20          SDDNLFNLLDLMetVEQSVLFTKALALTVEAAEVLEIMetAIRNTTAT---LNLCYDFPG
m2-9(j)        SDDNLFNLLDLMetVEQSVLFTKALALTVEAAEVLEIMetAIRNTTAT---LNLCYDFPG
m2-21          SDDNLFNLLDLMetVEQSVLFTKALALTVEAAEVLEIMetAIRNTTAT---LNLCYDFPG
m2-7(h)        SDDNLFNLLDLMetVEQSVLFTKALALTVEADEVLEIMetAIRNTTAT---LNLCYDFPG
m2-1           SDDNLFNLLDLMetVEQSVLFTKALALTVEAAEVLEIMetAIRNTTAT---LNLCYDFPG
m2-10(k)       SDDNLFNLLDLMetVEQSV---------------------------------------
m2-3(b)        SDDNLFNLLDLMetVEQSVLFTKALALTVEGAEVIEIMetAIRNTTAT---LNLCYDFPG
m2-8(i)        TDDNLFNLLDLMetVEQSVLFTKALALSVEAAEVLEIMetAIRNTTAT---LNLCYDFPG
m2-2(a)        TDDNLFNLLDLMetVEQSVLFTKALALSVEAAEVLEIMetAIRNTTAT---LNLCYDFPG
m2-25          TDDNLFNLLDLMetVEQSVLFTKALALSVEAAEVLEIMetAIRNTTAT---LNLCYDFPW
m2-14          TDDNLFNLLDLMetVEQSVLFTKALALSVEAAEVLEIMetAIRNTTAT---LNLCYDFPG
m2-6(f)        TDDNLFNLLDLMetVEQSVLFTKALALSVEAAEVLEIMetAIRNTTAT---LNLCYDFPG
m2-5(e)        TDDNLFNLLDLMetVEQSVLFTKALALSVEAAEVLEIMetAIRNTTAT---LNLCYDFPG
m2-13          TDDNLFNLLDLMetVEQSVLFTKALALSVEAAEV-------------------------
m2-23          SDDNLFNLLDLMetVEQSVLFTKALALTVEAAEVLEIMetAIRNTTAT---LNLCYDFPG
m2-19          SDDNLFNLLDLMetVEQSVLFTKALALTVEADEVLEIMetAIRNTTAT---XNLCYDFPG
m2-17          TDDNLFNLLDLMetVEQSVLFTKALALTVEGAEVIEIMetAIRNTTAT---LNLCYDFPS
m2-16          TDDNLFNLLDLMetVEQSVLFTKALALSVEAAEVLEIMetAIRNTTATStopNLCYDFPG
m2-15          TDDNLFNLLDLMetVEQSVLFTKALALSVEAAEVLEIMetAIRNTTDT---LNLCYDFPG
m2-4(d)        TDDNLFNLLDLMetVEQSVLFTKALALSVEAAEVLEIMetAIRNTTAT---LNLCYDFPG
m2-22          TDDNLFNLLDLMetVEQSVLFTKALALTVEAAEVLEIMetAIRNTTAT---LNLCYDFPG
m2-18          TDDNLFNLLDLMetVEQSVLFTKALALTVEAAEVLEIMetAIRNTTAT---LNLCYDFPG
m2-12(p)       TDDNLFNLLDLMetVEQSVLFTKALALTVEAAEVLEIMetAIRNTTAT---LNLCYDFPG
m2-24          TDDNLFNLLDLMetVEQSVLFTKALALTVEAAEVLEIMetAI-------------------
                .**:*********: ***

m2-26          MetELVKLNISPYTCAGIGGSVIGITKGHANLQLS-YKLKLGLNYRFRSNAVAYIGTSYQ
m2-27          MetELVKLNISPYTCAGIGGSVIGITKGHANLQLS-YKLKLGLNYRFRSNAVAYIGTSYQ
m2-11(n)       ------------------------------------------------------------
m2-20          MetKLVKLNISPYTCAG----------------------------------------------
m2-9(j)        MetK-----------------------------------------------------------
m2-21          MetKLVKLNISPYTCAGIGGSVIGITKGHANFQLS-YKLKLGLNYHFRSNAVAYIGTAYQ
m2-7(h)        MetELVKLNISPYTCAGIGGSVIGITKGHANFQLS-YKLKLGLNYHFRSNAVAYIGTAYQ
m2-1           L--KLVKLNISPYTCAGIGGSVIGITKGHANFQLS-YKLKLGLNYHFRSNAVAYIGTAYQ
m2-10(k)       ------------------------------------------------------------
m2-3(b)        L--ELVKLNISPYTCAGIGGSVIGITKGHANLQLS-YKLKLGLNYRFRSNAVAYIGTSYQ
m2-8(i)        L--ELVKLNISPYTCAGIGGSVIGITKGHANLQLS-YKLKLGLNYRFRSNAVAYIGTSYQ
m2-2(a)        L--ELVKLNISPYTCAGIGGSVIGITKGHANLQLS-YKLKLGLNYRFRSNAVAYIGTSYQ
m2-25          P--GVG---------------------------------------------------------
m2-14          L--ELVKLNISPYTCAGIGGSVIGITKGHANFQLS-YKLKLGLNYRFRSNAVAYIGTAYQ
m2-6(f)        L--ELVKLNISPYTCAGIGGSVIGITKGHANLQLS-YKLKLGLNYRFRSNAVAYIGTAYQ
m2-5(e)        L--ELVKLNISPYTCAGIGGSVIGITKGHANLQLS-YKLKLGLNYRFRSNAVAYIGTSYQ
m2-13          ------------------------------------------------------------
m2-23          MetKFQKENIAPHATGAFMetIRPLGMetEGAAGITFRAFYAIMetL----------------
m2-19          MetELVKLNISPYTCAXFMetIRPLGMetEGAA------------------------------
m2-17          MetELVKLNISPYTCAGIGGSVIGITKGHANLQLS-YKLKLGLNYRFRSNAVAYIGTSYQ
m2-16          L--ELVKLNISPYTCAGIGGSVIGITKGHANLQLS-YKLKLGLNYRFRSNAVAYIGTAYQ
m2-15          L--ELVKLNISPYTCAGIGGSVIGITKGHANLQLS-YKLKLGLNXGFR--------------
m2-4(d)        L--ELVKLNISPYTCAGIGGSVIGITKGHANLQLS-YKLKLGLNYRFRSNAVAYIGTAYQ
m2-22          L--ELVKLNISPYTCAG----------------------------------------------
m2-18          L--ELVKLHISP---------------------------------------------------
m2-12(p)       L--ELVKLNISPYTCAGIGGSVIGITKGHANLQLS-YKLKLGLNYRFRSNAVAYIGTAYQ
m2-24          ------------------------------------------------------------

m2-26          KVLGSEYYNVPLKRL
m2-27          KVLGSEYYNVPLKRL
m2-11(n)       ---------------
m2-20          ---------------
m2-9(j)        ---------------
m2-21          KVLGSEYYNVPLKRL
m2-7(h)        KVLGSEYYNVPLKRL
m2-1           KVLGSEYYNVPLKRL
m2-10(k)       ---------------
```

```
m2-3(b)        KVLGSEYYNVPLKRL
m2-8(i)        KVLGSEYYNVPLKRW
m2-2(a)        KVLGSEYYNVPLKRL
m2-25          ---------------
m2-14          KVLGSEYYNVPLKRL
m2-6(f)        KVLGSEYYNVPLKRL
m2-5(e)        KVLGSEYYNVPLKRL
m2-13          ---------------
m2-23          ---------------
m2-19          ---------------
m2-17          KVLGSEYYNV-----
m2-16          KVLGSEYY-------
m2-15          ---------------
m2-4(d)        KVLGSEYYNVPLKRL
m2-22          ---------------
m2-18          ---------------
m2-12(p)       KVLGSEYYNVPLKRL
m2-24          ---------------
```

Table 45: Assigned accession numbers of the new *msp2* sequences

Accession number	*msp2* variant	Sample	Animal species
KX395921	m2-6(f)	971	cat
KX395922	m2-6(f)	80	hedgehog
KX395923	m2-6(f)	S146	red fox
KX395924	m2-13	8074	dog
KX395925	m2-13	14333	dog
KX395926	m2-14	S958/09	dog
KX395927	m2-21	BwReh14	roe deer
KX395928	m2-16	263	hedgehog
KX395929	m2-15	209	hedgehog
KX448790	m2-25	22	hedgehog
KX448791	m2-23	G14	roe deer
KX448792	m2-22	D31	roe deer
KX448793	m2-20	BwReh13	roe deer
KX448794	m2-19	BwReh12	roe deer
KX448795	m2-18	BwRo9	red deer
KX448796	m2-24	M11	fallow deer

1.2. Combination of sequence variants

Table 46: Samples with 4/4 partial gene variants

Sample	Animal species	Origin	Partial gene combination			
			16S rRNA	***msp2***	***msp4***	***groEL***
S146	Red fox	Thuringia	16S-1(A)	m2-6(f)	m4-20	g-1(A)

14333	Dog	Slowenia	16S-1(A)	m2-13	m4-20	g-1(A)
6933	Dog	Germany	16S-1(A)	m2-5(e)	m4-2(B/C)	g-1(A)
680200	Dog	Germany	16S-1(A)	m2-6(f)	m4-2(B/C)	g-1(A)
22	Hedgehog	Germany	16S-1(A)	m2-25	m4-20	g-1(A)
80	Hedgehog	Germany	16S-1(A)	m2-6(f)	m4-20	g-1(A)
S 1729/08	Horse	Germany	16S-1(A)	m2-6(f)	m4-2(B/C)	g-1(A)
7444	Dog	Italy	16S-2(B)	m2-5(e)	m4-20	g-2(B)
8074	Dog	Slowenia	16S-2(B)	m2-13	m4-2(B/C)	g-2(B)
9774	Dog	Spain	16S-2(B)	m2-2(a)	m4-2(B/C)	g-2(B)
633200.1	Dog	Germany	16S-2(B)	m2-2(a)	m4-2(B/C)	g-2(B)
630300	Dog	Germany	16S-2(B)	m2-2(a)	m4-2(B/C)	g-2(B)
6380	Dog	Germany	16S-2(B)	m2-2(a)	m4-2(B/C)	g-2(B)
5937	Dog	Germany	16S-2(B)	m2-2(a)	m4-2(B/C)	g-2(B)
7641	Dog	Germany	16S-2(B)	m2-2(a)	m4-2(B/C)	g-2(B)
S 1710/05	Dog	Sweden	16S-2(B)	m2-2(a)	m4-2(B/C)	g-2(B)
50.Bella	Dog	Germany	16S-2(B)	m2-2(a)	m4-2(B/C)	g-2(B)
S 654/04	Horse	Germany	16S-2(B)	m2-2(a)	m4-2(B/C)	g-2(B)
S 1829/04	Horse	Germany	16S-2(B)	m2-2(a)	m4-2(B/C)	g-2(B)
S 1741/07	Horse	Germany	16S-2(B)	m2-2(a)	m4-2(B/C)	g-2(B)
S 2614/07	Horse	Germany	16S-2(B)	m2-2(a)	m4-2(B/C)	g-2(B)
S 2630/07	Horse	Germany	16S-2(B)	m2-5(e)	m4-2(B/C)	g-2(B)
S 1071/08	Horse	Germany	16S-2(B)	m2-5(e)	m4-2(B/C)	g-2(B)
S 1220/08	Horse	Germany	16S-2(B)	m2-5(e)	m4-2(B/C)	g-2(B)
S 1379/06	Horse	Germany	16S-2(B)	m2-5(e)	m4-2(B/C)	g-2(B)
S1074/09	Horse	Germany	16S-2(B)	m2-5(e)	m4-2(B/C)	g-2(B)
S1025/09	Horse	Germany	16S-2(B)	m2-5(e)	m4-2(B/C)	g-2(B)
S1085/09	Horse	Germany	16S-2(B)	m2-5(e)	m4-2(B/C)	g-2(B)
S 1201/05	Horse	Germany	16S-3(D)	m2-4(d)	m4-3(D)	g-13(L)

971	Cat	unknown	16S-13(O)	m2-6(f)	m4-2(B/C)	g-2(B)
S 1523/07	Horse	Germany	16S-16(S)	m2-6(f)	m4-18(S)	g-13(L)
D31	Roe deer	Germany	16S-19(V)	m2-22	m4-35	g-4(D)
BwReh9	Roe deer	Germany	16S-20(W)	m2-17	m4-42	g-24
11167	Cattle	Germany	16S-20(W)	m2-3(b)	m4-14(O)	g-3(C)
11047	Cattle	Germany	16S-20(W)	m2-3(b)	m4-14(O)	g-3(C)
11168	Cattle	Germany	16S-20(W)	m2-3(b)	m4-14(O)	g-3(C)
11023	Cattle	Germany	16S-20(W)	m2-3(b)	m4-14(O)	g-3(C)
11208	Cattle	Germany	16S-20(W)	m2-3(b)	m4-14(O)	g-3(C)
11078	Cattle	Germany	16S-20(W)	m2-3(b)	m4-14(O)	g-3(C)
11046	Cattle	Germany	16S-20(W)	m2-3(b)	m4-14(O)	g-3(C)
10946	Cattle	Germany	16S-20(W)	m2-3(b)	m4-14(O)	g-3(C)
11516	Cattle	Germany	16S-20(W)	m2-3(b)	m4-14(O)	g-3(C)
7_19.06.	Cattle	Germany	16S-20(W)	m2-26	m4-49	g-18(X)
14_04.06.	Cattle	Germany	16S-20(W)	m2-26	m4-49	g-18(X)
22_30.05.	Cattle	Germany	16S-20(W)	m2-26	m4-49	g-18(X)
46_22.05.	Cattle	Germany	16S-20(W)	m2-26	m4-49	g-18(X)
57_04.06.	Cattle	Germany	16S-20(W)	m2-26	m4-49	g-18(X)
58_17.05.	Cattle	Germany	16S-20(W)	m2-26	m4-50	g-18(X)
61_22.06.	Cattle	Germany	16S-20(W)	m2-26	m4-49	g-18(X)
BwReh13	Roe deer	Germany	16S-21(X)	m2-20	m4-13(N)	g-30
BwReh14	Roe deer	Germany	16S-	m2-21	m4-13(N)	g-5(E)

			21(X)			
M8 Okt	Goat	Switzerland	16S-21(X)	m2-2(a)	m4-1(A)	g-5(E)
BwReh12	Roe deer	Germany	16S-22(Y)	m2-19	m4-13(N)	g-7(G)

Table 47: Samples with 3/4 partial gene variants

Sample	**Animal species**	**Origin**	**Partial gene combination**			
			16S rRNA	***msp2***	***msp4***	***groEL***
152595	Dog	Germany	16S-1(A)	m2-6(f)		g-1(A)
69	Dog	Albania	16S-1(A)		m4-20	g-1(A)
438	Dog	Albania	16S-1(A)		m4-20	g-1(A)
S1360/09	Dog	Germany	16S-2(B)		m4-20	g-2(B)
150580	Dog	Germany	16S-2(B)		m4-2(B/C)	g-2(B)
153681	Dog	Germany	16S-2(B)		m4-20	g-2(B)
M19	Fallow deer	Germany	16S-2(B)		m4-36	g-8(H)
S226	Red fox	Germany	16S-2(B)		m4-20	g-2(B)
BwReh16	Roe deer	Germany	16S-7(I)		m4-39	g-7(G)
Rothirsch3	Red deer	Austria	16S-8(J)		m4-4(J)	g-8(H)
Reh26	Roe deer	Austria	16S-16(S)		m4-13(N)	g-7(G)
Steinbock6	Ibex	Austria	16S-16(S)		m4-13(N)	g-6(F)
Rothirsch14	Red deer	Austria	16S-16(S)		m4-8(H)	g-7(G)
Steinbock5	Ibex	Austria	16S-16(S)		m4-13(N)	g-6(F)
BwSi3	Sika deer	Germany	16S-16(S)		m4-31	g-34
Rotwild2	Red deer	Austria	16S-17(T)		m4-9(G)	g-6(F)
Gams16	Chamois	Austria	16S-20(W)		m4-10(K)	g-10(J)
Gams15	Chamois	Austria	16S-20(W)		m4-10(K)	g-9(I)
Gams24	Chamois	Austria	16S-20(W)		m4-5(I)	g-8(H)
Mufflon3	Mouflon	Austria	16S-20(W)		m4-5(I)	g-8(H)
Gams25	Chamois	Austria	16S-20(W)		m4-5(I)	g-16(O)
BwRo2	Red deer	Germany	16S-20(W)		m4-7(F)	g-32

BwRo4	Red deer	Germany	16S-20(W)		m4-23	g-33
BwSi12	Sika deer	Germany	16S-20(W)		m4-32	g-13(L)
89	Hedgehog	Germany	16S-20(W)		m4-20	g-1(A)
K41	Roe deer	Germany	16S-20(W)		m4-45	g-25
Z 31-27	Cattle	Germany	16S-20(W)		m4-15(P)	g-16(O)
Z 32-4	Cattle	Germany	16S-20(W)		m4-16(Q)	g-19(Y)
Z 32-6	Cattle	Germany	16S-20(W)		m4-17(R)	g-18(X)
35_17.05.	Cattle	Germany	16S-20(W)	m2-26	m4-49	
22_03.10.	Cattle	Germany	16S-20(W)		m4-51	g-15(N)
28_3.10.	Cattle	Germany	16S-20(W)		m4-51	g-15(N)
30_17.05.	Cattle	Germany	16S-20(W)		m4-50	g-18(X)
35_17.05.	Cattle	Germany	16S-20(W)	m2-26	m4-49	
36_19.05.	Cattle	Germany	16S-20(W)	m2-26	m4-49	
49_19.05.	Cattle	Germany	16S-20(W)		m4-51	g-15(N)
52_14.09.	Cattle	Germany	16S-20(W)	m2-27	m4-51	
53_22.06.	Cattle	Germany	16S-20(W)		m4-49	g-18(X)
Reh24	Roe deer	Austria	16S-21(X)		m4-13(N)	g-6(F)
Reh19	Roe deer	Austria	16S-21(X)		m4-12(M)	g-12(K)
Reh20	Roe deer	Austria	16S-21(X)		m4-13(N)	g-7(G)
BwReh10	Roe deer	Germany	16S-21(X)		m4-13(N)	g-20
A31	Mouflon	Germany	16S-21(X)		m4-21	g-21
Reh27	Roe deer	Austria	16S-22(Y)		m4-13(N)	g-7(G)
Mufflon4	Mouflon	Austria	16S-22(Y)		m4-5(I)	g-8(H)
Reh25	Roe deer	Austria	16S-22(Y)		m4-13(N)	g-7(G)

Reh23	Roe deer	Austria	16S-22(Y)		m4-13(N)	g-6(F)
BwReh8	Roe deer	Germany	16S-22(Y)		m4-13(N)	g-7(G)
BwReh24	Roe deer	Germany	16S-22(Y)		m4-13(N)	g-31
B11 Juni	Goat	Switzerland	16S-22(Y)		m4-1(A)	g-4(D)
G14	Roe deer	Germany	16S-22(Y)	m2-23	m4-13(N)	
L12	Roe deer	Germany	16S-22(Y)		m4-48	g-7(G)
M20	Fallow deer	Germany	16S-22(Y)		m4-40	g-29
BwRo9	Red deer	Germany	16S-25	m2-18	m4-27	
Rotwild23	Red deer	Austria	16S-28		m4-38	g-13(L)
BwSi11	Sika deer	Germany	16S-29		m4-45	g-28

Table 48: Samples with 2/4 partial gene variants

Sample	Animal species	Origin	Partial gene combination			
			16S rRNA	*msp2*	*msp4*	*groEL*
1	Hedgehog	Germany	16S-1(A)			g-1(A)
20	Hedgehog	Germany	16S-1(A)			g-1(A)
79	Hedgehog	Germany	16S-1(A)			g-1(A)
98	Hedgehog	Germany	16S-1(A)			g-1(A)
107	Hedgehog	Germany	16S-1(A)			g-1(A)
126	Hedgehog	Germany	16S-1(A)			g-1(A)
151	Hedgehog	Germany	16S-1(A)			g-1(A)
153	Hedgehog	Germany	16S-1(A)			g-1(A)
154	Hedgehog	Germany	16S-1(A)			g-1(A)
160	Hedgehog	Germany	16S-1(A)			g-1(A)
197	Hedgehog	Germany	16S-1(A)			g-1(A)
261	Hedgehog	Germany	16S-1(A)			g-1(A)
262	Hedgehog	Germany	16S-1(A)			g-1(A)
7225	Dog	Spain	16S-1(A)		m4-2(B/C)	
7996	Dog	Germany	16S-1(A)		m4-	

					2(B/C)	
S 2070/03	Dog	Germany	16S-1(A)	m2-6(f)		
155480	Dog	Germany	16S-1(A)			g-1(A)
8210	Dog	Germany	16S-1(A)			g-1(A)
S133	Red fox	Germany	16S-1(A)			g-1(A)
413	Dog	Albania	16S-1(A)			g-1(A)
Rotwild22	Red deer	Austria	16S-2(B)		m4-37	
BwSi23	Sika deer	Germany	16S-2(B)		m4-21	
144881	Dog	Germany	16S-2(B)			g-2(B)
8340	Dog	Germany	16S-2(B)			g-2(B)
Nike	Dog	Switzerland	16S-2(B)		m4-20	
B17	Mouflon	Germany	16S-2(B)		m4-22	
K26	Mouflon	Germany	16S-2(B)			g-13(L)
BwReh6	Roe deer	Germany	16S-7(I)		m4-23	
BwReh31	Roe deer	Germany	16S-7(I)		m4-23	
BwReh33	Roe deer	Germany	16S-7(I)			g-7(G)
BwSi14	Sika deer	Germany	16S-8(J)		m4-33	
150516	Dog	Germany	16S-9(K)		m4-2(B/C)	
Rothirsch6	Red deer	Austria	16S-16(S)			g-7(G)
Mufflon12	Mouflon	Austria	16S-16(S)			g-21
BwSi7	Sika deer	Germany	16S-16(S)		m4-44	
BwRo1	Red deer	Germany	16S-16(S)		m4-43	
BwRo5	Red deer	Germany	16S-16(S)		m4-25	
BwSi13	Sika deer	Germany	16S-16(S)		m4-46	
BwSi19	Sika deer	Germany	16S-16(S)		m4-34	
BwRo14	Red deer	Germany	16S-16(S)		m4-28	
H24	Mouflon	Germany	16S-16(S)		m4-5(I)	
K37	Mouflon	Germany	16S-16(S)			g-8(H)
Gams38	Chamois	Austria	16S-20(W)			g-27
BwRo6	Red deer	Germany	16S-20(W)		m4-26	
BwDa3	Fallow deer	Germany	16S-			g-24

			20(W)		
BwRo15	Red deer	Germany	16S-20(W)	m4-29	
BwRo18	Red deer	Germany	16S-20(W)	m4-30	
E09	Mouflon	Germany	16S-20(W)	m4-8(H)	
L35	Mouflon	Germany	16S-20(W)	m4-5(I)	
L38	Mouflon	Germany	16S-20(W)		g-13(L)
N04	Mouflon	Germany	16S-20(W)		g-34
O32	Mouflon	Germany	16S-20(W)		g-23
22_09.07.	Cattle	Germany	16S-20(W)	m4-50	
42_04.09.	Cattle	Germany	16S-20(W)	m4-49	
57_04.09.	Cattle	Germany	16S-20(W)	m4-49	
BwSi17	Sika deer	Germany	16S-21(X)	m4-47	
BwReh32	Roe deer	Germany	16S-21(X)		g-6(F)
B10 Juni	Goat	Switzerland	16S-21(X)	m4-2(B/C)	
E25	Roe deer	Germany	16S-21(X)	m4-41	
O08	Roe deer	Germany	16S-21(X)		g-26
BwReh15	Roe deer	Germany	16S-22(Y)	m4-12(M)	
A04	Roe deer	Germany	16S-22(Y)		g-7(G)
P10	Roe deer	Germany	16S-22(Y)		g-20
59_04.09.	Cattle	Germany	16S-22(Y)	m4-13(N)	
A8 Juli	Goat	Switzerland	16S-23(Z)	m4-2(B/C)	
S220	Red fox	Germany	16S-27	m4-20	
BwRo3	Red deer	Germany	16S-30	m4-24	
BwReh4	Roe deer	Germany		m4-2(B/C)	g-6(F)
212	Hedgehog	Germany		m4-20	g-1(A)

61_09.07.	Cattle	Germany	m2-27	m4-51	

1.3. Strain variation in *A. phagocytophilum* genes

Table 49: Variation within the four analyzed genes in ruminants

		cattle	goat	roe deer	red deer	sika deer	fallow deer	mouflon	chamois	ibex
	No. of variants	2	3	9	9	7	5	6	2	1
16S rRNA	Mean no. of variants					4,89				
	Empirical variance					8,32				
	No. of variants	5	2	11	7	5	3	6	5	1
groEL	Mean no. of variants					5,00				
	Empirical variance					7,78				
	No. of variants	8	2	10	17	10	2	4	2	1
msp4	Mean no. of variants					6,22				
	Empirical variance					15,76				
	No. of variants	3	1	6	3	0	1	0	0	0
msp2	Mean no. of variants					1,56				
	Empirical variance					3,80				

Table 50: Variation within the four analyzed genes in non-ruminants

		dog	horse	cat	hedgehog	red fox	wild boar
	No. of variants	4	4	1	5	3	1
16S rRNA	Mean no. of variants			3,00			
	Empirical variance			2,33			
	No. of variants	2	3	1	1	2	0
groEL	Mean no. of variants			1,50			
	Empirical variance			0,92			
	No. of variants	4	3	1	1	1	0
msp4	Mean no. of variants			1,67			
	Empirical variance			1,89			
	No. of variants	5	4	1	4	1	0
msp2	Mean no. of variants			2,50			
	Empirical variance			3,58			

Table 51: Variation within the four analyzed genes in wild animal species

		hedgehog	Red fox	Wild boar	Roe deer	Red deer	Sika deer	Fallow deer	mouflon	chamois	ibex
	No. of variants	5	3	1	9	9	7	5	6	2	1
16S rRNA	Mean no. of variants					4,80					
	Empirical variance					8,16					
	No. of variants	1	2	0	11	7	5	3	6	5	1
groEL	Mean no. of variants					4,10					
	Empirical variance					10,29					
	No. of variants	1	1	0	10	17	10	2	4	2	1
msp4	Mean no. of variants					4,80					
	Empirical variance					15,20					
	No. of variants	4	1	0	6	3	0	1	0	0	0
msp2	Mean no. of variants					1,50					
	Empirical variance					4,05					

Table 52: Variation within the four analyzed genes in domestic animal species

		dog	horse	cat	cattle	goat
	No. of variants	4	4	1	2	3
16S rRNA	Mean no. of variants			2,80		
	Empirical variance			1,36		
	No. of variants	2	3	1	5	2
groEL	Mean no. of variants			2,60		
	Empirical variance			1,84		
	No. of variants	4	3	1	8	2
msp4	Mean no. of variants			3,60		
	Empirical variance			5,84		
	No. of variants	5	4	1	3	1
msp2	Mean no. of variants			2,80		
	Empirical variance			2,56		

1.4. Comparison with sequences from the NCBI GenBank

Table 53: Accession numbers of the *16S rRNA* gene of *A. phagocytophilum* from the GenBank used for comparison

Animal species	Country	Accession numbers
Domestic animals		
Dog	USA	AY741099, AY741098, AY741097, AY741096, AY741095
	Europe	KF242628, KF242627, KF242626, KF242625, KF242624, KF242623, KF242622, KF242621, KF242620, KF242619, KF242618, KF242617, KF242616, KF242615, KF242543, KF242542, KF242541, KF242540, KF242539, GU236707, GU236706, GU236705, GU236704, GU236703, GU236702, GU236701, GU236700, GU236699, GU236698, GU236697, GU236696, GU236695, GU236694, GU236693, GU236692, GU236691, GU236690, GU236689, GU236688, GU236687, GU236686, GU236685, GU236684, GU236683, GU236682, GU236681, GU236680, GU236679, GU236678, GU236677, GU236676, GU236675, GU236674, GU236673. GU236672, GU236671. KC740448, GU236670
	Other (Brasil, Tunesia)	KF576219, EU781707, EU781706
Cat	Europe	HM138366, FJ515308, GU236717, FJ515308
Equids (horse, donkey)	USA	AF172167, AF172165, AF172164
	Europe	EU839852, AY527214, AY527213, KF242656, KF242655, KF242654, KF242653, KF242652, KF242651, KF242650, KF242649, KF242648, KF242647, KF242646, KF242645, KF242644, KF242643, KF242642, KF242641, KF242640, GU236716, GU236715, GU236714, GU236713, GU236712, GU236711, GU236710, GU236709, GU236708, AF482761, JN247407
Cattle	Europe	GU236585, GU236584, GU236583, GU236582, KC740449, KC776921, KC776920
Goat	Europe	KF242659
	Asia	HQ872465, HQ872464, HQ872463
Sheep	Europe	GU236652, GU236651, GU236650, GU236649, GU236648, GU236647, GU236646, GU236645, GU236644, GU236643, GU236642, GU236641, KC740450, GQ428333
	Asia	GQ412338
Wild animals		
Bison	Europe	GU236600, GU236599, GU236598, GU236597, GU236596, GU236595, GU236594, GU236593, GU236592, GU236591, GU236590, GU236589,

		GU236588, GU236586
Chamois	Europe	KF242538, KF242537, KF242536
Fox	Europe	KC763011, KC763010
Moose	Europe	KC800985, KC800983, KC800983
Mufflon	Europe	EU839851
Red deer	Europe	EU839850, AF481855, AF481853, AF481852, KF242657, KF242596, KF242555, KF242554, KF242553, KF242552, GU236581, GU236580, GU236579, GU236578, GU236577, GU236576, GU236575, GU236574, GU236539
Roe deer	USA	AF384214, AF384213, AF384212
	Europe	EU839848, EU839847, AF481854, AF481851, AF481850, KF242658, KF242560, KF242559, KF242558, KF242557, KF242556, GU236573, GU236572, GU236571, GU236570, GU236569, GU236568, GU236567, GU236566, GU236565, GU236564, GU236563, GU236562, GU236561, GU236560, GU236559, GU236558, GU236557, GU236556, GU236555, GU236554, GU236553, GU236552, GU236551, GU236550, GU236549, GU236548, GU236547, GU236546, GU236545, GU236544, GU236543, GU236542, GU236541, GU236540, GU236538, GU236537, GU236536, GU236535, GU236534
Wild boar	Europe	GU391320, GU391319, GU391318, GU391317, GU391316, GU391315, GU391314, GU391313, GU391312, KF242572, KF242571, KF242570, KF242569, KF242568, KF242567, KF242566, KF242565, KF242564, KF242563, KF242562, KF242561
Water deer	Asia	GU556625, GU556624, GU556623, GU556622, GU556621
Small mammals		
Rodent	USA	AY144728, AY094353
	Europe	KF481948, KF481947, KF481946, KF481945, KF481944, KF481943, KF481942, KF481941, KF481940, KF481939, KF481938, KF481937, KF481935, KF481933, KF481932, KF481931, KF481929, KF481928, AY082656, KC740432, KC740441, KC740440, KC740439, KC740438, KC740437, KC740436, KC740435, KC740434, KC740433, KC740431, KC740430, KC740429, KC740428, KC740427, KC740426, KC740425, KC740424, KC740423, KC740422, KC740421, KC740420, KC740419, KC740418, KC740444, KC740443, KC740442
	Asia	DQ458808, DQ458807, DQ458805, DQ342324, KC470064, GQ412339, GQ412337, FJ968659, FJ968655, DQ449945, KC583437, KC583436, KC583435, HQ630625, HQ630624, HQ630623, HQ630622, HQ630621

Rabbit	Asia	DQ458806
Human	USA	NR044762, AY886761, M73224, M73223, M73220, AF093789, AF093788
	Europe	EU839857, KF242614, KF242613, KF242612, KF242611, KF242610, KF242609, KF242608, KF242607, KF242606, KF242605, KF242604, KF242603, KF242602, KF242601, KF242600, KF242599, KF242598, KF242597, KF242551, KF242550, KF242549, KF242548, KF242547, KF242546, KF242545, KF242544, GU236664, GU236663, GU236662, GU236661, GU236660, GU236659, GU236658, GU236657, GU236656, GU236655, GU236654, GU236653, KC740447, KC740446, KC740445

Table 54: Accession numbers of the *groEL* gene of *A. phagocytophilum* from the GenBank used for comparison

Animal species	Country	Accession number
Domestic animals		
Dog	USA	JF494833, AY219849
	Europe	EU982549, EU381151, EU381151, AY848752, AY848751, AY848750
Cat	USA	DQ680012
Equids (horse, donkey)	USA	JF494840, EF647585, AF172161, AF172162, AF172160, AF172158
	Europe	AY848749, AY848748, AY848747, AY529490, AY529489
Sheep	Europe	EU860089, EU860088, AF548386, AF548385
Wild animals		
Cervid (deer)	Europe	AF478564, EU157921, EU157920, DQ779568
Moose	Europe	KC800986, KC800984, KC800984
Roe deer	USA	AF383227, AF383226, AF383225
	Europe	JN005748, JN005747, JN005746, JN005745, JN005744, JN005743, AY220469, AY220468, AY220467, AF478564, AF478563, AF478562, AF478561, AF478560, AF478559, AF478558, AF478557, AF478556, AF478555, AF478554, AF478553, AF478552, AF478551
Sika deer	Asia	JN055360, JN055359
Wild boar	Europe	EU184703
Water deer	Asia	HM752098
Small mammals		
Rodent	USA	JF494841, JF494838, JF494837, JF494836, JF494835, JF494834, AY626252, DQ088133
	Europe	KF383231, KF383230, KF383229, KF383228,

		KF383227, AF192796
	Asia	KC583433, KC583432, KC583431, KC753762, HQ630619, HQ630618, HQ630617, HQ630616, HQ630615, HQ630614
Human	USA	JF494839, AF172163, AF172159, AF033101
	Europe	EU860090

Table 55: Accession numbers of the *msp4* gene of *A. phagocytophilum* from the GenBank used for comparison

Animal species	Country	Accession number
Domestic animals		
Equids (horse, donkey)	Europe	AY702925
Cattle	Europe	EU857670, EU857669, EU857668, EU857667, EU857666, EU857665, AY829456, AY829455
Sheep	USA	AY706391
	Europe	EU857673, EU857672, EU857671, EU436164, JQ522935, HQ661163, HQ661162
Wild animals		
Bison	Europe	AY706389, AY706388, AY706387
Red deer	Europe	EU180065, EU180064, EU180063, EU180061, EU180059, EU180058
Reindeer	Europe	JX841250
Roe deer	Europe	JN005737, JN005734, JN005733, JN005728, JN005726, JN005725, EU180066, EU180062, EU180060, AY829457, EF067343
Small mammals		
Rodent	Europe	KF420102, KF420101, KF420100, KF420099, KF420098, KF420097, KF420096, KF420095, KF420094, KF420093, KF420092
	Asia	EU008082, GQ412348, GQ412347, GQ412346
Human	USA	AY530194
	Europe	EU857674

Table 56: Accession numbers of the *msp2* gene of *A. phagocytophilum* from the GenBank used for comparison

Animal species	Country of origin	Accession number
Domestic animals		
Dog	USA	FJ467335, FJ467334, FJ467333, DQ519568

Equids (horse, donkey)	USA	FJ467336
Cattle	USA	AY706392
Sheep	USA	DQ519569
	Europe	AY706393, AY541004
Wild animals		
Bear	USA	DQ519567
Small mammals		
Rodent	USA	DQ519570
Human	USA	AY164494, AY164493, AY164492, AY164491, AY164490, AY541007, AY541006, AY541005

XIII. ACKNOWLEDGEMENTS

First of all, I would like to thank my supervisor **PD Dr. Cornelia Silaghi** for her excellent and extremely helpful mentoring during this study. I am very grateful for her immense support and advice throughout this doctorate. Especially the weekly "Doktoranden Meetings" and her accurate and reliable corrections of manuscripts were very productive and motivating for me. With her patience and understanding, she enabled the finishing of this work. I am very grateful for having had the opportunity to collaborate with such a great mentor like her!

My sincere thanks to all the researchers of the previous studies for providing animal and/or DNA samples, amplicons or sequences of *A. phagocytophilum* available for the analysis in the present study: Thank you to **Prof. Dr. Barbara Kohn** (Klinik für Kleine Haustiere, Freie Universität Berlin) for some of the dog samples**, Dr. Anna Obiegala** (Institut für Öffentliches Veterinärwesen, Universität Leipzig) for the rodent samples, **Marion Nieder** (Tierärztin) for the cattle samples, **Dr. Evelyn Overzier** (Institut für Infektionsmedizin und Zoonosen, Ludwig-Maximilians-Universität München) for some of the roe deer sequences and samples, **PD Dr. Steffen Rehbein** (Merial GmbH, Forschungszentrum Kathrinenhof, Rohrdorf) for some of the wild animal samples, **Dr. Dietmar Hamel** (Merial GmbH, Forschungszentrum Kathrinenhof, Rohrdorf) for some of the dog samples, **Dr. Elisabeth Louise Meyer-Kayser** (Tierärztin) for the fox samples and to **Melanie Kauffmann** (Tierärztin) for some of the mouflon and roe deer sequences.

A special thanks to **Matthias Aßenmacher** from the Statistical Consulting Unit (StaBLab, Institute of Statistics) of the Ludwig-Maximilians-University Munich for the advice of the statistical analysis.

For financial support I would like to thank the European Union within the FP7-project 261504 EDENext.

I would especially like to express my gratitude to **Claudia Thiel**, who taught me the necessary skills of molecular techniques in the laboratory and also supported me in the lab at busy workdays. I would also like to thank **all the members of the Institute of Comparative Tropical Medicine and Parasitology** for such a

heartily welcome throughout my time as a doctoral candidate. I am very grateful for this supportive and motivating atmosphere.

My deepest and warmest thank you to **Dr. Marie Franziska Sommer**, with whom I shared my room at the Institute of Comparative Tropical Medicine and Parasitology. Without the motivation, support and inspiring discussions with her, this work would not have been accomplished, of course not to forget the countless laughs and chocolate bars we shared. A special thanks also to my fellow doctoral candidates **Dr. Anna Obiegala**, **Alexandra Kaspar**, **Dr. Carina Schüle** and **Dr. Miriam Wächter.** We shared great moments together at the institute.

It would not have been possible to accomplish this thesis without the help and support of my **family and friends**. Therefore, I owe my deep gratitude to my parents, who supported me throughout my life, and my sisters and brother. A very warm and special thank you to my twin sister **Sarah Fröhlich**, who always found the right words and helpful support to motivate me. Also, I would like to thank my friends, the "6-Pack ohne Butterbrot" with **Carolin Schillmeier**, **Nadine Keller**, **Melanie Zerbst** and **Anna-Marlen Weissenhorn**, with whom I shared several unique moments in my life and experienced real friendship with motivation, support and very welcoming distraction during this doctorate.