University of South Bohemia
Faculty of Science

2008

Master Thesis

Arsenophonus, an emerging clade of intracellular symbionts with a broad host distribution

Eva Nováková

Supervisor: Václav Hypša

Nováková, E., 2008: *Arsenophonus*, an emerging clade of intracellular symbionts with a broad host distribution. Mgr. Thesis, In English.- xx p., Faculty of Science, The University of South Bohemia, České Budějovice, Czech Republic.

Anotace:

The thesis brings new phylogenetic data on the rapidly growing bacterial genus *Arsenophonus*, and summarizes the current evolutionary picture of its symbiotic association with insects and other hosts. The study is prepared in form of a regular publication and will be submitted in microbiological journal.

Prohlašuji, že svoji diplomovou práci jsem vypracovala samostatně pouze s použitím pramenů a literatury uvedených v seznamu citované literatury.

Prohlašuji, že v souladu s § 47b zákona č. 111/1998 Sb. v platném znění souhlasím se zveřejněním své, a to v nezkrácené podobě elektronickou cestou ve veřejně přístupné části databáze STAG provozované Jihočeskou univerzitou v Českých Budějovicích na jejích internetových stránkách.

V Českých Budějovicích 6.5.2008

Eva Nováková

Contents:

Introduction 1

Materials and Methods 4

Samples 4

DNA extraction, PCR and sequencing 4

Alignments 5

Phylogenetic analyses 6

Results 8

Sequences and alignments 8

Phylogeny 8

Discussion 9

Phylogenetic patterns and the stability of the information 9

Relevance of the sampling 10

Cophylogeny vs. horizontal transfers: possible sources of phylogenetic incongruence 12

LGT and putative role of bacteriophages 13

Incongruence and intragenomic heterogeneity 14

Conclusion and taxonomic implications 15

Acknowledgements 16

References 16

Figures 22

Introduction

Bacterial genus *Arsenophonus* is currently the most rapidly growing group of insect intracellular symbionts. In the last several years there has been not only a remarkable flux of new *Arsenophonus* sequences, (Fig. 1) but also an ever increasing complexity in its evolutionary traits, namely the host-distribution and coevolutionary patterns. This is a particularly instructive situation since the genus *Arsenophonus* is not new to the insect-symbiont research. First records of these bacteria date back to the pre-molecular era; based on ultrastructural features, several authors described a transovarially transmitted infection associated with the son-killer phenomenon in the parasitoid wasp *Nasonia vitripennis* (Huger et al., 1985; Skinner, 1985; Werren et al., 1986). Later, they were formally assigned to a new genus within the family Enterobacteriacea with a single species, *Arsenophonus nasoniae* (Gherna *et al.*, 1991). The same authors also proposed a close relationship of *Arsenophonus* to free-living bacteria of the genus *Proteus*.

Independently on these reports, other microscopic studies revealed morphologically similar symbionts from various tissues of triatomine bugs (Louis et al., 1986; Hypša, 1993). However, only a decade later could these bacteria be determined on the molecular grounds as a member of the same bacterial clade and described as *Arsenophonus triatominarum* (Hypša and Dale, 1997). Interestingly, the next record on symbiotic bacteria closely related to *A. nasoniae* was brought by a phytopathological study investigating marginal chlorosis of strawberry (Zreik et al., 1998). Since the available sequence data were then insufficient for a reliable phylogenetical placing, the phloem-inhabiting pathogen was described as a new genus *Phlomobacter*, with a single species *P. fragariae* (Zreik et al., 1998). Additional studies revealed that the causative agent of another plant disease, the basses richesses syndrome of sugar beet, clustered within the same lineage (Semetey et al., 2007).

Since these descriptions, the number of *Arsenophonus* records has steadily been increasing. This led to two important changes in the view on *Arsenophonus* evolution and significance. First, the known host spectrum has been considerably extended with various insect groups and even non-insect taxa. Till now, members of the genus *Arsenophonus* have been determined from parasitic wasps, triatomine bugs, psyllids, whiteflies, aphids, ticks, ant lions, hippoboscids, streblids, bees, lice, and two plant species (Gherna et al., 1991; Zreik et al., 1998; Subandiyah et al., 2000; Thao et al., 2000; Spaulding and von Dohlen, 2001; Zchori-Fein and Brown, 2002; Grindle et al., 2003; Russell et al., 2003; Thao and Baumann,

2004b; Dunn and Stabb, 2005; Dale et al., 2006; Trowbridge et al., 2006; Allen et al., 2007; Babendreier et al., 2007; Hansen et al., 2007; Semetey et al., 2007; Sorfova et al., 2008). The second consequence of this extension is an emergence of ever growing diversity of symbiotic types contained within the genus. This dramatically changes the original perception of *Arsenophonus* as a bionomically homogeneous group of typical S-symbionts undergoing frequent horizontal transfers among phylogenetically distant hosts. For example, recent findings indicate that some insect groups harbour monophyletic clusters of *Arsenophonus*, possibly playing a role of typical P-symbionts. Such groups were reported from the dipteran families Hippoboscidae and Streblidae (Trowbridge et al., 2006) and most recently from several lice species (Sasaki-Fukatsu et al., 2006; Allen et al., 2007; Perotti et al., 2007).

This arrangement, intertangling different types of symbiotic bacteria, is not entirely unique among insect symbionts. The conception of S- and P-symbionts itself was coined before the molecular era and was later proved to be inapplicable to many insect-symbiont associations. Its original purpose was to distinguish evolutionary original and nutritionally essential bacteria inhabiting specialized host cells (P-symbionts) from incidental bacterial infections with nonspecific or even deleterious impact on the host (S-symbionts). In concord with this delimitation, it has been generally assumed that P-symbionts possess an exclusive vertical transmition from parents to progeny, and a strictly congruent phylogeny with the host. In contrast, S-symbionts are usually associated with their arthropod host for considerably shorter evolutionary periods and undergo frequent horizontal transfers. During the modern era of molecular biology, several genomic features were added to these characteristics. Particularly, typical P-symbionts are known to suffer considerable genome degeneration manifested mainly by loss of genes and decrease of guanine-cytosine content in their DNA (Moran, 1996).

However, with the increasing amount of knowledge on the heterogeneity and evolutionary dynamics of the symbiotic associations, it is becoming clear that there are no distinct boundaries between these two ecological types. Thus, in their strict meaning, the terms P- and S-symbionts have recently become insufficient, especially for more complex situations, such as studies exploring bacterial diversity within a single host species (Zchori-Fein and Brown, 2002; Dunn and Stabb, 2005). Furthermore, this terminological framework was shown not to reflect phylogenetic position of the symbiotic lineages. Remarkable versatility of the symbiotic associations can be observed in the entire subdivision of gamma

Proteobacteria as well as within the individual clusters, such as *Arsenophonus* or *Sodalis* (Dale et al., 2006; Fukatsu et al., 2007).

The ability of symbiotic bacteria to switch readily between different strategies was experimentally confirmed by a functional replacement of P-symbiotic bacteria with S-symbionts (Koga et al., 2003). A similar picture has been observed within the genus *Arsenophonus*. Apart from many secondary lineages, this genus gave rise to several P-symbiont clusters (Trowbridge et al., 2006; Allen et al., 2007; Perotti et al., 2007). Unfortunately, this heterogeneity introduced an annoying degree of phylogenetic instability and nomenclatory confusion. For example, an unpredictable clustering is well known for the sequences of long-branched P-symboints within Enterobacteriace (Herbeck et al., 2005), and the same behaviour can be seen in the louse-specific clade of *Arsenophonus*. As consequence, these symbionts may not always be recognized as part of the genus *Arsenophonus* and were originally described as a new bacterial genus *Riesia* (Sasaki-Fukatsu et al., 2006). In addition, *Arsenophonus* cluster is the only monophyletic group of symbiotic bacteria possessing at least four highly different phenotypes, including son-killer phenomenon (Gherna et al., 1991), phytopathogenicity (Zreik et al., 1998), P-symbionts living in a close association with the host (Trowbridge et al., 2006; Allen et al., 2007; Perotti et al., 2007), and apparently non-specific horizontally transmitted bacteria (Thao and Baumann, 2004b).

These characteristics indicate that the genus *Arsenophonus* represents an important and widespread lineage of symbiotic bacteria which significance has long been underestimated. Similarly to *Wolbachia* or *Sodalis* endosymbionts, possessing anologous mixture of horizontal transfer and co-speciation history, *Arsenophonus* may serve as a valuable source of data and suitable model for examining molecular evolution of bacteria-arthropod associations.

In this study, we add 34 new records on symbionts into the known spectrum of *Arsenophonus* lineages. We explore and summarize the current picture of *Arsenophonus* evolution by analyzing all available data on this rapidly growing clade. To investigate the position, stability and evolutionary trends of the *Arsenophonus* cluster in the frame of Enterobacteriacea, we complete the sample with related symbionts and free-living bacteria. Finally, we explore molecular characteristics and informative value of the 16S rRNA gene as the most frequently used phylogenetic marker.

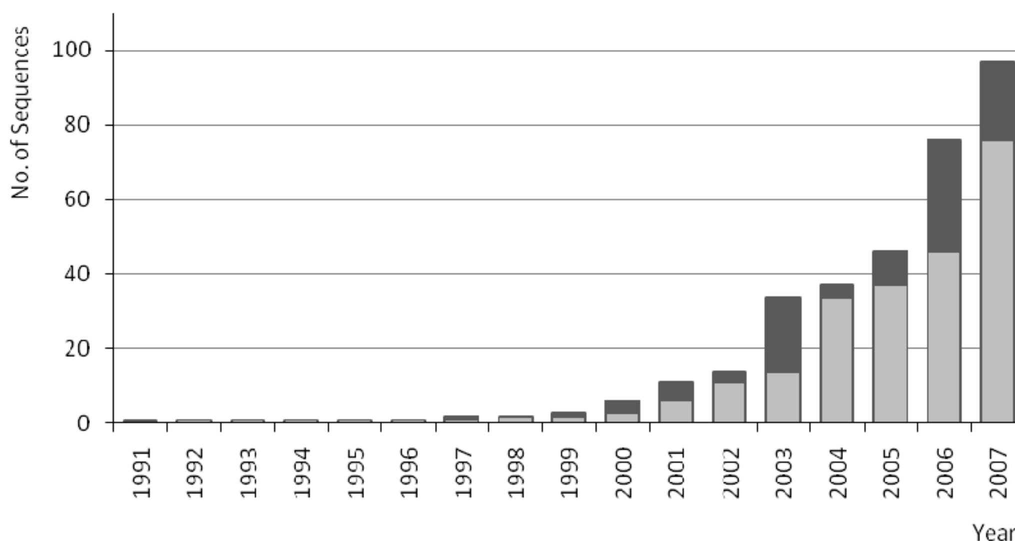

Figure 1. An increase of records on *Arsenophonus* bacteria from various insect groups. The bars show cumulative numbers of sequences deposited into the Genbank; dark tops represent new records added in the given year. The sequences are identified with the following accession numbers: 1991 - M90801; 1997 - U91786; 2000 - AF263561, AF263562, AF286129, AB038366; 2001 - AF400474, AF400480, AF400481, AF400478, AY057392; 2002 - AY136168, AY136153, AY136142; 2003 - AY265341-AY265348, Y264663-AY264673, AY264677; 2004 - AY587141, AY587142, AY587140; 2005 - DQ068928, DQ314770-DQ314774, DQ314777, DQ314768, DQ115536; 2006 - DQ538372-DQ538379, DQ508171-DQ508186, DQ517447, DQ508193, DQ837612, DQ837613; 2007 - EU039464, EU043378, EF110573, EF110574, DQ076660, DQ076659, EF110572, EF647590, AB263104.

Materials and Methods

Samples

The host species used in this study were acquired from several sources. All of the nycteribid samples were obtained from Radek Lucan. Most of the hippoboscids were provided by Jan Votypka. Ant species were collected by Milan Janda at Papua New Guinea. The rest of samples are from the authors‘ collection. Taxonomic position and geografic origin of the hosts are summarized in the Tab. 1.

DNA extraction, PCR and sequencing

The total genomic DNA was extracted from individual samples using DNEasy Tissue Kit (QIAGEN; Hilden, Germany). Primers F40 and R1060 designed to amplify approx. 1020 bp of 16S rDNA, particularly within Enterobacteriaceae (Hypša and Křížek, 2007) were used for all samples. PCR was performed under standard conditions using HotStart Taq polymerase

(HotStarTaqi DNA Polymerase, Qiagen). The PCR products were analyzed by gel electrophoresis and cloned into pGEM-T Easy System 1 vector (Promega). Inserts from selected colonies were amplified using T7 and SP6 primers and sequenced in both directions, with the exception of 3 fragments sequenced in one direction only (sequences from *Aenictus huonicus* and *Myzocalis sp.*). DNA sequencing was performed on automated sequencer model 310 ABI PRISM (PE-Biosystems, Forter City, California, USA) using the BigDye DNA sequencing kit (PE-Biosystems). For contig construction and sequence editing we used the SeqMan program from the DNASTAR platform (Dnastar, Inc. 1999). Identification of the sequences was done using BLAST, NCBI (http://www.ncbi.nlm.nih.gov/blast/Blast.cgi).

Alignments

To analyse thoroughly the behaviour of *Arsenophonus* 16S rDNA and assess its usefulness as phylogenetic marker, we prepared several matrices and performed an array of phylogenetic analyses on each of them.

The *Basic matrix* was composed of the 34 new sequences, all *Arsenophonus* sequences available in the GenBank and addional 45 sequences of various P-symbionts, S-symbiont and 5 free living bacteria as outgroups (Tab. 1). To show the impact of random or restricted sampling on the resulting topology, five different matrices labelled $Sampling_i$ (i.e. *Sampling1*, *Sampling2*, etc.) were prepared from *Basic matrix* by removing various taxa and including additional/alternative outgroups. The matrices *Sampling1* to *Sampling4* were composed of various numbers of non-*Arsenophonus* symbiotic taxa (ranging from 35 to 3), three sequences of free living bacteria, and arbitrary selected set of all *Arsenophonus* lineages. Matrix designated as *Sampling5* was restricted to a lower number of taxa, including 5 ingroup sequences and alternative lineages of symbiotic and free-living bacteria.

All matrices were alligned in server-based program MAFFT (http://align.bmr.kyushu-u.ac.jp/mafft/online/server/), using the E-INS-i algorhitm with default parameters. The program BioEdit (Hall, 1999) was used to manually correct the resulting matrices and to calculate the GC contant of the sequences.

To test an effect of unreliably aligned regions on the phylogenetic analysis, we further prepared the *Conservative matrix*, by removing variable regions from the *Basic matrix*. For this procedure, we used the program Gblocks (Castresana, 2000) available as server-based application on the web page http://molevol.ibmb.csic.es/Gblocks_server.html.

Finally, the *Clock matrix*, composed of 12 bacterial sequence (see Tab. 1), was designed to calculate time divergencies for several nodes within the *Arsenophonus* topology.

Phylogenetic analyses

The matrices were analyzed under the maximum parsimony (MP), maximum likelihood (ML) criteria and by calculation of Bayesian probability (due to its computing demands, the Bayesian analyses could only be performed on several selected matrices). For the analyses, we used the following programs and procedures. MP analysis was carried out in TNT program (Goloboff et al., 2004) using the *Traditional search* option, with 100 replicates of heuristic search, under the assumptions of Ts/Tv ratio 1 and 3. ML analysis was done in the Phyml program (Guindon and Gascuel, 2003) under the *GTR+Γ+inv* model, with parameters estimated from the data. Bayesian analysis was performed in Mr.Bayes ver. 3.1.2. with following parameter settings: nst=6, rates= invgamma, ngen=3000000, samplefreq=100, and printfreq=100 . The program Phylowin (Galtier et al., 1996) was employed for the ML analysis under the nonhomogeneous model of the substitutions (Galtier and Gouy, 1998).

A calculation of divergence time was performed in the program Beast (Drummond et al., 2002) which implements MCMC procedure to sample target distribution of the posterior probabilities. GTR + invgamma model of molecular evolution, determined as best fitting by the Modeltest (Posada and Crandall, 1998), was used to estimate the trees and branch lenghts. The gamma distributon was approximated by 6 categories of substitutional rates. Relaxed molecular clock (uncorrelated lognormal option) was applied to model the rates along the lineages. To obtain a time-framework for the tree, we used the estimate on lice divergence (approximately 5.6 mya; (Allen et al., 2007). Since the resulting estimate was considerably lower that that reported previously with *Escherichia-Salmonella* calibration (Sorfova et al., 2008), we prepared additional matrix and used the *Escherichia-Salmonella* split (Ochman and Wilson, 1987; Battistuzzi et al., 2004) as alternative calibration; taxa icluded according to Sorfova et al., (2008). All analyses were performed in three independent runs, each taking 5 million generations.

Host Species	Symbiont designation	Origin	Abb.
Hippobosca camelina	*Hippobosca camelina* 1	Kenya	HippCaC
	Hippobosca camelina 2	Kenya	HippCaB
	Hippobosca camelina 3	Kenya	HippCaA
Ornithomyia avicularia	*Ornithomyia avicularia* 1	CZ	OrniAvA
	Ornithomyia avicularia 2	CZ	OrniAv3
	Ornithomyia avicularia 3	CZ	OrnAvc2
	Ornithomyia avicularia 4	CZ	OrnAvc8
	Ornithomyia avicularia 5	CZ	OAc29Ars
	Ornithomyia avicularia 6	CZ	OrnAc24
	Ornithomyia avicularia 7	CZ	OrnAc20
	Ornithomyia avicularia 8	CZ	OrnA2c4
	Ornithomyia avicularia 9	CZ	OrnA2c6
Ornithomyia biloba	*Ornithomyia biloba* 1	CZ	OrniBi1
	Ornithomyia biloba 2	CZ	OrnBi12
	Ornithomyia biloba 3	CZ	12c8Ar
	Ornithomyia chloropus	CZ	23ArsD
Lipoptena fortisetosa	*Lipoptena fortisetosa* 1	CZ	LipForB
	Lipoptena fortisetosa 2	CZ	LipForC
	Lipoptena fortisetosa 3	CZ	LipFo14
Lipoptena cervi	*Lipoptena cervi* 1	CZ	LipCe7c
	Lipoptena cervi 2	CZ	LipCe7i
Lipoptena sp.	*Lipoptena* sp.	CZ	LipSp13
Melophagus ovinus	*Melophagus ovinus*	CZ	MelOvi
Nycteribia kolenati	*Nycteribia kolenatii* 1	CZ	NycKo2B
	Nycteribia kolenatii 2	CZ	NyKo2A
Nycteribia sp	*Nycteribia* sp.	CZ	NyctSp2
Penicilidia monoceros	*Penicilidia monoceros*	CZ	PenMo01
	Penicilidia sp.	Syria	Nyctdae
	Trichobius sp.	Venezuela	X50
Technomyrmex albipes	*Technomyrmex albipes* 1	P. N. Guinea	AntPNG1
	Technomyrmex albipes 2	P. N. Guinea	AntPNG2
Aenictus huonicus	*Aenictus huonicus*	P. N. Guinea	1028MJ
Myzocalis sp.	*Myzocalis* sp. 1	USA, AZ	MyzSpMM
	Myzocalis sp. 2	USA, AZ	MyzSpII
	Sequences retrieved from GenBank		
Dermacentor variabilis	*Dermacentor variabilis* 1	AY265348	ADerV01
	Dermacentor variabilis 2	AY265347	ADerV02
	Dermacentor variabilis 3	AY265346	ADerV03
	Dermacentor variabilis 4	AY265345	ADerV04
	Dermacentor variabilis 5	AY265344	ADerV05
	Dermacentor variabilis 6	AY265343	ADerV06
	Dermacentor variabilis 7	AY265342	ADerV07
	Dermacentor variabilis 8	AY265341	ADerV08
Lipoptena cervi	*Lipoptena cervi 3*	DQ314777	DLipC02
	Lipoptena cervi 4	DQ314778	DLipC01
Trichobius longipes	*Trichobius longipes 1*	DQ314773	DTriL06
	Trichobius longipes 2	DQ314772	DTriL07
	Trichobius longipes 3	DQ314771	DTriL08
	Trichobius longipes 4	DQ314770	DTriL09
Trichobius parasiticus	*Trichobius parasiticus* 1	DQ314774	DTriP05
	Trichobius parasiticus 2	DQ314775	DTriP04
Trichobius caecus	*Trichobius caecus 1*	DQ314769	DTriC10
Trichobius caecus	*Trichobius caecus 2*	DQ314768	DTriC11
Australiococcus greville	*Australiococcus greville*	AY264673	AustGre
Dialeurodes hongkongensis	*Dialeurodes hongkongensis*	AY264667	DialHon
Tetraleurodes acaciae	*Tetraleurodes acaciae*	AY264670	TetraAc
Trialeurodes hutchingsi	*Trialeurodes hutchingsi*	AY587140	TrialHu
Trialeurodes vaporariorum	*Trialeurodes vaporariorum*	AY264672	TrialVa
Tetraleurodes mori	*Tetraleurodes mori*	AY264671	TetraMo
Aleyrodes proletella	*Aleyrodes proletella*	AY587141	AleyrPr
Aleyrodes elevatus	*Aleyrodes elevatus*	AY264666	AleyrEl
Aleuroplatus gelatinosus	*Aleuroplatus gelatinosus*	AY264665	AleurGe
Neomaskellia andropogonis	*Neomaskellia andropogonis*	AY264668	NeomaAn
Aleurodicus dispersus	*Aleurodicus dispersus*	AY264664	AleuDis
Aleurodicus dugesii	*Aleurodicus dugesii* 1	AY587142	AleurDu
	Aleurodicus dugesii 2	AF286129	AleuDug
Acanthaleyrodes styraci	*Acanthaleyrodes styraci*	AY264663	AcantSt
Bemisia tabaci	*Bemisia tabaci* 1	AF400474	BemTa74
	Bemisia tabaci 2	AF400480	BemTa80
	Bemisia tabaci 3	AF400481	BemTa2
	Bemisia tabaci 4	AF400478	BemTa78
	Bemisia tabaci 5	AY264677	BemiTab
Heteropsylla texana	*Heteropsylla texana*	AF263562	HeteTex
Siphoninus phillyreae	*Siphoninus phillyreae*	AY264669	SiphoPh
Nasonia vitripenis	*Arsenophonus nasoniae*	M90801	ArseNas
Triatoma infestans	*Arsenophonus triatominarum* 1	U91786	ArsTr97
	Arsenophonus triatominarum 2	DQ508185	ArsTria
Diaphorina citri	*Diaphorina citri*	AB038366	DiapCit
Pseudolynchia canariensis	*Arsenophonus insecticola*	DQ115536	ArseIns
Wahlgreniella nervata	*Wahlgreniella nervata*	AY136168	WahlNer
Myzocallis sp.	*Myzocallis* sp. 3	AY136153	MyzocSp.
Aphis spiraecola	*Aphis spiraecola*	AY136142	ApSArs
Beta vulgaris	*Beta vulgaris*	AY057392	SugBeet
Fragaria vesca	*Phlomobacter fragariae 1*	DQ538372	Strawb2
	Phlomobacter fragariae 2	DQ538377	Strawb3
	Phlomobacter fragariae 3	DQ538374	Strawb4
	Phlomobacter fragariae 4	DQ538375	Strawb5
	Phlomobacter fragariae 5	DQ538376	Strawb6
	Phlomobacter fragariae 6	DQ538377	Strawb7
	Phlomobacter fragariae 7	DQ538378	Strawb8
	Phlomobacter fragariae 8	DQ538379	Strawb9
Myrmeleon mobilis	*Myrmeleon mobilis*	DQ068928	AntLion
Aphis mellifera	*Aphis mellifera* 1	DQ837613	HonBee1
	Aphis mellifera 2	DQ837612	HonBee2
Glycaspis brimblecombei	*Glycaspis brimblecombei* 1	EU043378	GlycBr3
	Glycaspis brimblecombei 2	EU039464	GlycBr1
Pediculus schaeffi	*Riesia pediculischaeffi*	EF110573	RiPsch
Pediculus capitis	*Riesia pediculicola 1*	DQ076660	PedHum
Pediculus humanus	*Riesia pediculicola 2*	DQ076659	PedCap
	Riesia pediculicola 3	EF110572	Riesi5
	Riesia pediculicola 4	EF647590	Riesi2
	Riesia pediculicola 5	AB263104	Riesi6
Pthirus pubis	*Riesia pthiripubis*	EF110574	Riphth
Triatoma rubrofasciata	*Triatoma rubrofasciata* 1	DQ508185	Trub16E
	Triatoma rubrofasciata 2	DQ508169	Trub16R
Eratyrus mucronatus	*Eratyrus mucronatus* 1	DQ508198	Erat16E
Eratyrus mucronatus	*Eratyrus mucronatus* 2	DQ508182	Erat16R
Triatoma melanosoma	*Triatoma melanosoma* 2	DQ508172	Mazo16R
	Triatoma melanosoma 1	DQ508188	Mazo16E
Meccus mazzoti	*Meccus mazzoti* 1	DQ517448	Tmel16E
	Meccus mazzoti 2	DQ517447	Tmel16R
Aphalaroida inermis	*Aphalaroida inermis*	AF263556	AphIn
Cuerna costalis	*Baumannia cicadellinicola*	AY676895	Bauma1
Psylla floccosa	*Psylla floccosa*	AF286128	PsyFlo
Bactericera cockerelli	*Bactericera cockerelli*	AF263557	BacCoc
Heteropsylla cubana	*Heteropsylla cubana*	AF286126	HetCub
Eulachnus pallidus	*Eulachnus pallidus*	AY136147	EulPaX
Blastopsylla occidentalis	*Blastopsylla occidentalis 1*	AF263558	BlaOc1
	Blastopsylla occidentalis 2	AF077608	BlaOc2
Cacopsylla pyri	*Cacopsylla pyri*	AY136145	CacPyr
Uroleucon ambrosiae	*Hamiltonella defensa 1*	AF293622	Hamil1
Aphis craccivora	*Hamiltonella defensa* 2	AY136136	Hamil2
Cacopsylla myrthi	*Cacopsylla myrthi*	AF263559	CacMyr
Psylla pyricola	*Psylla pyricola*	AF286125	PsyPys
Calophya schini	*Calophya schini*	AF263560	CalSch
Haematopinus apri	*Haematopinus apri*	DQ076665	HaemAp
Metamasius callizona	*Nardonella sp.*	AY126634	Nardo2
Trioza magnoliae	*Trioza magnoliae*	AF077607	TriMag
Haematomyzus elephantis	*Haematomyzus elephantis*	DQ076663	HemElJ
Glossina morsitans centralis	*Wigglesworthia glossinidia*	AF022878	Wiggles
Pseudolynchia canariensis	*Pseudolynchia canariensis*	DQ115535	PseCan
Planococcus ficus	*Planococcus ficus*	AF476108	PlaFic
Planococcus citri	*Planococcus citri 1*	AF322016	PlaCi1
	Planococcus citri 2	AF476107	PlaCi2
Paracoccus nothofagicola	*Paracoccus nothofagicola*	AF476109	ParNot
Cyphonococcus alpinus	*Cyphonococcus alpinus*	AF476102	CypAlp
Melanococcus albizziae	*Melanococcus albizziae*	AF476106	MelAlb
Australicoccus grevilleae	*Australicoccus grevilleae*	AF476099	AusGre
Antonina pretiosa	*Antonina pretiosa*	AF476101	AntPre
Antonina crawii	*Antonina crawii*	AB030020	AntCra
Amonostherium lichtensioides	*Amonostherium lichtensioides*	AF476100	AmoLic
Vryburgia amaryllidis	*Vryburgia amaryllidis*	AF476110	VryAma
Erium globosum	*Erium globosum*	AF476105	EriGlo
Dysmicoccus neobrevipes	*Dysmicoccus neobrevipes*	AF476104	DisNe1
Dysmicoccus brevipes	*Dysmicoccus brevipes*	AF476103	DisBre
Anomoneura mori	*Anomoneura mori*	AB013086	AnoMor
Camponotus fellah	*Blochmania fellah*	EF422835	Bloch1
Diuraphis noxia	*Buchnera aphidicola*	M63251	Buchne
Coptosoma parvipictum	*Ishikawaella capsulata*	AB244769	Ishik1
Bemisia tabaci	*Portiera aleyrodidarum* 1	AF400451	BemTa1
	Portiera aleyrodidarum 2	AF400458	BemTa3
Macrosiphoniella ludovicianae	*Regiella insecticola* 1	AF293619	RegIn1
Pemphigus betae	*Regiella insecticola* 2	AY136154	RegIn2
*	*Proteus mirabilis*	AB079370	ProtMir
*	*Photorhabdus luminescens*	BX571859	PhotLum
*	*Pasteurella multocida*	AE006065	Pasteu
*	*Pseudomonas aeruginosa*	EU515133	Pseudo1
*	*Chromatium okenii*	AJ223234	Chromo

Table 1. List of all sequences included into *Basic matrix*. Lighter tab field contains *Arsenophonus* records obtained from GenBank. Sequences of other symbionts and free-living bacteria, marked with asterisk, are placed into darker part of tab. Sequences included into the *Clock matrix* are underlined.

Results

Sequences and alignments

From 15 insect taxa, we obtained 34 sequences of 16S rDNA that exhibited high degree of similarity to bacterial genus *Arsenophonus* when identified by the BLAST program. The length of the PCR-amplified fragments varied from 632 to 1198 bp, with the guanine-cytosine (GC) content ranging from 46,22 to 54,84 % (Fig. 2, bars). For three specimens of the hippoboscid *Ornithomya avicularia*, two different sequences were obtained from each single individual. After completing with all *Arsenophonus* 16S rDNA sequences currently available in the GenBank, and several additional free-living and symbiotic bacteria, the data-set produced a 1222 bp long *Basic matrix*. This set exhibits several noticeable patterns. First, the alignment has a mosaic structure, discussed below. Second, it contains a large group of sequences with a high degree of similarity (divergence within the range of 0,1-7,3%), identified from several distant host taxa. Molecular characteristics of these sequences (i.e. GC content, sequence length, etc.) are similar to those found in various free-living enterobacteria. Apart from this homogeneous group, the set also includes several sequences with notable modifications typical for many proteobacterial symbionts, particularly the presence of long insertions within the variable regions and the decrease of GC content. Sequence distances among these taxa reach up to 17,8%.

Phylogeny

All phylogenetic analyses of the *Basic matrix* yielded monophyletic *Arsenophonus* clade (Fig. 2). The new 34 sequences (Fig. 2, arrows), identified by BLAST program as putative members or relatives of the genus *Arsenophonus*, clustered invariantly within the *Arsenophonus* clade. Their precise position was only partially determined by the host taxa. Some of the *Arsenophonus* sequences from hippoboscoid hosts clustered within monophyletic host-specific groups (Fig. 2, printed in red) while others were scattered across the tree as isolated lineages (Fig. 2, printed in dark orange). The two different sequences determined from each specimen of *O. avicularia* clustered at distant positions within the tree (Fig. 2, numbers with asterisks).

Considerable loss of phylogenetic information was observed in the *Conservative matrix*. In this case, the relationships among individual *Arsenophonus* lineages were highly

unstable, resulting in large polytomies of many short-branched taxa within the consensus trees (Fig. 3). Also the relationships among the long-branched lineages, although remained resolved, sharply differ from those derived from the *Basic matrix* data, and the genus *Proteus* was not recognized as the closest relative of *Arsenophonus*. Such picture indicates that the information contained in *Conservative matrix* (restricted to one third *Basic* dataset, i.e., 284bp) is insufficient for reliable phylogenetic placing of closely related taxa.

The analyses of taxonomically restricted matrices confirmed the expected dependence of the phylogenetic inference on the altering samplings (examples of topologies obtained are provided on Fig. 4, Fig. 7, and Fig. 8). The highest degree of susceptibility was observed with MP, particularly under Tv:Ts ratio set to 1. The most fundamental distortion occurred with the matrix *Sampling3*, where one lineage (composed of *Buchnera*, *Wigglesworthia*, *Blochmania*, and P-symbiont from *Trioza magnoliae*) clustered either as a sister group of *Riesia* clade or together with *Sodalis*. Thus, the relationships among individual lineages within the consensus tree produced a polytomy interfering with the monophyly of *Arsenophonus* clade (Fig.4).

The calculation of divergence time yielded substantially different results in dependence on the calibration dates. Using of the *Riesia* diversification as reference point suggested a recent origin of triatomine-associated *Arsenophonus* branch; median value of the estimate distribution was 2.6 mya. In contrast, the calibration by *Escherichia*-*Salmonella* returned considerably higher dates with median at 24.5 mya.

Discussion

Phylogenetic patterns and the stability of the information

Phylogenetic relationships of 110 *Arsenophonus* symbionts, including 34 new samples, revealed a remarkably complex arrangement with individual lineages displaying a broad spectrum of phylogenetic properties. The most typical are short-branched sequences with low degree of divergence and unstable positions within the *Arsenophonus* clade (Fig. 2, printed in dark orange). The opposite extreme is represented by well supported host-specific clusters of long branches, such as the louse symbiont *Riesia* or the symbionts described from several streblid species. An intermediate position is occupied by several putative host-specific but less robust clusters, such as the *Arsenophonus* lineages from triatomine bugs, some

hippoboscoids or homopterans (Fig. 2). In an analogy to previously analyzed symbiotic bacteria (Lefevre et al., 2004), the phylogenetic properties of the sequences were also reflected in their GC contents. In the short-branched taxa, the GC content varies from 51,72 to 54,84% , the values typical for S-symbiots and free-living bacteria (Heddi et al., 1998). In contrast, the sequences with low GC content, varying between 46,22 and 51,93%, represented the long-branched taxa clustering within the host-specific monophyletic lineages (e.g. the symbionts from *Ornithomyia*, *Lipoptena*, *Trichobius*, and the *Riesia* clade). Moreover, a comparison of the branch ordering within each of these subclusters to the host phylogeny indicates a cospeciation process within several lineages (discussed below). From the phylogenetic perspective, there is no strict split separating the set of *Arsenophonus* sequences into the ecologically distinct types. The position of the long-branched subclusters within the topology is not stable. Under the MP criterion with transition rate 1:3 and under the ML criterion they form a unique monophyletic cluster (Fig. 5a), while in other analyses the individual host-specific subclusters were scattered among the short-branched lineages (Fig. 5b).

The low resolution and instability of the trees inferred from the *Conservative matrix* suggests that substantial part of the phylogenetic information is located within the ”ambiguously” aligned regions that were removed by the GBlocks procedure. This fact is particularly important when considering the frequent occurrence of insertions/deletions within the sequences (Fig.6). This may lead to deletion of these critical fragments in many phylogenetic analyses. Interestingly, the monophyletic nature of *Arsenophonus* was preserved even in this highly *Conservative matrix*. This indicates that within the complete data set, the phylogenetic information underlying the *Arsenophonus* monophyly is sufficiently strong and is contained in the conservative regions of the sequences. In accordance with this presumption, several molecular synapomorphies can be identified in the *Basic* and *Conservative* matrices. The most pronounced is the motif GTC/GTT located in position 481-483 of *Basic matrix*, 159-161 respectively.

Relevance of the sampling

To test an effect of sampling on the phylogenetic inference within *Arsenophonus*, we examined five *Sampling* matrices with different taxa compositions (see MM). In addition to the MP, ML, and Bayesian analyses, we performed ML calculation under the nonhomogeneous model of the substitutions, designated as T92 (Tamura, 1992; Galtier and

Gouy, 1998). This model was previously used to test the monophyly/polyphyly of the P-symbiotic lineages and brought the first serious evidence for a possible independent origin of major P-symbiotic taxa (Herbeck et al., 2005). We were not able to apply the same approach to the *Basic* and *Conservative* matrices since the program Phylowin failed to process these large datasets under the ML criterion. The analyses of several taxonomically restricted matrices proved the sensitivity of phylogenetic signal to the sampling. In the most extreme case, shown in Fig. 4a, even the monophyly of the *Arsenophonus* clade was disrupted by other lineages of symbiotic bacteria. Considering the results of the extensive analysis of the *Basic matrix*, such arrangement is beyond doubt a methodological artifact. Since both *Riesia* and the P-symbiont lineage are long-branched taxa with rapid evolution of 16S rDNA, their affinity is very likely caused by Long Branch Attraction (LBA; for review see Bergsten, 2005) within the taxonomically compromised matrix. It is symptomatic that this topology was inferred by MP, the method known to be particularly prone to the LBA. To further test this distortion, one of the long-branched taxa was removed from the data set (matrix *Sampling4*). This approach restored the *Arsenophonus* monophyly and confirmed the effect of LBA phenomenon (Fig. 6).

The aim of these taxonomically restricted analyses was to “simulate” phylogenetic placement of newly determined symbionts. In such casual studies, the symbiotic lineages are rarely represented by all available sequences in the way we composed the *Basic matrix*. Rather, each symbiotic lineage is represented by few randomly selected sequences. Under such circumstances, incorrect topologies (e.g. the *Sampling5*-derived topology on the Fig. 8) can be obtained due to various methodological artifacts. This situation can be illustrated by empirical data: at least in two studies, louse-associated lineage of *Arsenophonus* was not recognized as member of the *Arsenophonus* clade (Sasaki-Fukatsu et al., 2006; Hypša and Křížek, 2007). Consequently, when more recent studies, based on better sampling, proved the position of *Riesia* within the *Arsenophonus* cluster (Allen et al., 2007; Perotti et al., 2007) the genus *Arsenophonus* became paraphyletic.

Interestingly, topologies inferred by likelihood analyses using the T92 evolution model (Galtier and Gouy, 1998) were influenced neither by the compromised sampling nor by the removal of unreliably aligned regions.

Cophylogeny vs. horizontal transfers: possible sources of phylogenetic incongruence

The phylogenetic tree of all *Arsenophonus* sequences exhibits both pattterns, the parallel evolution of symbionts and their hosts and the haphazard association of symbionts from different host taxa. Coincidentally, both arrangements can be demostrated on the newly sequenced symbionts from various hippoboscoid species. Some of hippoboscoid-associated *Arsenophonus* show possible host specificity; in few analyses they cluster within several monophyletic short-branched groups. Since relationships among the short-branched taxa are generaly not well resolved, these lineages, are scattered throughout the whole topology (Fig. 2). In contrast, relationships within the long-branched clusters of hippoboscoid-associated taxa are in agreement with the host phylogeny (the *Arsenophonus* clusters strictly reflecting the host phylogeny are designated by solid circle in the Fig. 2. Interestingly, a coevolutionary pattern was also identified for streblids of the genus *Trichobius* and their symbionts. In the original study published by Trowbridge et al. (2006), the distribution of *Trichobius* symbionts was apparently not consistent with the host phylogeny. Our analysis in a broad context indicates that this discrepancy might have been caused by different bacterial sampling and particularly by aberrant behavior of the sequence from *Trichobius yunkeri* (DQ314776). This sequence is likely to be a chimerical product of at least two distant lineages; according to our BLAST tests it shares 100% identity with S-symbiont of *Psylla pyricola* (AF286125) along a 1119 bp long region. Removal of this sequence from the dataset restored a complete phylogenetic congruence between *Trichobius* (phylogeny of the genus *Trichobius* was published by Dittmar et al., (2006) and its symbionts. This finding exemplifies the factual danger of chimerical sequences in the research of symbiotic bacteria. The presence of several symbiotic lineages within a single host is well known (Sandstrom et al., 2001; Thao et al., 2002; Zchori-Fein and Brown, 2002; Takiya et al., 2006). In this study, we demonstrate such possible case in *O. avicularia*. From three individuals of this species we obtained pairs of different sequences branching at distant positions within the tree. Consistent clustering of these sequences shows that they are well differentiated and do not produce chimerical forms in our data.

While the indentity between symbiont relationships and the host phylogeny is apparently a consequence of host-symbiont cophylogeny, the interpretation of the randomly scattered symbionts is less self-evident. Usually, such arrangement is explained as result of transient infections and frequent horizontal transfers among distant host taxa. This is typical, for example, of the *Wolbachia* symbionts in wide range of insect species (Werren, 1997).

Generally, the capability to undergo inter-host transfers is assumed for several symbiotic lineages and has even been demonstrated under experimental conditions (Heath et al., 1999; Russell and Moran, 2005). Since the *Arsenophonus* cluster contains bacteria from phylogenetically distant insect taxa and also bacteria isolated from plants, it is clear that horizontal transfers and/or multiple establishments of the symbiosis have to be assumed. However, several facts indicate that at least part of the incongruences could be caused by some type of methodological artifacts. The most conspicuous feature of the *Arsenophonus* topology is occurrance of monophyletic lineages bound to monophyletic groups of insect host without a co-speciation pattern. Although our study cannot present an exhaustive explanation of such picture, we want to point out two factors that might in theory take part in shaping the relationships among *Arsenophonus* sequences, lateral gene transfer (LGT) and intragenomic heterogeneity. Both of these phnomena have previously been determined as the causes of phylogenetic distortions and should be seriously considered in coevolutionary studies at a low phylogenetic level.

LGT and putative role of bacteriophages

An apparently "mosaic" structure of the *Arsenophonus* alignment raises the question whether various regions of this sequence could have undergone different evolutionary histories. Such properties of 16S rDNA were previously determined and demonstrated not to interfere with rRNA functions (Mylvaganam and Dennis, 1992; Wang et al., 1997). The occurrence of short segments with high number of non-random variations within ribosomal genes was reported for actinomycete species (Wang and Zhang, 2000). The authors estimated the lateral transfer of rRNA gene segments to be the most parsimonious explanation (Wang and Zhang, 2000). Later, Gogarten et al. (2002) suggested that, analogically to an entire bacterial genome, 16S rDNA possesses a mosaic character originated by LGT, respectively by transfer of gene subunits.

In our study, LGT could provide an explanation for some ambivalent patterns observed along the *Arsenophonus* sequences. Similarly to the preceding study of Wang and Zhang (2000), we determined an 41 bp long region with distribution inconsistent with the overall phylogenetic picture (Fig. 9). Such homoplasious states can arise even as molecular convergences or as result of LGT. We are not aware of any known process leading to such convergent evolution within the 16S rDNA. On the other hand, bacteriophages have been previously suggested to take part in LGT as possible vehicles of the DNA fragments

(Canchaya et al., 2003). Interestingly, phage-like particles have been reported from the *Arsenophonus* bacteria associated with triatomine bugs (Hypša, 1993). In the most recent studies several genes of the *Arsenophonus*-associated phage were shown to share a high sequence similarity with *Acyrtosiphon pisum* bacteriophage APSE (Hansen et al., 2007, our unpublished data), an active lysogenic phage infecting symbiotic bacteria *Hamiltonella defensa* (van der Wilk et al., 1999; Moran et al., 2005a; Degnan and Moran, 2008). Furthermore, considering the close relationship of *Arsenophonus* symbionts and bacterial genus *Proteus*, *Arsenophonus* most probably fall into the host spectrum of generalized transducing phage designated as SN-T (Winston and Thompson, 1979). This phage was proved to infect unrelated bacterial species across Enterobacteriaceae, including *Proteus vulgaris*, *Pseudomonas aeroginosa*, and *Shigella flexneri* from gamma subdivision (Jensen et al., 1998). Recently, the ability of SN-T to acquire 16S rRNA gene sequences from *P.aeroginosa* and *S. natans* was confirmed experimentally and the frequency of the acquisition was determined, implicating possible horizontal transfer of 16S rRNA genes among distinct bacterial species (Beumer and Robinson, 2005). Since the complete genome sequence is not yet available for *Arsenophonus* bacteria, no relevant molecular information is available on the associated phage. However, the properties of both supposedly related bacteriophages described above indicate that LGT might play a significant role in the evolution of small ribosomal subunit as well as the complete *Arsenophonus* genome.

Incongruence and intragenomic heterogeneity

As bacterial genomes often carry more than one rRNA operon, intragenomic heterogeneity of the rDNA copies is occasionally found to blur the phylogenetic picture (Pettersson et al., 1998; Yap et al., 1999; Marchandin et al., 2003; Lin et al., 2004). Although there is no direct information on the number of copies in *Arsenophonus* genome, Stewart and Cavanaugh (2007) showed bacterial genomes to encode in average 5 rRNA operons. The most closely related bacterium for which the complete genome is known, *Photorhabdus luminescens*, carries seven copies (NC_005126). *Arsenophonus*-focused studies indicate that at least two different forms of the rRNA operon are present in its genome (Thao and Baumann, 2004a; Sorfova et al., 2008). Furthermore, Sorfova et al. (2008) suggest that the variability among individual copies may cause the incongruence observed between triatomines and their *Arsenophonus* lineages. They point out that this process could, in principle, explain an otherwise problematic observation: in some hosts, such as triatomines or

some homopterans, the hosts and the *Arsenophonus* bacteria create reciprocally monophyletic clusters but do not show any cospeciation pattern. With the present knowledge on *Arsenophonus* genome, it is difficult to assess whether and how deeply can the rRNA heterogeneity affect the phylogenetic reconstruction. Sorfova et al. (2008) tried to use the estimation of divergence times as a guide for deciding between alternative coevolutionary scenarios. They used the *Escherichia-Salmonella* divergence (Ochman and Wilson, 1987; Battistuzzi et al., 2004) as a calibration point for calculating the divergence time among various *Arsenophonus* lineages from triatomine bugs. Applying the Multdiv method (Rutschmann, 2005), they placed the ancestor of triatomine-associated symbionts into a broad range of approx. 15 - 40 mya and concluded that this estimate is compatible or even exceeds the age estimates available for the tribe triatomine (according to Gaunt and Miles, 2002). Here, we took advantage of a new age-estimate for closely related bacteria, namely the louse-associated symbionts of the genus *Riesia* (Allen et al., 2007). Comparing the estimates based on the two calibration methods (*Escherichia-Salmonella* and *Riesia*), we found that due to the variability of evolutionary rates among the lineages, the results may differ in order of magnitude. Such marked variance among different bacterial lineages (including different symbiotic bacteria from the same host species) were previously reported for many bacterial groups (Werren, 1997; Heddi et al., 1998; Clark et al., 2000; Moran and Wernegreen, 2000; Dale et al., 2002; Degnan et al., 2004; Lefevre et al., 2004; Moran et al., 2005b; Takiya et al., 2006; Duron and Gavotte, 2007). We therefore conclude that this method cannot be directly used to assess the effect of intragenomic heterogeneity within the presented *Arsenophonus* tree.

Conclusion and taxonomic implications

With more than one hundred records, the genus *Arsenophonus* now belongs to the richest clusters of insect symbiotic bacteria. Considering its broad host spectrum and apparent ecological versatility, *Arsenophonus* may play an important role in the research of evolutionary trends in insect intracellular symbionts. The analysis of 110 available sequences of *Arsenophonus* 16S rDNA from 54 host taxa revealed several interesting evolutionary patterns, but it also demonstrated that a complete understanding of the *Arsenophonus* phylogeny can hardly be achieved with this genetic marker. A similar situation is, for example, found in another large symbiotic group, the genus *Wolbachia*, where other genes are often used as alternative source of phylogenetic information (Casiraghi et al., 2005; Baldo and

Werren, 2007). We also want to point out that the genus *Arsenophonus* is currently polyphyletic due to the two lineages described as separate genera *Riesia* and *Phlomobacter*. Two procedures can, in principle, solve this undesirable situation, splitting of the *Arsenophonus* cluster into several separate genera or classification of all its members within the genus *Arsenophonus*. Taking into account the phylogenetic arrangement of the individual lineages, the first approach would inevitably lead to establishment of many genera with low sequence divergences and very similar biology. We therefore consider the latter approach, (i.e. reclassification of all members of the *Arsenophonus* clade within a single genus) a more appropriate solution of the current situation.

Acknowledgements

This work was supported by the Grant Agency of Academy of Sciences of the Czech Republic (Grant IAA601410708) and the laboratory of Dr. N.A. Moran (University of Arizona). We thank all of our collaborators for providing the samples.

References

Allen, J.M., Reed, D.L., Perotti, M.A., Braig, H.R., 2007. Evolutionary relationships of "Candidatus *Riesia* spp.," endosymbiotic Enterobacteriaceae living within hematophagous primate lice. Applied and Environmental Microbiology 73, 1659-1664.

Babendreier, D., Joller, D., Romeis, J., Bigler, F., Widmer, F., 2007. Bacterial community structures in honeybee intestines and their response to two insecticidal proteins. Fems Microbiology Ecology 59, 600-610.

Baldo, L., Werren, J.H., 2007. Revisiting *Wolbachia* supergroup typing based on WSP: Spurious lineages and discordance with MLST. Current Microbiology 55, 81-87.

Battistuzzi, F.U., Feijao, A., Hedges, S.B., 2004. A genomic timescale of prokaryote evolution: insights into the origin of methanogenesis, phototrophy, and the colonization of land. Bmc Evolutionary Biology 4, 14.

Bergsten, J., 2005. A review of long-branch attraction. Cladistics 21, 163-193.

Beumer, A., Robinson, J.B., 2005. A broad-host-range, generalized transducing phage (SN-T) acquires 16S rRNA genes from different genera of bacteria. Applied and Environmental Microbiology 71, 8301-8304.

Canchaya, C., Fournous, G., Chibani-Chennoufi, S., Dillmann, M.L., Brussow, H., 2003. Phage as agents of lateral gene transfer. Current Opinion in Microbiology 6, 417-424.

Casiraghi, M., Bordenstein, S.R., Baldo, L., Lo, N., Beninati, T., Wernegreen, J.J., Werren, J.H., Bandi, C., 2005. Phylogeny of *Wolbachia* pipientis based on gltA, groEL and ftsZ gene sequences: clustering of arthropod and nematode symbionts in the F supergroup, and evidence for further diversity in the *Wolbachia* tree. Microbiology-Sgm 151, 4015-4022.

Castresana, J., 2000. Selection of conserved blocks from multiple alignments for their use in phylogenetic analysis. Molecular Biology and Evolution 17, 540-552.

Clark, M.A., Moran, N.A., Baumann, P., Wernegreen, J.J., 2000. Cospeciation between bacterial endosymbionts (*Buchnera*) and a recent radiation of aphids (*Uroleucon*) and pitfalls of testing for phylogenetic congruence. Evolution 54, 517-525.

Dale, C., Plague, G.R., Wang, B., Ochman, H., Moran, N.A., 2002. Type III secretion systems and the evolution of mutualistic endosymbiosis. Proceedings of the National Academy of Sciences of the United States of America 99, 12397-12402.

Dale, C., Beeton, M., Harbison, C., Jones, T., Pontes, M., 2006. Isolation, pure culture, and characterization of "Candidatus *Arsenophonus arthropodicus*," an intracellular secondary endosymbiont from the hippoboscid louse fly *Pseudolynchia canariensis*. Applied and Environmental Microbiology 72, 2997-3004.

Degnan, P.H., Lazarus, A.B., Brock, C.D., Wernegreen, J.J., 2004. Host-symbiont stability and fast evolutionary rates in an ant-bacterium association: Cospeciation of *Camponotus* species and their endosymbionts, Candidatus *Blochmannia.* Systematic Biology 53, 95-110.

Degnan, P.H., Moran, N.A., 2008. Evolutionary genetics of a defensive facultative symbiont of insects: exchange of toxin-encoding bacteriophage. Molecular Ecology 17, 916-929.

Dittmar, K., Porter, M.L., Murray, S., Whiting, M.F., 2006. Molecular phylogenetic analysis of nycteribiid and streblid bat flies (Diptera : Brachycera, Calyptratae): Implications for host associations and phylogeographic origins. Molecular Phylogenetics and Evolution 38, 155-170.

Drummond, A.J., Nicholls, G.K., Rodrigo, A.G., Solomon, W., 2002. Estimating mutation parameters, population history and genealogy simultaneously from temporally spaced sequence data. Genetics 161, 1307-1320.

Dunn, A.K., Stabb, E.V., 2005. Culture-independent characterization of the microbiota of the ant lion *Myrmeleon mobilis* (Neuroptera : Myrmeleontidae). Applied and Environmental Microbiology 71, 8784-8794.

Duron, O., Gavotte, L., 2007. Absence of *Wolbachia* in nonfilariid worms parasitizing arthropods. Current Microbiology 55, 193-197.

Fukatsu, T., Koga, R., Smith, W.A., Tanaka, K., Nikoh, N., Sasaki-Fukatsu, K., Yoshizawa, K., Dale, C., Clayton, D.H., 2007. Bacterial Endosymbiont of the Slender Pigeon Louse, *Columbicola columbae*, Allied to Endosymbionts of Grain Weevils and Tsetse Flies. Appl. Environ. Microbiol. 73, 6660-6668.

Galtier, N., Gouy, M., Gautier, C., 1996. SEAVIEW and PHYLO_WIN: Two graphic tools for sequence alignment and molecular phylogeny. Computer Applications in the Biosciences 12, 543-548.

Galtier, N., Gouy, M., 1998. Inferring pattern and process: Maximum-likelihood implementation of a nonhomogeneous model of DNA sequence evolution for phylogenetic analysis. Molecular Biology and Evolution 15, 871-879.

Gaunt, M.W., Miles, M.A., 2002. An insect molecular clock dates the origin of the insects and accords with palaeontological and biogeographic landmarks. Molecular Biology and Evolution 19, 748-761.

Gherna, R.L., Werren, J.H., Weisburg, W., Cote, R., Woese, C.R., Mandelco, L., Brenner, D.J., 1991. *Arsenophonus nasoniae* gen. nov., sp. nov., the causative agent of the son-killer trait in the parasitic wasp *Nasonia vitripennis*. International Journal of Systematic Bacteriology 41, 563-565.

Gogarten, J.P., Doolittle, W.F., Lawrence, J.G., 2002. Prokaryotic evolution in light of gene transfer. Molecular Biology and Evolution 19, 2226-2238.

Goloboff, P.A., Farris, J.S., Nixon, K.C., 2004. TNT. Cladistics-the International Journal of the Willi Hennig Society 20, 84-84.

Grindle, N., Tyner, J.J., Clay, K., Fuqua, C., 2003. Identification of *Arsenophonus*-type bacteria from the dog tick *Dermacentor variabilis*. Journal of Invertebrate Pathology 83, 264-266.

Guindon, S., Gascuel, O., 2003. A simple, fast, and accurate algorithm to estimate large phylogenies by maximum likelihood. Systematic Biology 52, 696-704.

Hall, A.T., 1999. BioEdit: a user-friendly biological sequence alignment editor and analysis program for Windows 95/98/NT. . Nuc Acid Symp Series 41 95-98.

Hansen, A.K., Jeong, G., Paine, T.D., Stouthamer, R., 2007. Frequency of secondary symbiont infection in an invasive psyllid relates to parasitism pressure on a geographic scale in California. Applied and Environmental Microbiology 73, 7531-7535.

Heath, B.D., Butcher, R.D.J., Whitfield, W.G.F., Hubbard, S.F., 1999. Horizontal transfer of *Wolbachia* between phylogenetically distant insect species by a naturally occurring mechanism. Current Biology 9, 313-316.

Heddi, A., Charles, H., Khatchadourian, C., Bonnot, G., Nardon, P., 1998. Molecular characterization of the principal symbiotic bacteria of the weevil *Sitophilus oryzae*: A peculiar G+C content of an endocytobiotic DNA. Journal of Molecular Evolution 47, 52-61.

Herbeck, J.T., Degnan, P.H., Wernegreen, J.J., 2005. Nonhomogeneous model of sequence evolution indicates independent origins of primary endosymbionts within the enterobacteriales (gamma-proteobacteria). Molecular Biology and Evolution 22, 520-532.

Huger, A.M., Skinner, S.W., Werren, J.H., 1985. Bactrerial infections associated with the son-killer trait in the parasitoid wasp, *Nasonia* (=*Mormoniella*) *vitripennis* (Hymenoptera, Pteromalidae). Journal of Invertebrate Pathology 46, 272-280.

Hypša, V., 1993. Endocytobionts of *Triatoma infestans*: distribution and transmission. Journal of Invertebrate Pathology 61, 32-38.

Hypša, V., Dale, C., 1997. In vitro culture and phylogenetic analysis of "Candidatus *Arsenophonus triatominarum*," an intracellular bacterium from the triatomine bug, *Triatoma infestans*. International Journal of Systematic Bacteriology 47, 1140-1144.

Hypša, V., Křížek, J., 2007. Molecular Evidence for Polyphyletic Origin of the Primary Symbionts of Sucking Lice (Phthiraptera, Anoplura). Microbial Ecology.

Jensen, E.C., Schrader, H.S., Rieland, B., Thompson, T.L., Lee, K.W., Nickerson, K.W., Kokjohn, T.A., 1998. Prevalence of broad-host-range lytic bacteriophages of *Sphaerotilus natans*, *Escherichia coli*, and *Pseudomonas aeruginosa*. Applied and Environmental Microbiology 64, 575-580.

Koga, R., Tsuchida, T., Fukatsu, T., 2003. Changing partners in an obligate symbiosis: a facultative endosymbiont can compensate for loss of the essential endosymbiont *Buchnera* in an aphid. Proceedings of the Royal Society of London Series B-Biological Sciences 270, 2543-2550.

Lefevre, C., Charles, H., Vallier, A., Delobel, B., Farrell, B., Heddi, A., 2004. Endosymbiont phylogenesis in the Dryophthoridae weevils: Evidence for bacterial replacement. Molecular Biology and Evolution 21, 965-973.

Lin, C.K., Hung, C.L., Chiang, Y.C., Lin, C.M., Tsen, H.Y., 2004. The sequence heterogenicities among 16S rRNA genes of *Salmonella* serovars and the effects on the specificity of the primers designed. International Journal of Food Microbiology 96, 205-214.

Louis, C., Drif, L., Vago, C., 1986. Mise en évidence et étude ultrastructurale de procaryotes de type rickettsien dans les glandes salivaires des Triatomidae (Heteroptera) =

Evidence and ultrastructural study of *Rickettsia*-like prokaryotes in salivary glands of Triatomidae (Heteroptera) Annales de la Société entomologique de France 22, 153-162.

Marchandin, H., Teyssier, C., de Buochberg, M.S., Jean-Pierre, H., Carriere, C., Jumas-Bilak, E., 2003. Intra-chromosomal heterogeneity between the four 16S rRNA gene copies in the genus *Veillonella*: implications for phylogeny and taxonomy. Microbiology-Sgm 149, 1493-1501.

Moran, N.A., 1996. Accelerated evolution and Muller's rachet in endosymbiotic bacteria. Proceedings of the National Academy of Sciences of the United States of America 93, 2873-2878.

Moran, N.A., Wernegreen, J.J., 2000. Lifestyle evolution in symbiotic bacteria: insights from genomics. Trends in Ecology & Evolution 15, 321-326.

Moran, N.A., Degnan, P.H., Santos, S.R., Dunbar, H.E., Ochman, H., 2005a. The players in a mutualistic symbiosis: Insects, bacteria, viruses, and virulence genes. Proceedings of the National Academy of Sciences of the United States of America 102, 16919-16926.

Moran, N.A., Tran, P., Gerardo, N.M., 2005b. Symbiosis and insect diversification: An ancient symbiont of sap-feeding insects from the bacterial phylum *Bacteroidetes*. Applied and Environmental Microbiology 71, 8802-8810.

Mylvaganam, S., Dennis, P.P., 1992. Sequence heterogeneity between the 2 genes encoding 16S ribosomal-RNA from the halophilic archeabacterium *Haloarcula marismortui*. Genetics 130, 399-410.

Ochman, H., Wilson, A.C., 1987. Evolution in bacteria: Evidence for a universal substitution rate in cellular genomes Journal of Molecular Evolution 26, 74-86.

Perotti, M.A., Allen, J.M., Reed, D.L., Braig, H.R., 2007. Host-symbiont interactions of the primary endosymbiont of human head and body lice. Faseb Journal 21, 1058-1066.

Pettersson, B., Bolske, G., Thiaucourt, F., Uhlen, M., Johansson, K.E., 1998. Molecular evolution of *Mycoplasma capricolum* subsp. capripneumoniae strains, based on polymorphisms in the 16S rRNA genes. Journal of Bacteriology 180, 2350-2358.

Posada, D., Crandall, K.A., 1998. MODELTEST: testing the model of DNA substitution. Bioinformatics 14, 817-818.

Russell, J.A., Latorre, A., Sabater-Munoz, B., Moya, A., Moran, N.A., 2003. Side-stepping secondary symbionts: widespread horizontal transfer across and beyond the Aphidoidea. Molecular Ecology 12, 1061-1075.

Russell, J.A., Moran, N.A., 2005. Horizontal transfer of bacterial symbionts: Heritability and fitness effects in a novel aphid host. Applied and Environmental Microbiology 71, 7987-7994.

Rutschmann, F., 2005. Bayesian molecular dating using PAML/multidivtime. A step-by-step manual. University of Zurich, Switzerland. Available at http://www.plant.ch.

Sandstrom, J.P., Russell, J.A., White, J.P., Moran, N.A., 2001. Independent origins and horizontal transfer of bacterial symbionts of aphids. Molecular Ecology 10, 217-228.

Sasaki-Fukatsu, K., Koga, R., Nikoh, N., Yoshizawa, K., Kasai, S., Mihara, M., Kobayashi, M., Tomita, T., Fukatsu, T., 2006. Symbiotic bacteria associated with stomach discs of human lice. Applied and Environmental Microbiology 72, 7349-7352.

Semetey, O., Gatineau, F., Bressan, A., Boudon-Padieu, E., 2007. Characterization of a gamma-3 proteobacteria responsible for the syndrome "basses richesses" of sugar beet transmitted by *Pentastiridius* sp. (Hemiptera, Cixiidae). Phytopathology 97, 72-78.

Skinner, S.W., 1985. Son-killer – A 3rd extrachromosomal factor affecting the sex-ratio in the parasitoid wasp, *Nasonia* (=*Mormoniella*) *vitripennis*. Genetics 109, 745-759.

Sorfova, P., Skerikova, A., Hypsa, V., 2008. An effect of 16S rRNA intercistronic variability on coevolutionary analysis in symbiotic bacteria: molecular phylogeny of *Arsenophonus triatominarum*. Systematic and Applied Microbiology. In Press.

Spaulding, A.W., von Dohlen, C.D., 2001. Psyllid endosymbionts exhibit patterns of co-speciation with hosts and destabilizing substitutions in ribosomal RNA. Insect Molecular Biology 10, 57-67.

Stewart, F.J., Cavanaugh, C.M., 2007. Intragenomic variation and evolution of the internal transcribed spacer of the rRNA operon in bacteria. Journal of Molecular Evolution 65, 44-67.

Subandiyah, S., Nikoh, N., Tsuyumu, S., Somowiyarjo, S., Fukatsu, T., 2000. Complex endosymbiotic microbiota of the citrus psyllid *Diaphorina citri* (Homoptera : Psylloidea). Zoological Science 17, 983-989.

Takiya, D.M., Tran, P.L., Dietrich, C.H., Moran, N.A., 2006. Co-cladogenesis spanning three phyla: leafhoppers (Insecta : Hemiptera : Cicadellidae) and their dual bacterial symbionts. Molecular Ecology 15, 4175-4191.

Tamura, K., 1992. Estimation of the number of nucleotide substitutions when there are strong transition-transversion and G+C-content biases. Molecular Biology and Evolution 9, 678-687.

Thao, M.L., Clark, M.A., Baumann, L., Brennan, E.B., Moran, N.A., Baumann, P., 2000. Secondary endosymbionts of psyllids have been acquired multiple times. Current Microbiology 41, 300-304.

Thao, M.L., Gullan, P.J., Baumann, P., 2002. Secondary (gamma-Proteobacteria) endosymbionts infect the primary (beta-Proteobacteria) endosymbionts of mealybugs multiple times and coevolve with their hosts. Applied and Environmental Microbiology 68, 3190-3197.

Thao, M.L., Baumann, P., 2004a. Evolutionary relationships of primary prokaryotic endosymbionts of whiteflies and their hosts. Applied and Environmental Microbiology 70, 3401-3406.

Thao, M.L.L., Baumann, P., 2004b. Evidence for multiple acquisition of *Arsenophonus* by whitefly species (Sternorrhyncha : Aleyrodidae). Current Microbiology 48, 140-144.

Trowbridge, R.E., Dittmar, K., Whiting, M.F., 2006. Identification and phylogenetic analysis of *Arsenophonus*- and *Photorhabdus*-type bacteria from adult Hippoboscidae and Streblidae (Hippoboscoidea). Journal of Invertebrate Pathology 91, 64-68.

van der Wilk, F., Dullemans, A.M., Verbeek, M., van den Heuvel, J., 1999. Isolation and characterization of APSE-1, a bacteriophage infecting the secondary endosymbiont of *Acyrthosiphon pisum*. Virology 262, 104-113.

Wang, Y., Zwang, Z.S., Ramanan, N., 1997. The actinomycete *Thermobispora bispora* contains two distinct types of transcriptionally active 16S rRNA genes. Journal of Bacteriology 179, 3270-3276.

Wang, Y., Zhang, Z.S., 2000. Comparative sequence analyses reveal frequent occurrence of short segments containing an abnormally high number of non-random base variations in bacterial rRNA genes. Microbiology-Sgm 146, 2845-2854.

Werren, J.H., Skinner, S.W., Huger, A.M., 1986. Male-killing bacteria in a parasitic wasp. Science 231, 990-992.

Werren, J.H., 1997. Biology of *Wolbachia*. Annual Review of Entomology 42, 587-609.

Winston, V., Thompson, T.L., 1979. Isolation and characterization of a bacteriophage specific for *Sphaerotilus natans* which contains an unusual base its deoxyribonucleic acid. Applied and Environmental Microbiology 37, 1025-1030.

Yap, W.H., Zhang, Z.S., Wang, Y., 1999. Distinct types of rRNA operons exist in the genome of the actinomycete *Thermomonospora chromogena* and evidence for horizontal transfer of an entire rRNA operon. Journal of Bacteriology 181, 5201-5209.

Zchori-Fein, E., Brown, J.K., 2002. Diversity of prokaryotes associated with *Bemisia tabaci* (Gennadius) (Hemiptera : Aleyrodidae). Annals of the Entomological Society of America 95, 711-718.

Zreik, L., Bove, J.M., Garnier, M., 1998. Phylogenetic characterization of the bacterium-like organism associated with marginal chlorosis of strawberry and proposition of a Candidatus taxon for the organism, 'Candidatus *Phlomobacter fragariae*'. International Journal of Systematic Bacteriology 48, 257-261.

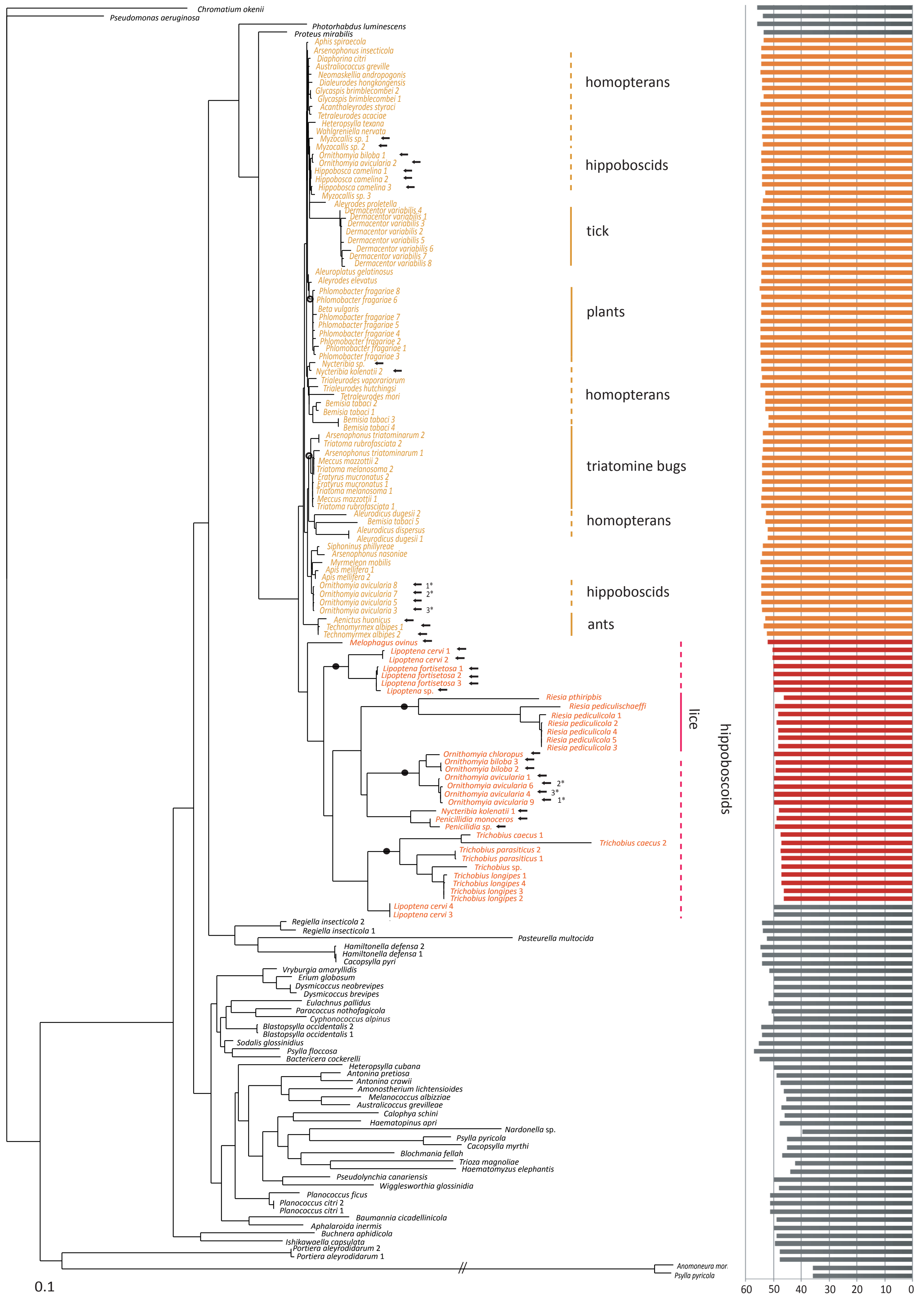

Figure 2. Phylogenetic tree derived from the *Basic matrix* under ML criterion. Names of the taxa clustering within the *Arsenophonus* clade are printed in colour: red for the long-branched taxa, dark orange for the short-branched taxa. The arrows point to the new sequences obtained in our study. Different types of sequences determined from the specimens of *O. avicularia* are designated by the numbers with asterisks. Solid circles on branches label the clusters strictly concordant with the host phylogenies. Open circles designate host-specific lineages without coevolutionary signal. Solid vertical lines indicate the reciprocally monophyletic groups of symbionts and hosts; dashed lines show paraphyletic symbiont clades restricted to monophyletic host groups.Bars represent GC content of each taxa.

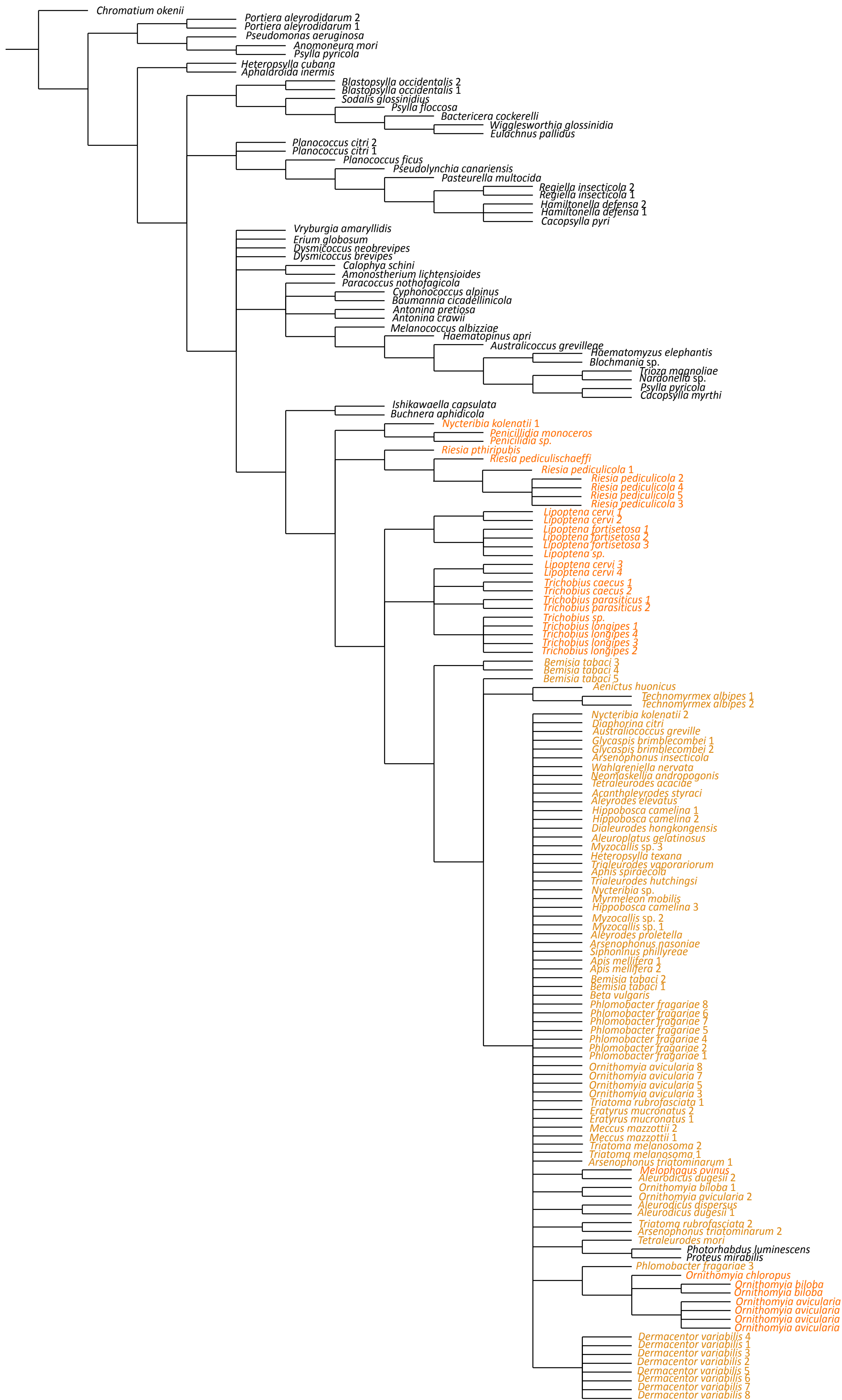

Figure 3. Consensus tree derived from the *Conservative matrix* under MP criterion with the transversion/transition ratio set to 1:3. Names of the taxa clustering within the *Arsenophonus* clade derived from the *Basic matrix* are printed in colour: red for the long-branched taxa, dark orange for the short-branched taxa.

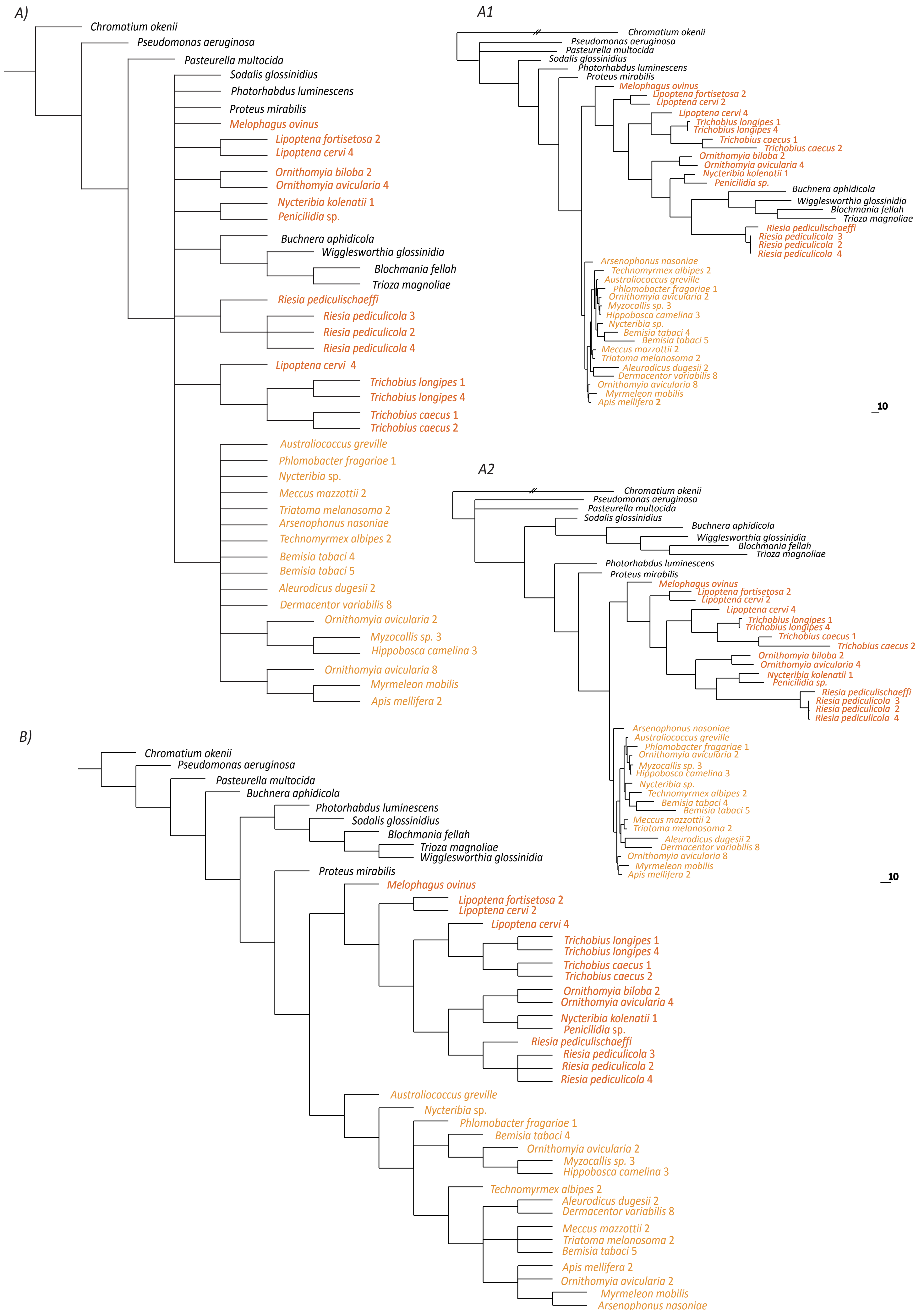

Figure 4. Topologies derived from *Sampling3* matrix: A) consensus of the trees and two tree examples A1 and A2, obtained under the MP criterion with Tv/Ts ratio set to 1:1 B) consensus of the trees obtained under the MP criterion with Tv/Ts ratio set to 1:3.

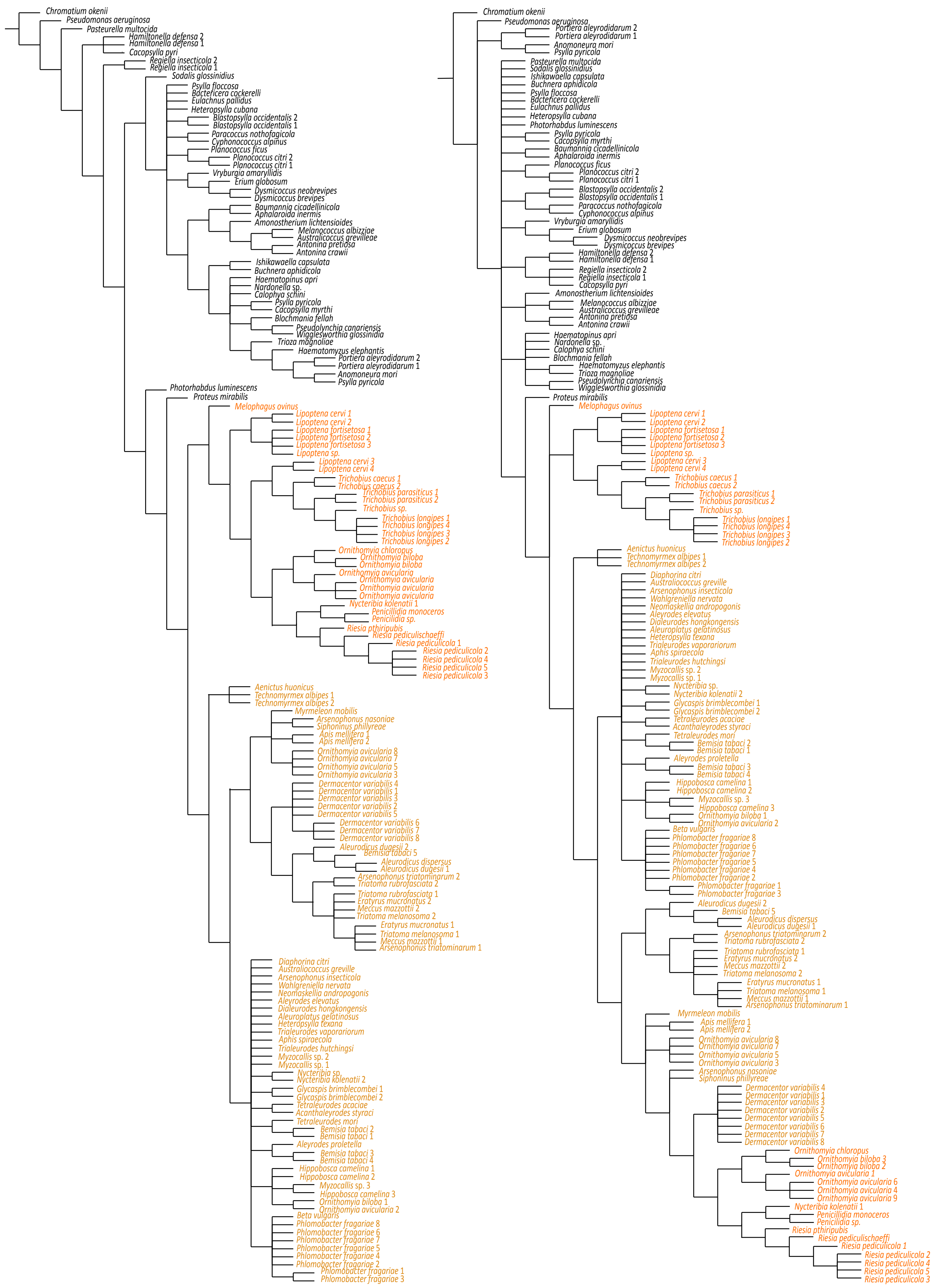

Figure 5**.** Topologies derived from the *Basic matrix*: A) consensus of the trees obtained under the MP criterion with transversion/transition ratio set to 1:3 and the ML criterion; B) consensus of the MP trees obtained with the transversion/transition ratio 1:1.

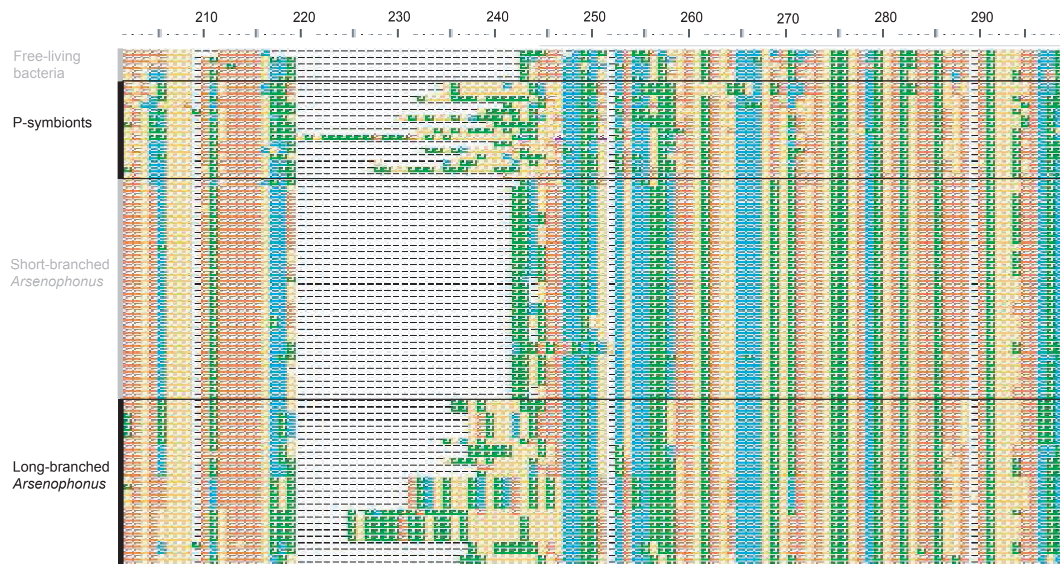

Figure 6. Insertions within the sequences of P-like symbionts.

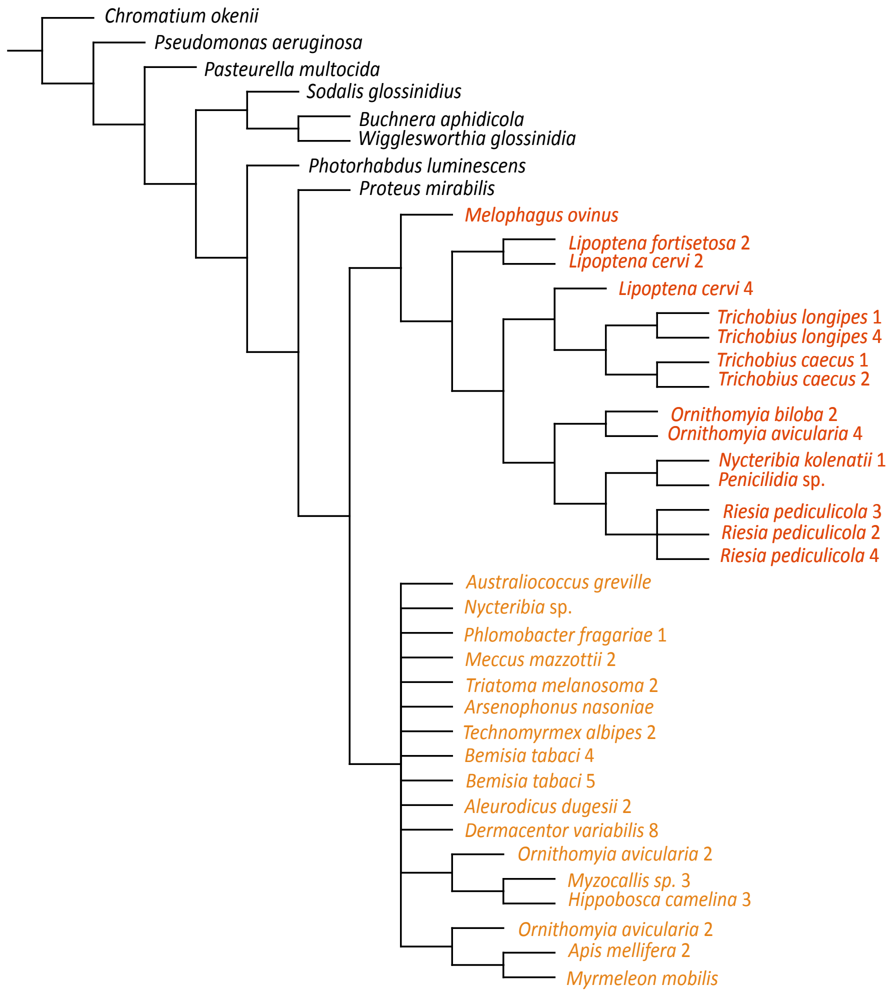

Figure 7. Tree consensus derived from *Sampling4* matrix under the MP criterion with transversion/transition ratio set to 1:1.

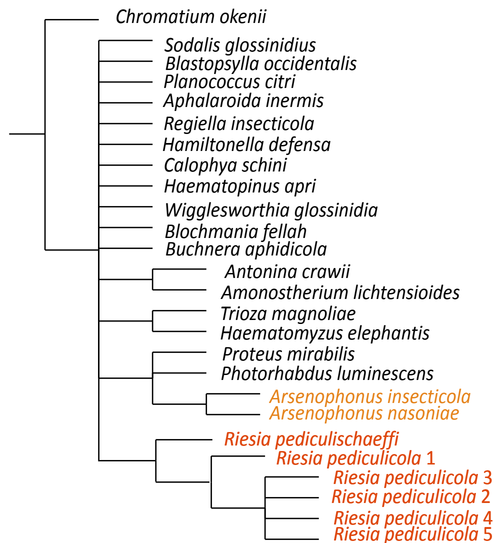

Figure 8. Tree consensus derived from *Sampling5* matrix under the MP criterion with transversion/transition ratio set to 1:1 .

Figure 9. The 41 bp long motif inconsistently distributed among distinct bacterial taxa. Position of the sequence in alignment and 16S rDNA secondary structure is indicated by the arows.