School of Doctoral Studies in Biological Sciences

University of South Bohemia in České Budějovice
Faculty of Science

Molecular phylogeny and genome evolution of insect symbiotic bacteria

Ph.D. Thesis

RNDr. Eva Nováková

Supervisor: prof. RNDr. Václav Hypša, CSc.
University of South Bohemia in České Budějovice, Faculty of Science

České Budějovice 2012

The thesis should be cited as:

Nováková E., 2012: Molecular phylogeny and genome evolution of insect symbiotic bacteria. PhD Thesis. University of South Bohemia, Faculty of Science, School of Doctoral Studies in Biological Sciences, České Budějovice, Czech Republic, 224 pp.

Annotation

Since the introduction of advanced molecular methods the research on insect bacterial symbioses underwent a major focus shift towards large scale phylogenetics and comparative genomics. These new fields provided answers to several fundamental questions of symbiont evolution, functional capabilities of the host-associated bacteria, and the role of symbionts in the host's biology. However, the vast diversity and complexity of symbiotic relationships still leaves gaps in our understanding to a rich mosaic of various symbiont types, effects and transitions from facultative association to obligate mutualism. The presented study focuses on distribution, diversity, phylogenetic patterns, evolutionary transitions and genome evolution of two less known but ecologically diverse bacterial genera, *Arsenophonus* and *Sodalis*. The thesis also takes advantage of the knowledge on a well established symbiotic model between aphids and *Buchnera* and reveals several evolutionary patterns in the host and symbiont.

Declaration [in Czech]

Prohlašuji, že svoji disertační 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 disertační práce, a to v úpravě vzniklé vypuštěním vyznačených částí archivovaných Přírodovědeckou fakultou elektronickou cestou ve veřejně přístupné části databáze STAG provozované Jihočeskou univerzitou v Českých Budějovicích na jejích internetových stránkách, a to se zachováním mého autorského práva k odevzdanému textu této kvalifikační práce. Souhlasím dále s tím, aby toutéž elektronickou cestou byly v souladu s uvedeným ustanovením zákona č. 111/1998 Sb. zveřejněny posudky školitele a oponentů práce i záznam o průběhu a výsledku obhajoby kvalifikační práce. Rovněž souhlasím s porovnáním textu mé kvalifikační práce s databází kvalifikačních prací Theses.cz provozovanou Národním registrem vysokoškolských kvalifikačních prací a systémem na odhalování plagiátů.

České Budějovice, 30.4. 2012 .....

Eva Nováková

This thesis originated from a partnership of Faculty of Science, University of South Bohemia, and Institute of Parasitology, Biology Centre of the ASCR, supporting doctoral studies in the Parasitology study program.



Financial support

GA AV, number IAA601410708

GA ČR, numbers 206/09/H026, P505/10/1401

GA JU, numbers 110/2010/P, 135/2010/P

MŠMT ČR, numbers LC06073H, MSM 6007665801

Fulbright Commission, grantee number 15102546

COST Action FA701

Acknowledgements

First, I would like to thank both of my academic advisers, Václav Hypša and Nancy Moran for their support, help, guidance and freedom with my research they both granted me with in their laboratories. Second, I have to thank Alistair Darby and Greg Hurst who have introduced me to the field of genomics. Many thanks also belong to all my fellow labbies from the Czech Republic and the US. I am grateful for our friendship with each of them.

My special thanks belongs to my family, my big and a little man, with whom I established a vital symbiosis back in New Haven, underwent the transition to the Czech Republic and prepared my dissertation. During our 35 week lasting intimate relationship, he successfully co-adapted to his mother organism and developed several effective mechanisms to cope with stress and occasional insanity periods. This thesis is devoted to our little man, Oskar Justen Leopold.

List of papers and author's contribution

1. Hypša, V. and Nováková, E. (2009): Insect symbionts and molecular phylogenetics. In: Insect Symbiosis, Bourtzis, K and Miller, T (eds), CRC press, pp. 1-32
E.N. participated in the general design of the chapter, writing of particular sections and preparing graphical parts.

2. Nováková, E., Hypša, V. and Moran, N.A. (2009): *Arsenophonus*, an emerging clade of intracellular symbionts with a broad host distribution. *Bmc Microbiology* 9, DOI: 10.1186/1471-2180-9-143 (IF=2.96)
E.N. participated in designing the study, performed all the laboratory work, participated substantially in phylogenetic analyses and their interpretation as well as writing the paper.

3. Nováková, E. and Moran, N.A. (2012): Diversification of genes for carotenoid biosynthesis in aphids following an ancient transfer from a fungus. *Molecular Biology and Evolution*. 29:313-323, DOI: 10.1093/molbev/msr206 (IF=5.5)
E.N. performed all the laboratory work and data analyses, and participated in result interpretation as well as writing the paper.

4. Wilkes, T.E, Duron, O., Darby, A.C., Hypša, V., Nováková E. and Hurst, G.D.D. (2011): The genus *Arsenophonus*. In: E Zchori-Fein & K Bourtzis ed(s). Manipulative Tenants. New York, CRC Press, pp. 225-241
E.N. participated in summarizing the knowledge on Arsenophonus triatominarum*, writing the part on* A. triatominarum *and reviewing the whole MS.*

5. Chrudimský, T., Husník, F., Nováková, E. and Hypša, V. *Candidatus* Sodalis melophagi sp. nov.: phylogenetically independent comparative model to the tsetse fly symbiont *Sodalis glossinidius.* MS submitted to PLOS ONE: under major revision.
E.N. generated genomic sequence data for the symbiont, participated in their assemblage and annotation, evolutionary inferences and writing the paper.

6. Nováková, E., Hypša, V., Klein, J., Foottit, R.G., von Dohlen, C.D. and Moran, N.A. Evolutionary history of aphids (Hemiptera: Aphididae): suitability of the P-symbiont *Buchnera aphidicola* DNA for reconstruction of the host phylogeny. MS prepared for publication.
E.N. participated in the study design, performed all the phylogenetic analyses and their interpretation, and wrote the first draft.

7. Nováková, E., Hypša, V., Moran, N.A., Hurst, G.D.D. and Darby, A.C.: A symbiotic process: The evolution of inherited symbiont genomes. MS in preparation.
*E.N. prepared the theoretical background and based on it selected a model organism for our part of the study (*Melophagus ovinus*), generated genomic data for the* M. ovinus *symbionts, participated in genome assemblage, annotation and comparative studies, prepared a concept of MS and coordinated its writing.*

1. BACKGROUND

1.1. Insect-bacteria associations in the molecular era

Symbioses between bacteria and insects are considered a significant driver of insect and microbe evolution and have been extensively studied (Buchner, 1965; Moran and Telang, 1998; Baumann et al., 2000; Baumann, 2005; Moran, 2006; Dale and Moran, 2006; Moran et al., 2008; Moya et al., 2008; McCutcheon and Moran, 2011). The most important advancements in insect-microbe interactions research are attributed to the introduction and application of new methods, allowing reasearchers to address previously unaccessible questions (electron microscopy, molecular phylogenetics, DNA-based visualization, etc.). Recently, modern molecular and computational approaches, especially large-scale sequencing and genome reconstruction, allowed for another shift in investigations on symbiont evolution, functional capabilities of the host-associated bacteria and their symbiotic role in its host's biology. Our current knowledge on insect-bacteria symbioses has grown immensely during the few last years and shows a rich mosaic of various symbiont effects and roles, host-symbiont relationships (including parasitism, facultative association and obligate mutualism), and phylogenetic patterns (from strict cophylogeny to frequent horizontal switches). All of these characteristics represent a surprisingly wide range of symbiotic phenotypes that in turn are reflected in the size and structure of the corresponding genomes. While the compact genomes of strict obligate mutualists are highly degraded and some can even surpass organelle genome size, facultative associates possess rather dynamic genomes similar to those of free living bacteria, retaining extensive metabolic capacities and various mechanisms for gene exchange, e.g. plasmids and phages (reviewed in McCutcheon and Moran, 2011). However, this is a simplified view that only highlights the two extremities found within symbiont diversity and does not encompass the variety of intermediate evolutionary stages that obviously arise in many bacterial lineages. To put genomic characteristics of these different ecological forms into a proper evolutionary framework, a reliable knowledge is essential not only on the genome's structure but also on phylogenetic relationships among the various lineages. Current state of the art insect-symbiont research combines advanced phylogenetic methods, often based on the genome-wide molecular data, with comparative genomics allowing for functional and ecological inferences.

1.2. Methodological issues

An apparent versatility of bacterial lineages that can adopt fundamentally distinct life strategies introduces a variety of methodological problems. Even closely related bacteria can rapidly evolve either into highly specialized mutualists or parasites/commensals loosely associated with their hosts (Hansen et al., 2012). Particular sources of phylogenetic artifacts thus do not only affect investigations on a broad phylogenetic scale but remain relevant even for reconstructions among closely related

genetic lineages. Genetic changes coincident with evolutionary shifts towards an obligate mutualism consist of high evolutionary rates along with relaxed selection, which lead to mutation accumulation and biased sequence composition (Moran 1996; Heddi et al. 1998; Lambert and Moran 1998, Rocha and Danchin, 2002). Most typically, the genomes of specialized mutualists are AT biased, although GC bias has also been reported for obligate symbionts (McCutcheon et al., 2009; McCutcheon and Dohlen, 2011). As a result, datasets composed of orthologous gene sequences from free-living bacteria and various symbionts that differ in ecological lifestyles tend to suffer from artifacts due to different nucleotide compositions.

This **sequence non-homogeneity** is a major phylogenetic obstacle and often leads to artificial clustering which reflects nucleotide composition (and therefore ecological characteristics) rather than real phylogenetic relationships. The main reason for this artifact is violation of the assumptions built in standard evolutionary models and methods used in phylogenetic analyses. Several approaches have been developed to overcome this problem. Early methods were based on modified distance measures, such as the paralinear (Logdet) distance method (Lake 1994; Lockhart et al. 1994) and the nonhomogeneous distance model (Galtier and Gouy 1995). Later, the nonhomogeneous approach was also implemented into the ML model (a nonhomogeneous T92 model; Tamura 1992) and used to readdress the persistent questions on origin of insect bacteria symbiosis. For instance, based on the nonhomogeneous model, Herbeck and colleagues (2005) brought the first strong evidence favouring P-symbionts polyphyly. The same approach was also used for inferring evolutionary relationships among *Arsenophonus* symbionts in **MS #2**. Although an important step towards eliminating the effect of nucleotide bias, these methods, similar to typical ML models, work with user-defined global parameters. Recently, the rapidly improving sequencing/genomic techniques allowed for accumulation of large multigene datasets. This in turn lead to the development of new category (CAT) models that do not rely on global parameters but rather estimate several parameter categories from the data (Lartillot and Philippe, 2004). Unlike the former methods, the CAT-like models require a substantial amount of data and were thus only adopted in the genome-based **MS #6**. Furthermore, we were able to use amino acid recoding (Lartillot et al., 2009) which was found effective in reducing artifacts when analyzing symbiotic lineages within the Enterobacteriacae (Husník et. al, 2011).

Compositional bias is not the only manifestation of progressive changes in symbiotic DNA. Mutations, expressed as insertions/deletions, disrupt open reading frames and larger deletions may even remove non-essential chromosomal regions (Ochman and Davalos, 2006; Toh et al., 2006; Burke and Moran, 2011). The genomes of symbiotic bacteria experience ongoing erosion through inactivation and loss of particular genes that are not essential for the symbiotic lifestyle. Under these conditions **missing data** for particular genes in some lineages may hamper the generation of a complete phylogenetic dataset. On a gene sequence level, occasional insertion/deletions pose a

problem for determining proper homologies and creating reliable alignments. For this reason, insertion/deletion regions are often removed from final alignment matrices. Such treatment of unreliably aligned regions and their impact on phylogentic inference has been illustrated in **MS #2**. When deletions occur on a genome level and cause loss of whole genes, absence of the markers in obligate symbionts may pose additional limitation on the phylogenetic analyses. This phenomenon may hamper analyses of highly heterogeneous datasets composed of various free-living and symbiotic phenotypes, but may also appear in supposedly uniform symbiotic lineages. This has been illustrated by the degraded tryptophan biosynthetic pathway in *Buchnera* genomes, particularly the trpB gene that has been lost in several members of the group (Gosalbes et al., 2008; also **MS #6**)

Apart from the methodological artifacts typical of bacterial symbionts, several more **general phylogentic issues** may also influence the molecular analyses. These issues, including the major methodological problems of phylogenetic and coevolutionary studies in insect-bacteria associations are reviewed in **MS #1**. The primary issue is determining the suitability of the selected gene as a meaningful phylogentical marker. In general, this requires two main prerequisites the gene has to fulfill; not to possess any paralogs within the genome and not to be acquired by horizontal gene transfer. Paralogs, which commonly arise from duplications, may be randomly sampled from different lineages in the phylogenetic analysis, and thus can result in serious topological inaccuracies. Frequently, this issue arises in low-level phylogenetic studies based on 16S rRNA gene. Intragenomic heterogeneity of the 16S-23S-5S rRNA operons and its influence on phylogenetic reconstruction has previously been shown in free/living bacteria (e.g. Yap et al. 1999; Marchandin et al. 2003; Boucher et al. 2004; Lin et al. 2004; Naum et al., 2008, Pei et al, 2010) and similar conditions were also observed in symbiotic lineages (Laguerre et al., 1997; Thao and Baumann, 2004; Trowbridge et al., 2006; Sorfova et al., 2008; **MS #2**).

Regardless of this potential risk in generating phylogenetic artifacts, 16S rDNA is the most commonly used gene marker in prokaryotes. In closely related lineages of obligate mutualist symbionts, it is less problematic since redundant operon copies were lost or even the entire operon was fragmented to isolated rDNA genes throughout the degradation process (Munson et al., 1993; van Ham et al, 2003). In these symbionts, 16S rDNA as a single copy marker served as an effective tool for deriving clear coevolutionary patterns (Chen et al. 1999; Clark et al. 2000; Sauer et al. 2000; Thao et al. 2000). In accordance to these findings, 16S rDNA was used as one of the markers for *Buchnera* derived phylogenies (**MS #6**). In a few other cases rRNA genes, as the gene with the highest number of sequence records in public databases, have been used to describe diversity and more general phylogenetical patterns within polymorphic symbiotic genera, e.g., *Rickettsia* (Perlman et al., 2006), *Serratia* (Lamelas et al., 2008), *Arsenophonus* (Trowbridge, 2006; Allen et al., 2007; **MS #2**), and *Sodalis* (Fukatsu et al., 2007, Novakova and Hypsa, 2007; **MS #5**).

Despite the possible problems discussed above, molecular data from bacterial symbionts are in many ways easier to use for phylogenetic reconstructions then the data derived from their insect hosts. This is a natural consequence of much lower complexity of the bacterial genomes than eukaryotic genomes. For example, **gene duplication** is usually much more extensive and therefore a stronger source of artifacts in eukaryotes then in Bacteria (Friedman and Hughes, 2001; Abi-Rached et al., 2002; McLysaght et al., 2002, Cui et al., 2006, Zhou et al., 2010). The reduced genome complexity of a symbiont may be of a particular use in a system where we encounter difficulty reconstructing a host phylogeny (due to gene duplication and/or other sources of artifacts) but we have a strong indication for strict cospeciation process between the host and symbiont. Aphid- *Buchnera* symbiosis provides such example. Recently, the genome sequence of *Acyrthosiphon pisum* confirmed a vast amount of gene duplications affecting more than 2,000 gene families (The International Aphid Genomic Consortium, 2010). In such cases, selection of proper markers may be an extremely demanding or impossible task and alternative approaches for understanding the host phylogeny, i.e. using symbiont derived genes may be more suitable (**MS #6**).

Many of the phylogenetic problems can be reduced or even overcome if complete genomes are available. Such data provide high amounts of phylogenetic information allowing proper use of advanced techniques (e.g., CAT-like models), thus making it possible to find gene duplications, losses, modifications, etc. Currently, genome sequencing is still a very demanding task in eukaryotes. However, the relatively small genomes of bacteria make the phylogenetic usage of complete genomes much more feasible, and this approach has been used in many studies (e.g., Eisen et al., 2000; Lerat et al., 2003; Comas et al., 2007; Wu et al., 2009; Williams et al., 2010). Nevertheless, investigations relying on *de novo* sequencing of symbionts introduce an entirely new set of technical considerations.

Naturally, successful genome sequencing depends on the purity of the DNA template being used. In complex symbiotic systems, acquiring a sufficient amount of high quality DNA may represent a major obstacle as illustrated by several authors (e.g., Burke and Moran, 2011; also **MS #7**). Furthermore, particular attention has to be paid to selection of a proper sequencing strategy, i.e. sequencing platforms and library preparation techniques that would recover data sufficient for the whole genome assembly. This can be particularly challenging in the case of genomes with extreme nucleotide content (e.g. McCutcheon and Moran, 2010), or repetitive DNA. Dynamic symbiont genomes such as *Bartonella* species, *Orientia tsutsugamushi*, *Serratia symbiotica* str. Aps, *Arsenophonus nasoniae*, *Arsenophonus triatominarum* contain a high proportion of repetitive DNA, especially transposable and phage elements, and as a result hamper assembly efforts (Cho et al., 2007; Nakayama et al., 2008; Engel et al., 2010; Darby et al., 2010; Burke and Moran, 2011; **MS #7**). More specifically, repetitive regions exceeding the read length create gaps in *de novo* assembly. The short length of next generation

sequencing (NGS) reads results in fragmented assemblies or even complex misassambled rearrangments (reviewed in Treangen and Salzberg, 2011). The use of other strategies (e.g., Sanger sequencing) with longer read lengths, or implementation of long paired-end information from fosmids and bacterial artifical chromosomes (BAC) may represent a possible solution to produce finished genomes. However, such an approach requires generating an immense amount of different data types, puts high demands not only on resources but also on bioinformatic processing and, for most cases, does not guarantee any substantial advances in the research outcomes based on the completely closed genome sequences.

2. OBJECTIVES AND MODELS

The above overview shows that current research on bacterial symbionts of insects has expanded into an integrated network of evolutionary, functional and taxonomical questions requiring a broad array of laboratory and computational techniques. During the studies summarized in the manuscripts presented here, I focused on a subset of problems and models. Their selection was partly based on my previous activities, particularly experimental work during my bachelor and master theses, and participation in several research projects. It was also further influenced by collaboration with two prestigious laboratories in this research area, where I spent a total of 14 months during my PhD (Fulbright fellowship with Nancy Moran, Yale University, and COST Short Term Scientific Mission with Gregory Hurst, University of Liverpool). More specifically, I focused on phylogenetic issues in two emerging clades of symbionts, *Arsenophonus* and *Sodalis*, coevolutionary relationships in a model symbiont system, *Buchnera aphidicola*, and genomic comparisons as a tool for revealing evolutionary processes within the symbiotic genomes.

In the first area I utilized standard phylogenetic approaches to explore evolutionary patterns in several symbiotic groups. I focused on two ecologically rich groups of bacterial endosymbionts that become popular models for investigating symbiont diversity and evolution: *Sodalis* and *Arsenophonus*. I screened for new lineages in several hosts, mainly blood-feeding dipterans of the family Hippoboscidae. Potential importance of this model rests in its close ecological and phylogenetic relationship to the association between the tsetse fly and its symbionts. I performed complex phylogenetic reconstructions to be used as a background for evolutionary interpretation and identification of suitable candidates for subsequent genomic comparisons.

In the second area I took advantage of the previously established collaboration with laboratory of Nancy Moran (Yale University, USA) and access to the well-established model of host-symbiont cospeciation, aphids and mutualistic bacteria of the genus *Buchnera*.

The third area of my research was based on genomic rather then phylogenetic approaches. Based on the results from *Arsenophonus* phylogeny, I selected a proper model of *Arsenophonus* for complex comparative study. This study was performed together with Gregory Hurst, Alistair Darby (University of Liverpool), and Nancy Moran (Yale University). The main aim of this study was to compare basic characteristics of four related symbionts (all members of the *Arsenophonus* clade) with significantly different lifestyles and presumably different evolutionary stages in the symbiogenesis.

Research in these three areas resulted in the seven presented studies. Five of the studies are standard research papers (published or in preparation). Additionally, two are invited chapters that originated as the result of this research. The first of these chapters summarizes our experience with phylogenetic tools and specifities of their use in symbiont evolutionary inferences. The second chapter puts together results of several groups from all over the world, on the genus *Arsenophonus*. The chapter includes various aspects of the biology, diversity and phylogeny of these symbionts. In both cases, the chapters (**MS #1**and **MS #4**) are invited reviews based on our previous research.

3. RESEARCH OUTCOMES

3.1. Coevolution: relationship between symbiont and host phylogenies

Depending on the nature of a given symbiosis, the relationships between the host and symbiont phylogenies may vary from perfect identity to complete incongruency. These varied associations raise different phylogenetic questions and evolutionary problems. During the molecular era, it became clear that most of the seemingly non-specialized symbionts are not phylogenetically isolated offshoots of free-living bacteria but belong to several discrete bacterial lineages capable of infecting phylogenetically distant hosts. In these groups, host-symbiont phylogenies display a high degree of incongruence; overall diversity and distribution of the symbionts among host taxa are therefore common topics studied by molecular phylogeny tools (e.g. Kikuchi and Fukatsu, 2003; Russell et al., 2003; Perlman et al., 2006; Burke et al., 2009; Sirviö and Pamilo, 2010; Mouton et al. 2012).

In contrast, intimate associations have been discovered that are vital for both bacteria and host (i.e. strictly speaking, mutualistic relationships). As an essential component of the host life strategy these bacteria become indispensable and reach a point where they can be treated as “organelles” rather than independent organisms. Origins of such associations are now understood to be the result of unique symbiont acquisitions by the host ancestor followed by strict coevolution between the two counterparts ensured by transovarial transmission. As such typical coevolutionary (in a sense cospeciation) studies are passe and the focus in these associations has been diverted to more general

questions, including genomic evolution of the symbionts or possible genetic exchange with the host. Among obligate endosymbionts, the genera *Wigglesworthia*, *Baumannia*, and *Buchnera* are intensively studied (e.g., Baumann et al., 1995; Aksoy, 1995; Shigenobu et al. 2000; Akman et al. 2002, van Ham et al., 2003, Wu et al. 2006, Hansen and Moran, 2011), and promote adaptation to a new life strategy adopted by the hosts, typically blood or plant sap feeders, by supplementing their unbalanced diet (Buchner 1965, Nogge 1981).

Among these obligatory mutualistic associations, ***Buchnera***-aphid symbiosis has for long been the most extensively studied since the host group includes major worldwide pests. The symbiosis is thought to have originated via a single acquisition event around 150-200 MYA, followed by strict cospeciation between *Buchnera* and aphid hosts (Moran et al., 1995; Martinez-Torres et al., 2001). The significance of the symbiotic association has been corroborated by analysis of genomic and transcriptomic data from both *Buchnera* and the host, *A. pisum*, (Shigenobu et al. 2000, The International Aphid Genomics Consortium, 2010; Nakabachi et al. 2005, Hansen and Moran, 2011). These analyses have brought a deeper understanding on particular functionalities of the entire system which thus became the best-known and extensive studied symbiotic model. It is therefore paradox that the most interesting features of the host biology (e.g., structure and complexity of aphid life cycle, alteration of sexual and asexual reproduction, host plant alternations) still remain uncoupled with aphid evolutionary history since a robust phylogeny has not yet been proposed for the aphids. Even though recent phylogentic studies on aphids implemented multi-gene analyses, combining the data from nuclear and mitochondrial genes (e.g., Ortiz-Rivas and Martínez-Torres, 2010), a pronounced lack of sufficient phylogenetic signal still prevents conclusive evolutionary interpretation. A common approach of broadening the analysis by additional genes with better informative capacity is extremly demanding as a vast majority of the gene families in the aphid genome underwent duplications (The International Aphid Genomics Consortium, 2010).

In the **MS #4**, we proposed and evaluated an alternative approach for generating an aphid phylogeny, using *Buchnera* derived markers (see the methodological issues). In order to reconstruct phylogenetic relationships within the family Aphididae we based our analyses on more than 300 sequences derived from five genes of *Buchnera* symbionts associated with a taxonomically broad set of aphid species. We focused on questions of the monophyly of individual subfamiles/tribes, their relationships, evolutionary rates and rooting of the Aphididae tree. Our results showed that symbiotic DNA yields an informative phylogenetic signal and is a useful source of data. In comparison to the aphid-based analyses, *Buchnera* derived data brought stronger support for several topological patterns and did not reveal any major discrepancies. Both data sources contradicted only in weakly supported and unstable relationships. Similar to aphid genes, analyses of *Buchnera*-derived genes did not bring a conclusive view for deeper phylogenetical nodes, which most probably reflects the initial rapid diversification in

some lineages previously proposed by several authors (von Dohlen and Moran, 2000; Martínez-Torres et al., 2001).

In addition to the obligatory mutualistic associations, there is a broad array of bacterial lineages in different stages of symbiogenesis. Our knowledge on these symbionts is much more fragmented. This is mostly due to their "random" and virtually unknown host distribution and diversity. The spread of these bacteria through host populations does not rely exclusively on vertical transmisson from mothers to progeny but also involves horizontal transfers causing incongruences between the host and symbiont phylogenies. Probably the most famous example is alphaproteobacterial genus *Wolbachia*, one of the most diverse and widespread bacteria among insects (e.g. Lo et al., 2002; Bordenstein et al., 2008; Augustinos et al., 2012). Within Gammaproteobacteria, similar pictures of rich and diverse symbiotic groups are provided particularly by two genera, *Arsenophonus* and *Sodalis*.

For the bacterial genus ***Arsenophonus*** there are only few valid described species: *Arsenophonus nasoniae*, the type species known as a son-killer parasite of pteromalid wasp, *Nasonia vitripennis* (Werren et al., 1986; Gherna et al., 1991); *Candidatus* Arsenophonus triatominarum, a bacterial associate of triatomine bugs (Hypša and Dale, 1997), *Candidatus* Arsenophonus arthropodicus described from a hippoboscid fly *Pseudolynchia canariensis* (Dale et al., 2006), and *Candidatus* Arsenophonus phytopathogenicus identified from the planthopper *Pentastiridius leporinus* (Bressan et al., 2011). The genus currently includes two other members originally designated as distinct genera: an obligate mutualist of human body louse, *Candidatus* Riesia pediculicola, and plant pathogenic bacteria transmitted by planthopper species, *Candidatus* Phlomobacter fragariae (Zreik et al., 1998; Allen et al., 2007; Salar et al., 2009). In addition to this modest number of formally described species, a high number of symbiotic bacteria identified exclusivelly by 16S rRNA sequences have been reported to fall into the *Arsenophonus* clade. Their host spectrum spans diverse insect groups, several non-insect taxa as well as a few plant species, and their associations constitute highly different phenotypes (e.g., Gherna et al., 1991, Dale et al., 2006; Trowbridge et al., 2006; Allen et al., 2007; Semetey et al., 2007; Duron et al., 2008; Hosokawa et al., 2012). These characteristics implicate the genus *Arsenophonus* as an important and widespread lineage of symbiotic bacteria and potentially valuable model for examining the evolution of insect symbiosis. In **MS #2**, we summarized the picture of *Arsenophonus* evolution and provide a unified phylogenetic framework for the further evolutionary and genomic investigations. The analysis, comprising more than hundred 16S rDNA sequences available at the time for the *Arsenophonus* clade, revealed at least two transitions from faculatitive symbiosis to obligate mutualism. We also showed that different life strategies proposed for particular lineages were clearly correlated to sequence properties, i.e. the GC content and a branch length. While short branched taxa represented bacterial associates with neutral or slightly negative effects on the host, e.g., *A. nasonie*, that apparently undergo horizontal transfers, long branches were associated with

obligate mutualists showing a strict pattern of maternal transmission. Furthermore, the study explored molecular characteristics and informative value of the 16S rRNA gene as the most frequently used phylogenetic marker (discussed in methodological issues).

Similarly to *Arsenophonus* bacteria, the genus ***Sodalis*** adopts several different types of symbiosis with a broad range of insect hosts, including tsetse flies, weevils, chewing lice, hippoboscid louse flies, ants, scale insects, aphids, stinkbugs, and cerambycid beetles (Dale and Maudlin, 1999; Sameshima et al., 1999; Fukatsu et al. 2007; Nováková and Hypša, 2007; Burke et al., 2009; Gruwell et al., 2010; Grünwald et al., 2010; Kaiwa et al., 2010; Toju et al., 2010, Toju and Fukatsu, 2011). Despite this considerable diversity, the knowledge of *Sodalis* as a bacterial genus is seriously limited. The most attention has been paid to the tse-tse fly symbiont, *Sodalis glossinidius* (Dale and Maudlin, 1999), which constitutes only a part of its host's symbiotic community. In addition to *S. glossinidius*, tse-tse harbours a strict mutualist, *Wigglesworthia glossinidia* (Aksoy, 1995), and more loosely associated bacteria of the genus *Wolbachia* (Cheng et al., 2000). Genomic studies focused on processes of metabolic complementation within the community and reveal an interaction between *Sodalis* and *Wigglesworthia* via thiamine biosynthesis (Belda et al., 2010). Furthermore, the genome sequence of *S. glossinidius* supported the hypothesis that these bacteria represent an intermediate stage of symbiosis and *S.glossinidius* became a general model for investigation on evolutionary changes in symbiogenesis. Our research has been focused on exploring *Sodalis* diversity within blood feeding hosts. To date we have described two members of the genus *Sodalis*, both phylogenetically distinct from the tsetse symbiont (Nováková and Hypša, 2007; **MS #5**). In the current manuscript, we present basic molecular and morphological characterization of *Candidatus* Sodalis melophagi isolated from a complex bacterial community of the sheep ked, *Melophagus ovinus*. Based on draft genomic data, we further characterized the type threee secretion system (T3SS) which mediates establishment of close relationships between the bacterium and and the eukaryotic hosts (Coombes, 2009) and thus plays a major role in pathogenesis and symbiosis. We also propose *Melophagus* and *Glossina* hosts to serve as suitable models for compartaive genomics focused on symbiosis in blood feeding organisms. Analyses comparing possible adaptive changes bound to evolutionary replacement of the obligate symbiont in closely related hosts with similar but unique biology will provide important insight into the symbiogenetic processes.

3.2. Comparative genomic of symbiotic bacteria

Current understanding of the genomic evolutionary processes shaping symbiogenesis is mainly based on comparisons between phylogenetically unrelated symbiotic forms and/or comparisons of symbiotic lineages to free-living bacteria. This naturally leaves gaps in our insight into the transitional processes and various intermediate steps of symbiosis. On the other hand, comparisons between related

symbiotic lineages, which could potentially bring important data on the gradual genomic modifications, have so far been done only on the symbionts of few genera, e.g., *Buchnera* and *Blochmannia* (Tamas et al., 2002; Degnan et al., 2005), with highly reduced genomes that do not allow for assessing the changes along all stages of the symbiosigenesis. Such dramatic reduction of symbiotic genomes due to massive gene loss is typical for these mutualistic relationships. Since the main role of the symbionts is to provide host with particular nutrients, these processes have always provoked a question how intimate is the genetic integration between the symbiont and the host. The highly reduced genomes of symbionts could be considered analogous to organelles. Therefore attention was paid to potential transfer of symbiont genes to the host genome as gene transfer has been found to occur between both mitochondria and plant chloroplasts and their hosts. So far, this issue was addressed in two symbiotic systems for which the host genomes became recently available, the aphid *Acyrthosiphon pisum* with symbiont *Buchnera aphidicola*, and the human body louse *Pediculus humanus* with *Ca.* Riesia pediculicola (Kirkness et al., 2010; The International Aphid Genomics Consortium, 2010). In the pea aphid genome, bacterial origin was identified for 12 genes or gene fragments. While the majority was *Wolbachia*-derived, there were only two pseudogenes of *Buchnera* origin (Nihoh et al., 2010). The louse genome completely lacked any evidence of bacterial gene acquisition (Kirkness et al., 2010). Although, these results suggest that gene transfer from the bacterial to the host chromosome is not common in insect symbiosis, they do not rule out possibility of such events in systems in which the symbiont has undergone greater levels of genome erosion than those observed in *B. aphidicola* or *R. pediculicola* (e.g., *Ca.* Sulcia muelleri, *Ca.* Zinderia insecticola, *Ca.* Carsonella ruddii, *Ca.* Hodgkinia cicadicola, or *Ca.* Tremblaya princeps with the smallest genomes sequenced so far). Furthermore, the investigations of *A.pisum* genome pointed out the importance of gene transfer for aphid biology. The genome does not only carries intact and highly expressed *Wolbachia*-derived genes with supposedly important roles in its symbiosis, it also gained fungal genes for carotenoid biosynthesis (Moran and Jarvik, 2010).

In **MS #3**, we focused on this unique trait and explored the evolution of carotene desaturase gene family within aphids. Analyses of more than 30 species spanning a wide taxonomic range of aphids and adelgids suggested that the gene transfer predated split of these two families. In contrast to the donor fungal species that carry a single copy of carotene desaturase, we showed that transferred genes underwent an extensive duplication that lead to accumulation of up to seven copies in the aphid genomes. The copies include paralogs of ancient as well as recent origin for which patterns of pseudogenizations, recombination and occasional positive selection were identified. We also showed a possible relation of carotene desaturase copy numbers and color polymorphisms that often occurs among aphids from the Macrosiphini subfamily. Altogether, our results indicate the aphid evolutionary history to be accompanied by ongoing evolution of carotenogenic genes that yield a wide variety of carotenoid profiles in different aphid species.

Recently, several studies with broader sampling brought an extended knowledge on the symbiont's phylogeny sufficient for selection of proper comparative models. In particular, bacterial genera *Wolbachia*, *Sodalis*, *Serratia* and *Arsenophonus* are prime examples of heterogenous lineages, which can provide a common framework for tracking various evolutionary changes behind the shifts in symbiotic lifestyle. (Nováková and Hypša, 2007; Bordenstein et al., 2009; Burke et al., 2009; also **MS #2**, **MS #4**, **MS #5**).

Based on our previous research, we focused our genomic investigations on the selected *Arsenophonus* symbionts. In the **MS #7**, we present a comparative study on the evolution of symbiosis based on four genomes covering such different forms as parasites, loosely associated facultative bacteria, and obligate mutualists in different stages of adaptation. We sequenced the genomes of *Arsenophonus triatominarum* (has been done by our collaborators from the University of Liverpool) and *Ca.* Arsenophonus melophagi (our own data), and verified the position of *Ca.* Riesia pediculicola within the genus *Arsenophonus*. The previously published genome of *A. nasoniae* (Darby et al., 2010) was used as a fourth lineage in this comparative study. Our results revealed several interesting patterns for the shifts within the hypothetical parasite-mutualist continuum. Among them, the most interesting finding is a significant shift in the nucleotide composition towards high AT content at the first codon position along with a biased codon usage, and exceptionally strong selection pressure in the obligate mutualist, *Ca.* Riesia pediculicola. Particularly the latter phenomenon is unexpected since the most derived mutualistic symbionts generally possess degraded genomes due to relaxed selection. Our tentative explanation involves effect of both, the repair mechanisms and selection strength. The process of genome modification along the spectrum from free-living bacterium to obligatory mutualists is thus likely to be more complex then believed so far. In this study, we also confirmed and summarized some general changes that often occur in bacteria with a host-restricted lifestyle (e.g., changes in the coding capacity, reduction of the genome size due to the gene losses of particular functional categories in different symbiotic stages, advanced elimination of mobile elements, limitations on metabolic, secretion and sensing capacity, erosion of reparation and recombination machinery).

4. CONCLUSION AND FUTURE PROSPECTS

Two main conlcusions can be drawn from the presented studies: 1) Bacterial symbiont lineages display a perplexing versatility of their life strategies. A single cluster of closely related lineages can produce a broad variety of ecological forms. In our studies we demonstrated this particularly on the diverse symbiotic cluster of *Arsenophonus spp*. Individual members of this group are associated with a broad range of hosts and even closely related members can adopt dramatically different lifestyles. 2)

Genomic comparisons show that the evolutionary changes accompanying the lifestyle shifts are complex and cannot be explained by a consistent set of simple rules. For example, the patterns we observed indicate that the typical manifestations of genome degradation, that is, elevated mutation rate, nucleotide bias and relaxed selection, do not work in concert across the different ecological forms, but each follows its own evolutionary trajectory.

These conclusions point in the same direction: rapidly growing information on the symbionts phylogeny and genomics reveal new unexpected phenomena and stress the complexity of the symbiogenetic process. The investigations on evolution and genomics in insect-bacteria symbioses are clearly in an analytical phase and will require many additional models with firmly established phylogenies, complete knowledge on their biology and rigorous comparisons of their genomes, before any meaningful synthesis can be reached.

In addition, limited attention has been paid to the fact, that the insect-bacteria symbiosis often involves more than one symbiont. In many cases, the host harbors a community of phylogenetically and ecologically distinct symbionts. So, the proper comparative study should take in the account not only the interactions between individual symbionts and the host, but also between the symbionts themselves. We have started this kind of research by selecting the *Melophagus ovinus* model which contains four different bacterial symbionts, the obligate mutualist of the genus *Arsenophonus* and three more loosely associated commensals/parasites of the genera *Sodalis*, *Bartonella* and *Wolbachia*.

In my future research I plan to further focus on *Arsenophonus* as a model organism. First, I would like to extend the knowledge on *Arsenophonus* diversity and establish a firm phylogenetic framework uncovering additional newly described lineages. Based on this background, I will be able to establish proper models to pursue further genomic research, tackling two different areas of my main interest. The first one involves investigations on various stages in symbiogenesis and the second addresses questions on molecular mechanisms underlying the origin of linear genomes in bacteria. To achieve this, genome sequencing will be performed on several lineages of distinct symbiotic forms evolving under different environmental conditions (preferably members of more complex symbiotic communities), as well as those closely related to *Ca.* Riesia pediculicola, a bacterium with a highly degenerated linear chromosome. Although genomic sequences provide insight into genetic and evolutionary changes, and prediction on regulation mechanism and metabolic capacity, it does not allow for detection of gene functionality. Thus, in order to further shift my research into functional genomics, I will integrate transcriptomic approaches into my future investigations.

5. RFERENCES

Abi-Rached L, Gilles A, Shiina T, Pontarotti P and Inoko H. (2002). Evidence of en bloc duplication in vertebrate genomes. *Nat Genet* **31:** 100-105.

Allen JM, Reed DL, Perotti MA, Braig HR. (2007). Evolutionary relationships of "*Candidatus* Riesia spp., " endosymbiotic *Enterobacteriaceae* living within hematophagous primate lice. *Appl Environ Microbiol* **73:**1659-1664.

Akman L, Yamashita A, Watanabe H, Oshima K, Shiba T, Hattori M, Aksoy S. (2002). Genome sequence of the endocellular obligate symbiont of tsetse flies, *Wigglesworthia glossinidia*. *Nat Genet* **32:** 402-407.

Aksoy S. (1995). *Wigglesworthia* gen. nov. and *Wigglesworthia glossinidia* sp. nov., taxa consisting of the mycetocyte-associated, primary endosymbionts of tsetse flies. *Int J Syst Bacteriol* **45:** 848-851.

Aksoy S, Chen X, Hypša V. (1997). Phylogeny and potential transmission routes of midgut-associated endosymbionts of tsetse (Diptera: Glossinidae). *Insect Mol Biol* **6:** 183-190.

Augustinos AA, Santos-Garcia D, Dionyssopoulou E, Moreira M, Papapanagiotou A, Scarvelakis M, Doudoumis V, Ramos S, Aguiar AF, Borges PAV, Khadem M, Latorre A, Tsiamis G, Bourtzis K. (2012). Detection and characterization of *Wolbachia* infections in natural populations of aphids: Is the hidden diversity fully unraveled? *PLoS One* **6:** e28695.

Baumann P, Baumann L, Lai C-Y, Rouhbakhsh D, Moran NA, Clark MA. (1995). Genetics, physiology, and evolutionary relationships of the genus *Buchnera*: intracellular symbionts of aphids. *Annu Rev Microbiol* **41:** 55-94.

Baumann P, Moran NA, Baumann L. (2000). Bacteriocyte-associated endosymbionts of insects, p. *403*-438. In M. Dworkin (ed.), The prokaryotes. *Springer, New York, NY.*

Baumann, P. (2005). Biology of bacteriocyte-associated endosymbionts of plant sap-sucking insects. *Annu Rev Microbiol* **59:** 155-189.

Belda E, Moya A, Bentley S, Silva FJ. (2010). Mobile genetic element proliferation and gene inactivation impact over the genome structure and metabolic capabilities of *Sodalis glossinidius*, the secondary endosymbiont of tsetse flies. *BMC Genomics* **11:** 449.

Bordenstein SR, Paraskevopoulos C, Hotopp JC, Sapountzis P, Lo N, Bandi C, Tettelin H, Werren JH, Bourtzis K. (2009). Parasitism and mutualism in *Wolbachia*: what the phylogenomic trees can and cannot say. *Mol Biol Evol* **26:** 231-241.

Boucher Y, Douady CJ, Sharma AK, Kamekura M, Doolittle WF. (2004). Intragenomic heterogeneity and intergenomic recombination among haloarchaeal rRNA genes. *J Bacteriol* **186:** 3980-3990.

Bressan A, Terlizzi F, Credi R. (2011). Independent origins of vectored plant pathogenic bacteria from arthropod-associated *Arsenophonus* endosymbionts. *Microbiol Ecol* **63:** 628-38.

Buchner P. (1965). Endosymbiosis of animals with plant microorganisms. Interscience Publishers, Inc., New York, NY.

Burke GR, Normark BB, Favret C, Moran NA. (2009). Evolution and diversity of facultative symbionts from the aphid subfamily Lachninae. *Appl Environ Microbiol* **75:** 5328-5335.

Burke GR, Moran NA. (2011). Massive genomic decay in *Serratia symbiotica*, a recently evolved symbiont of aphids. *Genome Biol Evol* **3:** 195-208.

Chen XA, Li S, Aksoy S. (1999). Concordant evolution of a symbiont with its host insect species: Molecular phylogeny of genus *Glossina* and its bacteriome-associated endosymbiont, *Wigglesworthia glossinidia*. *J Mol Evol* **48:** 49-58.

Cheng Q, Ruel TD, Zhou W, Moloo SK, Majiwa P, O'Neill SL, Aksoy S. (2000). Tissue distribution and prevalence of *Wolbachia* infections in tsetse flies, *Glossina* spp. *Med Vet Entomol* **14:** 44-50.

Cho NH, Kim HR, Lee JH, Kim SY, Kim J, Cha S, Kim SY, Darby AC, Fuxelius HH, Yin J, Kim JH, Kim J, Lee SJ, Koh YS, Jang WJ, Park KH, Andersson SG, Choi MS, Kim IS. (2007). The *Orientia tsutsugamushi* genome reveals massive proliferation of conjugative type IV secretion system and host-cell interaction genes. *Proc Natl Acad Sci USA* **8:** 7981-7986.

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

Comas I, Moya A, Gonzalez-Candelas F. (2007). From phylogenetics to phylogenomics: the evolutionary relationships of insect endosymbiotic γ-Proteobacteria as a test case. *Syst Biol* **56**: 1-16.

Coombes BK. (2009). Type III secretion systems in symbiotic adaptation of pathogenic and non-pathogenic bacteria. *Trends Microbiol* **17:** 89-94.

Cui LY, Wall PK, Leebens-Mack JH, Lindsay BG, Soltis DE, Doyle JJ, Soltis PS, Carlson JE, Arumuganathan K, Barakat A, Albert VA, Ma H, dePamphilis CW. (2006). Widespread genome duplications throughout the history of flowering plants. *Genome Res* **16:** 738-749

Dale C, Maudlin I. (1999). *Sodalis* gen. nov . and *Sodalis glossinidius* sp. nov., a microaerophilic secondary endosymbiont of the tsetse fly *Glossina morsitans morsitans*. *ISME J* **49:** 267-275.

Dale C, Moran NA. (2006). Molecular interactions between bacterial symbionts and their hosts. *Cell* **126:** 453-65.

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*. *Appl Environ Microbiol* **72:** 2997-3004.

Darby A, Choi JH, Wilkes T, Hughes M, Werren J, Hurst G, Colbourne J. (2010). Characteristics of the genome of *Arsenophonus nasoniae*, son killer bacterium of the wasp *Nasonia*. *Insect Mol Biol* **19:** 75-89.

Degnan PH, Lazarus AB, Wernegreen JJ. (2005). Genome sequence of *Blochmannia pennsylvanicus* indicates parallel evolutionary trends among bacterial mutualists of insects. *Genome Res* **15:** 1023-1033.

Duron O, Bouchon D, Boutin S, Bellamy L, Zhou L, Engelstädter J, Hurst GD. (2008). The diversity of reproductive parasites among arthropods: *Wolbachia* do not walk alone. *BMC Biol* **6:** 27.

Eisen, JA. (2000). Assessing evolutionary relationships among microbes from whole-genome analysis. *Curr Opin Microbiol* **3:** 475-480.

Friedman R, Hughes AL. (2001). Gene duplication and the structure of eukaryotic genomes. *Genome Res* **11:** 373-381.

Engel P, Salzburger W, Liesch M, Chang C-C, Maruyama S, Lanz C, Clateau A, Lajus A, Médigue C, Schuster SC, Dehio C. (2010). Parallel evolution of a Type IV secretion system in radiating lineages of *Bartonella*. *PLoS Genet* **7:** e1001296.

Fukatsu T, Koga R, Smith WA, Tanaka K, Nikoh N, Sasaki-Fukatsu K, Yoshizawa K, Dale C, Clayton DH. (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. (1995). Inferring phylogenies from DNA sequences of unequal base compositions. *Proc Natl Acad Sci USA* **92:** 11317-11321.

Gherna RL, Werren JH, Weisburg W, Cote R, Woese CR, Mandelco L, Brenner R. (1991). *Arsenophonus nasoniae*, genus novel, species novel, causative agent of son killer trait in the parasitic wasp, *Nasonia mtripennis*. *Int J Syst Bacteriol* **41:** 563-565.

Gosalbes MJ, Lamelas A, Moya A, Latorre A. (2008). The striking case of tryptophan provision in the cedar aphid *Cinara cedri*. *J Bacteriol* **190:** 6026-6029.

Gruwell ME, Hardy NB, Gullan PJ, Dittmar K. (2010). Evolutionary relationships among primary endosymbionts of the mealybug subfamily phenacoccinae (hemiptera: Coccoidea: Pseudococcidae). *Appl Environ Microbiol* **76:** 7521-7525.

Grünwald S, Pilhofer M, Höll W. (2010). Microbial associations in gut systems of wood- and bark-inhabiting longhorned beetles [Coleoptera: Cerambycidae]. *Syst Appl Microbiol* **33:** 25-34.

Hansen AK, Moran NA. (2011). Aphid genome expression reveals host-symbiont cooperation in the production of amino acids. *Proc Natl Acad Sci USA* **108:** 2849-2854.

Hansen AK, Vorburger C, Moran NA. (2012). Genomic basis of endosymbiont-conferred protection against an insect parasitoid. *Genome Res* **22:**106-14.

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. *J Mol Evol* **47:** 52-61.

Herbeck JT, Degnan PH, Wernegreen JJ. (2005). Nonhomogeneous model of sequence evolution indicates independent origins of primary endosymbionts within the enterobacteriales (alpha-proteobacteria). *Mol Biol Evol* **22:** 520-532.

Hosokawa T, Nikoh N, Koga R, Satô M, Tanahashi M, Meng XY, Fukatsu T. (2012). Reductive genome evolution, host-symbiont co-speciation and uterine transmission of endosymbiotic bacteria in bat flies. *ISME J* **6:** 577-87.

Hypša V, Dale C. (1997). In vitro culture and phylogenetic analysis of "*Candidatus* Arsenophonus triatominarum," an intracellular bacterium from the triatomine bug, *Triatoma infestans*. *Int J Syst Bacteriol* **47:** 1140-1144.

Husník F, Chrudimský T, Hypša V. (2011). Multiple origin of endosymbiosis within the Enterobacteriaceae (gamma-Proteobacteria): convergency of complex phylogenetic approaches. *BMC Biology* **9:** 87.

Kaiwa N, Hosokawa T, Kikuchi Y, Nikoh N, Meng XY, Kimura N, Ito M, Fukatsu T. (2010). Primary gut symbiont and secondary, *Sodalis*-allied symbiont of the Scutellerid stinkbug *Cantao ocellatus*. *Appl Environ Microbiol* **76:** 3486-3494.

Kikuchi Y, Fukatsu T. (2003). Diversity of *Wolbachia* endosymbionts in heteropteranbugs. *Appl Environ Microbiol* **69:** 6082-6090

Kirkness EF, Haas BJ, Sun W, Braig HR, Perotti MA, Clark JM, Lee SH, Robertson HM, Kennedy RC, Elhaik E, Gerlach D, Kriventseva EV, Elsik CG, Graur D, Hill CA, Veenstra JA, Walenz B, Tubío JM, Ribeiro JM, Rozas J, Johnston JS, Reese JT, Popadic A, Tojo M, Raoult D, Reed DL, Tomoyasu Y, Kraus E, Mittapalli O, Margam VM, Li HM, Meyer JM, Johnson RM, Romero-Severson J, Vanzee JP, Alvarez-Ponce D, Vieira FG, Aguadé M, Guirao-Rico S, Anzola JM, Yoon KS, Strycharz JP, Unger MF, Christley S, Lobo NF, Seufferheld MJ, Wang N, Dasch GA, Struchiner CJ, Madey G, Hannick LI, Bidwell S, Joardar V, Caler E, Shao R, Barker SC, Cameron S, Bruggner RV, Regier A, Johnson J, Viswanathan L, Utterback TR, Sutton GG, Lawson D, Waterhouse RM, Venter JC, Strausberg RL, Berenbaum MR, Collins FH, Zdobnov EM, Pittendrigh BR. (2010). Genome sequences of the human body louse and its primary endosymbiont provide insights into the permanent parasitic lifestyle. *Proc Natl Acad Sci USA* **107:** 12168-12173.

Lake JA. (1994). Reconstructing evolutionary trees from DNA and protein sequences: paralinear distances. *Proc Natl Acad Sci USA* **91:** 1455-1459.

Lambert JL, Moran NA. (1998). Deleterious mutations destabilize ribosomal RNA of endosymbionts. *Proc Natl Acad Sci USA* **95:** 4458-4462.

Lamelas A, Perez-Brocal V, Gomez-Valero L, Gosalbes MJ, Moya A, Latorre A. (2008). Evolution of the secondary symbiont "*Candidatus* Serratia symbiotica" in aphid species of the subfamily Lachninae. *Appl Environ Microbiol* **74**: 4236-4240.

Lartillot N, Philippe H. (2004). A Bayesian mixture model for across-site heterogeneities in the amino-acid replacement process. *Mol Biol Evol* **21:**1095-1109.

Lartillot N, Lepage T, Blanquart S. (2009). PhyloBayes 3: a Bayesian software package for phylogenetic reconstruction and molecular dating. *Bioinformatics* **25:** 2286-2288.

Lerat E, Daubin V, Moran NA. (2003). From gene trees to organismal phylogeny in prokaryotes: the case of the γ-Proteobacteria. *PLoS Biol* **1**: e19.

Lin CK, Hung CL, Chiang YC, Lin CM, Tsen HY. (2004). The sequence heterogenicities among 16S rRNA genes of *Salmonella* serovars and the effects on the specificity of the primers designed. *Int J Food Microbiol* **96:** 205-214.

Lo N, Casiraghi M, Salati E, Bazzocchi C, Bandi C. (2002). How many *Wolbachia* supergroups exist? *Mol Biol Evol* **19:** 341-346.

Lockhart PJ, Steel MA, Hendy MD, Penny D. (1994). Recovering evolutionary trees under a more realistic model of sequence. *Mol Biol Evol* **11:** 605-612.

Marchandin H, Teyssier C, de Buochberg MS, 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. *Microbiol-Sgm* **149:** 1493-1501.

Martínez-Torres D, Buades C, Latorre A, Moya A, (2001). Molecular systematics of aphids and their primary endosymbionts. *Mol Phylogenet Evol* **20:** 437-449.

McCutcheon JP, McDonald BR, Moran NA. (2009). Origin of an alternative genetic code in the extremely small and GC-rich genome of a bacterial symbiont. *PLoS Genet* **5:** e1000565.

McCutcheon JP, Moran NA. (2010). Functional convergence in reduced genomes of bacterial symbionts spanning 200 million years of evolution. *Genome Biol Evol* **2:** 708-718.

McCutcheon JP, von Dohlen CD. (2011). An interdependent metabolic patchwork in the nested symbiosis of mealybugs. *Curr Biol* **21:** 1366-1372.

McCutcheon JP, Moran NA. (2011). Extreme genome reduction in symbiotic bacteria. *Nat Rev Microbiol* **10:** 13-26.

McLysaght A, Hokamp K, Wolfe KH. (2002). Extensive genomic duplication during early chordate evolution. *Nat Genet* **31:** 200-204.

Moran NA, von Dohlen CD, Baumann P. (1995). Faster evolutionary rates in endosymbiotic bacteria than in cospeciating insect hosts. *J Mol Evol* **41:** 727-731.

Moran NA. (1996). Accelerated evolution and Muller's ratchet in endosymbiotic bacteria. *Proc Natl Acad Sci USA* **93:** 2873-2878.

Moran NA, Telang A. (1998). Bacteriocyte-associated endosymbionts in insects. *Bioscience* **48:** 295-304.

Moran NA. (2006). Symbiosis. *Curr Biol* **16:** R866-71.

Moran NA, McCutcheon JP, Nakabachi A. (2008). Genomics and evolution of heritable bacterial symbionts. *Annu Rev Genet* **42:** 165-190.

Moran NA, Jarvik T. (2010). Lateral transfer of genes from fungi underlies carotenoid production in aphids. *Science* **328:** 624-627.

Mouton L, Thierry M, Henri H, Baudin R, Gnankine O, Reynaud B, Zchori-Fein E, Becker N, Fleury F, Delatte H. (2012). Evidence of diversity and recombination in *Arsenophonus* symbionts of the *Bemisia tabaci* species complex. *BMC Microbiology* **12**: S10.

Moya A, Pereto J, Gil R, Latorre A. (2008). Learning how to live together: genomic insights into prokaryote-animal symbioses. *Nat Rev Genet* **9:** 218-229.

Munson MA, Baumann L, Baumann P. (1993). *Buchnera aphidicola* (a prokaryotic endosymbiont of aphids) contains a putative 16S rRNA operon unlinked to the 23S rRNA-encoding gene: sequence determination and promoter and terminator analysis. *Gene* **137:** 171-178.

Nakayama K, Yamashita A, Kurokawa K, Morimoto T, Ogawa M, Fukuhara M, Urakami H, Ohnishi M, Uchiyama I, Ogura Y, Ooka T, Oshima K, Tamura A, Hattori M, Hayashi T. (2008). The whole-genome sequencing of the obligate intracellular bacterium *Orientia tsutsugamushi* revealed massive gene amplification during reductive genome evolution. *DNA Res* **15:** 185-199.

Nakabachi A, Shigenobu S, Sakazume N, Shiraki T, Hayashizaki Y, Carninci P, Ishikawa H, Kudo T, Fukatsu T. (2005). Transcriptome analysis of the aphid bacteriocyte, the symbiotic host cell that harbors an endocellular mutualistic bacterium, Buchnera. *Proc Natl Acad Sci USA* **102:** 5477-82.

Naum M, Brown EW, Mason-Gamer RJ. (2008). Is 16S rDNA a reliable phylogenetic marker to characterize relationships below the family level in the Enterobacteriaceae? *J Mol Evol* **66:** 630-642.

Nikoh N, McCutcheon JP, Kudo T, Miyagishima SY, Moran NA, Nakabachi A. (2010). Bacterial genes in the aphid genome: absence of functional gene transfer from *Buchnera* to its host. *PLoS Genet* **6:** e1000827.

Nogge G. (1981). Significance of symbionts for the maintenance of an optimal nutritional state for successful reproduction in hematophagous arthropods. *Parasitology* **82:** 299-304.

Nováková E, and Hypša V. (2007). A new *Sodalis* lineage from bloodsucking fly *Craterina melbae* (Diptera, Hippoboscoidea) originated independently of the tsetse flies symbiont *Sodalis glossinidius*. *Fems Microbiol Letters* **269:** 131-135.

Ortiz-Rivas B, Martínez-Torres D. (2010). Combination of molecular data support the existence of three main lineages in the phylogeny of aphids (Hemiptera: Aphididae) and the basal position of the subfamily Lachninae. *Mol Phylogenet Evol* **55:** 305-317.

Ochman H, Davalos LM. (2006). The nature and dynamics of bacterial genomes. *Science* **311:** 1730-1733.

Pei AY, Oberdorf WE, Nossa CW, Agarwal A, Chokshi P, Gerz EA, Jin Z, Lee P, Yang L, Poles M, Brown SM, Sotero S, Desantis T, Brodie E, Nelson K, Pei Z. (2010). Diversity of 16S rRNA genes within individual prokaryotic genomes. *Appl Environ Microbiol* **76:** 3886-97.

Perlman SJ, Hunter MS, Zchori-Fein E. (2006). The emerging diversity of *Rickettsia*. *Proc Biol Sci* **273:** 2097-2106.

Rocha EP, Danchin A. (2002). Base composition bias might result from competition for metabolic resources. *Trends Genet* **18:** 291-294.

Russell JA, LaTorre AL, Sabater-Munoz B, Moya A, Moran NA. (2003). Side-stepping secondary symbionts: widespread horizontal transfer across and beyond the Aphidoidea. *Mol Ecol* **12:** 1061-1075.

Salar P, Sémétey O, Danet JL, Boudon-Padieu E, Foissac X. (2009). '*Candidatus* Phlomobacter fragariae' and the proteobacterium associated with the low sugar content syndrome of sugar beet are related to bacteria of the *Arsenophonus* clade detected in hemipteran insects. *Eur J Plant Pathol* **126:** 123-127.

Sameshima S, Hasegawa E, Kitade O, Minaka N, Matsumoto T. (1999). Phylogenetic comparison of endosymbionts with their host ants based on molecular evidence. *Zool Sci* **16:** 993-1000.

Sauer C, Stackebrandt E, Gadau J, Holldobler B, Gross R. (2000). Systematic relationships and cospeciation of bacterial endosymbionts and their carpenter ant host species: proposal of the new taxon *Candidatus* Blochmannia gen. nov. *Int J Syst Evol Microbiol* **50:** 1877-1886.

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.

Sirviö A, Pamilo P. (2010). Multiple endosymbionts in populations of the ant *Formica cinerea*. *BMC Evol Biol* **10:**335.

Shigenobu S, Watanabe H, Hattori M, Sakaki Y, Ishikawa H. (2000). Genome sequence of the endocellular bacterial symbiont of aphids *Buchnera* sp APS. *Nature* **407:** 81-86.

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*. *Syst Appl Microbiol* **31:** 88-100.

Tamas I, Klasson L, Näslund K, Canbäck B, Eriksson AS, Wernegreen JJ, Sandström JP, Moran NA, Andersson SGE. (2002). 50 million years of genomic stasis in endosymbiotic bacteria. *Science* **296:** 2376-2379.

Tamura K. (1992). Estimation of the number of nucleotide substitutions when there are strong transition-transversion and G+C-content biases. *Mol Biol Evol* **9:** 678-687.

Thao ML, Moran NA, Abbot P, Brennan EB, Burckhardt DH, Baumann P. (2000). Cospeciation of psyllids and their primary prokaryotic endosymbionts. *Appl Environ Microbiol* **66:** 2898-2905.

Thao MLL, Baumann P. (2004). Evidence for multiple acquisition of *Arsenophonus* by whitefly species (Sternorrhyncha : Aleyrodidae). *Curr Microbiol* **48:** 140-144.

The International Aphid Genomics Consortium. (2010). Genome Sequence of the Pea Aphid *Acyrthosiphon pisum*. *PLoS Biol* **8:** e1000313.

Toh H, Weiss BL, Perkin SA, Yamashita A, Oshima K, Hattori M, Aksoy S. (2006). Massive genome erosion and functional adaptations provide insights into the symbiotic lifestyle of *Sodalis glossinidius* in the tsetse host. *Genome Res* **16:** 149-156.

Toju H, Hosokawa T, Koga R, Nikoh N, Meng XY, Kimura N, Fukatsu T. (2010). "*Candidatus* Curculioniphilus buchneri," a novel clade of bacterial endocellular symbionts from weevils of the genus *Curculio*. *Appl Environ Microbiol* **76:** 275-282.

Toju H, Fukatsu T. (2011). Diversity and infection prevalence of endosymbionts in natural populations of the chestnut weevil: relevance of local climate and host plants. *Mol Ecol* **20:** 853-868.

Treangen TJ, Salzberg SL. (2011). Repetitive DNA and next-generation sequencing: computational challenges and solutions. *Nat Rev Genet* **13:** 36-46.

Trowbridge RE, Dittmar K, Whiting MF. (2006). Identification and phylogenetic analysis of *Arsenophonus*- and *Photorhabdus*-type bacteria from adult Hippoboscidae and Streblidae (Hippoboscoidea). *J Invertebr Pathol* **91:** 64-68.

van Ham RC, Kamerbeek J, Palacios C, Rausell C, Abascal F, Bastolla U, Fernández JM, Jiménez L, Postigo M, Silva FJ, Tamames J, Viguera E, Latorre A, Valencia A, Morán F, Moya A. (2003). Reductive genome evolution in *Buchnera aphidicola*. *Proc Natl Acad Sci USA* **100:** 581-6.

von Dohlen CD, Moran NA. (2000). Molecular data support a rapid radiation of aphids in the Cretaceous and multiple origins of host alternation. *Biol J Linn Soc* **71:** 689-717.

Werren JH, Skinner SW, Huger AM. (1986). Male killing bacteria in a parasitic wasp. *Science* **231:** 990-992.

Williams KP, Gillespie JJ, Sobral BW, Nordberg EK, Snyder EE, Shallom JM, Dickerman AW. (2010). Phylogeny of Gammaproteobacteria. *J Bacteriol* **192:** 2305-2314.

Wu, D., S.C. Daugherty, S.E. Van Aken, G.H. Pai, K.L. Watkins, H. Khouri, L.J. Tallon, J.M. Zaborsky, H.E. Dunbar, P.L. Tran, N.A. Moran, and J.A. Eisen. 2006. Metabolic complementarity and genomics of the dual bacterial symbiosis of sharpshooters. *Plos Biology* 4: 1079-1092.

Wu D, Daugherty SC, Van Aken SE, Pai GH, Watkins KL, Khouri H, Tallon LJ, Zaborsky JM, Dunbar HE, Tran PL, Moran NA, Eisen JA. (2009). A phylogeny-driven genomic encyclopaedia of bacteria and archaea. *Nature* **462:** 1056-1060.

Yap WH, Zhang ZS, 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. *J Bacteriol* **181:** 5201-5209.

Zhou X, Lin Z, Ma H. (2010). Phylogenetic detection of numerous gene duplications shared by animals, fungi and plants. *Genome Biol* **11:** R38.

Zreik L, Bové JM, 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*'. *Int J Syst Bacteriol* **48:** 257-261.

Manuscript #1

Hypša, V. and **Nováková, E.** (2009): Insect symbionts and molecular phylogenetics. In: Insect Symbiosis, Bourtzis, K and Miller, T (eds), CRC press, pp. 1-32

Insect symbionts and molecular phylogenetics. In: Insect Symbiosis, Bourtzis, K and Miller, T (eds),

Hypša, V. and **Nováková, E.**

CRC press (2009), pp. 1-32

Printed in the United States of America on acid-free paper
10 9 8 7 6 5 4 3 2 1
International Standard Book Number-13: 978-1-4200-6410-0 (Hardcover)

Manuscript #2

Nováková, E., Hypša, V. and Moran, N.A. (2009): *Arsenophonus*, an emerging clade of intracellular symbionts with a broad host distribution. *Bmc Microbiology* 9, DOI: 10.1186/1471-2180-9-143

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

Nováková E., Hypša V., Moran N.A.

BMC Microbiology (2009) 9:143

Background: The genus *Arsenophonus* is a group of symbiotic, mainly insect-associated bacteria with rapidly increasing number of records. It is known from a broad spectrum of hosts and symbiotic relationships varying from parasitic son-killers to coevolving mutualists. The present study extends the currently known diversity with 34 samples retrieved mainly from hippoboscid (Diptera: Hippoboscidae) and nycteribiid (Diptera: Nycteribiidae) hosts, and investigates phylogenetic relationships within the genus.

Results: The analysis of 110 *Arsenophonus* sequences (incl. *Riesia* and *Phlomobacter*), provides a robust monophyletic clade, characterized by unique molecular synapomorphies. On the other hand, unstable inner topology indicates that complete understanding of *Arsenophonus* evolution cannot be achieved with 16S rDNA. Moreover, taxonomically restricted *Sampling* matrices prove sensitivity of the phylogenetic signal to sampling; in some cases, *Arsenophonus* monophyly is disrupted by other symbiotic bacteria. Two contrasting coevolutionary patterns occur throughout the tree: parallel host-symbiont evolution and the haphazard association of the symbionts with distant hosts. A further conspicuous feature of the topology is the occurrence of monophyletic symbiont lineages associated with monophyletic groups of hosts without a co-speciation pattern. We suggest that part of this incongruence could be caused by methodological artifacts, such as intragenomic variability.

Conclusion: The sample of currently available molecular data presents the genus *Arsenophonus* as one of the richest and most widespread clusters of insect symbiotic bacteria. The analysis of its phylogenetic lineages indicates a complex evolution and apparent ecological versatility with switches between entirely different life styles. Due to these properties, the genus should play an important role in the studies of evolutionary trends in insect intracellular symbionts. However, under the current practice, relying exclusively on 16S rRNA sequences, the phylogenetic analyses are sensitive to various methodological artifacts that may even lead to description of new *Arsenophonus* lineages as independent genera (e.g. *Riesia* and *Phlomobacter*). The resolution of the evolutionary questions encountered within the *Arsenophonus* clade will thus require identification of new molecular markers suitable for the low-level phylogenetics.

Manuscript #3

Nováková, E. and Moran, N.A. (2012): Diversification of genes for carotenoid biosynthesis in aphids following an ancient transfer from a fungus. *Molecular Biology and Evolution* 29:313-323, DOI: 10.1093/molbev/msr206

Diversification of genes for carotenoid biosynthesis in aphids following an ancient transfer from a fungus.

Nováková, E. and Moran, N.A.

Molecular Biology and Evolution (2012) 29:313-323

The pea aphid genome was recently found to harbor genes for carotenoid biosynthesis, reflecting an ancestral transfer from a fungus. To explore the evolution of the carotene desaturase gene family within aphids, sequences were retrieved from a set of 34 aphid species representing numerous deeply diverging lineages of aphids, and analyzed together with fungal sequences retrieved from databases. All aphids have at least one copy of this gene, and some aphid species have up to 7, whereas fungal genomes consistently have a single copy. The closest relatives of aphids, adelgids, also have carotene desaturase; these sequences are most closely related to those from aphids, supporting a shared origin from a fungal to insect transfer predating the divergence of adelgids and aphids. Likewise, all aphids, and adelgids, have carotenoid profiles that are consistent with their biosynthesis using the acquired genes of fungal origin, rather than derivation from food plants. The carotene desaturase was acquired from a fungal species outside of Ascomycota or Basidiomycota, and closest to Mucoromycotina among sequences available in databases. In aphids, an ongoing pattern of gene duplication is indicated by the presence of both anciently and recently diverged paralogs within genomes, and by the presence of a high frequency of pseudogenes that appear to be recently inactivated. Recombination among paralogs is evident, making analyses of patterns of selection difficult, but tests of selection for a non-recombining region indicates that duplications tend to be followed by bouts of positive selection. Species of Macrosiphini, which often show color polymorphisms, typically have a larger number of desaturase copies relative to other species sampled in the study. These results indicate that aphid evolution has been accompanied by ongoing evolution of carotenogenic genes, which have undergone duplication, recombination, and occasional positive selection to yield a wide variety of carotenoid profiles in different aphid species.

Manuscript #4

Wilkes, T.E, Duron, O., Darby, A.C., Hypša, V., **Nováková E.** and Hurst, G.D.D. (2011): The genus *Arsenophonus*. In: E Zchori-Fein & K Bourtzis ed(s). Manipulative Tenants. New York, CRC Press, pp. 225-241

The genus *Arsenophonus*. In: E Zchori-Fein & K Bourtzis ed(s). Manipulative Tenants. New York,

Wilkes, T.E, Duron, O., Darby, A.C., Hypša, V., **Nováková E.** and Hurst, G.D.D.

CRC Press (2011), pp. 225-241

Version Date: 20110705
International Standard Book Number-13: 978-1-4398-2750-5 (eBook - PDF)

Manuscript #5

Chrudimský, T., Husník, F., **Nováková, E.** and Hypša, V. *Candidatus* Sodalis melophagi sp. nov.: phylogenetically independent comparative model to the tsetse fly symbiont *Sodalis glossinidius.* MS submitted to PLOS ONE: under major revision.

Candidatus* Sodalis melophagi sp. nov.: phylogenetically independent comparative model to the tsetse fly symbiont *Sodalis glossinidius.

Chrudimský T., Husník F., **Nováková E.**, Hypša V.

Plos One (resubmitted)

Bacteria of the genus *Sodalis* live in symbiosis with various groups of insects. The best known member of this group, a secondary symbiont of tsetse flies *Sodalis glossinidius*, has become one of the most important models in investigating establishment and evolution of insect-bacteria symbiosis. It represents a bacterium in the early/intermediate state of the transition towards symbiosis, which allows for exploring such interesting topics as: usage of secretory systems for entering the host cell, tempo of the genome modification, and metabolic interaction with a coexisting primary symbiont. In this study, we describe a new *Sodalis* species which provides a useful comparative model to the tsetse symbiont. It lives in association with *Melophagus ovinus*, an insect related to tsetse flies, and resembles *S. glossinidius* in several important traits. Similar to *S. glossinidius*, it cohabits the host with another symbiotic bacterium, the bacteriome-harbored primary symbiont of the genus *Arsenophonus*. As a typical secondary symbiont, *Candidatus* Sodalis melophagi infects various host tissues, including bacteriome. Here, we provide basic morphological and molecular characteristics of the symbiont and show that these traits also correspond to the early/intermediate state of the evolution towards symbiosis. Particularly, we demonstrate the ability of the bacterium to live in insect cell culture as well as in cell-free medium. We also provide basic characteristics of type three secretion system and using three reference sequences (16S rDNA, *groEL* and *spaPQR* region) we show that the bacterium branched within the genus *Sodalis*, but originated independently of the two previously described symbionts of hippoboscoids We propose the name *Candidatus* Sodalis melophagi for this new bacterium.

Manuscript #6

Nováková, E., Hypša, V., Klein, J., Foottit, R.G., von Dohlen, C.D. and Moran, N.A. Evolutionary history of aphids (Hemiptera: Aphididae): suitability of the P-symbiont *Buchnera aphidicola* DNA for reconstruction of the host phylogeny. MS prepared for publication.

Evolutionary history of aphids (Hemiptera: Aphididae): suitability of the P-symbiont *Buchnera aphidicola* DNA for reconstruction of the host phylogeny.

Nováková, E., Hypša, V., Klein, J., Foottit, R.G., von Dohlen, C.D. and Moran, N.A

MS prepared for publication.

Reliable phylogenetic reconstruction, as a background for evolutionary inference, may be difficult to achieve in some groups of organisms. Particularly, for the diverse taxa that experienced rapid diversification, lack of sufficient information may lead to inconsistent and unstable results and a low degree of resolution. Coincidentally, such rapidly diversifying taxa are often among the biologically most interesting groups. Aphids provide such a typical example. Due to rapid adaptive diversification, they feature variability in many interesting biological traits, but they are also difficult group from the phylogenetic point of view. Particularly within the family Aphididae, many interesting evolutionary questions remain unanswered due to the phylogenetic uncertainties. In this study, we show that molecular data derived from the symbiotic bacteria of the genus *Buchnera* can provide a more powerful tool then the aphid-derived sequences. We analyze 255 *Buchnera* gene sequences and compare the resulting trees to the phylogenies previously retrieved from the aphid sequences. We show that the host and symbiont data do not contradict in any major phylogenetic elements. Also, we demonstrate that the symbiont-derived phylogenies support some previously questionable relationships and even reveal new aspects of the aphid phylogeny and evolution.

Manuscript #7

Nováková, E., Hypša, V., Moran, N.A., Hurst, G.D.D. and Darby, A.C.: Comparative genomics of inherited *Arsenophonus* bacteria across the symbiotic continuum. MS in preparation.

A symbiotic process: The evolution of inherited symbiont genomes. MS in preparation.

Nováková, E., Hypša, V., Moran, N.A., Hurst, G.D.D., Lehane M. and Darby, A.C.

MS in preparation.

Background. Arthropods commonly have symbiotic relationship with maternally inherited bacteria. These relationships are diverse, and include obligate mutualism, facultative benefit to the host, and reproductive parasitism. Our understanding of patterns of genomic evolution under different forms of symbiosis has generally relied on contrast to either free living relatives or distantly related symbionts. The genus *Arsenophonus* is unusual in containing symbionts whose interactions with their hosts are diverse in themselves, and therefore allows comparison of genome form between closely related symbionts that share a recent common ancestor but display very different interactions with their host.

Results. We here report properties of draft genome sequences of the secondary symbiont *Arsenophonus triatominarum* and the obligate *Arsenophonus* symbiont from the bloodfeeding diptera *Melophagus ovinus*. We compare these genomes to those previously reported for *A. nasoniae*, a male-killing bacterium that retains infectious transfer ability, and *R. pediculicola*, a louse symbiont that is obligately required by its host. We observed that the genome of *A. nasoniae* was largest, had the most complex metabolism, and largely intact secretion system. *A. triatominarum* had degenerated compared to the reproductive parasite, but encoded substantially greater number of pathways than the obligate symbionts. It is clear that in this clade there is a pattern of degeneration on evolution towards mutualism.

*Conclusions.*We demonstrate that *Arsenophonus* lineages adapted to different lifestyles also show substantial differences in important genomic traits. Summarized, these differences reveal general tendencies along the hypothetical evolutionary continuum from almost “standard” bacterial genomes in parasites and facultative symbionts to highly modified and degenerated genomes in obligate mutualists. The most typical manifestation of these tendencies are massive gene loss, compositional bias and economization of the genome with decreasing frequency of horizontal gene transfer. In contrast to these parallel processes, some traits are not clearly coupled with the transition towards symbiosis (i.e. the mutation rates together with selection) and their investigation will require data sets extended with additional lineages of *Arsenophonus*.

CURRICULUM VITAE

Eva Novakova, M.Sc.

University of South Bohemia, Faculty of Science and Laboratory of Molecular Phylogeny and Evolution of Parasites, Institute of Parasitology, Biology Centre ASCR, v.v.i., Branisovska 31, 370 05 Ceske Budejovice, Czech Republic
tel.: +420-38-777-5409
e-mail: novaeva@paru.cas.cz

Date of Birth: October 18, 1984
Place of Birth: Rokycany, Czech Republic

Education:
***2006*:** Bachelor degree in Biology. Faculty of Biological Sciences, University of South Bohemia
***2008*:** Master degree in Parasitology. Faculty of Science, University of South Bohemia
2008 - present: Doctoral student of Parasitology. Faculty of Science, University of South Bohemia

Professional Experience:
2008 - present: Research Scientist, Biology Centre, Institute of Parasitology, Academy of Sciences of the Czech Republic, České Budějovice, Czech Republic and Faculty of Science, University of South Bohemia, České Budějovice, Czech Republic

Abroad stays and fellowships:
***2007*:** Visiting Researcher, Laboratory of Professor Nancy Moran, The University of Arizona, Tucson, AZ, USA
2010: COST Short Term Scientific Mission (STSM), Laboratory of Professor Gregory Hurst, University of Liverpool, Great Britain
***2010/2011*:** Fulbright Grantee, Laboratory of Professor Nancy Moran, Yale University, New Haven, CT, USA

Teaching and mentoring experience:
2008 – 2010: Supervising of high school students and bachelor degree students enrolled in Department of Parasitology, Faculty of Science, University of South Bohemia.

2008 – 2 010: Teaching assistant, Faculty of Science, University of South Bohemia, České Budějovice. *Biochemistry laboratory for Biology and Environmental studies*; *Biochemistry Laboratory 1.*

Successful grant applications or participations:
GAJU 110/2010/P (2010, Nováková E.) Evolutionary patterns and phylogenetic relationships among *Arsenophonus* symbionts associated with triatomine bugs. Applicant

GA ČR P505/12/1620 (2012 – 2015, Hypša V.): Population genetics, demography and molecular evolution in interspecific associations: comparative study of two complex parasitic/symbiotic systems. Member of the research team.

GAČR P505/10/1401 (2010 – 2013, Hypša V.) Molecular evolution of *Arsenophonus*, an emerging group of symbiotic bacteria with a broad host distribution. Member of the research team.

GAJU 135/2010/P (2010 – 2013, Hypša V.) „Vznik diverzity modelových skupin živočichů: od fylogeografie po fylogenezi". Member of the research team.

GAAV IAA601410708 (2007-2009, Hypša V.) An effect of rDNA intragenomic variability on coevolutionary analysis in insect-symbiont associations. Member of the research team.

Publications:

Novakova E., and Hypsa V., 2007. A new *Sodalis* lineage from bloodsucking fly *Craterina melbae* (Diptera, Hippoboscoidea) originated independently of the tsetse flies symbiont *Sodalis glossinidius*. FEMS Microbiology Letters, 269:131-135.

Hypsa V., and **Novakova E**., 2008. Insect symionts and molecular phylogenetics. In: Insect Symbiosis, Volume 3. Eds. K. Bourtzis and T. A Miller. CRC Press, pp 1-22.

Novakova E., Hypsa V., Moran N.A., 2009. *Arsenophonus*, an emerging clade of intracellular symbionts with a broad host distribution. BMC Microbiology, 9:143.

Wilkes, T.E, Duron, O., Darby, A.C., Hypša, V., **Nováková E**. and Hurst, G.D.D., 2011. The genus *Arsenophonus*. In: E Zchori-Fein & K Bourtzis ed(s). Manipulative Tenants. New York, CRC Press, pp 225-241.

Nováková, E. and Moran, N.A., 2012. Diversification of genes for carotenoid biosynthesis in aphids following an ancient transfer from a fungus. Molecular Biology and Evolution. 29:313-323

Invited Lectures:

Nováková E., Hypša V. 2010. Evolution and diversity of the genus *Arsenophonus*,the largest group of inherited symbionts within gamma Proteobacteria. COST, Action FA0701, Bad Bevensen, Germany.

Nováková E., Moran N.A. July 2012. Evolutionary significance of horizontal gene transfer in aphids. 7th International Symbiosis Society Congress "The earth's vast symbiosphere". Krakow, Poland.