UPPSALA
UNIVERSITET

Wolbachia genes in *Drosophila ananassae*

A study of laterally transferred bacterial genes in a fruit fly genome

Sofi Bäcklin

Degree project in biology, Master of science (2 years), 2010
Examensarbete i biologi 45 hp till masterexamen, 2010
Biology Education Centre and Department of Evolution, Genomics and Systematics, Molecular Evolution, Uppsala University
Supervisor: Lisa Klasson

Summary

Obligate intracellular bacteria from the genus *Wolbachia* are vertically transmitted reproductive parasites that infect a great number of insect species. *Wolbachia* can cause various alterations to the host's reproductive system, thereby increasing their own rate of transmission to the next generation. One such alteration is cytoplasmic incompatibility, which in its simplest form makes crosses between infected males and uninfected females sterile. If females are also infected, the effects of the cytoplasmic incompatibility are "rescued", making the cross fertile. The fruit fly *Drosophila ananassae* is infected with the *w*Ana strain of *Wolbachia*, known to cause cytoplasmic incompatibility.

In 2007, Dunning Hotopp *et al.* (Science **317**:1753–1756, 2007) reported the presence of several *Wolbachia* genes inserted into the genomes of four strains of *D. ananassae*. It seemed like almost the entire bacterial chromosome had been inserted. It is not known whether the inserted genes are functional or not, and their evolutionary history is unclear.

There were two main aims in the current study. The first aim was to investigate whether the inserted genes had any functional role in the induction or rescue of cytoplasmic incompatibility. This was done by carrying out crosses between *D. ananassae* strains of different infection and insert status and measuring the mortality of the resulting offspring. The second aim was to try to clarify the evolutionary history of the inserts. This was done by PCR-amplification and sequencing of a number of genes in the different *D. ananassae* strains. The sequences were then compared in search for differences that could be used to infer the evolutionary history.

No significant results were obtained from the study of cytoplasmic incompatibility since not enough crosses could be carried out. The sequencing results, however, showed the presence of polymorphisms in all sequenced *Wolbachia* genes in one of the *D. ananassae* strains, indicating the presence of at least two different variants of each gene. These variants could not be separated by crosses and backcrosses to a *D. ananassae* strain lacking the insert. Also, the inserted genes showed a peculiar pattern of inheritance that could not be easily explained.

The tentative conclusion from these results is that there seem to be several copies of the studied genes inserted into this *D. ananassae* strain, and that these copies are positioned close enough to each other to avoid being broken up by recombination over two generations.

Introduction

Wolbachia: discovery and brief description

Wolbachia is a genus of small intracellular alphaproteobacteria[1] belonging to the Rickettsiales order. The first description of a member of the *Wolbachia* genus was published by Hertig and Wolbach in 1924 (Hertig & Wolbach 1924). In their paper, they report the presence of a rodlike gram-negative rickettsia in the testes and ovaries of *Culex pipiens*, a very common species of mosquito. They also found the same organism in the eggs of the mosquito and, in lower amounts, in the larvae. Based on these observations, they suggested that the microorganism is vertically transmitted to all offspring via the egg, and once the larvae pupate, the microorganisms in the gonads rapidly increase in number. In 1936, Hertig gave the microorganism the name *Wolbachia pipientis* (Hertig 1936).

The *Wolbachia* genus is now known to contain many strains infecting a great number of insect species (Jeyaprakash & Hoy 2000, Hilgenboecker *et al*. 2008) and most species of pathogenic filarial nematodes (Bandi *et al*. 1998). All *Wolbachia* strains are obligate intracellular bacteria and, as Hertig & Wolbach (1924) suggested, they are vertically transmitted from generation to generation through the cytoplasm of the egg. The species in the *Wolbachia* genus are divided into eight supergroups (A–H) based on phylogenies created with, among others, the *gltA* (coding for citrate synthase), *groEL* (coding for chaperone Hsp60) and *ftsZ* (coding for a cell division protein) gene sequences (Casiraghi *et al*. 2005, Lo *et al*. 2007). All supergroups are monophyletic (Werren *et al*. 2008). The known *Wolbachia* supergroups are listed in Table 1 along with their respective hosts. In filarial nematodes, *Wolbachia* act mutualistically and seem sometimes to be required for reproduction (Bandi *et al*. 2001). In arthropods, on the other hand, *Wolbachia* are reproductive parasites, altering different aspects of the host's reproductive system leading to increase in their own transmission rate.

Table 1. *Wolbachia* supergroups and their hosts[a]

Supergroup	Host
A	Arthropods
B	Arthropods
C	Filarial nematodes
D	Filarial nematodes
E	Springtails
F	Arthropods and filarial nematodes
G	Spiders
H	Termites

[a] Data from Lo *et al*. 2007.

Cytoplasmic incompatibility and other host effects

Hardly anything was published about *Wolbachia* in the decades that followed Hertig's 1936 paper, but quite a few articles were published describing a phenomenon of reproductive incompatibility that was observed in the *Wolbachia* host species *Culex pipiens*. Laven (1951), for instance, described how certain crosses between *C. pipiens* strains from different locations

[1] Alphaproteobacteria is a class in the phylum of Proteobacteria. This class contains symbionts of animals and plants as well as some pathogens.

in Europe produced eggs of which only a small proportion, or none at all, hatched. Laven was not able to explain why this was the case, but in 1953 he concluded that the reproductive incompatibility status was inherited from mother to offspring, and he therefore, assuming that it is inherited through the egg cytoplasm, named the phenomenon cytoplasmic incompatibility (see Yen & Barr 1971 for reference).

Twenty years later, Yen & Barr (1971) published an article in Nature in which they suggested that the phenomenon of cytoplasmic incompatibility (CI), as observed in *C. pipiens*, could be caused by *Wolbachia*. They based this suggestion on their observation of large numbers of *Wolbachia pipientis* bacteria in the eggs of *C. pipiens*. Two years later, they were able to confirm this hypothesis by showing that the CI disappeared when the mosquitoes were treated with tetracycline, an antibiotic that kills the *Wolbachia* (Yen & Barr 1973).

In 1989, two papers were published in the same issue of the Journal of Invertebrate Pathology, both reporting observations of "*Wolbachia*-like organisms" (Binnington & Hoffman 1989) or "*Wolbachia*-type procaryotes [*sic*]" (Louis & Nigro 1989) in the testes and ovaries of certain strains of *Drosophila simulans*. The presence of *Wolbachia* was shown to coincide with the presence of CI.

Since then, the phenomenon of CI caused by *Wolbachia* has been well described in most insect orders including Coleoptera (beetles), Diptera (flies) and Hymenoptera (bees, wasps etc.) as well as in some mite species (Stouthamer *et al.* 1999). In the simplest case of CI, a cross between a *Wolbachia*-infected male and an uninfected female produces fertilised eggs that die. This is referred to as unidirectional CI. There are also more complex forms of CI where crosses between individuals infected with different strains of *Wolbachia* are incompatible. This is known as bidirectional CI. Also, CI may be complete (all offspring die) or partial (only a portion of the offspring die). Since *Wolbachia* is transmitted through the female line, the result of all types of CI is an increased proportion of infected offspring resulting in a quick spread of *Wolbachia* through the population (Serbus *et al*. 2008). The different types of CI are illustrated in Figure 1 (showing only complete CI).

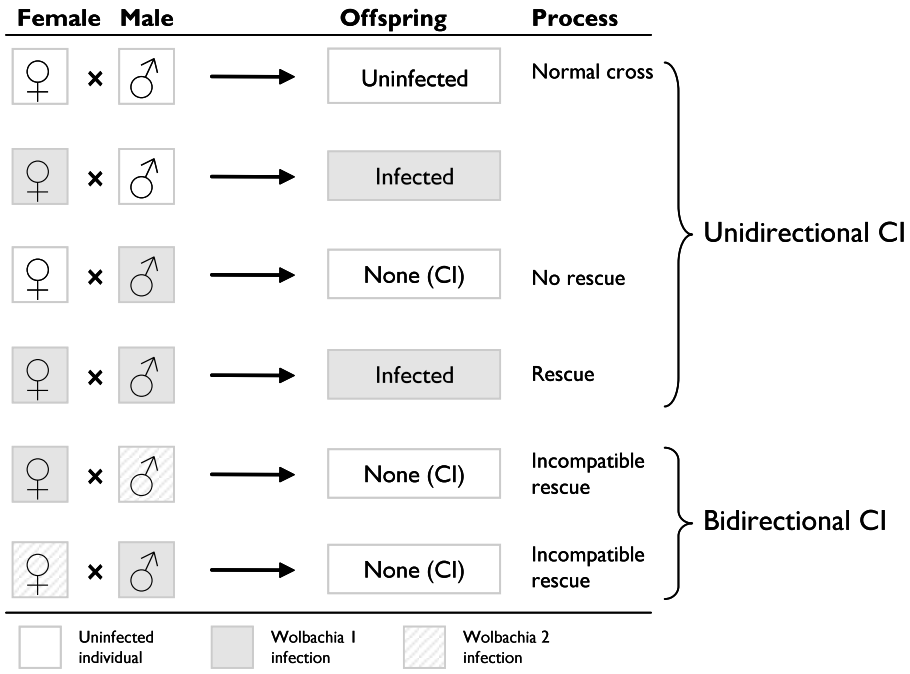

Figure 1. Schematic illustration of uni- and bidirectional CI caused by *Wolbachia*. Wolbachia 1 and 2 refer to two hypothetical strains of *Wolbachia*. Data from Sinkins & Gould 2006.

Although CI is the most common and well-studied *Wolbachia*-induced host effect (Stouthamer *et al*. 1999), the underlying molecular mechanisms causing this phenomenon are yet to be clarified. The so called mod/resc model proposed by Werren (1997) is a widely used model that does not imply very much about the mechanisms of CI but provides a good framework for understanding the principles. In this model, the sperm of an infected male is modified (mod) by *Wolbachia* before maturation so that a cross with an uninfected female becomes incompatible. If the male instead mates with an infected female, the egg carries a rescue function (resc) that is able to suppress the mod function, thus resulting in viable, but infected, offspring (see Figure 1).

Due to space limitations, I am not able to go into any detail about the observations that have been made in CI embryos, but the topic is reviewed by Serbus *et al*. (2008). In short, it seems like the *Wolbachia* mod-function somehow alters the paternal chromatin causing abnormal condensation during metaphase of the cell cycle. It also seems like there is a delay in the activation of Cdk1 (cyclin-dependent kinase 1) in the male pronucleus compared to the female, causing disruptions of the cell cycle.

In addition to CI, *Wolbachia* can cause other effects on the host's reproductive system. These include male killing (where male offspring from infected females die), feminisation of genetic males and induction of parthenogenesis (where unfertilised eggs develop into females) (Werren *et al*. 2008). As with CI, all of these effects result in an increased transmission rate of *Wolbachia* to the next generation.

Wolbachia genomics

The genome of *w*Mel, the *Wolbachia* strain infecting *Drosophila melanogaster*, was sequenced and assembled in 2004 (Wu *et al*. 2004). The *w*Mel genome is 1.27 Mbp in size, of which 85.4% is coding DNA. When this genome was analysed, it was found to contain large amounts of repetitive DNA as well as DNA that seemed to be mobile genetic elements, features that had not previously been found in the genomes of intracellular bacteria (Wu *et al*. 2004). Part of the putative mobile genetic elements are insertion sequence (IS) elements. The *w*Mel genome was also found to contain three prophage elements, and it shows signs of rearrangements and duplications in addition to the genome reduction that has taken place (Wu *et al*. 2004).

The genome of *w*Pip, infecting *Culex pipiens*, is larger than that of *w*Mel (1.48 Mbp compared to 1.27Mbp) and there are numerous rearrangements between the two genomes (Klasson *et al*. 2008). *w*Pip also contains large numbers of repeats and mobile genetic elements, including 116 IS-elements.

The genome of *w*Ri, infecting *Drosophila simulans*, is also larger in size compared to the *w*Mel genome (1.45 Mbp) and a much higher fraction of its genome is made up of repeated sequences (Klasson *et al*. 2009b). *w*Ri also contains many IS-elements and four prophage segments.

The features of repetitive DNA and mobile genetic elements in *w*Mel and *w*Ri are compatible with recombination (Baldo *et al*. 2006) and both intra- and intergenic recombination has been shown to take place in *Wolbachia* housekeeping genes as well as in other genes and in IS-elements (Baldo *et al*. 2006).

Lateral gene transfer from *Wolbachia* to the host

Lateral gene transfer (LGT) is the transfer and incorporation of genetic material from the genome of one organism to the genome of another through processes other than inheritance (which is defined as vertical gene transfer). LGT is very well documented and frequent in bacteria, where genetic exchange can take place through conjugation, transduction or transformation. The current consensus opinion is that LGT has played a great role in the evolution of bacteria, although the exact extent of LGT between bacteria is difficult to determine (Gribaldo & Brochier 2009). One very well known example illustrating the importance of LGT in bacteria is the spread of antibiotic resistance between different bacteria strains (Aminov & Mackie 2007).

LGT in eukaryotes is much less reported than in bacteria and, with the exception for transfer from organelles of endosymbiotic origin to eukaryotic nuclei, it was long considered exceptionally rare (Andersson 2005, Keeling & Palmer 2008). In the last few years, however, reports of LGT from bacteria to eukaryotes have increased steadily (Keeling & Palmer 2008). Most of these reports involve LGT from bacteria to protists (unicellular eukaryotes). Confirmed cases of LGT in animals are still rare, which could be explained by the fact that animals have segregated germlines into which the genetic material would have to be incorporated. *Wolbachia* and other vertically transmitted intracellular bacteria that are present in the germcells of their hosts are the most likely donors in LGT events involving animals (Blaxter 2007), and there are now several examples of *Wolbachia* genes or genome fragments that have been transferred laterally into host genomes.

In 2002 Kondo *et al*. (2002) reported the discovery of a *Wolbachia* genome fragment in the adzuki bean beetle genome. This beetle was thought to be triple-infected with three separate strains of *Wolbachia*, but Kondo *et al*. (2002) discovered that one of the strains (identified based on one gene sequence) in fact was not a bacterial strain but a fragment of a *Wolbachia* genome that had been laterally transferred to the X chromosome of the beetle. A few years later Nikoh *et al*. (2008) reported that they had been able to amplify 57 out of 205 tested *Wolbachia* genes by the polymerase chain reaction (PCR) in a tetracycline-cured[2] line of the beetle *Callosobruchus chinensis*. They also used quantitative reverse transcription (RT)-PCR to measure the expression levels of some of these genes and found them so low that their biological significance is questionable.

Another example of LGT between *Wolbachia* and their host species is the salivary gland secretion (SGS) genes in the mosquitoe *Aedes aegypti*. The SGS genes are found only in mosquitoes and have no known homologs in other eukaryotes, but they do show sequence similarity to a gene found in *w*Mel and *w*Pip. Phylogenetic analyses show that an LGT event involving *Wolbachia* and *Ae*. *aegypti* almost certainly has taken place, but the direction of the event is less clear (Klasson *et al*. 2009a, Woolfit *et al*. 2009).

Laterally transferred *Wolbachia* genes in *Drosophila ananassae*

Drosophila ananassae is a cosmopolitan[3] species of fruitfly found mostly in tropical regions. It belongs to the *ananassae* subgroup in the *melanogaster* species group within the *Drosophila* genus (Singh & Singh 2003). *D. ananassae* has been used extensively for genetic studies since the 1930s, mainly because it has a high mutability but also because it exhibits

[2] A cured strain refers to a strain that has lost its *Wolbachia* infection through tetracycline treatment.
[3] In biology, a cosmopolitan distribution means a world-wide distribution.

different unusual genetic features such as spontaneous meiotic crossing over in males (Singh 1996). *D. ananassae* is infected with a *Wolbachia* strain called *w*Ana (Salzberg *et al*. 2005) that causes partial CI (Bourtzis *et al*. 1996). The genome of *D. ananassae* from Hawaii has been sequenced with ~8x coverage and the sequences (whole-genome shotgun sequences) are publicly available from the NCBI Trace Archive (http://www.ncbi.nlm.nih.gov/Traces/trace.cgi). The genome of *w*Ana was assembled in 2005 (Salzberg *et al*. 2005) from *Wolbachia* sequences retrieved from the *D. ananassae* traces in the NCBI Trace Archive using the already sequenced *Wolbachia* strain *w*Mel as a probe. Salzberg *et al*. (2005) also assembled the genome of *w*Sim, infecting *D. simulans*, in the same manner. *w*Ana and *w*Sim are very similar with 99.8% nucleotide sequence identity (Salzberg *et al*. 2005) and they both show high sequence similarity to *w*Ri, a *Wolbachia* strain also infecting *D. simulans* (Iturbe-Ormaetxe *et al*. 2005).

Two years later, Dunning Hotopp *et al*. (2007) reported the discovery of several *Wolbachia* genes inserted into the genome of *D. ananassae*. They had downloaded whole genome shotgun sequences for 26 arthropod and nematode species and scanned these for *Wolbachia* sequences with the specific purpose of detecting LGT events from *Wolbachia* to the host species. This way, they found several sequences containing junctions between *Wolbachia* genes and *D. ananassae* (Hawaii) retrotransposons. They then continued with a PCR assay covering 45 *Wolbachia* genes spread out in the *Wolbachia* genome. By running this assay on several tetracycline-cured strains of *D. ananassae*, they confirmed the presence of between 23 and 44 *Wolbachia* genes in four different strains of *D. ananassae*. The highest number of genes was found in the sequenced *D. ananassae* Hawaii strain (see Table 2). This also showed that large parts of the *Wolbachia* genome, if not the entire genome, had been inserted into the *D. ananassae* genome. This also means that the genome of *w*Ana, which was assembled two years earlier, contains not only bacterial sequences but also sequences from the *Wolbachia* genes inserted into *D. ananassae*.

Table 2. Inserted *Wolbachia* genes found in *D. ananassae* strains[a]

Strain	Origin	Number of inserted genes	Notes
A13	Hawaii, USA	44	Sequenced strain
A31	Mumbai, India	29	
A33	Selangor, Malaysia	23	
A34	Java, Indonesia	34	

[a] Number of inserted *Wolbachia* genes out of the 45 genes tested by Dunning Hotopp *et al*. (2007). Strain names (A13, A31, A33 and A34) are based on the stock numbers given to the strains by the UC San Diego Drosophila Species Stock Center.

Through reverse transcription PCR (RT-PCR) and sequencing, and quantitative RT-PCR, Dunning Hotopp *et al*. (2007) also showed that 28 inserted *Wolbachia* genes out of 1206 assayed are transcribed, albeit at much lower rates than *D. ananassae*'s highly transcribed actin gene (*act5C*). It is unclear whether these low-level transcripts are biologically significant. Furthermore, they demonstrated that the insert is paternally inherited in a Mendelian fashion to both sons and daughters, indicating that it is located on an autosome rather than on a sex chromosome. Comparisons of sequences of PCR-products from the inserted genes in the different *D. ananassae* strains showed that the inserts are highly similar to each other and to *w*Mel, the *Wolbachia* strain infecting *Drosophila melanogaster*.

Comparisons between the *Wolbachia* strain *w*Ri and the *Wolbachia* sequences found in *D. ananassae* Hawaii showed that these sequences are almost identical (Klasson *et al*. 2009b). Further comparative studies between *w*Ri and the *D. ananassae Wolbachia* inserts were carried out by Lisa Klasson (unpublished). Since *D. ananassae* Hawaii (A13) was not tretracycline-cured before being sequenced, the resulting sequences contain *w*Ana sequences from the infecting bacterium in addition to the *D. ananassae* sequences and the *Wolbachia* insert in the *D. ananassae* genome. By comparing these sequences to the genome of *w*Ri, Klasson discovered several polymorphic sites that either contained or lacked insertion sequence (IS) elements in the *Wolbachia* sequences from the *D. ananassae* genome. She also discovered that several *Wolbachia* genes in the *D. ananassae* genome contain insertions and deletions (indels) that result in frameshifts compared to the same genes in the *w*Ri genome. Since these genes most likely are functional in *w*Ri and presumably also in *w*Ana, it is probable that the indels are located in the inserted genes.

It seems like almost the entire *Wolbachia* genome has been inserted into the genomes of four strains of *D. ananassae* from different geographical locations. It is not known when this LGT event took place, if it was a single event with the insert subsequently spreading to other *D. ananassae* strains or if there have been several, independent events. Also, it is still unclear whether these inserts have any biological function in the *D. ananassae* genomes. Since *w*Ana has been shown to induce CI, one hypothesis is that the insert might provide a selective advantage if some of the genes contained in it are involved in the rescue of CI induced by *w*Ana.

Aims

There were three aims of the present study. The first aim was to study the IS-elements found in the *D. ananassae Wolbachia* sequences with the purpose of finding markers to discriminate between the insert and the bacteria. This was done by PCR amplification and sequencing.

The second aim was to investigate whether the genes inserted into these four *D. ananassae* strains are involved in the induction or rescue of CI. This was done by crossing *D. ananassae* flies of different *Wolbachia* infection/insert status and measuring the mortality of the resulting eggs.

The third aim was to use the frameshift indels observed in the *Wolbachia* genes in the *D. ananassae* genome as a basis for a comparison between the four inserts. These genes were PCR amplified and sequenced in all strains, and the sequences were searched for differences between the strains. The purpose of this comparison was to clarify the evolutionary history of the inserts.

Results

Genetic differences between the inserts and bacteria

Six different strains of the *Drosophila ananassae* fruit fly were used in this study. Four of these (A13, A31, A33 and A34) are the strains found to contain *Wolbachia* inserts by Dunning Hotopp *et al*. (2007). These strains are also infected with *w*Ana. One strain is infected with *w*Ana but lacks the insert (A16) and the last strain lacks both endosymbiont and insert (A12). Tetracycline-cured strains of A13, A31, A33 and A34 were established along with a tetracycline-cured strain of A16 as a control. Details of the strains used can be found in Table 3.

Table 3. Details of *D. ananassae* strains used in the study[a]

Strains	Origin	Inserted genes	*w*Ana infection	Notes
A12	Florida, USA	-	-	
A13	Hawaii, USA	44	+	Sequenced strain
A13tet		44	-	
A16	Nayarit, Mexico	-	+	
A16tet		-	-	Control strain[b]
A31	Mumbai, India	29	+	
A31tet		29	-	
A33	Selangor, Malaysia	23	+	
A33tet		23	-	
A34	Java, Indonesia	34	+	
A34tet		34	-	

[a] "Inserted genes" refers to observations of the number of inserted *Wolbachia* genes out of the 45 tested by Dunning Hotopp *et al*. (2007). "-tet" refers to tetracycline-cured strains.
[b] This strain was used as a control for the tetracycline treatment.

Nine polymorphic IS-sites were studied. At these sites, differences between the *w*Ri genome and the *Wolbachia* sequences in the *D. ananassae* genome had been observed (Lisa Klasson, unpublished). Most of these differences were deletions of insertion sequence (IS) elements in some of the *D. ananassae* sequences compared to the *w*Ri genome. In one case, the *D. ananassae* sequences contained an IS element that is not present in *w*Ri (Table 4).

Table 4. IS-sites[a]

IS-site	Position in *w*Ri genome[b]	*w*Ri	A13	Size (b.p)
IS_1	391.666-393.148	+	–	1500
IS_2	582.810-583.710	+	–	1000
IS_3	607900-609380	+	–	1500
IS_4	766.366-767.281	+	–	1000
IS_5	837.000-839.000	+	–	1500
IS_7	877.152-878.770	+	–	1500
IS_8	996.580-998.063	+	–	1500
IS_9	1.192.855-1.194.336	+	–	1000
IS_10	1.210.962	–	+	1500

[a] IS-site in *Wolbachia* sequences of endosymbiotic *w*Ri and in genomic seqqunces in *D. ananassae* A13.
[b] http://www.ncbi.nlm.nih.gov/nuccore/NC_012416.1?report=fasta&log$=seqview&format=text

Primer pairs for each IS-site were designed based on the *w*Ri genome placed before and after the position of the IS site in *w*Ri and used in a PCR-assay to investigate in which *D. ananassae* strains the IS-elements were present or absent. PCR products from strains where the IS-element is present are expected to be longer than PCR products from strains where the same element is absent. The expected size differences correspond to the sizes of the IS-elements (see Table 4). These differences were visualised by running the PCR products on an

agarose gel, and the results are presented in Table 5. All PCRs were re-run several times to confirm the results, partly because results were sometimes inconsistent (Figure 2).

Table 5. Occurence of IS-elements in *w*Ri and *D. ananassae* strains[a]

	*w*Ri	A13	A13tet	A16	A31	A31tet	A33	A33tet	A34	A34tet
IS_1	+	–	–	–	–	–	–/+	–	–	–
IS_2	+	–	–	–	–	–	–	–	–	–
IS_3	+	–/+	–	–	–	–	–	–	–	–
IS_4	+	–	–	–	–	–	–	n/a	–	–
IS_5	+	–	–	–	–	–	–	–	–	–
IS_7	+	–/+	–	–	–/+	–	–/+	–	–	–
IS_8	+	–	–	–	–/+	–	–/+	–	–	–
IS_9	+	–	–	–	–	–	–	–	–	–
IS_10	–	–/+	+	–	–/+	+	–/+	+	–	–

[a] + indicates the presence of the IS-element, – indicates the absence of the IS-element and –/+ indicates that both variants are present, as seen by size of PCR amplicons spanning insertion site. n/a means that no PCR-product could be obtained. A13, A31 etc. refer to the untreated *D. ananassae* strains containing the endosymbiont while A13tet, A31tet etc refer to the tetracycline-treated strains that have lost the endosymbiont.

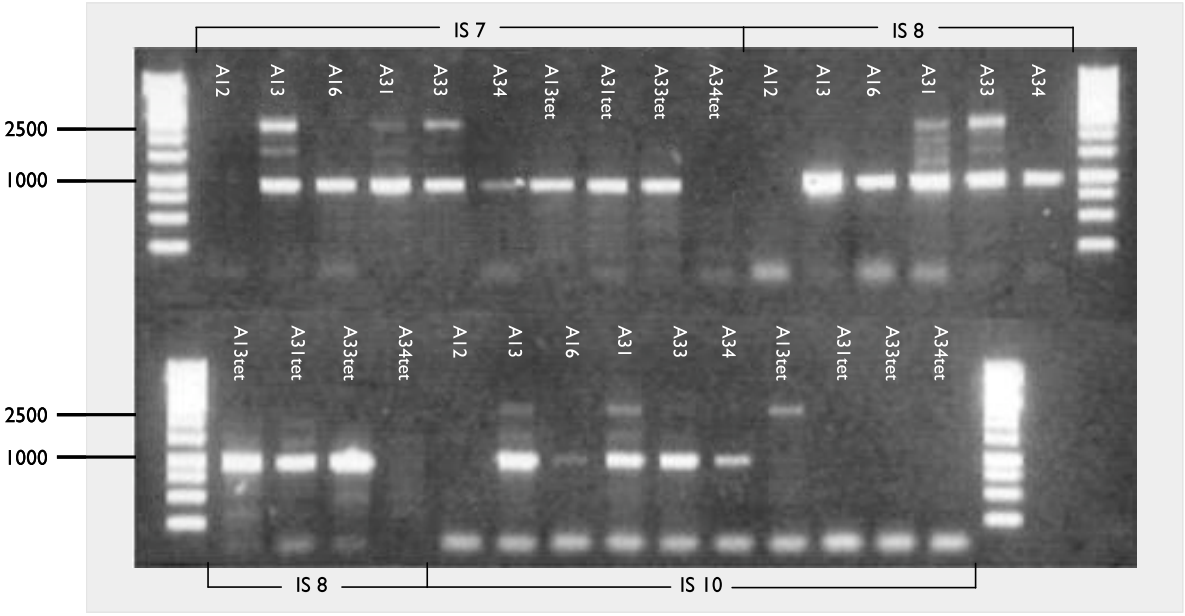

Figure 2. IS_7, IS_8 and IS_10 in different *D. ananassae* strains. PCR amplicons covering IS 7, 8 and 10 in different *D. ananassae* strains as shown above each lane were analyzed by agarose gel electrophoresis.

Consistent differences between the *w*Ana bacterial genome and the *Wolbachia* inserts in the different strains were observed only for IS_7, IS_8 and IS_10 (see Table 5). For IS_7 and IS_8, the uncured *D. ananassae* strains (except A13 for IS_8 and A34 for both), containing both insert and bacteria, showed double bands, indicating polymorphisms. The tetracycline-cured strains, on the other hand, showed only single bands of sizes indicating the absence of the IS-elements. For IS_10, the uncured strains, except A34, gave double bands while the tetracycline-treated strains gave single bands of the longer length, indicating the presence of the IS-element in the inserts. This shows that at these three IS-sites, *w*Ana is identical to *w*Ri while the inserts differ.

The differences observed between the free bacterium and the bacterial inserts with respect to IS_7, IS_8 and IS_10 can be used to detect the presence of the insert and also to confirm the absence of bacteria in tetracycline-cured strains.

The exception was A34, which did not show double bands even though it is supposed to have inserted *Wolbachia* genes according to Dunning-Hotopp *et al*. (2007). Also, A34tet showed single bands in the beginning of the project, but within a few weeks I was unable to obtain any PCR-amplicons of *Wolbachia* genes from this strain. Therefore, I investigated the insert status of A34 in more detail.

The disappearance of the A34 insert

D. ananassae strain A34 tested positive for 34 out of 45 tested genes in Dunning Hotopp *et al*.'s PCR assay (Dunning Hotopp *et al*. 2007). In the assay of the IS-elements described in the previous section, however, A34 never gave double bands where the other strains containing both bacteria and insert did, and although single bands were obtained for the IS-sites from A34tet in the beginning of the study, these bands could not be obtained after a few weeks. A new A34tet strain was then established and DNA was extracted and tested from several individuals from this strain. No PCR-amplicons of the tested *Wolbachia* genes could be obtained from any of these individual DNA preparations (Figure 3). The positive control gave a band at the expected length (data not shown).

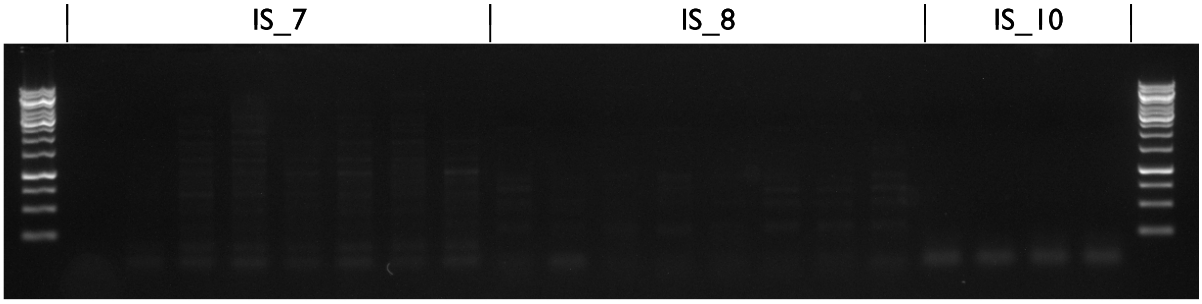

Figure 3. Lack of insert in A34tet. DNA spanning insertions IS7, IS8 and IS10 in A34tet was amplified by PCR, and amplicons subjected to agarose gel electrophoresis. Lanes 1-8: PCR-amplicons for IS_7 from eight different A34tet individuals; lanes 9-16: PCR-amplicons for IS_7 from eight different A34tet individuals; lanes 17-20: PCR-amplicons of IS_10 from four different A34tet individuals.

The most probable explanation for this is that the A34 *D. ananassae* strain that was shipped to the lab from the UC San Diego Stock Center never contained any inserted *Wolbachia* genes. The PCR amplicons of *Wolbachia* genes that were obtained from the tetracycline-cured A34 strain during the first few weeks after treatment could have come from bacterial contaminants that had not been completely cleared from the population.

Test of cytoplasmic incompatibility induction and rescue

To test whether the inserted *Wolbachia* genes in the different *D. ananassae* strains were able to induce cytoplasmic incompatibility (CI) or rescue the CI induced by *w*Ana, single-pair crosses were set up between *D. ananassae* strains of different infection and insert status according to the crossing scheme outlined in Table 6. Since A34 apparently did not contain any *Wolbachia* insert, this strain was excluded from the crossing scheme, and due to time limitations A33tet was also excluded. Egg mortality was used as an estimate of CI. Different types of control crosses were designed to estimate the egg mortalities in crosses between different strains of *D. ananassae* without CI. The cross between male A16 and female A16tet was designed as a control of the entire CI test. Since *w*Ana is known to induce CI (Bourtzis *et*

al. 1996), this cross should give a significantly higher egg mortality compared to the cross between male A16 and female A16.

Table 6. Crossing scheme for testing the ability of the inserted *Wolbachia* genes to induce or rescue cytoplasmic incompatibility

Male	Female	Purpose of cross[a]	Male status[b]	Female status[b]
A12	A12	control	–	–
A12	A13tet	mod control A13 insert	–	I
A12	A31tet	mod control A31insert	–	I
A16	A16	control	*w*Ana	*w*Ana
A16	A16tet	mod test *w*Ana	*w*Ana	–
A16	A13tet	resc test A13 insert	*w*Ana	I
A16	A31tet	resc test A31 insert	*w*Ana	I
A16tet	A16	mod control *w*Ana	–	*w*Ana
A16tet	A16tet	control	–	–
A13tet	A12	mod test A13 insert	I	–
A13tet	A16	resc control A13 insert	I	*w*Ana
A13tet	A13tet	control	I	I
A31tet	A12	mod test A31 insert	I	–
A31tet	A16	resc control A31 insert	I	*w*Ana
A31tet	A31tet	control	I	I

[a] mod, test for induction of CI; resc, test for rescue of CI
[b] *w*Ana, flies carrying endosymbiotic *w*Ana; I, inserted -*Wolbachia* genes; –, no *Wolbachia* sequences

Eggs were counted 24 hours after the crosses and egg mortality was calculated 48 hours later by counting the number of eggs that had not hatched. Unfortunately, I was not able to complete all planned crosses due to time constraints, and for some of the crosses that I did carry out, the number of females was too low to give a good estimate. Nevertheless, the results from the crosses are presented in Table 7.

Table 7. Cytoplasmic incompatibility in single-pair crosses between different *D. ananassae* strains.

Cross		Females (number)	Eggs (number)	Mortality (percent ± SEM)	Comparison[a]
1	♂A12 × ♀A12	16	305	11.5±3.4	1 vs 7; $p < 0.01$
2	♂A16 × ♀A16	11	422	28.3±6.9	
3	♂A13tet × ♀A12	10	446	12.9±5.1	3 vs 1 N.S.
4	♂A31tet × ♀A12	14	514	31.8±8.1	4 vs 1; $p < 0.05$ 4 vs 7 N.S.
5	♂A16 × ♀A16tet	3	129	36.2±16.0	5 vs 2 N.S. 5 vs 6 N.S.
6	♂A16tet × ♀A16tet	5	202	19.9±13.3	
7	♂A31tet × ♀A31tet	10	362	45.6±9.5	

[a] Mortalities in pairs of crosses were compared using Student's t-test (two-tailed, unpaired). N.S. = not significant.

The mortality of the eggs resulting from the cross between male A31tet and female A12 (mod test A31 insert) was found to be significantly higher than the egg mortality in the A12xA12 cross (control), but on the other hand it was not significantly different from the mortality in the A31tet × A31tet cross (control). The ideal control for the cross between male A31tet and female A12 would have been the male A12 × female A31tet cross, but unfortunately there was not enough time to complete this cross. Therefore, no conclusions regarding the ability of the inserts to induce or rescue CI can be drawn from these crosses.

PCR assay of frameshift indels in the *Drosophila ananassae* strains

When comparing the *Wolbachia D. ananassae* A13 sequences to the *w*Ri genome, 42 genes in the A13 sequences had been found to contain indels that cause frameshifts (Lisa Klasson, unpublished). These genes are referred to as ns_1 – ns_42 from here on. Details about these genes can be found in Appendix I. These genes are relatively evenly spread out when mapped to the *w*Ri genome (see Appendix II) and the presence or absence of 41 of them was confirmed by PCR-amplification in the different *D. ananassae* strains. The numbering refers to their position in the *w*Ri genome. I was never able to amplify ns_32 in any of the *D. ananassae* strains, and it is therefore left out of the assay. Each gene was PCR amplified at least twice at different times in each *D. ananassae* strain.

Table 8. PCR assay of ns-genes in *Drosophila ananassae* strains[a]

Gene	A13	A16	A31	A33	A34	A13tet	A31tet	A33tet	A34tet
ns_1	+	+	+	+	+	+	+	+	–
ns_2	+	+	+	+	+	+	+	+	–
ns_3	+	+	+	+	+	–	–	–	–
ns_4	+	+	+	+	+	+	+	+	–
ns_5	+	+	+	+	+	+	+	+	–
ns_6	+	+	+	+	+	+	+	+	–
ns_7	+	+	+	+	+	+	+	+	–
ns_8	+	+	+	+	+	+	+	+	–
ns_9	+	+	+	+	+	+	+	+	–
ns_10	+	+	+	+	+	+	+	+	–
ns_11	+	+	+	+	+	+	+	+	–
ns_12	+	+	+	+	+	+	+	–	–
ns_13	+	+	+	+	+	+	+	+	–
ns_14	+	+	+	+	+	+	+	+	–
ns_15	+	+	+	+	+	+	+	+	–
ns_16	+	+	+	+	+	+	+	+	–
ns_17	+	+	+	+	–	+	+	+	–
ns_18	+	+	+	+	+	+	+	+	–
ns_19	+	+	+	+	+	+	+	+	–
ns_20	+	+	+	+	+	+	+	+	–
ns_21	+	+	+	+	+	+	+	+	–
ns_22	+	+	+	+	+	+	+	+	–
ns_23	+	+	+	+	+	+	+	+	–
ns_24	+	+	+	+	+	+	+	+	–
ns_25	+	+	+	+	+	+	+	+	–
ns_26	+	+	+	+	+	+	+	+	–
ns_27	+	+	+	+	+	+	+	+	–
ns_28	+	+	+	+	+	+	+	+	–
ns_29	+	–	+	+	–	+	–	–	–
ns_30	+	+	+	+	+	+	+	–	–
ns_31	+	+	+	+	+	+	+	+	–
ns_33	+	–	–	–	–	+	–	–	–
ns_34	+	+	+	+	+	+	+	+	–
ns_35	+	+	+	+	+	+	+	–	–
ns_36	+	+	+	+	+	+	+	+	–
ns_37	+	+	+	+	+	+	+	+	–
ns_38	+	+	+	+	+	+	+	+	–
ns_39	+	+	+	+	+	+	+	+	–
ns_40	+	+	+	+	+	+	+	+	–
ns_41	+	+	+	+	+	+	+	+	–
ns_42	+	+	+	+	+	+	+	+	–

[a] +, the gene is present; –, the gene is absent (or at least cannot be PCR amplified with the primers used)

A majority of the 41 tested genes were found to be present in all *D. ananassae* strains except in A34tet which, as discussed above, seems to lack the inserted genes. As expected, A13 tested positive for all the genes and A13tet lacked only two. A31tet and A33tet also tested positive for a surprisingly high number of genes.

Comparison of the frameshift indel sequences from the *Drosophila ananassa*e strains

All the ns-gene PCR products obtained in the PCR assay described above were sequenced and compared between all strains. For each ns-gene, the sequences from all strains were compared at the sequence trace chromatogram level. The trace chromatograms constitute the raw data that are obtained as electronic output files in dye-terminator sequencing. These files were visually examined in the Consed/Autofinish software.

In the chromatograms, each base in the sequence is represented by a peak of a base-specific colour. The height of a peak depends on the signal intensity, which can differ from base to base. Also, sometimes there can be two peaks (double peaks) at the same position in the sequence. These double peaks are a result of polymorphisms within the sample at this position (i.e. a single nucleotide polymorphism – SNP) caused by heterozygosity. Potential polymorphisms can be detected by comparing the sequences in text format (FASTA format). In the case of an indel polymorphism within the sequenced sample, the chromatogram will have single peaks up to the point of the mutation, after which there will be double peaks caused by the shift in the sequence.

No SNPs were detected either between or within the strains in any of the ns-genes. As expected, indels were found in all sequenced ns-genes in A13 (see Figure 4). Indels were also found in a subset of the ns-genes in A31, A33 and A34. The initial hypothesis was that these indels were located in the inserted genes while the genes in *w*Ana would be identical to those in *w*Ri (i.e. lacking the mutations). The difference between the inserted genes and the bacterium would then account for the presence of double peaks after the indel in the ns-genes in uncured *D. ananassae* strains containing both insert and bacteria (A13, A31, A33 and A34). Surprisingly however, for most ns-genes where double peaks had been observed in the uncured strains, the same peaks were also observed in the tetracycline-cured strains. This indicates the presence of two variants of these genes (one containing the indel and one lacking it) within the inserts. There were only a few exceptions to this pattern. ns-genes 5, 6 and 7 had double peaks in A13 but not in A13tet, and ns_41 had double peaks in A31 but not in A31tet. In these genes within these strains, the insert contains the indels with no exception while the infecting *w*Ana does not. The results of the sequence analysis are summarised in Table 9.

Although differences were observed between the inserts in the different *D. ananassae* strains in terms of variation in the number of ns-genes containing indel polymorphisms, these differences were not numerous enough to carry out an evolutionary analysis. Instead, the focus was shifted towards investigating the large number of polymorphic ns-genes found in the A13 insert.

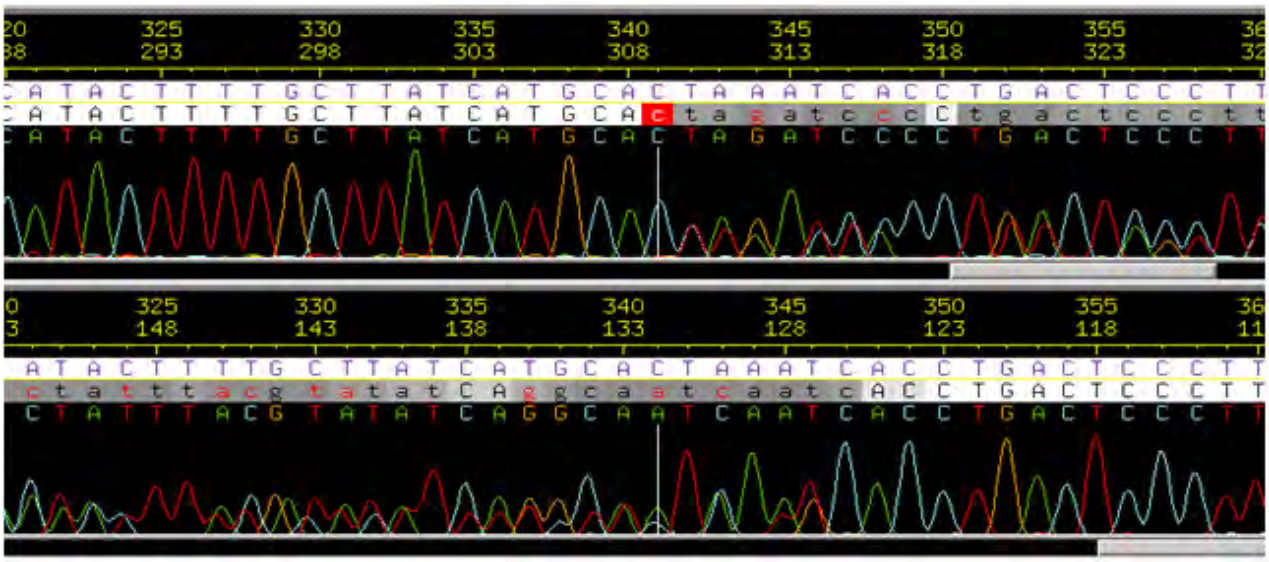

Figure 4. Chromatogram for forward and reverse ns_39 sequences for A13tet. Raw sequence data from the sequencing of ns_39 in A13tet are shown as viewed in the Consed/Autofinish software.

Table 9. ns-gene sequencing results[a]

	A13	A13tet	A16	A31	A31tet	A33	A33tet	A34
ns_1	D	D	S	S	S	S	S	S
ns_2	D	D	S	S	S	S	S	S
ns_3	D	n/a	S	b.s.	n/a	S	n/a	b.s.
ns_4	D	D	S	D	D	S	S	S
ns_5	D	Ins	S	S	S	S	S	S
ns_6	D	Del	S	S	S	S	S	S
ns_7	D	Del	S	S	S	S	S	b.s.
ns_8	D	D	S	S	S	S	S	S
ns_9	D	D	S	S	S	S	S	S
ns_10	D	D	S	S	S	S	S	S
ns_11	D	D	S	S	S	S	S	S
ns_12	D	D	S	D	D	S	n/a	S
ns_13	D	D	S	S	S	S	S	S
ns_14	D	D	S	S	S	S	S	S
ns_15	D	D	S	S	S	S	S	S
ns_16	D	D	S	S	b.s.	S	S	S
ns_17	D	D	S	S	S	S	S	n/a
ns_18	D	D	S	S	S	S	S	S
ns_19	D	D	S	S	S	S	S	S
ns_20	D	D	S	S	S	S	S	S
ns_21	D	D	S	S		S	S	S
ns_22	D	D	S	D	D	D	D	S
ns_23	D	D	S	S	S	S	S	S
ns_24	D	D	S	S	S	S	S	S
ns_25	D	D	S	S	S	S	S	S
ns_26	D	D	S	D	D	S	D	S
ns_27	D	D	S	D	D	S	S	S
ns_28	D	D	S	S	S	S	S	S
ns_29	D	D	n.s.	S	n.s.	S	n.s.	n.s.
ns_30	D	D	S	n.s.	n.s.	S	n/a	n.s.
ns_31	D	D	S	D	D	D	D	S
ns_33	D	D	n/a	n/a	n/a	n/a	n/a	n/a
ns_34	D	D	S	S	S	S	S	S
ns_35	D	D	S	D	D	D	n/a	D
ns_36	D	D	S	S	S	S	S	S

ns_37	D	D	S	S	S	S	S	S
ns_38	D	D	S	S	S	S	S	S
ns_39	D	D	S	S	S	S	S	S
ns_40	D	D	S	S	D	S	D	D
ns_41	D	D	S	D	Del	D	D	S
ns_42	D	D	S	S	S	S	S	S

[a] D, double peaks in the sequence chromatograms; S, single peaks; b.s., bad sequence that could not be analysed; n/a, no PCR product could be obtained.

Investigating the polymorphisms in the A13 insert

The sequencing of the ns-genes in the different *Drosophila ananassae* strains showed the presence of polymorphisms in almost all ns-genes in A13tet. The presence of polymorphisms in these genes indicates that there are at least two different copies of the genes in A13tet. This was a rather surprising finding which deserved further investigation.

The first step in this investigation was to sequence a subset of the ns-genes in several individual A13tet flies. DNA was extracted from four males and four females and ns-genes 2, 10, 16, 25, 39 and 42 were PCR-amplified and sequenced. Analysis of the sequence data showed that there was no variation between individuals, i.e. all individuals had the same polymorphisms in the tested ns-genes.

There are at least three potential causes for the observed intraindividual variation in A13tet. If the tetracycline-treatment had not been entirely successful, bacteria remaining in A13tet could be the cause. Another possible explanation is that the A13 flies are homozygous for the insert, i.e. the insert is present in both chromosomes in a chromosome pair but that the two copies of the insert are different. The third possibility is that the there are two (or more) different copies of the insert on the same chromosome. Flies could be hemizygous for these inserts, i.e. the inserts are present in only one of the chromosomes in a chromosome pair, or they could be homozygous.

In order to investigate these different possibilities, crosses were set up between male A13tet and female A12 flies. Since *Wolbachia* is only transmitted to the offspring through the egg cytoplasm, this crossing design eliminates the possibility of bacterial contaminants from A13tet in the first generation of offspring (F1 generation). Depending on whether the flies were homozygous or hemizygous for the insert and on the copy number of the insert, different outcomes of this cross were possible (see Figure 5).

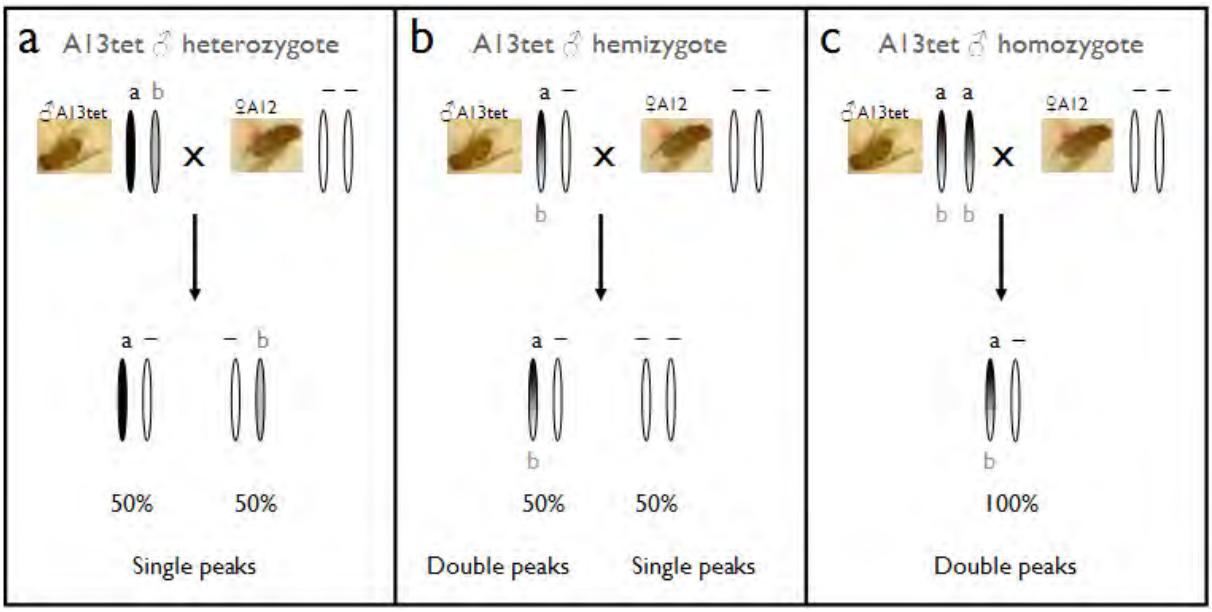

Figure 5. A13tet crosses. Three different outcomes of A13tet x A12 crosses are possible depending on whether the A13tet male is (a) heterozygous, (b) hemizygous or (c) homozygous for the two variants of the ns-genes. The illustration is very schematic, showing only the potential configuration at one chromosome pair. a and b within the figures refer to the two variants of the the ns-genes (with and without indel) and – refers to absence of the insert. Percentages refer to the percentage of offspring of each type expected from the crosses. Single and double peaks, respectively, refer to the expected observations in the chromatograms sequenced ns-genes in the offspring.

DNA was extracted from six male F1 offspring and five female F1 offspring from two of the A13tet x A12 crosses. DNA was also extracted from the fathers in the two crosses. Four ns-genes (ns_2, ns_10, ns_16 and ns_25) were PCR amplified in these individuals. The results of the PCR assay are presented in Table 10. PCR products were obtained for all four genes in both fathers, confirming the presence of the insert. In the F1 generation however, the results were inconclusive since PCR products were obtained for all four genes from all F1 in cross 2 but not in cross 1. In the F1 of cross 1, no PCR products were obtained from four individuals while all genes except ns_2 could be detected in two individuals. These inconsistencies, combined with the fact that only a fraction of the entire F1 generations were tested, meant that no firm conclusions regarding the zygosity for the insert could be drawn based only on the PCR-results.

Table 10. Results from PCR assay of A13tet x A12 crosses 1 and 2[a]

		ns_2	ns_10	ns_16	ns_25
Cross 1	Father	+	+	+	+
	F1♂1	+	+	+	+
	F1♂2	–	–	–	–
	F1♂3	+	+	+	+
	F1♂4	+	+	+	+
	F1♂5	+	+	+	+
	F1♂6	–	–	–	–
	F1♀1	–	+	+	+
	F1♀2	–	+	+	+
	F1♀3	–	–	–	–
	F1♀4	–	–	–	–
	F1♀5	+	+	+	+
Cross 2	Father	+	+	+	+

F1♂1	+	+	+	+
F1♂2	+	+	+	+
F1♂3	+	+	+	+
F1♂4	+	+	+	+
F1♂5	+	+	+	+
F1♂6	+	+	+	+
F1♀1	+	+	+	+
F1♀2	+	+	+	+
F1♀3	+	+	+	+
F1♀4	+	+	+	+
F1♀5	+	+	+	+

[a] +, the gene was detected; –, the gene could not be detected.

The PCR products obtained from cross 2 were also sequenced and the sequences were visually examined as previously described. If the two variants of the insert are present on one chromosome each in a chromosome pair in the A13 flies (heterozygous), a cross such as the one carried out should result in hemizygous offspring. Since these offspring would carry only one of the variants, the ns-gene sequences should contain only single peaks (Figure 5a). If, on the other hand, the two variants of the insert are present on the same chromosome and are close enough to avoid recombination over one generation, the resulting ns-gene sequences should yield double peaks as observed for A13tet (Figures 5b and c). The difference between the hemi- and homozygous cases would be detected as a difference in the fraction of the offspring carrying the insert, 50% in hemizygous and 100% in homozygous cases, respectively.

The results from the analyses of the ns-sequences from cross 2 were in agreement with the flies being homozygous for the insert and multiple variants being present (Figure 5c). Double peaks were observed for all four ns-genes in all offspring, indicating the persistence of the polymorphism. This meant that there were at least two different inserts in A13, but it was still unclear whether these two inserts were located on the same chromosome and if so, whether they were positioned in tandem or further apart.

Crosses between A13tet males and A12 females were again set up but this time with the purpose of carrying out backcrosses between F1 males and A12 females. The idea behind the backcrosses was to see if the two variants could be separated by recombination over two generations. This would give some indication of how closely together the variants are positioned. Two backcrosses were carried out and several individuals from the resulting generations (F2 generations) were collected. DNA was extracted from these individuals as well as from the fathers of each backcross and from the father from the original cross. Four ns-genes, ns_2, ns_10, ns_39 and ns_40, were PCR amplified from all individuals. The results of this PCR assay were used to determine how the insert or inserts had been inherited. The results from the PCR study are presented in Table 11 and summarised in Table 12.

Table 11. Results of PCR assay on backcrosses 1 and 2[a]

		ns_2	ns_10	ns_39	ns_40
	Original father	+	+	–	+
Backcross 1	Father	+	+	+	+
	F2♂1	–	–	–	–
	F2♂2	+	+	+	+
	F2♂3	+	+	+	+
	F2♂4	–	–	–	–
	F2♂5	+	+	+	+
	F2♂6	–	+	+	+
	F2♂7	+	+	–	+
	F2♂8	–	+	+	+

	F2♂9	–	–	–	–
	F2♂10	–	–	–	–
	F2♀1	+	+	+	+
	F2♀2	–	+	–	+
	F2♀3	+	+	+	+
	F2♀4	–	+	–	+
	F2♀5	–	–	–	+
	F2♀6	+	–	–	–
Backcross 2	Father	+	+	+	+
	F2♂1	+	+	+	+
	F2♂2	+	+	+	+
	F2♂3	–	+	–	+
	F2♂4	–	–	–	+
	F2♂5	–	–	–	+
	F2♂6	–	–	–	–
	F2♂7	–	–	–	–
	F2♂8	–	–	–	–
	F2♂9	–	+	–	+
	F2♂10	–	–	–	–
	F2♂11	+	+	+	+
	F2♂12	–	+	+	+
	F2♂13	–	+	+	+
	F2♂14	+	+	+	+
	F2♀1	–	–	–	–
	F2♀2	+	+	+	+
	F2♀3	–	+	–	–
	F2♀4	–	+	–	+
	F2♀5	+	+	+	+
	F2♀6	+	+	+	+
	F2♀7	+	+	+	+
	F2♀8	+	+	+	+
	F2♀9	+	+	+	+
	F2♀10	–	+	–	+
	F2♀11	–	+	–	+
	F2♀12	+	+	+	+
	F2♀13	+	+	+	+
	F2♀14	+	+	+	+
	F2♀15	+	+	+	+

[a] +, the gene was detected; –, the gene was not detected.

Table 12. Summary of the results from the PCR assay of backcrosses 1 and 2

	Offspring (F2) where genes were detected (percent)					
	No genes detected	ns_2	ns_10	ns_39	ns_40	All genes
Backcross 1	25	44	63	44	69	31
Backcross 2	14	48	76	55	79	48
Total	18	47	71	51	76	42

Again, the results did not lend themselves to an easy interpretation. All four genes were present in the father in the two backcrosses. 42% of the F2 generation were positive for all four tested ns-genes, while in 18% of the F2, none of the genes could be detected. In a few individuals, some genes were detected while other genes were not. There was not really any clear pattern. In addition, one has to bear in mind that PCR amplification is not a definite way to determine the presence or absence of genes in an individual. The quality of the individual DNA extractions can differ and so can the primer binding efficiency. For instance, ns_39 could not be detected in the original father even though it must have been present since his

offspring (the fathers in the two backcrosses) inherited it. Nevertheless, it seems like the inheritance of the insert is not straightforwardly Mendelian and does not follow the expected results outlined in Figure 6. If one assumes that the F2 individuals in which some genes were detected in fact have all genes, but that they were not detected in the PCRs, this would mean that 37 of the total 45 F2 individuals inherited all four genes. With a probability of 0.5 for each individual, the binomial probability for this is 6.13×10^{-6}, well below the threshold for significance.

The obtained PCR products were also sequenced, and when the sequences were analysed they showed the same double peaks that had been observed previously. This was the case for all sequenced genes from all individuals, indicating that the two variants were still present, i.e. that they had not segregated over the two generations of the backcross. It therefore seems like the strain is homozygous for the insert, that there are at least two different copies of the *Wolbachia* insert in A13 and that the copies are positioned close enough to each other to remain linked over two generations.

Discussion

The PCR assay of the IS-elements in tetracycline cured and uncured *Drosophila ananassae* strains provided three stable markers (IS_7, IS_8 and IS_10) for discriminating between the *Wolbachia* insert and *w*Ana. These markers can be used to confirm the absence of *w*Ana from tetracycline-cured strains by a simple PCR assay.

The test of cytoplasmic incompatibility (CI) induced or rescued by the inserts unfortunately did not yield any conclusive results. The lack of results was not a consequence of the experimental setup since this was almost identical to previously used methods for testing CI (see e.g. Bourtzis *et al*. 1996). The real problem was lack of experience in carrying out this type of experiment, leading to the time reserved for it being too short. Completing the test with an adequate number of crosses would require more time than what would be possible within the timeframes of the present project.

Size of the insert in *Drosophila ananassae*

The PCR assay of 41 ns-genes in the *D. ananassae* strains showed that most genes were present in most of the inserts (Table 8). On the other hand, no *Wolbachia* genes could be detected in A34tet. These results are in contrast to those of Dunning Hotopp *et al*. (2007) who also carried out a PCR assay on the tetracycline-cured *D. ananassae* strains. The 45 genes assayed by Dunning Hotopp *et al*. (2007) are, in conformity with the ns-genes, rather evenly spread out when mapped on the *w*Ri genome. In their article, Dunning Hotopp *et al*. (2007) do not report which of the assayed genes were detected in which strains. The only possible comparison therefore, between my results and theirs, is a comparison between the percentage of assayed genes detected in each strain (see Table 13).

Table 13. Results from this study compared to literature data[a]

	genes detected[a] (percent)	
D. ananassae strain	ns-genes[b]	Dunning Hotopp genes[c]
A13tet	98	98
A31tet	93	64
A33tet	85	51
A34tet	0	76

[a] detection by PCR; numbers show the percent of tested genes with positive amplification
[b] this study
[c] Dunning Hotopp *et al*. (2007)

In A13tet, A31tet and A33tet most of the ns-genes were detected, albeit with small differences in the number of genes detected between the strains. In Dunning Hotopp *et al*.'s assay, these differences were bigger, but they followed the same trend. In both, most genes were detected in A13tet, followed by A31tet and then A33tet (not considering A34tet). Overall, a smaller fraction of genes were detected by Dunning Hotopp *et al*.-compared to the fraction of ns-genes detected in the present work. This could reflect a real difference, since I assayed a different set of genes compared to Dunning Hotopp *et al.* Many of the ns-genes, however, are positioned very close to the genes assayed by Dunning Hotopp *et al*. when mapped to the *w*Ri genome, and, unless the inserts in the *D. ananassae* strains are extremely fragmented, a higher degree of conformity between the assays would be expected. An alternative explanation for these differences could be false negatives in the Dunning Hotopp assay caused by the inherent unpredictability of PCR amplification. Dunning Hotopp *et al.* (2007) used primers that were designed based on sequences from other *Wolbachia* strains, and some of these primers were also degenerate. The primers used in my assay were designed specifically for the *D. ananassae* A13 *Wolbachia* insert. Also, I repeated the PCRs for all ns-

genes at least twice with different DNA extractions from each *D. ananassae* strain, which was necessary since there were sometimes inconsistencies between the runs. If Dunning Hotopp *et al*. (2007) had reported which of the 45 assayed genes they had detected in each strain, it would have been possible to make a more informative comparison including the presence and absence of genes in specific regions.

Since it is not known how the insertion of *Wolbachia* genes into *D. ananassae* took place – if the whole *Wolbachia* genome was inserted as one piece or if genome fragments were inserted separately – and since neither my PCR assay nor that of Dunning Hotopp *et al*. (2007) covered the entire *Wolbachia* genome, it is impossible to draw any firm conclusions regarding the insert sizes based on the PCR assay results alone, but most likely a large part of the bacterial chromosome has been inserted into the three strains A13, A31 and A33.

The most striking difference between my findings and those of Dunning-Hotopp *et al*. (2007) are the observations concerning A34tet. Considering the results in the other strains, it seems unlikely that this discrepancy reflects a true case of presence or absence depending on what genes were assayed. It also seems unlikely that the insert could have been lost from the entire A34 population during the short time between its arrival and the start of this assay. Most likely, the A34 strain used in my assay did not have the *Wolbachia* insert when it was shipped to us, i.e. it was in fact not A34.

Polymorphic ns-genes

The results expected from the sequencing of the ns-genes were that the uncured strains would show a polymorphism while the cured strains would show only the indel. This was also the case for ns_5, ns_6 and ns_7 in A13 vs A13tet and for ns_41 in A31 vs A31tet, but for all other ns-genes, the polymorphisms were present in both the uncured and the cured strains of A13. Similar polymorphisms were also found in a subset of ns-genes in A31 vs A31tet and in A33 vs A33tet. Crosses and backcrosses, carried out in an attempt to separate the two variants in A13tet, showed that the insert seems to be inherited in a non-Mendelian fashion. Also, it was not possible to separate the two variants.

The seemingly non-Mendelian inheritance of the insert in A13 contradicts the findings of Dunning Hotopp *et al*. (2007). It is possible that the number of F2 offspring screened in my study was too low to show the real inheritance pattern, but at the same time it is peculiar that some F2 individuals seemed to have only a few of the four genes tested. This type of inheritance is difficult to explain. If the insert in A13 is fragmented and the fragments are spread out over many chromosomes, this could explain the patchy inheritance. It can not explain, however, that the two variants are always inherited together, giving double peaks in the chromatograms. *D. ananassae* is one of few *Drosophila* species in which spontaneous meiotic recombination occurs in males (Matsuda *et al*. 1983). This phenomenon could potentially explain the strange inheritance pattern of the insert, but it fails to explain the persistence of the polymorphisms.

The only reasonable explanation for the polymorphisms is that there are several *Wolbachia* inserts in *D. ananassae* A13 and that these inserts are positioned close enough to avoid being broken up by recombination. This, however, leaves the puzzling inheritance unexplained. As discussed above, PCR assays are not flawless and it is possible that all four tested genes in fact were present in the individual F2 offspring that tested positive for only a few of them. If this was the case, it would mean that 82% of the F2 inherited all genes and, given an individual probability of 0.5, the binomial probability for this is only 6.13×10^{-6}. If this indeed reflects reality, it would indicate the presence of some sort of meiotic drive increasing the probability of inheritance of the insert above 0.5.

Evolutionary history of the inserts

The differences observed between the inserts were not numerous enough to carry out a real phylogenetic analysis. The main difference observed was between the A13 insert and the inserts in A31 and A33. In A13, the insert is polymorphic at almost all ns-genes while this is not the case in A31 and A33. However, it is important to remember that the ns-genes were chosen because they were polymorphic in the *D. ananassae* A13 data set. Also, I only studied 41 genes, which, presumably, is only a small subset of all the *Wolbachia* genes inserted into A13, A31 and A33. It is therefore possible that there are other loci where the inserts differ more.

Future perspectives

Several questions were asked at the beginning of this project. Some of these questions remain unanswered and there are also many new questions to be addressed. It would be very interesting to complete the cytoplasmic incompatibility test as this could provide important insight into the potential function of the inserts. The inheritance of the inserts also needs to be sorted out and this would require more extensive crosses where all offspring from all generations are tested with respect to a greater number of genes. The continuous presence of polymorphisms of the ns-genes in A13 is yet another interesting observation that deserves further investigation. Sequencing of more genes in the different inserts should also be carried out so that a more thorough comparison and perhaps a phylogenetic analysis could be carried out.

Materials and Methods

Drosophila ananassae strains

Six *Drosophila ananassae* strains were used in this study, all of which were ordered from the Drosophila Species Stock Center at UC San Diego. Details of the strains are given in Table 14. All flies were maintained in vials with fly food medium[4] (see Table 15 for ingredients). Approximately 30 flies, males and females, were transferred to new vials every two weeks. The flies were kept at 25°C and a 12h light/dark cycle.

Table 14. *Drosophila ananassae* strains

Strain	Stock center number	Origin	*Wolbachia* endosymbiont	*Wolbachia* insert
A12	14024-0371.12	Florida, USA	–	–
A13	14024-0371.13	Hawaii, USA	+	+
A16	14024-0371.16	Nayarit, Mexico	+	–
A31	14024-0371.31	Mumbai, India	+	+
A33	14024-0371.33	Selangor, Malaysia	+	+
A34	14024-0371.34	Java, Indonesia	+	+

Table 15. Fly food (for 100 vials)

Ingredient	Amount (g)	Amount (L)
ddH_2O		1.1
Molasses		0.1
Agar	12.1	
Cornmeal	82	
Yeast	33.8	
Methyl-p-hydroxybenzoic acid	1.8	
Ethanol 95%		0.0184
Proprionic acid		0.0066

Tetracycline-treatment of *D. ananassae*

To cure the flies from the *Wolbachia* endosymbiont, around 30 flies, males and females, from each strain were transferred to new food vials containing 0.025 % tetracycline. When eggs were visible in the vials, the adults were removed. The new generations were transferred to new food vials with tetracycline added at the same concentration as before. This procedure was then repeated once more so that the flies were treated with tetracycline for three generations. Tetracycline-treated strains are referred to as the name of the original strain followed by tet, e.g. A13tet for the tetracycline-treated A13.

DNA extractions

DNA extractions were carried out in principle as described by O'Neill *et al*. (1992). Individual flies were homogenised in 50 µl STE buffer (STE buffer: 100 mM NaCl, 10 mM TrisHCl pH 8.0, 1 mM EDTA pH 8.0) with a clean micro-pestle. The homogenate was then incubated with 2 µl proteinase K (2 mg/ml) for 30 minutes at 60°C followed by a 10-minute incubation at 95°C. The samples were then centrifuged briefly. 1µl of the supernatant was used as template in the PCR amplifications.

[4] Vials with fly food medium were supplied by Ted Morrow at the Department of Ecology and Evolution, Animal Ecology, Uppsala University.

PCR amplification

Primer sequences can be found in Tables 16 and 17. All PCR amplifications were carried out with AccuTaq™ LA DNA polymerase from Sigma Aldrich. Reactions were set up according to Table 18 and run in a thermal cycler according to Table 19.

Table 16. IS-element primer sequences

IS element	Forward primer (5'-3')	Backward primer (5'-3')
IS1	GAAGGAACTCACCATTCACC	GGAGGAAAACTGCTACTAGATTGTG
IS2	CCCTCCATCATCGTACTTTT	GACGGAAAAACAGTGAGCTT
IS3	GCACCATACCGAGATTGTTT	AGAAAAAGCGTGCGAGTATT
IS4	CAACTGCTCCTTTTACTTTGC	AATTATCAAGAAGTAGGGGCTGC
IS5	TCGTTAACAAGTGAGGCAAA	TATGCGATCCTCAACAGAAA
IS7	ATATAAAAGTGCGAGAGCGAG	CGGACAATGTTACGCTAATTT
IS8	AAAGCACTTGAGGAAGCCTA	GAACGGCTAATTCGAGACAC
IS9	TGTATAAAGAGGCCAAAGCC	ACGGGTTATGTATGGCAAGT
IS10	TAGCAGATGCCTGAGCTAGA	GCTGGTATCAAAGGCTCACT

Table 17. ns-gene primer sequences

ns gene	Forward primer (5'-3')	Backward primer (5'-3')
ns01	ACTGCCATTTCTATTACACGC	CTTATCACTTCTCCCACATCG
ns02	CCAAATGGACAAAAGAAAGC	GCATGAAATAATATCGAACCTA
ns03	CTTTCTACTATGCTTGCTGCC	CCAGTATGAAAGGAGATGCAG
ns04	GCAATGAAGATGGAGTAAACG	TGGCTGTCAGTACCAACTAAAG
ns05	TACTGCAACATACACACCAGC	GCGAGCAATCACTAGAACAAG
ns06	GAAGCAGAACGTCAGGTAAAG	CCCAACAAATCTGTACCAAAG
ns07	AGGGCATTTCATAAGATAGAGG	AGCTTCTAAGTCAAGGCCAAG
ns08	AAGATCAATTGGACACAGCTC	ACTGATGAACTCGAGCATTTC
ns09	AGACTACATAAAGCCCAGCG	GAAATAGTGTCGGTTGAGACG
ns10	GCAAAGACGGCTTATAGTGAG	CTCTAATTGCTGTGCATAAAGC
ns11	AAAGCCAGGTATTGGAGAAAG	GCTTGAATCCACTAGAGAACG
ns12_	ACCAAAGTGCGAGTTAAGAAG	CTGACCAAGTTGAAAAAGCTC
ns13	GAGTTGCCTGTGGTTTACATC	CATCAACGCTTGGTTCTTTAC
ns14	CTCTGTTTGAATTGCTCATCC	AAGTAGGGGTTAGCAAAATGC
ns15	ATTGAGTGTCACATACCCACC	TTGTTAGAGGTATTGGCGAAG
ns16	CAGTGAGCAATAGAAGAAGGC	TTTCTCGACTTATGTCACTTGG
ns17	ACCACCTCATCTATCATTTCG	ATATGTCAGTGGATGTACGCC
ns18	GCCTACGTAACATAAAAGGGC	AGTAATTGGCAACGAGAGATG
ns19	CAAAAGGACGTACTGATCGAC	AAGCTGGAAAAAGCACTCTC
ns20	AAAAGTCAATGTGGTGGAGC	ACTTCCGGATAGGTTGTTTTG
ns21	ATCTCAGCAACTTGGTTCTTG	CCCAACATATAAAGTGATGCC
ns22	TCCAGCTCCTTACTTTTTGTG	ATTTCGACTATCACTTGCAGC
ns23	TCAAAGCGGTAATAACGTAGG	GCGATATCATCACACGTTTTC
ns24	TCTCCTATTTCTTGCAACTCG	GGCTCAAGAAACTTAGGGAAG
ns25	ATCTAAAATGCATCAAAACACCG	TTTCGTTTTGATAATGATCCAGG
ns26	ACCTGTGTAGTTGCTGACCTC	TGCTCAGTAAGTTGCAAACAG
ns27	CATGACAGTTTTTGGCAACTC	ATGCACAATGCTCTTTCTAGG
ns28	CAGACATATTGTGCTCTGCTG	ATAGGTTCAAGTATGGGAGGG
ns29	CTCTCATCCTTGGTTAATTGG	CTGCAAGTCCTGAAAATTACC
ns30	ATAGCTAACACGTCATACCGC	GGCAGTATATAACGTTACGCC
ns31	AGCAACAGAGATTGTTCCAAG	GATCTCATGGGTTAGCACAAC
ns32	GCTGTTTTGTTCACTATTGGG	AACTGCAGATATTCCAATCCC
ns33	ATAATCAGCGAGCAGAAAGAG	CTCCTCAGTTTCTGGTATTGC
ns34	AATTGATGAATGCTCAAGGTC	GAAAGAATCCTCGTGTTAGGC
ns35	AAATTCCGTTCGGGTATACAG	TAAGCCTGCAGTAGAACTTGC
ns36	ATATAACTGCCTTGCACATCC	ATAGCAGTTAACGAAATCCCC
ns37	GGCACTATATGTACCGGAGTG	CAACTGCCTTTATGGGATATG
ns38_	CGGCCTCTTTCATTCTCTATATC	AGTCATCAAGCTTGGTTTCAC

ns39	GGTCCAAGAGTTGTTTTGTGAAG	GGCAGGGAAATAAAAATGTATCC
ns40	TCTAAAATATCGGGTTGTTTATG	CGCTGAAGATCAGTATTCCTC
ns41	CTCAACTGACTACGATACCGC	CTTTTTCCCCATCACATACAG
ns42	GATTGTCTTACCCACCAAGTG	TTGCTGAGGCTAGAAAATACG

Table 18. AccuTaq™ PCR reaction set-up

Component	Volume (µl)
ddH_2O	12.25
AccuTaq™ 10x Reaction Buffer (Sigma)	2.5
dNTP mix 2mM each (Fermentas)	5
Primer F 10 µM	2
Primer R 10 µM	2
AccuTaq™ LA DNA polymerase (Sigma)	0.25
Template DNA	1
Total volume	25

Table 19. AccuTaq™ PCR programme

Temperature	Duration
108°C	Lid temperature
96°C	30 s
94°C	30 s[a]
Annealing temperature – see primer list for specific temperatures	45 s[a]
68°C	5 min[a]
68°C	10 min
4°C	Hold

[a] these three steps were repeated 33 times

PCR products were run in a 1 % agarose gel prepared in 1xTAE buffer (40 mM Tris acetate pH 8.0, 1 mM EDTA) with 1 µg/ml ethidium bromide. MultiScreen $PCR_{\mu 96}$ Filter Plates from Millipore were used, following the manufacturer's protocol, to clean the PCR products for sequencing.

Sequencing

PCR products were Sanger-sequenced using the BigDye® Terminator v3.1 Cycle Sequencing Kit from Applied Biosystems. Details of how the reactions were set up and run in a thermal cycler are outlined in Tables 18 and 19 respectively.

Table 20. Sequencing reaction set-up

Component	Volume (µl)
ddH2O	6.25
BigDye® Sequencing Buffer	1.75
Primer	0.5
BigDye® Terminator Ready Reaction Mix	0.5
Template DNA (clean PCR products)	1

Table 21. Sequencing reaction programme

Temperature	Duration
96°C	1 min
96°C	10 s[a]
50°C	5 s[a]
60°C	4 min[a]
4°C	Hold

[a] these three steps were repeated 25 times

Sequencing reactions were cleaned using Sephadex G50 columns (GE Healthcare Life Sciences) on 96-well Millipore plates as described in the online document *Post Sequencing Purification in 96-well Gel Filtration Column Plates.*[5] The sequences were read in an ABI 3730XL DNA sequencer from Applied Biosystems.

Analysis of sequence data

The phred software was used to read the DNA sequencing trace files and to call bases and assign quality values to each called base. Phrap was used to create assemblies which were then viewed and manually edited in Consed/Autofinish. The phred, phrap and Consed/Autofinish softwares can be downloaded from www.phrap.org/phredphrapconsed.html.

Fly crosses

Virgin flies were collected soon after eclosion (emergence from the pupae) and males and females were placed in separate vials. The flies were aged for 3-7 days before being used in the crosses. When the flies had reached an appropriate age, one virgin male and one virgin female were placed in a vial that was subsequently sealed with cotton. Mated females were placed on egg-laying plates containing egg-laying medium[6]. The plates were then changed three times, the first time after 48h and then two more times every 24h. The number of eggs on each plate was counted and the plates were then left at 25°C for 48h to allow the eggs to hatch. After 48h the number of un-hatched eggs were counted and the hatch-rate calculated. The total hatch-rate was calculated from the combined data from all three plates from each female.

A13tet backcrosses

Crosses and backcrosses with A12 were set up as shown in Figure 6 and as described in the previous paragraph. When eggs began to appear in the vial the parents were removed and DNA was extracted from the males. Two males from the F1 hybrid generation were used in two parallel backcrosses carried out in the same manner as the first cross. All offspring (the F2 generation) from this cross were collected and DNA was extracted to use for PCRs and sequencing.

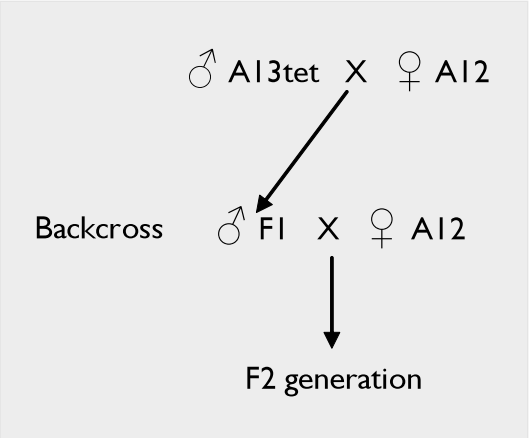

Figure 6. A13tet backcrosses

[5] Online documentation. URL: http://www.beckmancoulter.com/literature/Bioresearch/T-1874A-t.pdf. Retrieved 28/7 2009.

[6] For 500 mL egglaying medium, autoclave 8.75 g agar in 250 mL ddH_2O. Dissolve 6.25 g sucrose in 125 mL ddH_2O while heating; then add 125 mL apple juice. Combine this with the agar solution and finally add 10 mL Nipagin (sodium methylparaben, from Clariant) and 0.2-0-3 ml food colouring.

Acknowledgements

First of all I want to thank my supervisor Lisa Klasson for giving me the opportunity to do this project and for her constant encouragement and high spirits. I would also like to thank Siv Andersson for her great enthusiasm and positivity and for letting me do this project in her group. Thanks also to Ann-Sofie Eriksson, Kristina Näslund and Kirsten Ellegaard for answering so many questions about the practical work and for being such good company in the lab. Thank you also to everyone else at Molecular Evolution for creating such a nice work environment. Anna Knöppel – thanks for all the inspiration, support, hard work and coffee during our writing sessions! It would have been so much harder without it. Mathias – thank you for valuable comments on the text and finally I would like to thank Karin Carlson for her patience and her many valuable comments on the manuscript.

References

Aminov, R. I. & Mackie, R. I. 2007. Evolution and ecology of antibiotic resistance genes. FEMS Microbiol. Lett. **271**:147–161.

Andersson, J. O. 2005. Lateral gene transfer in eukaryotes. Cell. Mol. Life Sci. **62**:1182–1197.

Baldo, L., S. Bordenstein, J. J. Wernegreen & J. H. Werren. 2006. Widespread recombination throughout *Wolbachia* genomes. Mol. Biol. Evol. **23**:437–449.

Bandi, C., C. G. Anderson, C. Genchi & M. L. Baxter. 1998. Phylogeny of *Wolbachia* in filarial nematodes. Proc. R. Soc. Lond. B **265**:2407–2413.

Bandi C., A. J. Trees and N. W. Bratting. 2001. *Wolbachia* in filarial nematodes: evolutionary aspects and implications for the pathogenesis and treatment of filarial diseases. Vet. Parasitol. **98**:215–238.

Baxter, M. 2007. Symbiont genes in host genomes: fragments with a future? Cell Host Microbe **2**:211–213.

Binnington, K. C. & A. A. Hoffmann. 1989. *Wolbachia*-like organisms and cytoplasmic incompatibility in *Drosophila simulans*. J. Invertebr. Pathol. **54**:344–352.

Bourtzis, K., A. Nirgianaki, G. Markakis & C. Savakis. 1996. *Wolbachia* infection and cytoplasmic incompatibility in *Drosophila* species. Genetics **144**:1063–1073.

Casiraghi, M., S. R. Bordenstein, L. Baldo, N. Lo, T. Beninati, J. J. Wernegreen, J. H. Werren & C. Bandi. 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 **151**:4015–4022.

Dunning Hotopp, J. C., M. E. Clark, D. C. S. G. Oliveira, J. M. Foster, P. Fisher, M. C. Muños Torres, J. D. Giebel, N. Kumar, N. Ishmael, S. Wang, J. Ingram, R. V. Nene, J. Shepard, J. Tomkins, S. Richards, D. J. Spiro, E. Ghedin, B. E. Slatko, H. Tettelin and J. H. Werren. 2007. Widespread lateral gene transfer from intracellular bacteria to multicellular eukaryotes. Science **317**:1753–1756.

Gribaldo, S. & C. Brochier. 2009. Phylogeny of prokaryotes: does it exist and why should we care? Res. Microbiol. **160**:513–521.

Hertig, M. & S. B. Wolbach. 1924. Studies on Rickettsia-like micro-organisms in insects. J. Med. Res. **44**:329–372.

Hertig, M. 1936. The Rickettsia, *Wolbachia pipientis* (gen. et sp.n.) and associated inclusions of the mosquito, *Culex pipiens*. Parasitology **28**:453 – 486.

Hilgenboecker, K., P. Hammerstein, P. Schlattmann, A. Telschow & J. H. Werren. 2008. How many species are infected with *Wolbachia*? – a statistical analysis of current data. FEMS Microbiol. Lett. **281**:215–220.

Iturbe-Ormaetxe, I., M. Riegler & S. L. O'Neill. 2005. New names for old strains? *Wolbachia w*Sim is actually *w*Ri. Genome Biol. **6**:401.

Jeyaprakash, A. & M. A. Hoy. 2000. Long PCR improves *Wolbachia* DNA amplification: *wsp* sequences found in 76% of sixty-three arthropod species. Insect Mol. Biol. **9**:393-405.

Keeling, P. J. & J. D. Palmer. 2008. Horizontal gene transfer in eukaryotic evolution. Nat. Rev. Genet. **9**:605–618.

Klasson, L., T. Walker, M. Sebaihia, M. J. Sanders, M. A. Quail, A. Lord, S. Sanders, J. Earl, S. L. O'Neill, N. Thomson, S. P. Sinkins & J. Parkhill. 2008. Genome evolution of *Wolbachia* strain *w*Pip from the *Culex pipiens* group. Mol. Biol. Evol. **25**:1877–1887.

Klasson, L., Z. Kambris, P. E. Cook, T. Walker & S. P. Sinkins. 2009a. Horizontal gene transfer between *Wolbachia* and the mosquito *Aedes aegypti*. BMC Genomics **10**:33. doi: 10.1186/1471-2164-10-33.

Klasson, L., J. Westberg, P. Sapountzis, K. Näslund, Y. Lutnaes, A. C. Darby, Z. Veneti, L. Chen, H. R. Braig, R. Garrett, K. Bourtzis & S. G. E. Andersson. 2009b. The mosaic genome structure of the *Wolbachia w*Ri strain infecting *Drosophila simulans*. Proc. Natl. Acad. Sci. USA **106**:5725–5730.

Kondo, N., Nikoh, N., Ijichi, N., Shimada, M. & Fukatsu, T. 2002. Genome fragment of *Wolbachia* endosymbiont transferred to X chromosome of host insect. Proc. Natl. Acad. Sci. USA. **99**:14280–14285.

Laven, H. 1951. Crossing experiments with Culex strains. Evolution **5**:370–375.

Lo, N., C. Paraskevopoulos, K. Bourtzis, S. L. O'Neill, J. H. Werren, S. R. Bordenstein and C. Bandi. 2007. Taxonomic status of the intracellular bacterium *Wolbachia pipientis*. Int. J. Syst. Evol. Micr. **57**:654–657.

Louis, C. & L. Nigro. 1989. Ultrastructural Evidence of *Wolbachia* Rickettsiales in *Drosophila simulans* and Their Relationships with Unidirectional Cross-Incompatibility. J. Invertebr. Pathol. **54**:39–44.

Matsuda, M, H. T. Imai & Y. N. Tobari. 1983. Cytogenic analysis of recombination in males of *Drosophila ananassae*. Chromosoma **88**:286–292.

NCBI Trace Archive [online database]. URL: http://www.ncbi.nlm.nih.gov/Traces/trace.cgi. Date visited 2010-05-22.

Nikoh, N., K. Tanaka, F. Shibata, N. Kondo, M, Hizume, M. Shimada & T. Fukatsu. 2008. *Wolbachia* genome intergrated in an insect chromosome: Evolution and fate of laterally transferred endosymbiont genes. Genome Res. **18**:272–280.

O'Neill, S. L., R. Giordano, A. M. E. Colbert, T. L. Karr & H. M. Robertson. 1992. 16S rRNA phylogenetic analysis of the bacterial endosymbionts associated with cytoplasmic incompatibility in insects. Proc. Natl. Acad. Sci. USA. **89**:2699–2702.

Salzberg, S. L., J. C. Dunning Hotopp, A. L. Delcher, M. Pop, D. R. Smith, M. B. Eisen & W. C. Nelson. 2005. Serendipitous discovery of *Wolbachia* genomes in multiple *Drosophila* species. Genome Biol. **6**:R23.

Serbus, L. R., C. Casper-Lindley, F. Landmann and W. Sullivan. 2008. The genetics and cell biology of *Wolbachia*-host interactions. Annu. Rev. Genet. **42**:683–707.

Singh, B. N. 1996. Population and behaviour genetics of *Drosophila ananassae*. Genetica **97**:321–329.

Singh, S. R. & B. N. Singh. 2003. Behavioral genetics of *Drosophila ananassae*. Genet. Mol. Res. **2**:394–409.

Sinkins, S. P. & F. Gould. 2006. Gene drive systems for insect disease vectors. Nat. Rev. Genet. **7**:427–435.

Stouthamer, R., J. A. J. Breeuwer & G. D. D. Hurst. 1999. *Wolbachia pipentis*: microbial manipulator of arthropod reproduction. Annu. Rev. Microbiol. **53**:71–102.

Werren, J. H. 1997. Biology of *Wolbachia*. Annu. Rev. Entomol. **42**:587–609.

Werren, J. H., L. Baldo & M. E. Clark. 2008. *Wolbachia*: master manipulators of invertebrate biology. Nat. Rev. Microbiol. **6**:741–751.

Woolfit, M., I. Iturbe-Ormaetxe, E. A. McGraw & S. L. O'Neill. 2009. An ancient horizontal gene transfer between mosquito and the endosymbiont bacterium *Wolbachia pipientis*. Mol. Biol. Evol. **26**:367–374.

Wu, M., L. V. Sun, J. Vamathevan, M. Riegler, R. Deboy, J. C. Brownlie, E. A. McGraw, W. Martin, C. Esser, N. Ahmadinejad, C. Wiegand, R. Madupu, M. J. Beanan, L. M. Brinkac, S. C. Daugherty, A. S. Durkin, J. F. Kolonay, W. C. Nelson, Y. Mohamoud, P. Lee, K. Berry, M. Brook Young, T. Utterback, J. Weidman, W. C. Nierman, I. T. Paulsen, K. E. Nelson, H. Tettelin, S. L. O'Neill & J. A. Eisen. 2004. Phylogenomics of the reproductive parasite *Wolbachia pipientis w*Mel: a streamlined genome overrun by mobile genetic elements. PLoS Biol. **2**:0327–0341. DOI: 10.1371/journal.pbio.0020069.

Yen, J. H. & A. R. Barr. 1971. New hypothesis of the cause of cytoplasmic incompatibility in *Culex pipiens* L. Nature **232**:657–658.

Yen, J. H. & A. R. Barr. 1973. The Etiological Agent of Cytoplasmic Incompatibility in *Culex pipiens*. J. Invertebr. Pathol. **22**:242–250.

Appendix I: Description of ns-genes

Data from Lisa Klasson (unpublished).

ns_1. wRi pos. 62.646, 5 bp ins/variation in rplM, non-sense, polymorphic
ns_2. wRi pos. 97.064, 1 bp del in glycyl-tRNA-synthetase, non-sense, polymorphic
ns_3. wRi pos. 141.930, 7bp non-sense del in gyrB (WRi_001420), one bad qual read overlap
ns_4. wRi pos. 146.051, 4bp non-sense ins in WRi_001470, aconitate hydratase 1, polymorph
ns_5. wRi pos. 180.591, 16 bp ins, seq diff in ANK WRi_001790, one read same as wRi
ns_6. wRi pos. 253.688, 10bp non-sense del in WRi_002440 (miaB-like tRNA mod enzyme)
ns_7. wRi pos. 274.416, 4 bp del in WRi _002610 (competence lipoprotein ComL), non-sense, polymorphic
ns_8. wRi pos. 296.486, 1bp non-sense ins in hypo WRi_002820, polymorphic
ns_9. wRi pos. 303.535, 1 bp del in WRi_002890 (major facilitator), non-sense, polymorphic
ns_10. wRi pos. 306.572, 11bp ins in the few last bp's of WRi_002910, polymorphic
ns_11. wRi pos. 313.075, 8 bp del/variation in WRi _002960 (malonyl-coa decarboxylase),non-sense, variation same position wMel, polymorphic
ns_12. wRi pos. 338.923, 1bp ins in beginning WRi_003130 (glycosyl transferase), polymorph
ns_13. wRi pos. 354.685, 16 bp ins non-sense in WRi_003250 (pyruvate dehydrogenase, alpha sub), 1 read like wRi, 2 reads with insert.
ns_14. wRi pos. 437.725, 1 bp del in WRi_004080 (bicyclomycin resistance protein), non-sense, polymorphic
ns_15. wRi pos. 457.415, 2 bp non-sense ins in WRi_004260 (ispE), polymorpic
ns_16. wRi pos. 480.977-986, 7bp del/variation non-sense in WRi_004490 (adenylosuccinate synthase, purA), polymorphic
ns_17. wRi pos. 584.683, 7bp ins/variation non-sense in WRi_005450 ANK (WD0637), polymorphic
ns_18 and ns19. wRi pos. 611.393, 1bp ins and 1 subst non-sense in WRi_005720 (helicase, SNF2 fam), polymorphic.
ns_20. wRi pos. 620.810, 5bp del non-sense in WRi_005780 (ABC transporter), polymorphic
ns_21. Ri pos. 625.198, 2 bp ins/variation non-sense in WRi_005810 (Major facilitator family), polymorphic.
ns_22. wRi pos. 633.535, 29bp del over the start of WRi_005890 (hypo, duplicated gene in breakpoint), polymorphic.
ns_23 .wRi pos. 689.846, 4bp del non-sense in WRi_006380 (peptidyl-prolyl cis-trans isomerase D, put) polymorphic.
ns_24. wRi pos. 717.204, 5bp del non-sense in WRi_006640 (hypo, not in wMel or pseudo?) in breakpoint, polymorphic.
ns_25. wRi pos. 727652, 1 bp del in WRi_006710 (hypothetical), non-sense, polymorphic
ns_26. wRi pos. 758.562, 11bp del non-sense in WRi_006940 (tail tape measure WO-C), polymorphic.
ns_27. wRi pos. 801.350, 3bp del in WRi_007390 (hypo, unique wRi?, low sim to wBm, but no gene), polymorphic.
ns_27. wRi pos. 801.690, 1 bp del in WRi_007390 (hypo, unique wRi?), non-sense, polymorphic, not the same reads as above.
ns_28. wRi pos. 833.981, 1bp ins/variation non-sense in WRi_007710 (hydrolase alpha/beta fold fam), polymorph
ns_29. wRi pos. 872.940, 1 bp del inWRi_08110 (hypo, putative outer mebrane), non-sense, polymorphic
ns_30. wRi pos. 966.011, 1 bp del in WRi_008900 (cytochrome c oxidase assembly protein), non-sense, polymorphic
ns_31. wRi pos. 973.513, 2bp del non-sense in WRi_008970 (amino acid permease family protein, this gene is tandemly duplicated in wPip and wBm), polymorphic.
ns_32. wRi pos. 985.924, 4bp ins non-sense in WRi_009120 (nuoK), 1 hq read and 1 lq read with ins.
ns_33. wRi pos. 1.010.833, 5bp ins non-sense in WRi_009380 (hypo, wRi-specific, SignalP 2xTM), polymorphic.
ns_34. wRi pos. 1.017.082, 1bp del non-sense in WRi_009450 (iscS cysteine desulfurase), polymorphic.
ns_34. wRi pos. 1.017.252, 2bp ins non-sense in WRi_009450 (iscS cysteine desulfurase), not the same reads as above, polymorphic.
ns_34. wRi pos. 1.017.277, 4bp del non-sense in WRi_009450 (iscS), not the same reads as above, polymorphic.
ns_35. wRi pos. 1.032.571, 4bp del non-sense in WRi_009600 (mnmG aka gidA), polymorphic.
ns_35. wRi pos. 1.032.717, 25bp del non-sense in WRi_009600, same reads as above, polymorphic.
ns_36. wRi pos. 1.046.165, 10bp del non-sense in WRi_009730 (hypo), polymorphic.
ns_37. wRi pos. 1.197.422, 8bp ins non-sense in WRi_011070 (petB, cytochrome b), tandem duplication?, polymorphic.

ns_38. wRi pos. 1.307.332, 1 bp del in WRi_012150 (pyrD, dihydroorotate dehydrogenase), non-sense, polymorphic
ns_39. wRi pos. 1.309.568, 8bp del non-sense in WRi_012170 (hypo, for other gene order see Contig5384), polymorphic
ns_40. wRi pos. 1.336.275, 11bp ins/variation non-sense in WRi_012430 (penicillin-binding protein, check Contig5402), not the same reads that become unaligned downstream, polymorphic
ns_40. wRi pos. 1.336.510, 1bp del non-sense in WRi_012440 (lytB- essential gene in E.coli), polymorphic
ns_41. wRi pos. 1.410.376, 8bp del/variation non-sense in WRi_013200 (carB), polymorphic
ns_42. wRi pos. 1.414.799, 17bp del/variation non-sense in WRi_013260 (rpoD), polymorphic, not the same reads that are unaligned

Appendix II: ns-genes and IS-elements mapped on the *w*Ri genome

ns-genes are represented by the pink lines, IS-elements by blue lines, both in the innermost circle.