universität wien

DIPLOMARBEIT

Titel der Diplomarbeit

Drosophila paulistorum trinity: Strain typing, immunity, mitochondrial transmission.

Verfasser

Wolfgang Däuble

angestrebter akademischer Grad

Magister der Naturwissenschaften (Mag.rer.nat.)

Wien, 2013

Studienkennzahl lt. Studienblatt:	A 490
Studienrichtung lt. Studienblatt:	Molekulare Biologie
Betreuerin / Betreuer:	Prof. Dr. Michael Jantsch

Mit größtem Dank an Wolfgang Miller, der mir mehr beibrachte als jeder andere, und Daniela Schneider, ohne deren Rat und Hilfe ich nie soweit gekommen wäre. Und an Olli P, dessen Prügel mir durch den Tag halfen. Sie werden weiter hupen!

Auf dem Fliegenplaneten

Auf dem Fliegenplaneten,
da geht es dem Menschen nicht gut:
Denn was er hier der Fliege,
die Fliege dort ihm tut.

An Bändern voll Honig kleben
die Menschen dort allesamt,
und andere sind zum Verleben
in süßliches Bier verdammt.

In einem nur scheinen die Fliegen
dem Menschen vorauszustehn:
Man bäckt uns nicht in Semmeln,
noch trinkt man uns aus Versehn.

- Christian Morgenstern -

Index of Abbreviations

A	Adenosine
A28	Classic strain of described *D. paulistorum* Amazonian semispecies
AB	*D. paulistorum* Andean-Brazilian semispecies
Abs	Absolute
Adh	*Alcohol dehydrogenase* gene
AM	*D. paulistorum* Amazonian semispecies
Amd	*Alpha methyldopa* gene
Amp	Ampicillin
AMP	Antimicrobial Peptide
AttA	*Attacin A* gene
AttC	*Attacin C* gene
AttD	*Attacin D* gene
bp	Base pairs
C	Cytosine
C°	Degree Celsius
C2	Classic strain of described *D. paulistorum* Centroamerican semispecies
CA	*D. paulistorum* Centroamerican semispecies
CI	Cytoplasmic incompatibility
cm	Centimetre
CO1	*Cytochrome c oxidase subunit 1* gene
CO2	*Cytochrome c oxidase subunit 2* gene
dCTP	Deoxycytidine-5´-triphosphate
DH5α	*Escherichia coli* strain for competent cells
Dipt	*Diptericin* gene
Dipt-B	*Diptericin B* gene
DNA	Deoxyribonucleic acid
Dro	*Drosocin* gene
EDTA	Ethylenediaminetetraacetic acid
dNTP	2´-deoxyribonucleoside-5´-triphosphate
G	Guanosine
g	Gram
hb	*Hunchback* gene
Hmu	*Hemomucin* gene
hrs	Hours
i.e	Id est
I1	Classic strain of described *D. paulistorum* Interior semispecies
IN/INT	*D. paulistorum* Interior semispecies
IPTG	Isopropyl-β-D-1-thiogalactopyranoside
kb	Kilobases
KCl	Potassium chloride
lacZ	Gene encoding for the enzyme *β-galactosidase*
l	Litre
M	Molar
m	Milli
M13	Sequencing primer

MS	Mesitas, classic strain of described *D. paulistorum* Andean-Brazilian semispecies
mf	Multi fly
MgCl2	Magnesium chloride
min	Minute(s)
N	Normal
O11	Classic strain of described *D. paulistorum* Orinocan semispecies
o.n.	Over night
OR	*D. paulistorum* Orinocan semispecies
NaOH	Sodium hydroxide
Per	*Period* gene
PBS	Phosphate buffered saline
PEG	Polyethylenglycole
PCR	Polymerase chain reaction
pH “	“Potentia hydrogenii”
RFLP	Restriction fragment length polymorphism
RNase	Ribonuclease
rpm	Rounds per minute
SDS	Sodiumdodecylsulfate
sec	Second(s)
SM	Santa Marta, classic strain of described *D. paulistorum* Transitional semispecies
SOC	”Salt optimized + carbon”
SrC1	*Scavenger receptor class C subunit 1* gene
SSC	Sodium salt citrate
T	Thymidine
TAE	Tris[aminomethyl]aminoethane
TE-Buffer	Tris-EDTA-Buffer
TR	*D. paulistorum* Transitional semispecies
Tris	Tris(hydroxymethyl)aminomethane
U	Unit
UV	Ultraviolet (light)
V	Volts
w1118	*Drosophila melanogaster* white mutant strain
wsp	*Wolbachia* surface protein
w/v	Weight to volume
X-Gal	5-bromo-4-chloro-3-indolyl- β-D-Galactopyranoside

Abstract

Drosophila paulistorum is a neotropical fruit fly in *statu nascendi*, i.e. under on-going evolutionary segregation into multiple new species. Six described *D. paulistorum* semispecies are currently assigned to this species complex based on mating compatibilities. The distinction and the phylogenetic classification of these semispecies provide a big challenge, as they are morphologically indistinguishable and no reliable molecular marker system has been published yet. In addition, the existing phylogeny of *D. paulistorum* is conflicting at nuclear and mitochondrial level. All of the described semispecies harbour distinct strains of *Wolbachia*, maternally transmitted intracellular bacteria, living in mutualistic symbiosis and expressing reproductive phenotypes in their hosts. These phenotypes induce pre- and postmating isolation of sympatric populations of different semispecies. Male hybrids of the semispecies, hatching only under laboratory conditions, are completely sterile. Antibiotic treatment of the flies reduces the *Wolbachia* titer to a minimum and can partially rescue those reproductive phenotypes. Under those conditions, hybrid lines of *D. paulistorum* semispecies were generated in previous studies. Remarkably, such intersemispecies hybrids showed signatures of paternal *Wolbachia* transmission.

The first aim of this thesis was to find and examine highly informative nuclear and mitochondrial genes to improve the phylogeny of the *D. paulistorum* species complex and to resolve the conflicts between the published nuclear and mitochondrial data. For this purpose, genes for antimicrobial peptides (AMPs) were chosen at nuclear level, as previous studies showed that their rate of evolution is elevated in *Drosophila* species. As a mitochondrial marker gene, *NADH dehydrogenase subunit 4* (ND4) was selected. The second aim was to create with those sequences a genetic strain typing system in order to group newly collected and unclassified *D. paulistorum* strains in the system of described semispecies. Additionally, the gene for *Attacin D*, an AMP potentially acting intracellularly, was examined for signs of coevolution with *Wolbachia.* Finally, we wanted to investigate the mitochondrial transmission mode in hybrid lines mentioned above, as their paternally inherited *Wolbachia* are potentially hitchhiking with the paternal mitotype.

Sequence analysis of the six AMP genes *Attacin A, -C, -D, Diptericin, Diptericin-B* and *Drosocin* revealed high levels (2.4% - 7.9%) of nucleotide diversity within the *D. paulistorum* semispecies. Three of those AMP genes showed multiple paralogues in some semispecies. The highest nucleotide divergence and paralogue abundance was found in *Attacin D.* Although K_a/K_s ratio analysis of this gene provided signs of negative selection, the maintenance of the divergent paralogues and the high level of nucleotide divergence of this gene were interpreted as an adaptation to a divergent set of symbionts. *ND4* sequencing and Restriction fragment length polymorphism (RFLP) analysis via P^{32}-labelled Southern blot hybridization showed that some semispecies harbour multiple mitochondrial haplotypes. By using the *ND4* data and the nuclear *Attacin C,* and - *D* sequences, a molecular marker system was established. It allowed the unambiguous integration of three unclassified strains into

the system of described semispecies. Phylogenetic trees established in this study provide a higher resolution than the previously published *D. paulistorum* phylogenies. However, they are also inconsistent at nuclear and mitochondrial level. The finding of multiple mitotypes and the proof of paternal transmission of mitochondria in *D. paulistorum* hybrids, produced in this work, serve as a possible explanation for these incongruence.

Zusammenfassung

Drosophila paulistorum ist eine neotropische Fruchtfliegenart in *statu nascendi,* d.h., sie spaltet sich derzeit in mehrere neue Spezies auf. Bisher werden auf Grund von Kreuzungsinkompabilitäten sechs beschriebene Semispezies zu dem *D. paulistorum* Spezieskomplex gezählt. Deren Unterscheidung und phylogenetische Klassifizierung stellen eine große Herausforderung dar, da sie keine morphologischen Eigenheiten besitzen und bisher keine zuverlässigen molekularen Marker veröffentlicht wurden. Darüber hinaus weisen die publizierten mitochondrialen und nukleären Phylogenien Widersprüche auf. Jede der beschriebenen Semispezies besitzt einen eigenen Stamm an *Wolbachia* maternal vererbte, intrazelluläre Bakterien, die in mutualistischer Symbiose mit den Fliegen leben und reproduktive Phänotypen in ihren Wirten exprimieren. Diese führen zu sexueller Isolation sympatrischer Populationen unterschiedlicher Semispezies. Männliche Hybride, welche ausschliesslich unter Laborbedingungen durch Kreuzung verschiedener Semispezies erzeugt werden können, sind steril. Durch Behandlung der Fliegen mit Antibiotika können der *Wolbachia* Titer auf ein Minimum gesenkt und somit die reproduktiven Phänotypen teilweise umgangen werden. Unter solchen Bedingungen wurden in vorrangegangenen Studien stabile Hybridlinien von *D. paulistorum* Semispezies generiert, die Signaturen einer paternalen Vererbung der *Wolbachien* aufwiesen.

Das erste Ziel dieser Arbeit war es, hochinformative nukleäre und mitochondriale Marker zu finden um die Phylogenie des *D. paulistorum*-Komplexes zu verbessern und die Widersprüche zwischen den publizierten mitochondrialen und nukleären Daten zu klären. Zu diesem Zweck wurden Gene für antimikrobielle Peptide (AMPs) als nukleäre Marker ausgesucht, da vorangegangene Studien gezeigt haben, dass diese Gene erhöhte Evolutionsraten in Drosophila-Spezies aufweisen. Als mitochondrieller Marker wurde das Gen für die *NADH Dehydrogenase Untereinheit 4* ausgewählt. Das zweite Ziel dieser Arbeit war das Erstellen eines genetischen Typisierungssystems, um mit Hilfe der oben beschriebenen Markergene neu gesammelte oder unklassifizierte *D. paulistorum* Stämme in das System der sechs beschriebenen Semispezies genotypisch einzuordnen. Zudem sollte das Gen für *Attacin D*, ein AMP welches potentiell intrazellulär wirkt, auf Hinweise für eine Koevolution mit *Wolbachien* untersucht werden. Als letztes Ziel sollte die mitochondrielle Vererbung in den oben erwähnten stabilen Hybridlinien untersucht

werden, da *Wolbachien* gemeinsam mit Mitochondrien vererbt werden, was eine paternale Übertragung der Mitochondrien in diesen Hybriden vermuten lässt.

Die Sequenzanalyse der sechs AMP-Gene *Attacin A, -C, -D,* und *Diptericin-B* und *Drosocin* zeigte eine hohe Rate (2.4 - 7.9%) an unterschiedlichen Nukleotiden innerhalb der *D. paulistorum* Semispezies. Für drei dieser AMPs wurden mehrere Paraloge in manchen Semispezies entdeckt. Die unterschiedlichsten Nukleotidsequenzen und die meisten Paraloge fanden sich für *Attacin D.* Obwohl eine Analyse der K_a/K_s-Verhältnisse dieses Gens Hinweise auf negative Selektion hervorbrachte, kann die Aufrechterhaltung der hohen Nukleotiddiversität und der unterschiedlichen Paralogen als eine Adaption an Semispezies-spezifische Symbionten gedeutet werden. Die Sequenzierung und Restriktionsfragmentslängen-polymorphismus- (RFLP-) Analyse durch P^{32}- markierte Southern-Blot-Hybridisierung des mitochondriellen *ND4*-Gens ergab, dass manche Semispezies mehrere mitochondrielle Haplotypen besitzen. Mit Hilfe der *ND4-, Attacin C-* und *–D-*Datensätze wurde ein molekulares Markersystem erstellt, welches die eindeutige Zuordnung dreier unklassifizierter Stämme in das System der sechs Semispezies erlaubte. Darüberhinaus konnten in dieser Studie phylogenetische Bäume mit höherer Auflösung erstellt werden als die der bisher veröffentlichten *D. paulistorum* Phylogenien. Es traten aber ebenfalls Widersprüche zwischen nukleären und mitochondriellen Daten auf. Die Entdeckung von mehreren Mitotypen und der in dieser Studie erbrachte Beweis der paternalen Vererbung von Mitochondrien in *D. paulistorum*-Hybriden kann als eine mögliche Erklärung für diese Widersprüche herangezogen werden.

Table of Contents

1. Introduction

1.1 The species complex of *Drosophila paulistorum* among its relatives

Drosophila paulistorum is a neotropical fruit fly, whose habitat stretches from Central America to southern Brazil (Dobzhansky and Pavan in Burla et al., 1949; Dobzhansky and Spassky, 1959). As a member of the *willistoni* subgroup of *Drosophila*, *D. paulistorum* is morphologically almost indistinguishable from its sibling species *D. equinoxialis*, *D. insularis*, *D. pavlovskiana*, *D. tropicalis* and *D. willistoni* (Dobzhansky, 1946; Burla et al., 1949; Dobzhansky et al., 1957; Kastritsis and Dobzhansky, 1966; and shown in **Fig. 1.1**). Those close relatives can nevertheless be assigned to their distinct species by differences in the chromosome strands in the nuclei of their larval salivary glands (Burla et al., 1949), slight differences in the male genitalia (Spassky, 1957) and crossing experiments (Dobzhansky et al., 1957), as crossmating of the species give rise to sterile hybrids of both sex if forced to mate under laboratory conditions. As an exception, Spassky reported crossbreedings between *D. paulistorum* and *D. pavlovskiana,* which led to sterile male but fertile female hybrids (Kastritsis and Dobzhansky, 1966). More recent differentiation methods use the identification of allozymes of acid phosphatase-1 through electrophoresis (Garcia et al. 2006). As difficult as it is to determine the taxonomy of this group, the differentiation into subspecies provides an even bigger challenge: *D. equinoxialis*, *D. willistoni*, and *D. tropicalis* divide into two subspecies, respectively (Townsend, 1954; Ayala and Tracey, 1973; Ayala et al., 1974).

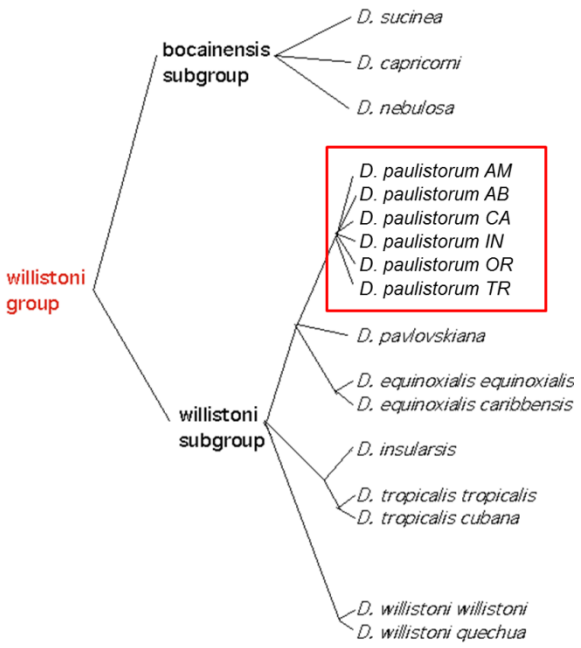


Fig.1.1: Consensus tree of the *willistoni* group, established with data obtained by analysis of polytene chromosomes, mating behaviour and initial DNA sequences (after Powell, 1997). *D. paulistorum* species complex is highlighted in red square.

The taxonomic complexity of the *willistoni* group is taken to a further level with the *D. paulistorum* superspecies, which comprises at least six semispecies: Amazonian (AM), Andean-Brazilian (AB), Centroamerican (CA), Interior (IN), Orinocan (OR) and Transitional (TR, Dobzhansky and Spassky, 1959; Pérez-Salas et al., 1970). Semispecies are defined after Ernst Mayr as "*natural populations which have not yet fully completed the process of speciation*", as gene exchange among the semispecies is still possible but not as freely as among ordinary conspecific populations. They have thereby "*acquired some but not all properties of species*" (Mayr, 1970). In the case of *D. paulistorum*, this constrained gene exchange is primarily caused by geographic isolation, but even in those territories shared by two or more semispecies, intrinsic isolation mechanisms are operative. Those mechanisms, divided into premating isolation (Dobzhansky and Pavlovsky ,1971), mediated by females rejecting males of other semispecies, and postmating isolation, caused by a high embryonic mortality (Daniels and Ehrmann, 1974; Bates et al., 1977) and hybrid male sterility (Dobzhansky and Pavlovsky, 1971) are the main mechanisms to differentiate *D. paulistorum* semispecies. Theoretically, however, gene exchange is still possible, as strains of the Transitional semispecies were reported in the 1960s to represent bridging populations, which are capable of producing fertile hybrids with other semispecies under laboratory conditions (Ehrmann, 1962). Due to the on-going speciation, the *D. paulistorum* superspecies complex is an ideal system for studies on evolutionary mechanisms and speciation genetics in statu nascendi.

The six semispecies are represented by strains collected by Lee Ehrmann in the 1950s, which were kept as inbreeding, isofemale lines in her laboratory and have served as reference ever since (**Table 1.1**).

Semispecies	Strain	Origin	Coll. Date
Amazonian	**A28**	Brazil	1952
Andean-Brazilian	**MS**	Colombia	1962
Centroamerican	**C2**	Honduras	1954
Interior	**I1**	Colombia	1958
Orinocan	**O11**	Guyana	1957
Transitional	**SM**	Colombia	1956

Table 1.1: "Classic" representative strains of *D. paulistorum* semispecies used in this thesis, collected by Lee Ehrmann.

1.1.1 *Drosophila paulistorum* phylogeny

Two studies, based on nuclear and mitochondrial DNA markers, investigated the phylogeny of *D. paulistorum* semispecies and their relationship with other members of the *D. willistoni* group (Gleason et al., 1998; Robe et al., 2010) with partially conflicting results. In both studies, however, incongruence between mitochondrial and nuclear sequence data were observed. Gleason et al. (1998) suggested, based on parts of the coding sequence of the nuclear *period* (*per*) locus that the Amazonian semispecies diverged first, followed by a clade comprising the Andean-Brazilian and Orinocan semispecies (**Fig. 1.2 A**). The most recent clade branches off into two lineages: the first clusters Centroamerican semispecies with *D. pavlovskiana*, the second one groups Interior with Transitional semispecies. However, bootstrap values of 68, 63 and 65 per cent do not provide high confidence for the topology of this tree. Their mitochondrial dataset included the gene for *cytochrome c oxidase subunit 1* (*CO1*) dataset (**Fig. 1.2 B**). Based on *CO1* analyses Gleason et al. showed that within the *willistoni* group, *D. paulistorum* semispecies group together, with the exception of the Amazonian semispecies. Those nodes, however, in most cases have bootstrap values below 50%. For the nodes separating Centroamerican and Interior semispecies, bootstrap values of 63 and 67 %, respectively, are indicated. The low resolution of this mitochondrial marker makes it difficult to compare the mitochondrial and nuclear topologies.

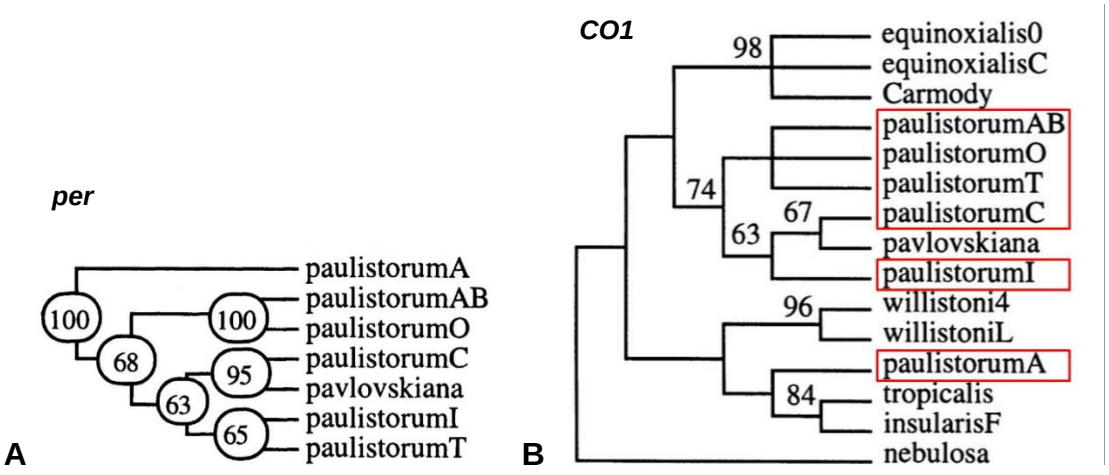


Fig. 1.2 A: Extract of bootstrap 50%-majority-rule consensus tree (maximum parsimony) based on *period* (*per*) nuclear data set showing *D. paulistorum* semispecies Amazonian (A), Andean-Brazilian (AB), Orinocan (O), Centroamerican (C), Interior (I) and Transitional (T). Bootstrap values for each node are shown in ellipses. **B:** Most-parsimonious tree based on *CO1* dataset showing *D. paulistorum* semispecies in red boxes among other members of the *willistoni* group. Bootstrap values greater than 50% of 100 replicates are shown at nodes. Figures after Gleason et al. (1998).

More recently, Robe et al. (2010) included the nuclear genes for *alcohol dehydrogenase* (*adh*)*, alpha methyldopa* (*amd*), *hunchback* (*hb*) and *period* (*per*) in their analysis and compared the results with data of mitochondrial genes *CO1* and *cytochrome c oxidase subunit 2* (*CO2*). Based on the results for the nuclear genes, they suggested, in accordance with Gleason et al. (1998) that the Amazonian semispecies branched off first, followed by the separation of the remaining clade into two lineages: The first clusters the Orinocan with Andean-Brazilian semispecies, the second consisting of the Centroamerican, Interior and Transitional (**Fig. 1.3 A**). The grouping of the Transitional and Centroamerican semispecies differs from the study by Gleason.

Combined data of mitochondrial genes *CO1* and *CO2*, analysed by Robe and colleagues (2010), however, show a very different clustering, integrating *D. equinoxialis* and *D. pavlovskiana* species in the clade of *D. paulistorum* (**Fig 1.3 B**). Importantly, the bootstrap values of the mitochondrial tree are very low for the basic nodes, and more stringent tree building methods do not provide any support for this topology (indicated by hyphens). The lowest bootstrap is given for the node separating Amazonian from Andean-Brazilian semispecies with 36% using maximum likelihood. Neighbour joining and maximum parsimony algorithms do not provide any support for this branching.

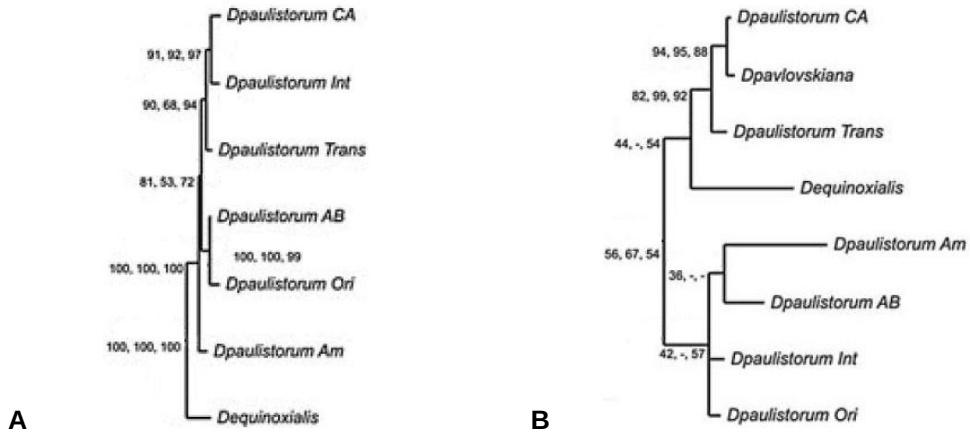


Fig. 1.3: Extract of Bayesian majority-rule consensus trees showing *D. paulistorum* semispecies Amazonian (Am), Andean-Brazilian (AB), Orinocan (Ori), Centroamerican (CA), Interior (Int), Transitional (Trans), *D. equinoxialis* and *D. pavlovskiana.* **A:** based on combined sequence data of nuclear *adh, amd, hb* and *per* genes, **B:** based on combined sequence data of mitochondrial *CO1 and CO2* genes. Numbers at the nodes are the bootstrap supports for each clade in the maximum likelihood, neighbour joining and maximum parsimony trees, respectively. Figures after Robe et al. (2010).

Robe and colleagues (2010) also examined four new strains of *D. paulistorum*; one from the Amazonas (Pa) and three collected in South Brazil (JAI, MLC, RP). The authors, however, designated these four new Brazilian strains exclusively on the basis of their geographic origin (collection site) and not on the basis of their respective mating compatibilities with any of the six classic semispecies and hence they designated the Pa strain as Amazonian and the three South Brazilian strains as Andean-Brazilian semispecies. As *D. paulistorum* semispecies are sympatric, this classification represents a conceptual mistake.

The unsolved phylogeny of the *D. paulistorum* superspecies, especially at mitochondrial level, and the conflicting results between nuclear and mitochondrial data demand further studies with mitochondrial genes, which provide a higher resolution. As *CO1* and *CO2* genes are not informative enough, one aim of this thesis was to examine the mitochondrial gene *NADH dehydrogenase subunit 4* (*ND4*) in order to contribute to the understanding of the relationships between the *D. paulistorum* semispecies. *NADH-dehydrogenase* subunits are known to be fast evolving mitochondrial polypeptides in *Drosophila* (Garesse, 1988; Moriyama and Powell, 1997) and should therefore be well suited for investigation of closely related species. Furthermore, a marker system will be established in order to classify newly collected strains.

1.2 The endosymbiont *Wolbachia*

Robe et al. (2010) raised the possibility that inconsistencies between mitochondrial and nuclear phylogenies in *D. paulistorum* might be associated with endosymbiotic bacteria *Wolbachia.* Those gram-negative α-proteobacteria were first discovered by Hertig and Wolbach (1924) and described as “rickettsia-like organisms” (RLOs) in the mosquito *Culex pipiens*. As obligate intracellular bacteria, they need a host system to survive and reproduce. Up to 76% of all arthropods (Jeyaprakash and Hoy, 2000), including insects, spiders and mites, but also filarial nematodes (Bandi et al., 1998) are infected with *Wolbachia*. As intracellular bacteria, their main transmission pathway within a species is via the maternal cytoplasm of the egg, using the same route as mitochondria. The widespread distribution of *Wolbachia* in arthropods, however, is hard to explain with exclusively vertical transfer. An increasing number of cases shows that horizontal transfer from one taxa to another and hybrid introgression of *Wolbachia*, as well as multiple infections with different *Wolbachia* strains can be observed in nature (Raychoudhury et al., 2011).

Furthermore, *Wolbachia* induce several reproductive phenotypes, which increase their frequency within an infected population: The most common of those phenotypes is cytoplasmic incompatibility (CI), first described by Laven in the 1950´s, although without knowledge of *Wolbachia* as causative agents (Laven, 1951). CI is mediated by sperm from *Wolbachia*-infected males, which is incompatible with eggs from

females that are not infected with the same type (or types) of the bacteria (Yen and Barr, 1971; reviewed by Werren, 2008). CI can be uni- or bidirectional (**Fig. 1.4**): unidirectional CI is mediated by a single *Wolbachia* infection within a population, causing CI if infected males mate with uninfected females but leaving the progeny of an uninfected male and an infected female unharmed. Bidirectional CI, however, is caused by multiple *Wolbachia* infections, if males infected with one strain mate with females harbouring the other.

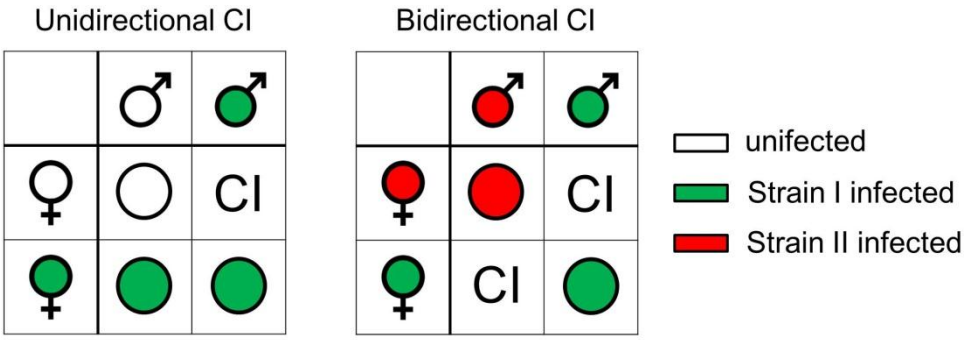


Fig. 1.4: Uni- vs. bidirectional Cytoplasmic Incompatibility (CI). Single *Wolbachi*a infections cause unidirectional CI if infected males mate with uninfected females. Bidirectional CI is caused by multiple strains of *Wolbachia* if both mating partners do not harbour the same strain. Picture adapted from the Wikimedia commons file "Uni bidirectional ci.png" (http://en.wikipedia.org/wiki/File:Uni_bidirectional_ci.png

A second phenotype associated with *Wolbachia* is parthenogenesis (Stouthamer et al., 1990) in species were males develop from unfertilized eggs (e.g. mites and wasps). Unfertilized eggs infected with *Wolbachia* develop to females, which are able to transmit the endosymbiont to their offspring. Other phenotypes also increase the amount of infected females in a population by feminization of genetic males (Rigaud et al., 1991; Bouchon et al. 1998) and male-killing during embryogenesis (Hurst et al., 1999). A summary of the *Wolbachia*-induced, reproductive phenotypes is shown in **Figure 1.5**.

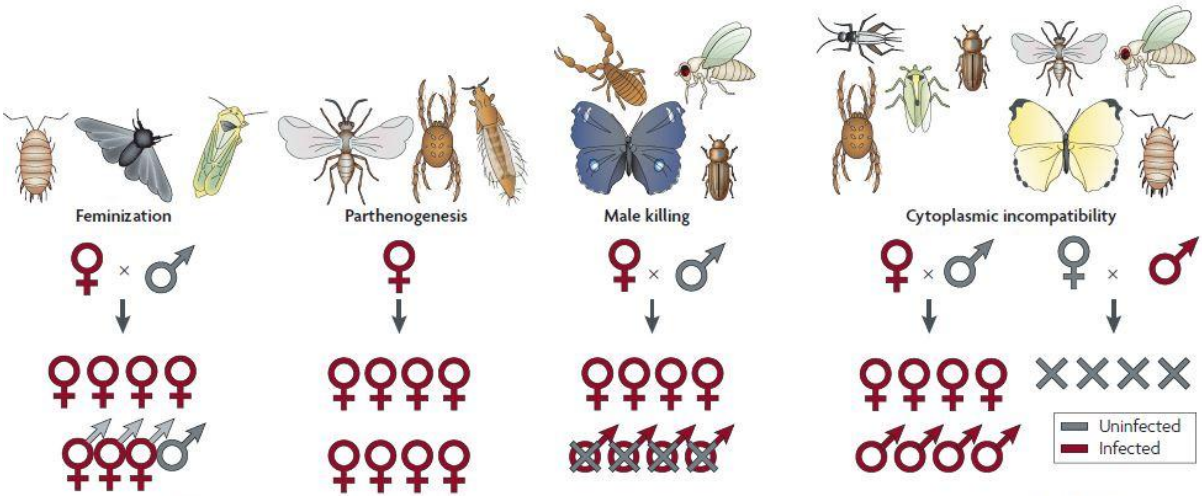


Fig. 1.5 *Wolbachia*-induced phenotypes: Feminization results in transformation of genetic males into females in the orders of Hemiptera, Isopoda and Lepidoptera; Parthenogenesis leads in the orders Acari, Hymenoptera and Thysanoptera to a 100% female progeny; Male killing of infected males during embryogenesis in the orders Coleoptera, Diptera, Lepidoptera and Pseudoscorpiones leaves all resources to the female progeny; Cytoplasmic incompatibility in the orders of Acari, Coleoptera, Diptera, Hemiptera, Hymenoptera, Isopoda, Lepidoptera and Orthoptera prevents reproduction of infected males with females that do not share the same infection. Figure after Werren et al. (2008).

1.2.1 Reproductive phenotypes in hybrids of *D. paulistorum* semispecies

If *D. paulistorum* semispecies are forced to mate under laboratory conditions, very few eggs develop. The high mortality of those F1 hybrids was found to be caused by an abnormal cell pole development in early embryonic stages of the hybrids (Daniels and Ehrman, 1974; Bates et al., 1977). Rarely hatching F1 hybrids, however, showed complete male sterility (Dobzhansky and Pavlovsky, 1971), except for hybrids with transitional semispecies, as described above. Transinfection studies were performed, transferring extracts from testes of male F1 hybrids to females of the parental strain (Williamson et al., 1971). If those females were crossed with males from their own semispecies, the progeny also showed male sterility. Although at a nuclear level nonhybrid, they expressed a similar phenotype as F1 hybrids. Those effects are observed in both reciprocal crosses between *D. paulistorum* semispecies and are not caused by nuclear factors, hence it can be described as bidirectional CI.

The first evidence to answer the question for the causative agent of both bidirectional CI and hybrid male sterility in *D. paulistorum* was found in 1970, when *mycoplasma-like organisms* (MLOs) were detected in ovaries, testes and early embryos (Kernaghan and Ehrman, 1970). Further experiments showed that in testes of hybrid males those MLOs overreplicate (Ehrman and Kernaghan, 1972). Attempts to create MLO-free *D. paulistorum* strains by antibiotic treatment failed, however, the MLO titre in hybrid testes could be diminished with low concentrations of antibiotics to a level where the hybrid male sterility was partially rescued (Kernaghan and Ehrman, 1970). Miller and co-workers showed recently that formerly described MLOs are most likely *Wolbachia*, of which each *D. paulistorum* semispecies harbours a distinct strain

(Miller et al., 2010). They are considered to have a crucial impact on the on-going speciation process in the *D. paulistorum* species complex by triggering pre- and postmating isolation.

1.3 Mitochondrial transmission

Wolbachia are co-transmitted with mitochondria, whose canonical transmission pathway is strictly maternal in animals. The resulting homoplasy and, therefore, lack of recombination of mitochondrial DNA is the premise for its use in phylogenetic analyses. Three main mechanisms are thought to inhibit paternal mitochondria from leakage and propagation into the offspring generations: First, the sheer number of maternal mitochondria in the egg dilutes the vastly outnumbered paternal mitotype in the zygote (Birky, 2001). Secondly, ubiquitination of paternal cellular structures leads to destruction of sperm mitochondria in the zygote (first observed in mammals by Sutovsky et al. 2000). Thirdly, a germ line bottleneck for mitochondria that discriminates against minor mitochondrial haplotypes (mitotypes) should prevent rare heteroplasy events to be passed on to further generations (Bergstrom and Pritchard, 1998). Furthermore, recent observations were made that sperm mitochondria in *Drosophila melanogaster* are disposed of DNA during sperm development even before fertilization (DeLuca and O'Farrell, 2012).

1.3.1 Paternal leakage and heteroplasy

Increasing evidence, however, showed that in different genera like bivalves (reviewed by Breton et al., 2007), *Drosophila* (Kondo et al. 1992), birds (Kvist et al., 2003) and mice (Gyllenstein et al., 1991), paternal transmission of mitochondria is extremely rare but possible. Furthermore, natural populations of East African *Drosophila simulans* showed unexpected high rates (6-12%) of heteroplasy for two distinct mitotypes (Satta et al., 1988; Matsuura et al., 1991; Dean et al., 2003), which, in some cases, could be maintained in isofemale lines for six years under laboratory conditions (Satta et al., 1988). In these populations, paternal leakage seems to be a regular feature of mitochondrial inheritance rather than the exception to the strict maternal rule. This assumption has been recently reinforced by crossing experiments of two distinctive *D. simulans* mitotypes, which showed a non-random (pair-specific) occurrence of paternal leakage with a frequency of 0.66% per generation (Wolff et al., 2012).

1.3.2 Association of *Wolbachia* with mitochondrial haplotypes

The "hitchhiking" of *Wolbachia* with mitochondria affects the prevalence of mitochondrial haplotypes if selective pressure is forced upon associated *Wolbachia* strains, as observed during *Wolbachia* sweeps of *D. simulans* laboratory strains

(Nigro and Prout, 1990) and natural populations (Turelli and Hoffmann, 1991). The question arises if the paternal leakage and heteroplasy of mitochondria described above for *D. simulans* can be adapted or even traced back to the infection with distinct *Wolbachia* strains. In *D. paulistorum*, paternal transmission of *Wolbachia* has been observed in hybrids of Amazonian and Orinocan as well as of Centroamerican and Amazonian semispecies (Miller and Schneider, unpublished data). Whether those hybrids also show paternal leakage of mitochondria has hence been examined in this thesis.

1.4 Insect immunity

The fact that *D. paulistorum* semispecies are infected with distinct obligate mutualistic *Wolbachia* strains (Miller et al., 2010) raises the question in what way their immune system interacts with those gram negative bacteria. The effector molecules of the *D. paulistorum* immune system have to differentiate between harmful pathogens and the mutualist essential for survival. The molecular defence against microbes of insects, like all invertebrates, relies solely on the innate immune system. Insects often inhabit, feed on or breed in microbe-rich environments like decomposing media and developed therefore a sophisticated innate immunity response. Its cellular component is mediated by hemocytes circulating in the hemolymph (reviewed by Nappi et al., 2004). Their function comprises phagocytosis and subsequent destruction by fusion with phagosomes containing lysozymes, reactive oxygen species and nitric oxide (Jones et al., 1999), nodulation by formation of multicellular aggregates (Miller et al., 1999) and encapsulation followed by melanisation and killing of the invader with reactive oxygen- and nitrogen species (Nappi et al., 1995).

The humoral response consists primarily of antimicrobial peptides (AMPs) that are mainly produced in the fatbody of insects, to a small amount also in the hemocytes and the epithelium, and secreted into the hemolymph (Ferrandon et al., 1998). More than 150 of those small, cationic effector molecules have been isolated and described in insects and were found to fight a broad range of microbial pathogens (reviewed by Zasloff, 2002). In *Drosophila melanogaster*, 20 AMPs have been identified, which can be classified by their structure and target into seven groups (Imler and Bulet, 2005; **Fig. 1.6**): Drosomycins and Metchnikowin, targeting fungi; Defensin, effective against gram-positive bacteria; Attacins, Cecropins, Diptericins and Drosocin, killing gram-negative bacteria.

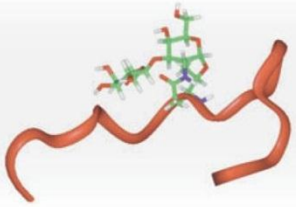
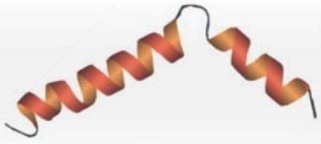
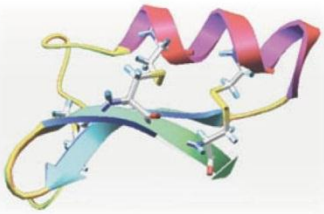
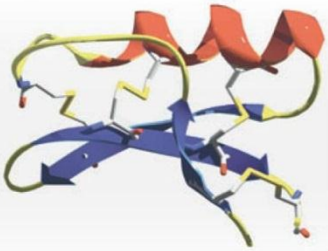
Peptides	# of genes	Main activity	Concentration	3-D structure
Diptericin	2	Gram-negative bacteria	0.5 µM	nd
Attacin	4	Gram-negative bacteria	nd	nd
Drosocin	1	Gram-negative bacteria	40 µM	
Cecropin	4	Gram-negative bacteria	20 µM	
Defensin	1	Gram-positive bacteria	1 µM	
Drosomycin	7	Fungi	100 µM	
Metchnikowin	1	Fungi	10 µM	nd

Fig. 1.6: The seven classes of *Drosophila melanogaster* antimicrobial peptides. The name of the AMP class, the number of genes in the genome, their main antimicrobial targets, the estimated concentration in the hemolymph after bacterial injection and their 3-D structure are shown; nd = not determined. Figure after Imler and Bulet (2005).

1.4.1 Divergence of AMPs and immune related genes

Besides their different modes of action, a key feature of all antimicrobial peptides is the direct interaction with their microbial targets. Therefore, strong selective pressure operating on such genes could be assumed, as the effector proteins have to adapt their ability to kill fast evolving pathogens or, on the other hand, to maintain the titre of mutualistic symbionts. Simultaneously, pathogens are selected for effective adaptation to evade the hosts' defence. Under this presumption, the potential for a coevolution of the peptide with its target microbe is given. This "arms race" (Dawkins and Krebs, 1979) of host and pathogen should be traceable to elevated rates of amino acid evolution in AMP coding genes. Intimate coevolution with fast evolving

mutualistic symbionts is expected to accelerate the evolutionary rate of interacting host-encoded immune genes in an adaptive manner. Surprisingly, AMP genes of *D. melanogaster* show in general little indication of adaptation on the amino acid sequence level (reviewed by Lazzaro, 2008). Members of the *Attacin* (Lazzaro and Clark, 2001) and *Diptericin* (Lazzaro and Clark, 2003) gene family, however, show high levels of polymorphism in *D. melanogaster* populations. Paralogous gene conversions, insertion/deletion events, gene duplication and gene loss are observed in high frequencies and with linkage disequilibrium for those gene families. Therefore, they are assumed to stand under positive selection and to have a functional effect on the hosts' capacity to fight infection or maintain mutualistic symbiosis.

Similar to AMPs, pattern recognition receptors mediate immune responses by direct contact with invading pathogens and are therefore also under potential selective pressure to coevolve with their microbial targets. *Hemomucin*, a hemocyte surface protein (Theopold et al., 1996), hemocyte-specific *class C scavenger receptor* (Pearson et al., 1995) and *Tehao*, a Toll-like transmembrane receptor (Luo et al., 2001) are three immune receptors which in earlier studies showed elevated rates of polymorphism within *D. simulans* populations (Schlenke and Begun, 2004). Due to their high levels of polymorphism in other *Drosophila* species, genes for *Attacins*, *Diptericins*, *hemomucin, class C scavenger receptor* and *tehao* were examined for the *D. paulistorum* species complex in this study.

2. Aims of the thesis

2.1 Selection of informative genes for *D. paulistorum* semispecies fingerprinting

The phylogeny of the *D. paulistorum* species complex was studied by Gleason et al. (1998), who used the nuclear gene *per* and the mitochondrial gene *CO1*, and by Robe et al. (2010), who combined the nuclear genes *adh, amd, hb* and *per* and mitochondrial genes *CO1* and *CO2*. Those genes were not informative enough to provide a phylogeny with a sufficient resolution to resolve the relationships between the six described semispecies of *D. paulistorum.* Genes for antimicrobial peptides of the *Attacin* (Lazzaro and Clark, 2001) and *Diptericin* (Lazzaro and Clark, 2003) family as well as genes for immunoreceptors *Hemomucin*, *Scavenger Receptor Class C1* and *Tehao* (Schlenke and Begun, 2005) showed a high degree of polymorphism within *Drosophila* species. Hence, (*i*) *tracing* their orthologous peptides and genes in the annotated genome of the *D. willistoni,* i.e. the sister species of *D. paulistorum,* (*ii*)designing appropriate PCR primer sets to amplify the corresponding genes in *D. paulistorum* semispecies by PCR and (*iii*) performing sequence analysis was the first aim of this thesis in order to help resolve the phylogeny of the semispecies. Furthermore, the mitochondrial gene *ND4,* known to be fast evolving in *Drosophila* (Garesse, 1988; Moriyama and Powell, 1997), was investigated to resolve the conflicts between nuclear and mitochondrial phylogeny of previous studies.

2.2 Generating highly informative genetic Marker system for *D. paulistorum* semispecies and strain typing of newly collected lines

Sequence analysis of the candidate genes for AMPs as well as mitochondrial gene *ND4* was used to determine diagnostic single nucleotide polymorphisms (SNPs) within the six *D. paulistorum* semispecies. A minimal set of diagnostic SNPs to create a reliable marker system was determined. Additionally, restriction fragment length polymorphism (RFLP) analysis was performed in order to examine candidate genes for semispecies-specific fingerprints. Those SNP and RFLP patterns were then used to identify recently collected strains.

2.3 Evolutionary pressure on *Attacin D*

The genes investigated in this thesis are coding for proteins that interact directly with their microbial targets. As microbes are generally fast evolving, proteins that interact with microbial structures need to coevolve in order to adapt and maintain their function, either as pathogen defence or as a regulating factor in mutualistic symbiosis. As each *D. paulistorum* semispecies harbours its distinct *Wolbachia* strain which acts as a mutualistic endosymbiont, signs of diversifying selection can be assumed in genes coding for proteins interacting with those bacteria. In this thesis,

the AMP Attacin D proteins were selected as prime cantidates for a potential direct interaction with *Wolbachia,* as they lack a signal- and propeptide, both considered as hallmarks for their intracellular functions (Hedengren, 2000). Attacin D was tested for domain regions with high proportions of non-synonymous substitution and K_a/K_s ratios, *i.e.* the ratio of the number of non-synonymous substitutions per non-synonymous site (K_a) to the number of synonymous substitutions per synonymous site (K_s), that hint at diversifying selection (Li et al., 1985).

2.4 Mitochondrial transmission modes in *D. paulistorum* hybrids

The observed paternal transmission of *Wolbachia* in hybrids of *D. paulistorum* semispecies (Miller and Schneider, unpublished) and the earlier reported hitchhiking effect of *Wolbachia* with their mitochondrial haplotype (Nigro and Prout, 1990; Turelli and Hoffmann, 1991) raise the question if those artificial hybrids can coinherit paternal mitochondria along with the endosymbiont from males, and if so, to what extent. In this thesis, the RFLP pattern of mitochondrial sequences were used to differentiate via Southern blot the maternal from the paternal mitotype and to examine whether stabilized *D. paulistorum* hybrids inherit solely one mitotype, or if both coexist in a state of heteroplasy.

3. Materials and methods

3.1 List of consumables

Consumable	Supplier
Acetic acid	Carl Roth GmbH+Co.KG
Agar, bacteriology grade	AppliChem
Agarose, standard quality	peqLab Biotechnologie GmbH
Agarose, high quality	Roche
Ampicillin	Carl Roth GmbH+Co.KG
CO_2 gas	Messer
Ethanol absolute	Merck
Ethididum bromide	Carl Roth GmbH+Co.KG
Ethylen diaminotetraacidic acid (EDTA)	Carl Roth GmbH+Co.KG
Glass wool	Serva
HCl	Carl Roth GmbH+Co.KG
IPTG	Roche
Isopropanol	Merck
KH_2PO_4	Merck
Methanol	Carl Roth GmbH+Co.KG
Na_2HPO_4	Merck
Orange G dye	Sigma Chemical Company
Parafilm “M”	Amersham
Potassium acetate	BDH Chemicals Ltd Poole England
Sephadex G50	Carl Roth GmbH+Co.KG
Sodium acetate	Sigma Chemical Company
Sodium chloride	Carl Roth GmbH+Co.KG
Sodium hydroxide pellets	Carl Roth GmbH+Co.KG
Sodium dodecyl sulphate (SDS)	Carl Roth GmbH+Co.KG
SDS ultra-pure	Carl Roth GmbH+Co.KG
Tris Base ultra Quality	Carl Roth GmbH+Co.KG
Tween 20	Sigma Chemical Company
Triton 100 X	Serva/ Carl Roth GmbH+Co.KG
Whatman paper 3MM	Schleicher & Schuell
X-Gal	Carl Roth GmbH+Co.KG
X-ray fixing solution	Kodak
X-ray developing solution	Kodak

Table 3.1: List of consumables

3.2 Fly Strains

Table 3.2 and **3.3** list all the fly strains used in the course of this thesis.

Strain	Geographic origin	Country	Coll. date	Origin
A28	Belem, PA	Brazil	1952	L. Ehrman
C2	Lancetilla	Honduras	1954	L. Ehrman
I1	Llanos	Colombia	1958	L. Ehrman
MS	Mesitas	Colombia	1962	L. Ehrman
O11	Georgetown	Guyana	1957	L. Ehrman
SM	Santa Marta	Colombia	1956	L. Ehrman
POA1	Porto Alegre, RS	Brazil	2003	A. Garcia
POA10	Porto Alegre, RS	Brazil	2003	A. Garcia
MLC	Morro da Lagoa da Conceicao, SC	Brazil	2003	M. Gottschalk
RP	Ribeirao Preto, SP	Brazil	1995	A. Garcia
MTS	Mario Totta Street, RS	Brazil	1991	A. Garcia
JB	Jardim Botanico, Porto Alegre, RS	Brazil	1991	A. Garcia
ZOO1	Dois Irmaos Park, Recife Municipality, Pernambuco	Brazil	2009	Atlantic Forest C. Rohde & A. Garcia
TRI-POU	Pousada, Triunfo Municipaly, Pernambuco	Brazil	2009	Atlantic Forest C. Rohde & A. Garcia
SET-RIA	Riacho, Serra Talhada Municipality, Pernambuco	Brazil	2008	Caatinga C. Rohde & A. Garcia
white	Lancetilla	Honduras	1960	L. Ehrman
yellow	Copan	Honduras	1960	L. Ehrman
prune	Copan	Honduras	1960	L. Ehrman

Table 3.2: *Drosophila paulistorum* strain list, after WJ Miller (2012)

Species	Strain	Geographic origin	Country	Coll. date	Origin
D. ananassae	ANA/SRI	Ilha das Rolas, Sao Tome	Sao Tome	2002	K. Wittman
D. equinoxialis	FS	Panama	Panama	1999	E. Bartel
D. melanogaster	w^{1118-}	Texas	USA	1968	R. Lewis
D. nebulosa		Sinaloa	Mexico	2004	Drosophila Species Stock Center
D. tropicalis		San Salvador	El Salvator	2011	V. Valente
D. willistoni	NEG-DED	Serra Negra, Bezerros Municipality, Pernambuco	Brazil	2009	Altitude Forest, C. Rhode & A. Garcia
D. willistoni	SAL-13	Biological Reserve of Saltinho, Tamandaré Municipality, Pernambuco	Brazil	2009	Atlantic Forest, C. Rhode & A. Garcia
D. willistoni	Wls/P98	Fort Sherman, Colon	Panama	1998	E. Bartel

Table 3.3: *Drosophila* outgroup strain list, after WJ Miller (2012)

3.2.1 Culturing

D. paulistorum and other *Drosophila* strains (see **Table 3.3**) were reared in 50 ml vials containing 10 ml Formula 4-24 Drosophila medium (Carolina, Burlington) and small pieces of heat-sterilized cellulose. The vials were stored at room temperature, except for *D. willistoni* Wls/P98 and *D. melanogaster* w^{1118-} strains, which were stored at 18°C. The flies at room temperature were transferred into new vials weekly, those at 18°C twice to three times a month. For DNA purification, at least week-old flies were used.

3.2.2 A28xO11^{R} and C2xA28^{R} hybrids (established by D. Schneider)

To produce fertile progeny of A28 ♀ x O11 ♂ and C2 ♀ x A28 ♂ crossbreedings, only the paternal strains (OR and AM) were kept on fly medium containing 0.2% of the antibiotic rifampicin (indicated by superscript R, i.e. A28^{R}, O11^{R}) at constant 24-25°C for many consecutive generations, minimizing their native *Wolbachia* titre. Naturally *Wolbachia*-infected females of the maternal strains (A28 and C2) were crossed with such antibiotic-treated males of the paternal strains of the reciprocal semispecies, emerging progeny was kept for at least three consecutive generations on medium with 0.2% rifampicin applied. Subsequently, the hybrids were transferred on normal medium without antibiotics. The flies used for this thesis were in the 28th and 29th filial generation (F28/F29) after the crossbreeding.

3.2.3 [A28xO11^{R}]xO11 and [C2xA28^{R}]xA28 introgression lines (established by D. Schneider)

To remove the nuclear genetic background of the maternal strain in the A28xO11^{R} and C2xA28^{R} hybrids (as mentioned above), females were backcrossed for five generations with their respective paternal strain (O11 for A28xO11^{R} and A28 for C2xA28^{R}). In each of the five filial generations, female progeny was afresh cultivated with males of the paternal strain.

3.3 DNA Purification

3.3.1 Puregene Extraction Kit

DNA of 10 female or 15 male flies was purified following the optimized manual for the Puregene DNA Purification Kit for purification (Qiagen, Germany) from 10 to 15 *Drosophila melanogaster*:

Cell Lysis

1. Add 100 µl chilled Cell Lysis Solution to the flies, homogenize thoroughly for 20 seconds using a microfuge tube pestle. Place sample back on ice until next step
2. Add another 200 µl chilled Cell Lysis Solution to the sample, homogenize thoroughly for another 10 seconds.
3. Incubate lysate at 65°C for 30 minutes

RNase Treatment

1. Add 1.5 µl RNase A Solution (4 mg/ml) to the cell lysate
2. Mix the sample by inverting the tube 25 times and incubate at 37°C for 30 minutes.

Protein Precipitation

1. Cool sample to room temperature
2. Add 100 µl Protein Precipitation Solution to the cell lysate
3. Vortex vigorously at high speed for 20 seconds to mix the Protein Precipitation Solution uniformly with the cell lysate. Incubate for 5 minutes on ice.
4. Centrifuge at 14,000 x g for 4 minutes. The precipitated proteins and tissue particulates will form a tight pellet. If protein pellet is not tight, repeat step 3 followed by incubation on ice for 5 minutes, then repeat step 4.

DNA Precipitation

1. Pour the supernatant containing DNA (leaving behind the precipitated protein pellet) into a clean 1.5 ml centrifuge tube containing 300 µl 100% Isopropanol (2-propanol)
2. Mix sample by inverting the tube gently 50 times.
3. Centrifuge at 14,000 x g for 5 minutes.
4. Pour off supernatant and drain tube on clean absorbent paper. Add 300 µl 70% Ethanol and invert tube several times to wash DNA pellet.
5. Centrifuge at 14,000 x g for 1 minute. Carefully pour off the ethanol. Pellet may be loose so pour slowly and watch pellet.
6. Invert and drain the tube on clean absorbent paper and allow to air dry 15 minutes at 37°C.

DNA Hydration

1. Add 50 µl DNA Hydration Solution (50 µl will give a concentration of 200 µg/ml if the total yield is 10 µg DNA).
2. Rehydrate the DNA by incubating sample for 1 hour at 65°C with 300 rpm agitation and/or overnight at room temperature. If possible, tab tube periodically to aid in dispersing the DNA.
3. If particulates are present in the rehydrated DNA sample, centrifuge at 14,000 x g for 5-10 minutes and then transfer the supernatant containing the DNA to a clean tube
4. Store DNA at -20°C

3.3.2 DNA Extraction for Southern blot after N. Junakovic (2004)

The following protocol was designed for DNA extraction of individual Drosophila flies, but was applied in this thesis for multifly extraction:

1. Place 10 female or 15 male flies in 1.5 ml reaction tubes containing 100 µl of homogenization buffer; lift the flies a few mm up the tube wall and place tubes on ice.
2. Insert the hand drill with the installed pestle into the tube and switch on rotation at approximately 1000 rpm. During the first 10 seconds of homogenization, press the tube tightly against the pestle, then start gently moving it up and down by approximately one mm for an additional 20 seconds. Remove lysate adhering to the rotating pestle by keeping it in touch with the tube wall while slowly withdrawing it. Immerse the pestle in a beaker containing distilled water; get rid of the droplet of water by wiping it on a piece of paper towel. Keep homogenates on ice until next step.
3. Incubate the tubes at 65°C for 1 hour; flick each tube individually after 30 minutes.
4. Leave at room temperature to cool down for a few minutes. Flick once to disperse the sediment, add 14 µl of 8 M potassium acetate and keep flicking until the liquid stops foaming.
5. Incubate on ice for 30 minutes.
6. Centrifuge at 10,000 rpm for 10 minutes in a bench top centrifuge.
7. Transfer the supernatant to fresh 1.5 ml reaction tubes.
8. Add 100 µl of isopropanol, flick the tube once and incubate one hour at room temperature to precipitate.
9. Centrifuge at 10,000 rpm for 10 minutes in a bench top centrifuge.
10. Remove the supernatant, add 200 µl of 70% ethanol at room temperature and flick the tube once.
11. Centrifuge at 13,000 rpm for 3 minutes in a bench top centrifuge.
12. Discard the supernatant and leave the tube open at room temperature for 15-30 minutes to allow the residual ethanol to evaporate.

13. Add 27 µl ddH_2O and incubate for 10 minutes at room temperature to hydrate, flick once and incubate another 10 minutes.

3.4 Polymerase Chain Reaction (PCR)

3.4.1 PCR Primer

Table 3.4 lists all primers used for cloning and direct sequencing.

Gene	Primer Name	Sequence
Attacin A	GK19632_F	AAGCGTTCTAATCTTGGCCTGC
	GK19632_R	ATGGGTAAGACCTAAACCGGCG
Attacin C	GK17822 long1 F	TCAGCTTGGCCGTTCTAGTGGC
	GK17822 new F	GGACAGGTGTTTGCTGCTGGC
	GK17822 new R	GCATGGGCAGACTGCTGGAAGC
Attacin D	GK12085_F	ATGGATTGTCAGGCCACAGG
	GK12085_R	GTATCGACCACTTATCCCGG
Diptericin	GK20931_F_long	GAATTGCAAGATGAAGGCTACAATTATC
	GK20931_R_long	GAGATTACATGTAATCAAAAGTTGGGG
Diptericin B	GK20932_F	ATGCAGTTTACTTTCGGTTTGC
	GK20932_R	CCTAAAAGTATATTGAGCTCCTGC
Drosocin	GK19342 F	ATGAAGTTCTCCATCATCTTTGCGG
	GK19342 R	TTGATCTGCTGGCCCTGTTCC
Hmu – Exon+Intron 1	GK11749-IE1 F	GCGCTGCAGTTAAGAATCATG
	GK11749-IE1 R	GCGTGGACAGATCAACTTGCC
Hmu - Exon+Intron 2	GK11749-IE2 F	GGCAAGTTGATCTGTCCACGC
	GK11749-IE2 R	GCATAATACTTTCGCCATGACC
SrC1 - Exon+Intron 1	GK23906-EI1 F	GCAGAGCTATAGAATTGAGGTGC
	GK23906-EI1 R	CCTCAGTCTCGAAATCGCACG
SrC1 - Exon+Intron 2	GK23906-EI2 F	CGTGCGATTTCGAGACTGAGG
	GK23906-EI2 R	CCAATGGTTACCCTGACTGC
Tehao - Exon 1	GK14848-E1 F	GTAGTGCGCTTTCACTGTCCGG
	GK14848-E1 R	ACGCAATTGCTTAAGTTCACGC
Tehao - Exon 2	GK14848-E2 F	GGGTGCAACGAATGACACG
	GK14848-E2 R	CGCTTCGAATCGTCTACAGTGCG
DesatF-ζ (designed by D. Schneider)	GLEANR17505_wil_1F	ATGCCACCGAATGCCGAA
	GLEANR17505_wil_981R	CCAGTGCGTATAACCCGT
DesatF-η (designed by D. Schneider)	GLEANR1159_wil_184F	CTATCGGGCTATGGTTTGTG
	GLEANR1159_wil777R	GTATTGGTTCATTGGTGGGA
ND4 (designed by D. Schneider)	ND4pa_F	CACCACAATGCTATAGCAGGTAT
	ND4pa_R	GTATCAACCTGAACGATTGCAAG

Table 3.4: Primer list. Genes listed in the first column refer to the target locus of the PCR fragment. All primers are given in 5´→3´orientation.

3.4.2 PCR protocols

3.4.2.1 Immune related genes

Table 3.5 lists the standard PCR reaction mix for a final reaction volume of 10 µl for all immune related genes examined in this thesis.

Component	Stock conc.	Final conc.	1 x Vol.
GreenGoTaq Reaction Buffer (Promega)	5 X	1 X	2 µl
$MgCl_2$ (Promega)	25 mM	2 mM	0.8 µl
fwd-Primer	10 µM	0.5 µM	0.5 µl
rev-Primer	10 µM	0.5 µM	0.5 µl
dNTP Mix (Fermentas)	2 mM	35 µM	0.18 µl
GoTaq DNA Polymerase (Promega)	5 U/µl	0.04 U/µl	0.08 µl
dd-H_2O	-	-	4.94 µl
DNA template	100-500 ng/µl	10-50 ng/µl	1 µl

Table 3.5: Components, stock concentrations, final concentrations and volumes for 10 µl PCRs for all immune related genes.

All PCRs were carried out on a Biometra T3 Thermocycler. **Table 3.6** shows the PCR programme for all immune related genes and ND4.

Step	Temperature	Time	Repetitions
1. Initial denaturation	94°C	2 min	
2. Denaturation	94°C	30 sec	34x steps 2 - 4
3. Annealing	55°C (50°C)	45 sec	
4. Elongation	72°C	1 min	
5. Final elongation	72°C	10 min	
6. Storage	15°C	∞	

Table 3.6: PCR programme for all immune related genes and *ND4*. *Attacin D* annealing temperature is indicated in brackets

3.4.2.2 NADH dehydrogenase subunit 4 (ND4)

Table 3.7 shows the PCR components for a final reaction volume of 10 µl for ND4, the corresponding PCR programme is shown in **Table 3.6**.

Component	Stock conc.	Final conc.	1 x Vol.
GreenGoTaq Reaction Buffer (Promega)	5 X	1 X	2 µl
$MgCl_2$ (Promega)	25 mM	3.5 mM	1.4 µl
fwd-Primer	10 µM	0.8 µM	0.8 µl
rev-Primer	10 µM	0.8 µM	0.8 µl
dNTP Mix (Fermentas)	2 mM	50 µM	0.25 µl
GoTaq DNA Polymerase (Promega)	5 U/µl	0.04 U/µl	0.08 µl
dd-H_2O	-	-	3.67 µl
DNA template	100-500 ng/µl	10-50 ng/µl	1 µl

Table 3.7: Components, stock concentrations, final concentrations and volumes for 10 µl PCRs for ND4.

3.4.2.3 Desaturases

In **Table 3.8** and **Table 3.9**, reaction components and PCR programme for those primers are shown.

Component	Stock conc.	Final conc.	1 x Vol.
GreenGoTaq Reaction Buffer (Promega)	5 X	1 X	2 µl
$MgCl_2$ (Promega)	25 mM	2 mM	0.8 µl
fwd-Primer	10 µM	0.5 µM	0.5 µl
rev-Primer	10 µM	0.5 µM	0.5 µl
dNTP Mix (Fermentas)	2 mM	35 µM	0.18 µl
GoTaq DNA Polymerase (Promega)	5 U/µl	0.04 U/µl	0.08 µl
dd-H_2O	-	-	4.94 µl
DNA template	100-500 ng/µl	10-50 ng/µl	1 µl

Table 3.8: Components, stock concentrations, final concentrations and volumes for 10 µl PCRs for *DesatF-ζ* and *DesatF-η*.

Step	Temperature	Time	Repetitions
1. Initial denaturation	94°C	2 min	
2. Denaturation	94°C	30 sec	35x steps 2 - 4
3. Annealing	56°C	45 sec	
4. Elongation	72°C	1.5 min	
5. Final elongation	72°C	10 min	
6. Storage	15°C	∞	

Table 3.9: PCR programme for *DesatF-ζ* and *DesatF-η*.

3.4.2.4 Colony – PCR

Successful transformation of chemically competent DH5α *E. coli* cells was tested via colony PCR using M13 forward (5´-TGTAAAACGACGGCCAGT-3´) and reverse (5´-CAGGAAACAGCTATGACC-3´) primer. Transformed white colonies, were picked from LB-Amp plates with a pipet tip, dabbed on a replica LB-Amp plate and placed in a 1.5 ml reaction tube filled with 50 µl 1 x GoTaq PCR buffer (Promega, U.S.A.). Tubes were boiled for 30 seconds and 1µl of the bacterial lysate was added to the PCR reaction mix described in **Table 3.10**.

Component	Stock conc.	Final conc.	1 x Vol.
GreenGoTaq Reaction Buffer (Promega)	5 X	1 X	2 µl
$MgCl_2$ (Promega)	25 mM	1.5 mM	0.6 µl
M13 fwd-Primer	10 µM	0.3 µM	0.3 µl
M13 rev-Primer	10 µM	0.3 µM	0.3 µl
dNTP Mix (Fermentas)	2 mM	35 µM	0.18 µl
GoTaq DNA Polymerase (Promega)	5 U/µl	0.025 U/µl	0.05 µl
dd-H_2O	-	-	5.58 µl
Colony dilution	-	-	1 µl

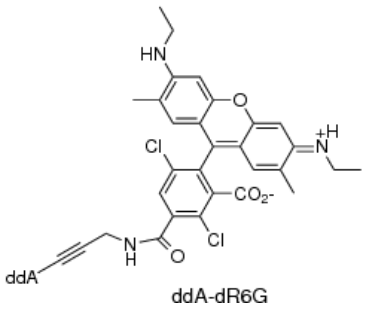
Table 3.10: Components, stock concentrations, final concentrations and volumes for 10 µl Colony-PCR samples

Step	Temperature	Time	Repetitions
1. Initial denaturation	94°C	2 min	
2. Denaturation	94°C	30 sec	35x steps 2 - 4
3. Annealing	52°C	30 sec	
4. Elongation	72°C	1 min 15 sec	
5. Final elongation	72°C	5 min	
6. Storage	15°C	∞	

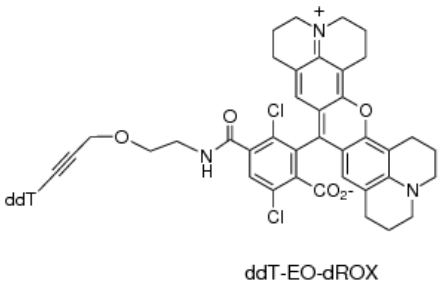
Table 3.9: PCR programme for Colony-PCR.

3.4.2.5 BigDye terminator labelling PCR for Sanger Sequencing

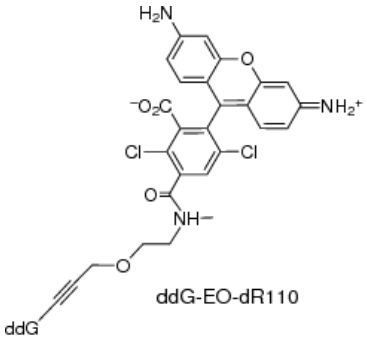
Sanger sequencing of PCR amplicons was performed by Dr. J. Rath, Department of Evolutionary Biology, University of Vienna, Austria. BigDye Terminator Cycle Sequencing Kit (Applied Biosystems) was used to label the fragments. **Fig. 3.1** shows the dRhodamine acceptor dyes and their respective terminators.



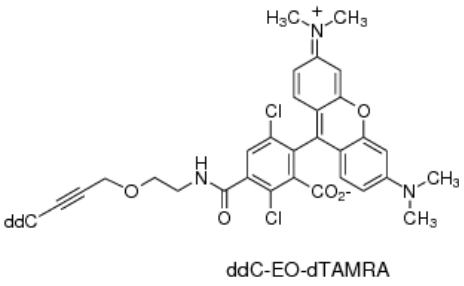
Adenine - dichloro[R6G]



Thymine - dichloro[ROX]



Guanine - dichloro[R110]



Cystosine - dichloro[TAMRA]

Fig. 3.1: dRhodamine terminators of BigDye Terminator Cycle Sequencing Kit (Alexander Binder, 2001)

The volumes for one reaction of 10 µl according to the manufacturer's protocol are listed in **Table 3.10**. The PCR programme for sequencing is shown in **Table 3.11**. M13 primers were used for amplicons cloned into the pTZ57R/T vector (Fermentas, Lithuania, see **3.4 cloning**). We used the same target-specific primers for direct sequencing as used to obtain the PCR fragments. PCR products for direct sequencing were purified using the wizard SV Gel and PCR clean up kit (Promega, U.S.A.) or the Illustra ExoStar PCR and Sequence Reaction clean up Kit (GE Healthcare, UK), following the manufacturer's protocol.

Component	**1 x Vol.**
Big Dye terminator reaction mix	1 µl
5 X Sequencing Buffer	2 µl
Primer (fwd or rev)	1 µl
dd-H_2O	5 - x µl
DNA (PCR fragments)	x ≥ 5 µl

Table 3.12: Components and volumes for a 10 µl reaction sample for dye terminator labelling performed with BigDye Terminator Cycle Sequencing Kit (Applied Biosystems)

Step	**Temperature**	**Time**	**Repetitions**
1. Initial denaturation	96°C	2 min	
2. Denaturation	96°C	20 sec	39x steps 2 - 4
3. Annealing	50-56°C	10 sec	
4. Elongation	60°C	4 min	
5. Storage	15°C	∞	

Table 3.13: PCR programme for dye terminator labelling. Annealing temperature depends on specific primer annealing temperature.

3.5 Sequence analysis

All obtained sequences were analysed with Geneious DNA and Protein analysis software (Biomatters ltd.). All alignments were built with the Geneious Alignment tool. Global alignment with a Cost Matrix of 51% similarity, gap open and gap extension penalty of 20 were chosen as setup. Ks/Ks ratio was determined using dnaSP 5.10.01 software (Julio Rozas et al., University of Barcelona). Sliding window analysis was performed manually, graphs were established with Microsoft Excel software.

3.6 Molecular cloning

3.6.1 pTZ57R/T vector

PCR amplicons were ligated into the pTZ57R/T cloning vector (Fermentas, Lithuania). An ampicillin resistance and a multiple cloning site (MCS) located in an IPTG dependent *lac*Z gene enabled blue/white screening for transformants on LB-AMP plates treated with 960 µg Isopropyl-β-D-thiogalactopyranosid (IPTG) and 800 µg 5-bromo-4-chloro-indolyl-β-D-galactopyranoside (X-gal). **Fig. 3.2** shows a map of pTZ57R/T as provided in Fermentas' InsTAclone PCR Cloning Kit protocol.

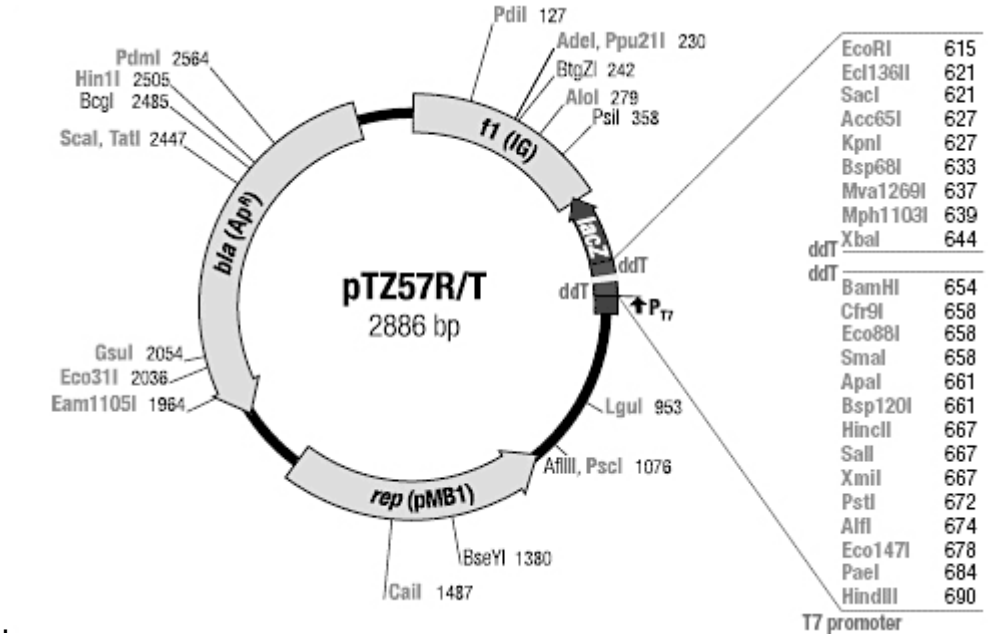


Fig.3.2: Vector map of pTZ57R/T (Fermentas, Lithuania). Ampicillin resistance gene, *lac*Z Sequence and multiple cloning site (MCS) are indicated

3.6.2 Ligation

The PCR products amplified for ligation into the pTZ57R/T vector (Fermentas, Lithuania) were purified with the PeqGOLD Gel extraction Kit (Peqlab, Germany) according to the manufacturer's protocol. Ligation was performed with the InsTAclone PCR Cloning Kit (Fermentas, Lithuania). Single 3'-A overhangs at both ends of the PCR products were ligated with the linearized vector's single 3'-ddT overhangs. A minimized reaction volume of 3 µl was prepared according to the formula in **Table 3.14** and incubated at 4°C over night.

Component	1 x Vol.
Ligation Buffer 10X	0.3 µl
pTZ57R/T Vector (55 ng/µl)	0.1 µl
PEG4000 50%	0.3 µl
T4 Ligase (5 U/µl)	0.1 µl
ddH_2O nuclease free	1.4 µl
Purified PCR fragments	0.8 µl

Table 3.14: Components and volumes for a 3 µl reaction sample for ligation of PCR fragments with pTZ57R/T vector.

3.6.3 Transformation DH5α *E. coli*

Chemically competent *E. coli* DH5α cells were transformed with the PCR-amplicon containing pTZ57R/T vector (Fermentas, Lithuania) according to the protocol below:

1. Thaw competent cells (400 – 600 µl aliquots) on ice. Incubate ligation reaction on ice.
2. Add 50 µl competent cells to each ligation mix (3µl).
3. Incubate 20 min on ice.
4. Heat shock cells at 42°C for 50 sec. and immediately place back on ice for 2 minutes.
5. Add 300 µl SOC – Medium (RT) to each tube and incubate for two hours at 37°C with 160 rpm agitation.
6. Prepare for each transformation one LB (amp) agar plate with 40 µl IPTG (24 mg/ml) and 40 µl X-Gal (20 mg/ml). Pre-warm the plates to 37°C.
7. Plate the whole transformation mix (~350 µl) on the plate, incubate o/n at 37°C.

3.6.4 Plasmid purification (miniprep)

Plasmid DNA was purified from 1-3 ml o/n – culture using the Invisorb Spin Plasmid Mini Two Kit (Stratec molecular, Germany). After centrifugation and resuspension of the bacterial pellet in 250 µl Solution A, the *E. coli* cells were lysed with 250 µl SDS containing Solution B. Proteins and phospholipids of the cell membrane were solubilized, RNA was digested by RNase. Subsequently, by addition of 250 µl Solution C, the proteins were denaturated, chromosomal DNA, cellular debris and SDS were precipitated and the plasmid-rich solution was added on a DNA-binding membrane in a spin column. By washing with 750 µl Washing Solution, the plasmid DNA was purified and eventually eluted in 50 µl elution solution.

3.7 Agarose gel electrophoresis

To analyse PCR products and/or process them for cloning, agarose gel electrophoresis was applied. The 1% (w/v) agarose gels were prepared in 1 x TAE buffer (20 mM Tris base, 5 mM sodium acetate, 1 mM EDTA pH 8.2). PCR amplicons were separated according to their size by applying constant 70 – 100 V for 1,5 -2,5 hours. GeneRuler 1 kb DNA Ladder (Fermentas, Lithuania) was used as size marker.

3.8 Genomic Southern blot analyses

3.8.1 Genomic Southern blot for P^{32}-labelling after Junakovich (2004)

3.8.1.1 Vertical gel electrophoresis

DNA extracted either with the Puregene Kit (Qiagen, Germany) or the method adapted from Junakovich (2004, see **3.2.**) was quantified with the ND4000 NanoDrop spectrophotometer (Peqlab, Germany). Concentrations were equalized before further analyses. Subsequently, DNA was digested with restriction enzymes suitable for the analysis of the respective gene, i.e., without restriction sites in the hybridizing fragment. For *Attacin D* analysis, *Hin*dIII (New England Biolabs, U.S.A.) was used, whereas *Psi*I was chosen for *ND4*/*CO1* analysis (New England Biolabs, U.S.A.). In a total reaction volume of 30 µl, 20 U *Hin*dIII (1µl) or 1.5 U of *Psi*I (0.3 µl) were added to each sample and incubated with the according reaction buffer at 37°C over night. The reaction was stopped by heat inactivating the enzyme at 65°C for 20 minutes.
Vertical agarose gels were prepared in custom-made electrophoresis chambers, according to Junakovich (2004): 38 x 19 cm glass plates and 3 mm rubber spacers determined the gel dimensions; the combs were equipped with 20 5 mm wide slots. For the gel stack at the bottom of the chamber (2% agarose in 1 x TBE Buffer), 100 ml were prepared, poured hot into the gel cast to a height of approximately 7 cm and left for polymerising at RT for at least 30 minutes. 250 ml separating gel (0.8% agarose in 1 x TBE) were then dissolved and poured on top of the solid stack. High temperature of the dissolved agarose and a quick handling was required to avoid air bubbles and quick polymerization. After insertion of the comb, additional separating gel solution was applied on the gel with a Pasteur pipette to compensate shrinking of the gel during polymerization. The gel was then allowed to completely polymerize for at least one hour at RT.

For electrophoresis, the apparatus was placed in a plastic tray and glass chamber plus tray were filled with 1xTBE buffer. A pump assured a constant circulation of TBE buffer from the tray into the electrophoretic chamber. The samples were mixed with 6 µl loading dye (6 x Orange G in 20% Ficoll) and loaded into the slots of the removed comb. As size marker, *Hind*III digested λ –phage DNA was used. Electrodes were installed and connected to a power supply. The whole set-up was covered with a

plastic bag in order to prevent extensive evaporation of buffer. The electrophoresis was performed at constant 80 V over night. The run was completed when the loading dye front had reached the gel stack. Subsequently, the apparatus was carefully disassembled and the gel was incubated for 30 minutes in 1 x TBE with 1 µg/ml ethidium bromide to stain the DNA. A photo was taken using UV transillumination.

3.8.1.2 Vacuum blotting

DNA was transferred onto a positively charged nylon membrane (Highbond-N+, GE Healthcare, UK) by vacuum blotting. 4 M NaOH was used as denaturating transfer solution. First, the rubber mask of the blotting apparatus was cut to the size of the gel. One layer of Whatman filter paper was placed on the blotter frit and soaked with 4 M NaOH. The edge length of the filter paper was cut 2 cm longer than the edge length of the gel. The rubber mask was then placed on the filter paper. The nylon membrane (Highbond-N+, GE Healthcare, UK) of the size of the gel was subsequently placed on the filter paper and the gel was placed on top of it. Air bubbles were removed by carefully rolling a 50 ml Falcon tube over each layer. The gel was covered with 4 M NaOH and vacuum was applied. A total of 300 ml 4 M NaOH was constantly poured on the gel so that it was fully covered during the whole transfer. Finally, the gel was removed, the membrane briefly rinsed with 2 x SSC and the DNA cross-linked on the membrane via UV-light (4 min on each side). For storage at 4°C, the membranes were shrink-wrapped in plastic foil. To control the transfer success, the gel was re-stained in ethidium bromide. Lacking of DNA on the gel suggested successful transfer onto the membrane.

3.8.1.3 Probe labelling

The DNA templates for the hybridization probe were obtained through PCR amplification and cloning as described in section **3.3** and **3.4**. *Attacin D* was labelled to investigate paralogous loci in the genome, *ND4* and *CO1* sequences were used for analysis of mitochondrial DNA. Radioactive [α^{32}P]-dCTP (PerkinElmer, USA) was incorporated in the newly synthesized probe via random priming using the Rediprime II DNA Labelling Kit (GE Healthcare, UK), according to the manufacturer's protocol. The plasmid templates were linearized (*Hin*dIII or *Pst*I) and quantified via the ND4000 NanoDrop (Peqlab, Germany). Approximately 700 ng DNA were diluted in 1 x TE Buffer to a final volume of 45 µl and denatured at 95°C for 5 minutes, followed by 5 minutes incubation on ice. The DNA was then added to the premixed labelling reaction containing Klenow fragment of DNA polymerase I, a mixture of dATP, dGTP and dTTP, random primers, and buffer salts. At last, 3-5 µl [α^{32}P]-dCTP solution were added. Synthesis of labelled DNA was then carried out at 37°C for 30 minutes. To stop the reaction, 5 µl of 0.2 M EDTA was added. To remove excessive [α^{32}P]-dCTP, the sample was cleaned up using a self-made Sephadex G50 column: a plug of glass wool was inserted into a short-tip Pasteur pipette and filled up with autoclaved

Sephadex G50 in 1 x TE buffer. The labelled probe was diluted in 200 µl 1 X TE and loaded on the column. 200 µl aliquots of 1 x TE buffer were then applied to elute the labelled probe. Each fraction was collected in a tube and the activity measured using a β-γ-detector (handheld contamination monitor LB122, Berthold Technologies, Germany). The two fractions with the highest activity were combined, incubated for 5 minutes at 95°C, snapped cool on ice for 5 minutes and used for hybridization in the following.

3.8.1.4 Hybridization

The hybridization reactions were carried out with the Belly Dancer Hybridization Water Bath (Stovall Life Science). Membranes (see **3.6.2**) were incubated (pre-hybridized) for 30 min in a plastic box with 70 ml Church buffer (0.5 M sodium phosphate buffer pH 7.5, 1 mM EDTA, 7% SDS (w/v); adjust to pH 7.2) at 65°C under constant agitation. 30 ml of Church buffer were filled in a 50 ml Falcon tube and incubated as well at 65°C. The labelled probe (**3.6.3**) was added to the tube, briefly mixed by inverting the tube and then poured on the pre-hybridized membrane. The membranes were incubated over night at 65°C under constant agitation. Subsequently, the membranes were washed at 65°C for 15 minutes using three different solutions: 2 x SSC, 0.1% SDS; 0.5 x SSC, 0.1% SDS; 0.1 x SSC, 0.1% SDS. The decreasing SSC concentration leads to an increasing stringency of the washing. Finally, the membranes were drained on a paper towel and shrink-wrapped in plastic foil.

3.8.1.5 Imaging / stripping

Membranes were exposed to photostimulable storage phosphor imaging plates for 30 minutes to two hours, and signals were then detected by a Typhoon phosphoimager (Amersham Biosciences, UK). Subsequently, the membranes were exposed to X-ray films (BioMax MS, Kodak, U.S.A.) in a lead-coated exposure box at -80°C for 13 hours up to three days. After developing the X-ray film, the membranes were stripped by subsequent washings in 0.2 M NaOH at 45°C. The following washing steps with 2x SSC were repeated until no radioactivity was detectable and the pH value was close to 7. Stripped membranes were sealed in plastic foil and stored at 4°C.

3.8.2 Genomic Southern blot for digoxigenin (DIG) labelling

3.8.2.1 Gel electrophoresis

DNA of ten females of each *D. paulistorum* strain was extracted with the Puregene Kit (Qiagen, Germany) and quantified with the ND4000 NanoDrop spectrophotometer (Peqlab, Germany). 5 µg DNA of each sample were subsequently digested with 2 U *Sau*3A restriction enzyme (New England Biolabs, U.S.A.) in a total reaction volume of 20 µl. The reaction was carried out over night and stopped by heat inactivation at 65 °C for 20 minutes. Standard agarose gel electrophoresis (see section **3.7**) was performed at 90 V for 4 hours to separate the digested DNA fragments by size.

3.8.2.2 Capillary blotting

To stain the DNA in the agarose gel, it was incubated for 30 minutes in 1 x TAE containing 1 µg/ml EtBr. For depurination, the gel was incubated in 300 ml 0.25 M HCl for 15 minutes. Subsequently, the gel was denatured in 250 ml 0.5 m NaOH for 45 minutes, then neutralized in 0.5 M Tris base, pH 7.3, 1.5 M NaCl for 30 minutes. Four layers of Whatman filter papers and a positively charged nylon membrane (Highbond-N+, GE Healthcare, UK) were equilibrated in 2 x SSC for 5 minutes, the blotting apparatus was then assembled as shown in **Fig. 3.3**. Blotting was performed over night, both sides of the membrane were subsequently illuminated with UV light for 4 minutes for crosslinking of the transferred DNA. If not hybridized immediately after, the membranes were shrink-wrapped in plastic foil and stored at 4°C.

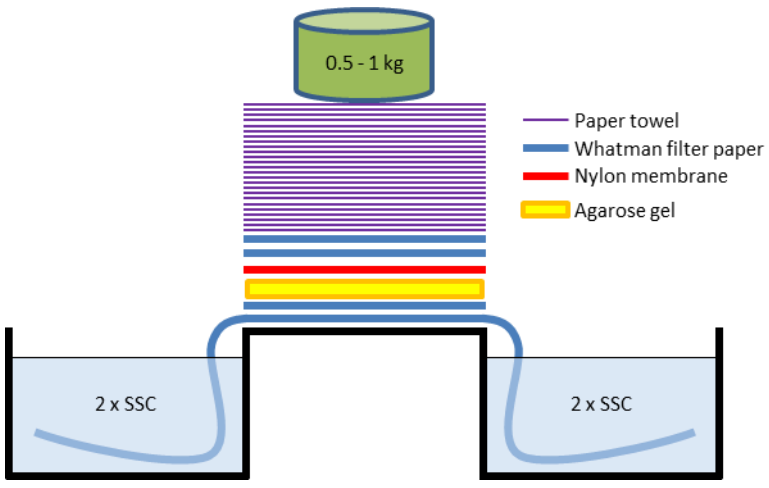


Fig. 3.3: Capillary blotting apparatus. As transfer solution, 2 x SSC was used, the DNA transfer from the agarose gel on the nylon membrane was established by capillary forces against gravity, mediated by Whatman filter papers and paper towels on top of the membrane.

3.8.2.3 Probe labelling

The *ND4* DNA templates used as hybridization probe were obtained through PCR amplification and cloning as described in section **3.3** and **3.4**. DNA probes were labelled with digoxigenin(DIG)-11-dUTP via random priming using the DIG-DNA labelling and detection kit (Roche, Switzerland) and following the manufacturer's protocol. 1 µg *ND4* - plasmid DNA was linearized with 10 U *Pst*I restriction enzyme for 1h at 37°C and subsequently purified with the Wizard SV Gel+PCR clean up kit (Promega, U.S.A.), using 30 µl ddH_2O for elution. 15 µl of the purified plasmid solution (0.5 µg DNA) was denatured in boiling water for 10 min., snapped cool and mixed with 2 µl 10 x Hexanucleotide solution, 2 µl dNTP labelling mix and 1 µl Klenow enzyme (2 U/µl) for incubation at 37°C over night. The reaction was stopped by adding 2 µl 0.2 M EDTA (pH 8.0) and heating to 65°C for 10 minutes.

3.8.2.4 Hybridization / washing

The hybridization reactions were carried out with the Belly Dancer Hybridization Water Bath (Stovall Life Science). Membranes were pre-hybridized for 1 hour in a plastic box with 50 ml pre-hybridization solution (50 ml Church buffer (0.5 M sodium phosphate buffer pH 7.5, 1 mM EDTA, 7% SDS (w/v); adjust to pH 7.2), 312.5 µl salmon sperm) at 50°C under constant agitation. The labelled probe was denatured for 5 min, poured on the pre-hybridized membrane and incubated for six hours at 50°C under constant agitation. Subsequently, stringency washes were performed with three different solutions at constant agitation: five minutes in 2 x SSC, 0.1% SDS at RT, 15 minutes in 0.5 x SSC, 0.1% SDS at 50 °C, 15 minutes in 0.1 x SSC, 0.1% at 50 °C.

3.8.2.5 Detection

All following steps were performed at room temperature. Washed membranes were briefly rinsed in washing buffer (0.1 M Maleic acid, 0.15 M NaCl, pH 7.5, 0.3% (v/v) Tween 20) and incubated for 30 minutes in 100 ml 1 x blocking solution (1% (w/v) Blocking reagent in maleic acid buffer (0.1 M maleic acid, 0.15 M NaCl, pH 7.5, 0.3% (v/v) Tween 20). Blocking solution was discarded and the membrane incubated for 30 minutes in 20 ml antibody solution (2 µl anti-digoxigenin-AP in 20 ml blocking solution). After two washing steps (2 x 15 minutes in 100 ml washing buffer), membranes were equilibrated for 2-5 minutes in 20 ml detection buffer (0.1 M Tris-HCl, 0.1 M NaCl, pH 9.5) and then incubated in 25 ml freshly prepared colorsubstrate solution (112.5 µl NBT, 87.5 µl BCIP in 25 ml NTMT buffer) in the dark until bands were visible. The colour reaction was stopped by washing the membrane with ddH_2O.

4. Results

4.1 Antimicrobial Peptides (AMPs) and immune-related genes in *Drosophila paulistorum* semispecies

Fast evolving antimicrobial peptides (AMPs) and immune-related genes are prime candidates for the development of molecular marker systems for strain typing, as they show high allelic divergence within insect species (Lazzaro and Clark, 2001; Schlenke and Begun, 2004). As no sequence data for AMPs and immune-related genes of *D. paulistorum* semispecies are available, described protein sequences of *D. melanogaster* AMPs and immune-related genes were used to perform protein BLAST (blastp algorithm) in *D. willistoni*, the closest related species to *D. paulistorum* with a sequenced genome (Drosophila twelve Genomes Consortium, 2007). Sequences highly homologous at the amino acid level to the *D. melanogaster* AMP genes *Attacin A*, *Attacin C*, *Attacin D*, *Diptericin*, *Diptericin B* and *Drosocin* (Åsling et al., 1995; Hedengren et al., 2000) were found in *D. willistoni.* Additionally, the genes for *D. willistoni* homologues of two immune receptors, Scavenger Receptor Class C1 and Tehao (Luo and Zheng, 2000; Pearsin et al., 1995) and the cell-surface protein Hemomucin (Theopold et al., 1996) could be identified. Using the *D. willistoni* sequence data as template, twelve sets of PCR primers (see primer list **Table 3.3**) were designed, targeting the coding region of those genes. The primer sets were applied for PCRs with DNA samples deriving from *D. paulistorum* semispecies in order to determine presence/absence of aforementioned genes. For *Attacin C,* two forward primers were designed, producing a larger (458 bp, “GK17822 long1 F”, **Fig. 4.1 A-b**) and a smaller (242 bp, “GK17822 new F”, **Fig. 4.1. A-c**) fragment, as a first set of primers, covering the whole gene, did not show any signal. For each PCR, *D. willistoni* DNA served as positive control and DNA isolated from distantly related *D. ananassae* was used as negative control to rule out random priming.

4.1.1 Presence of AMP homologues in *Drosophila paulistorum* semispecies

As shown in **Figure 4.1 A (a-f)**, PCRs on *D. paulistorum* produced clear signals in all six semispecies samples for *Attacin* and *Diptericin* genes. The corresponding gene maps, indicating the exon/intron structure of the *D. willistoni* genes as well as the expected size of the PCR fragments are shown in **Fig. 4.1 B**. *Drosocin*, however, could not be detected in the Orinocan strain, but was present in all other semispecies. The *Drosocin* PCR was repeated on O11 DNA of different, independent templates, including DNA and cDNA provided by D. Schneider (**Fig. 4.2**). All PCRs showed the same result, *i.e*., absence of a *Drosocin* signal from O11. Hence, the Orinocan semispecies is either lacking the gene for *Drosocin* due to a deletion or has acquired an insertion of a size exceeding the maximum length for a PCR fragment with the applied conditions.

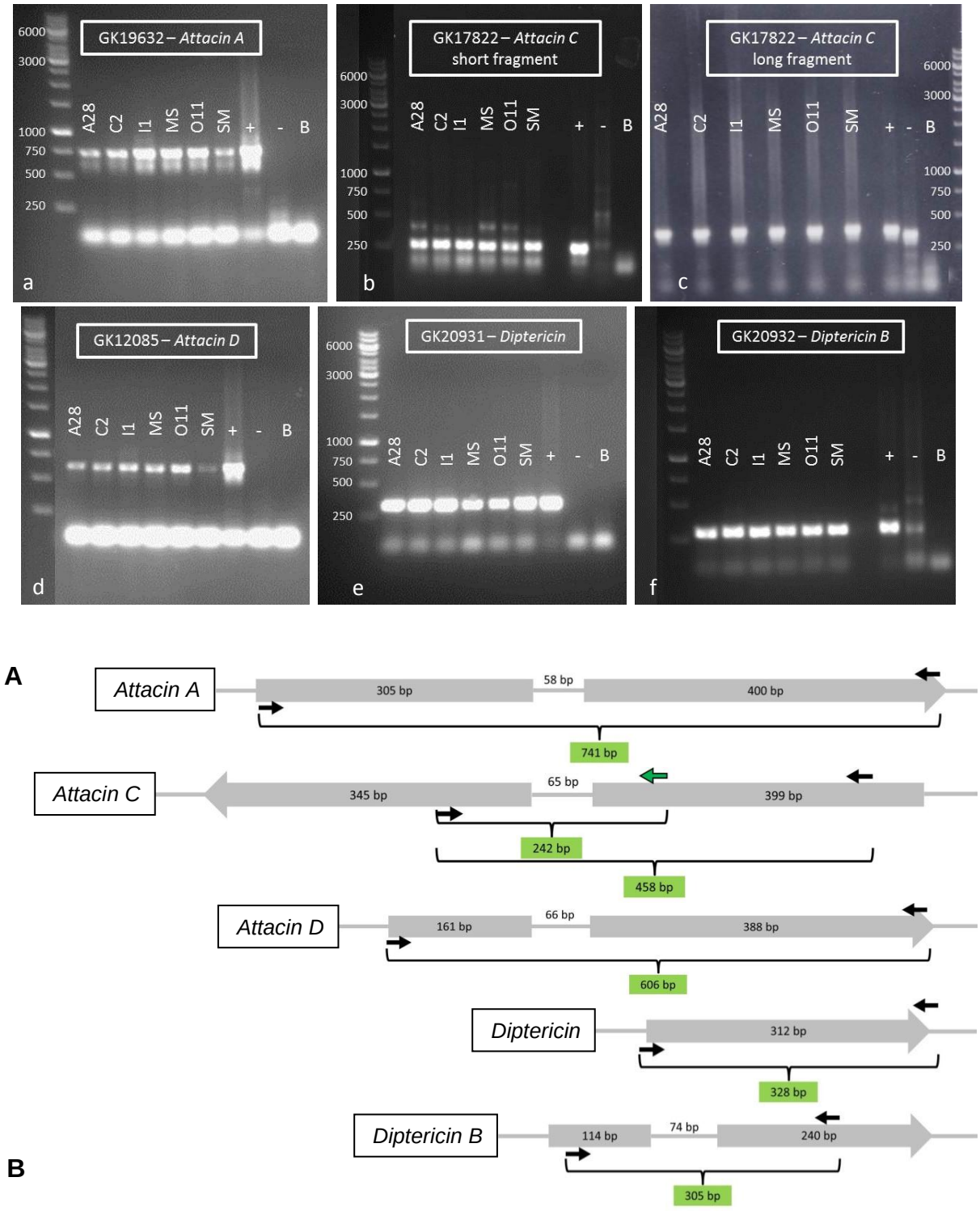


Fig. 4.1 A : PCRs for anti-microbial peptide (AMP) genes *Attacin A, -C, -D*, *Diptericin* and *Diptericin B* (**a-f**). DNA of *D. paulistorum* strains was used as template: A28 = Amazonian, C2= Centroamerican, I1 = Interior, MS = Andean Brazilian, O11 = Orinocan, SM = Transitional. Positive control (+) = *D. willistoni* DNA, negative control (-) = *D. ananassae* DNA, B = Blank. **B:** Gene maps of *D. willistoni* AMP homologues. PCR primer binding sites are indicated with black arrows, the PCR fragment size is highlighted in green. Green arrow for *Attacin C* indicates primer for the short fragment. Exon (thick grey arrows/fields) and intron (grey lines) sizes are indicated. *Attacin C* is coded on the negative strand, hence the gene orientation points to the left.

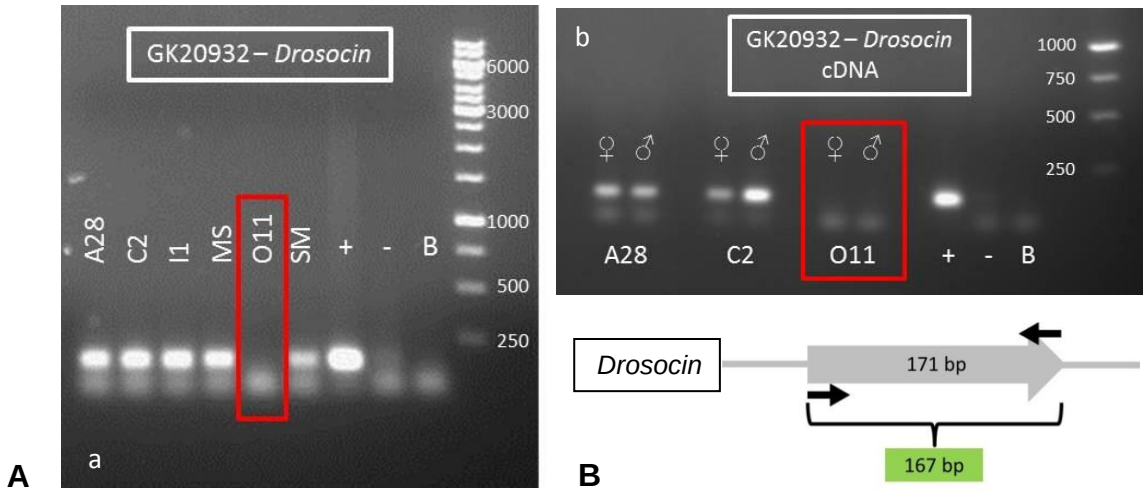


Fig. 4.2 A: PCR for *Drosocin*. **a** = DNA of *D. paulistorum* strains was used as template; A28 = Amazonian, C2= Centroamerican, I1 = Interior, MS = Andean-Brazilian, O11 = Orinocan, SM = Transitional. **b** = cDNA deriving from female and male *D. paulistorum* strains A28, C2 and O11 was used as template. **a & b:** positive control (+) = *D. willistoni* DNA, negative control (-) = *D. ananassae* DNA, B = Blank. Red squares point out the lacking signal of O11. **B:** Gene map of *D. willistoni Drosocin* homologue. PCR primer binding sites are indicated with black arrows, the PCR fragment size is highlighted in green.

4.1.2 Presence of *SrC1* homologues in *Drosophila paulistorum* semispecies

In *D. willistoni, the* immune receptor *Scavenger Receptor Class C1* (*SrC1*) gene consists of four exons of 223, 529, 792 and 376 bp length (**Fig. 4.3**), hence, three sets of PCR primers were designed: *SrC1*-I and *SrC1*-II to cover the first two exons and introns (intron sizes are 65 and 152 bp, respectively) and *SrC1*-III, covering 348 bp of the the fourth exon. The PCRs showed signals for all three fragments in all examined strains (**Fig. 4.4 a-c**).

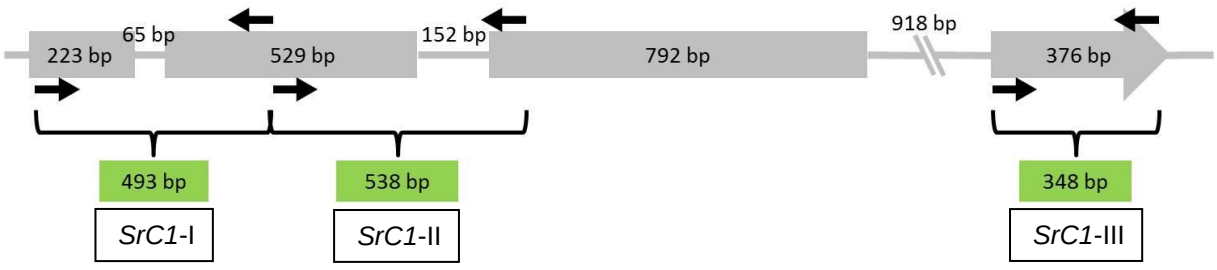


Fig. 4.3 Gene map of *D. willistoni Scavenger receptor Class I (SrC1)* homologue. PCR primer binding sites are indicated with black arrows, PCR fragment sizes are highlighted in green. Four Exons (thick grey arrows/fields) and three introns (grey lines) are indicated with their respective sizes. Three PCR fragments were examined, *SrC1*-I, -II and –III.

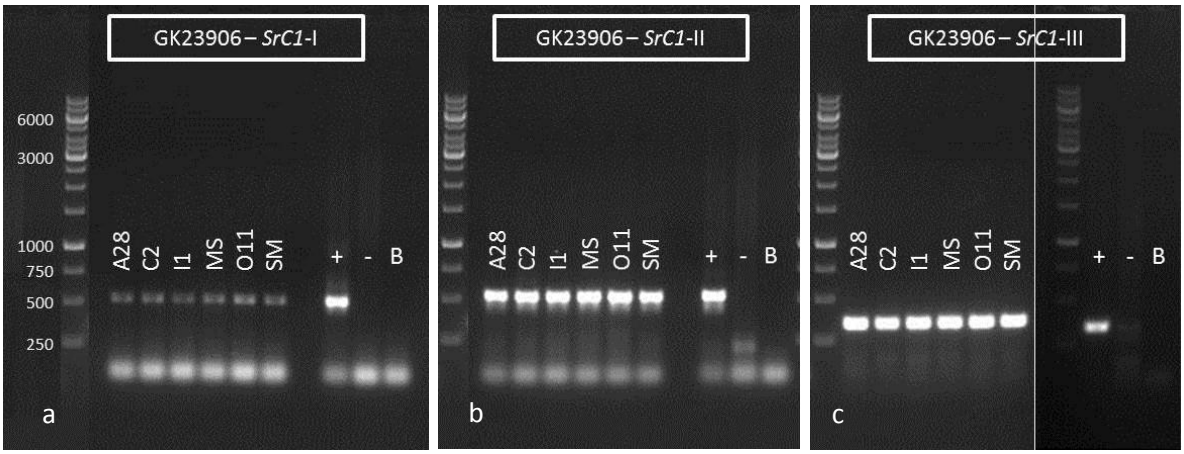


Fig. 4.4: PCRs for *Scavenger Receptor Class C1* (*SrC1*). **a:** *SrC1*-I = fragment covering exon 1, intron 1 and part of exon 2. **b:** *SrC1*-II = fragment covering part of exon 2, intron 2 and part of exon 3. **c:** *SrC1*-III = fragment covering exon 4. DNA of *D. paulistorum* strains was used as template: A28 = Amazonian, C2 = Centroamerican, I1 = Interior, MS = Andean-Brazilian, O11 = Orinocan, SM = Transitional. Positive control (+) = *D. willistoni* DNA, negative control (-) = *D. ananassae* DNA, B = Blank.

4.1.3 Presence of *Tehao* homologues in *Drosophila paulistorum* semispecies

The PCR primers for *Tehao* target 600 bp (*Tehao*-I) and 817 bp (*Tehao*-II) of the first and second exon, respectively (**Fig. 4.5**). All examined strains showed signals of expected size for both fragments (**Fig. 4.6 a, b**).

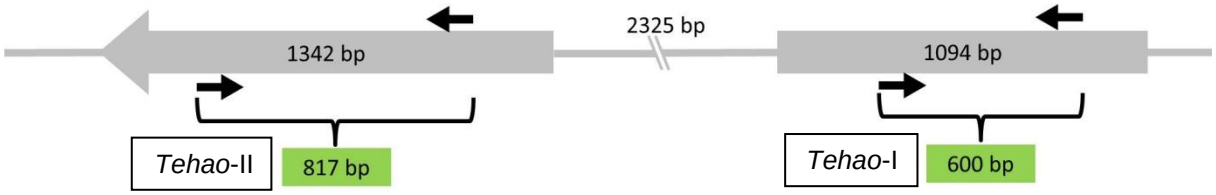


Fig. 4.5: Gene map of *D. willistoni* homologue of *Tehao*. The two exons are indicated by thick grey arrows/fields, the intron by a grey line. Black arrows represent PCR primer binding sites, the PCR-fragment sizes for *Tehao*-I and *Tehao*-II are highlighted in green. *Tehao* is coded on the negative strand.

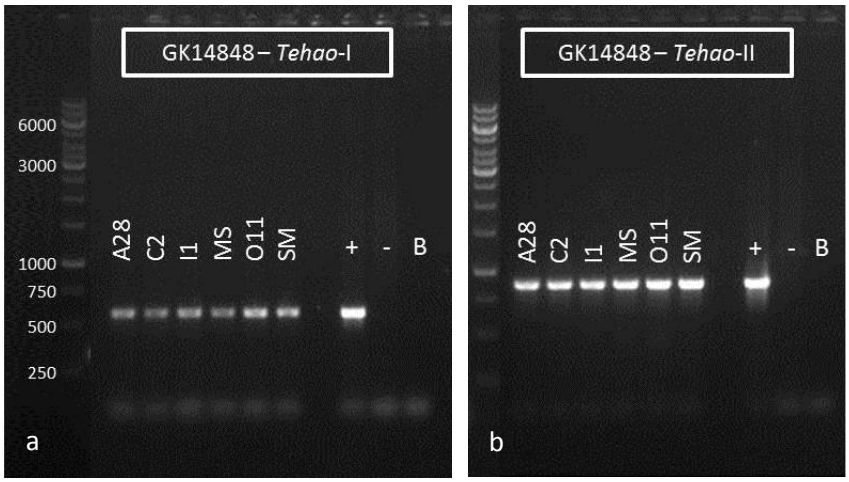


Fig. 4.6: PCRs for *Tehao*. **a**: *Tehao*-I fragment, partially covering exon 1, **b**: *Tehao*-II fragment, partially covering exon 2. DNA of *D. paulistorum* strains was used as template: A28 = Amazonian, C2 = Centroamerican, I1 = Interior, MS = Andean-Brazilian, O11 = Orinocan, SM = Transitional. Positive control (+) = *D. willistoni* DNA, negative control (-) = *D. ananassae* DNA, B = Blank.

4.1.4 Presence of *Hmu* homologues in *Drosophila paulistorum* semispecies

The *D. willistoni* homologue for the gene for the cell-surface protein *Hemomucin* (*Hmu*) comprises four exons of 117, 206, 301 and 1024 bp length (**Fig. 4.7**). Two sets of primers were designed: The first set (*Hmu*-I) targets the first, second and part of the third exon with the intermediate introns, covering 572 bp. The second set (*Hmu*-II) covers the other part of the third exon and half of the fourth exon, including the intron located in between. PCR with *Hmu*-I primers showed a clear signal in the *D. willistoni* positive control, but did not function in any of the *D. paulistorum* strains (**Fig. 4.8 a**). *Hmu*-II, however, showed clear signals for all *D. paulistorum* strains (**Fig. 4.8 b**).

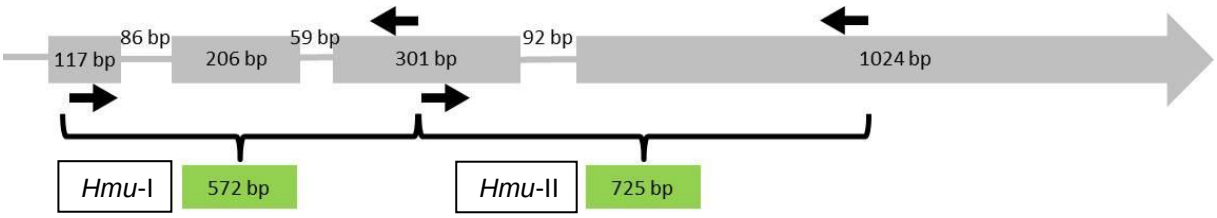


Fig. 4.7: Gene map of *D. willistoni* homologue of *Hemomucin*. Exons are indicated by thick grey arrows/fields, introns by a grey line. Black arrows represent PCR primer binding sites, the sizes of the resulting PCR-fragments *Hmu*-I and *Hmu*-II are highlighted in green.

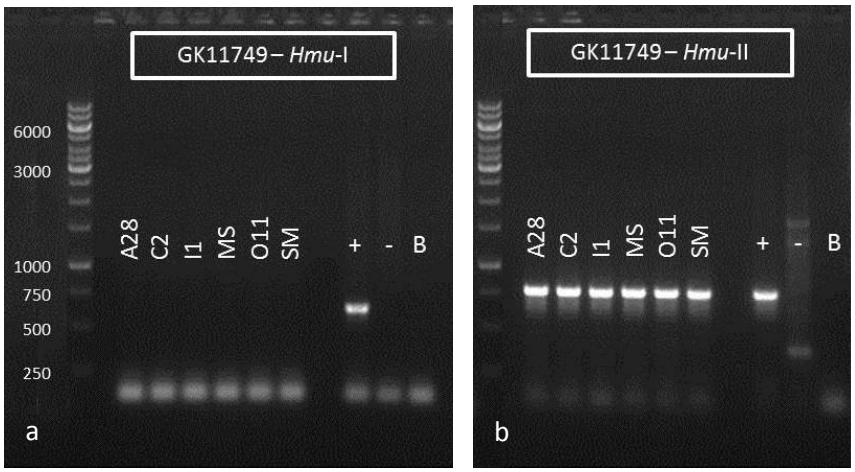


Fig. 4.8: PCRs for *Hemomucin*. **a:** *Hmu*-I fragment, covering exon1, 2 and partially exon 3, **b:** *Hmu*-II fragment, partially covering exon 3 and 4. DNA of *D. paulistorum* strains was used as template: A28 = Amazonian, C2 = Centroamerican, I1 = Interior, MS = Andean-Brazilian, O11 = Orinocan, SM = Transitional. Positive control (+) = *D. willistoni* DNA, negative control (-) = *D. ananassae* DNA, B = Blank.

4.1.5 Cloning and Sequencing of *D. paulistorum* AMP genes

Based on the PCR presence/absence screen (see section **4.1.1**), candidate genes for AMPs were considered the most interesting for more detailed studies, as the PCR fragments are small (167 – 741 bp) and cover in most cases large parts of the coding region. Hence, we excluded the immune-related genes *Hmu, SrC1 and Tehao* from further analyses and focused on the six AMP homologous for cloning and sequencing. In a first round, DNA from strains of the six semispecies of *D. paulistorum* and five potential outgroups (*D. equinoxialis, D. capricorni, D. nebulosa, D. tropicalis and D. saltans*) were used to amplify, clone and sequence the AMP genes. Sequences of *Attacin A* -, *Attacin C* -, *Attacin D* - and *Diptericin B* showed for some strains multiple types. To rule out PCR errors and to confirm potential paralogues, a second round of PCR-amplification, cloning and sequencing was performed with those four genes in the six *D. paulistorum* strains. For *Attacin C*, the inconsistencies could be solved as the second PCR confirmed only one sequence type for each strain. The potential paralogues of *Attacin A* and *-D*, *Diptericin* and *Diptericin B* for some strains, however, were confirmed by the cloning and sequencing of the second round of independent PCRs. **Table 4.1** summarizes the number of complete reads obtained for each strain and fragment. Not all sequencing reactions resulted in high-quality sequence reads or they were incomplete, thus causing variation in the number of complete sequences available for each strain.

	AttA	*AttC*	*AttD*	*Dipt*	*DiptB*	*Dro*
A28	4 / 2	4 / 2	4 / 2	4 / 1	8 / 2	3 / 1
C2	4 / 2	3 / 2	4 / 1	3 / 1	8 / 2	3 / 1
I1	4 / 2	3 / 2	4 / 2	4 / 1	6 / 2	4 / 1
MS	2 / 2	3 / 2	4 / 2	2 / 1	6 / 2	1 / 1
O11	3 / 2	4 / 2	4 / 2	4 / 1	8 / 2	-
SM	2 / 2	3 / 2	4 / 2	2 / 1	6 / 2	2 / 1
D. equinoxialis	3 / 1	2 / 1	3 / 1	3 / 1	3 / 1	2 / 1
D. capricorni	3 / 1	3 / 1	3 / 1	3 / 1	3 / 1	-
D. nebulosa	3 / 1	3 / 1	-	-	3 / 1	-
D. tropicalis	3 / 1	3 / 1	-	-	3 / 1	2 / 1
D. saltans	3 / 1	2 / 1	-	-	-	-

Table 4.1: Number of complete reads (first digit in each cell, each from an independent clone) obtained from sequencing of cloned AMP PCR products. The second digit in each cell displays the number of PCRs performed to obtain the sequenced clones. Abbreviations: *AttA*, *-C*, *-D* = *Attacin A, -C, -D*; *Dipt* = *Diptericin, DiptB* = *Diptericin B; Dro* = *Drosocin*. A28 = Amazonian, C2 = Centroamerican, I1 = Interior, MS = Mesitas, O11 = Orinocan, SM = Santa Marta. Hyphen indicates that no PCR fragment was obtained for the respective strain and AMP.

Direct sequencing was subsequently performed to support the evidence for potential paralogues obtained from cloned sequences of *Attacin D*. DNA provided by D. Schneider from the classic six *D. paulistorum* semispecies was included in the set to confirm the reads with independent templates and to rule out strain contaminations**.** **Figure 4.9** shows in an exemplary alignment of cloned and direct-sequencing reads the confirmation of the first single nucleotide polymorphism (SNP) in *Attacin D* for Amazonian semispecies at position 181. Direct- sequencing PCRs were performed twice for each template with the *Attacin D* forward and reverse primer, respectively. Altogether, eight direct sequencing reads were obtained for each strain to confirm potential paralogues, which were labelled with roman numerals (A28-I, -II etc.). The SNPs of *Attacin D* shown in the following chapter were all confirmed by direct sequencing.

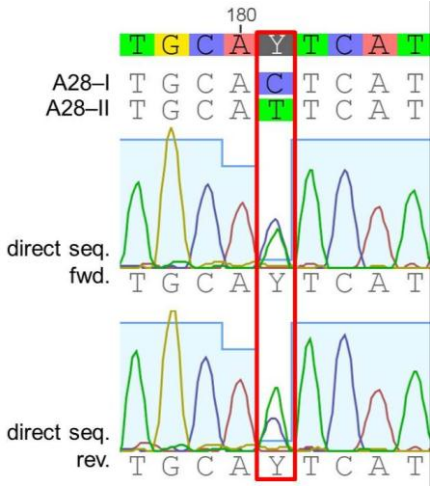


Fig. 4.9: Section of the alignment of *Attacin D* sequences of Amazonian *D. paulistorum* strain (A28). The consensus sequences of the cloned reads (A28-I, A28-II) and an exemplary forward and reverse direct sequencing read (direct seq. fwd, -rev.) were aligned using the Geneious software. Position 181 shows the first SNP to differentiate two potential A28 paralogues. Double peaks in the direct sequencing chromatogram confirm the C/T polymorphism observed with cloned sequences.

4.1.6 AMP sequence analysis

The obtained AMP-sequences of *D. paulistorum* were aligned with outgroups, using the Geneious alignment algorithm (Biomatters Ltd.). Variable nucleotide positions within the *paulistorum* group were examined and intron/exon structure as well as protein domains were determined, according to the *D. melanogaster* homologues (Hedengren, 2000). Through *in silico* splicing and in-frame translation, synonymous mutations could be distinguished from non-synonymous mutations. The non-synonymous mutations were then classified as chemically equivalent or non-equivalent, according to the chemical properties of the alternative amino acid. For all AMPs, polymorphism tables and substitution tables were established.

4.1.6.1 *Attacin A*

The protein structure of *Attacin A* with the corresponding nucleotide positions of the PCR fragment is shown in **Figure 4.10**. In total, we found 17 variable sites in the 696 - 697 bp long fragment (2.4% divergence without primers) within the six classic *D. paulistorum* strains. Out of these, eleven sites locate to the coding region, and six are present within the intron, including one deletion in I1 (**Table 4.2 A**). The signal- and propeptide (nt position 1-102) are almost completely conserved, only a single variable site was found in the signal peptide (nt position 48). The remaining ten variable sites are evenly distributed over the N-terminal (nt position 103-274), G1- (nt position 348-539) and G2-domain. In the C2 strain, we found two different types of the *Attacin A* fragment (0.4% divergence to each other). Two non-synonymous sites were found in the N-terminal region, leading to a chemically equivalent amino-acid substitution from serine to asparagine in aa-position 44 and a non-equivalent substitution in position 63 from lysine to asparagine (**Table 4.2 B**). Interestingly, an "opal" mutation resulting in a novel stop codon (TG**A** at position 662 in the alignment) was found in the G2 domain of the SM strain, leading to a truncated G2 domain.

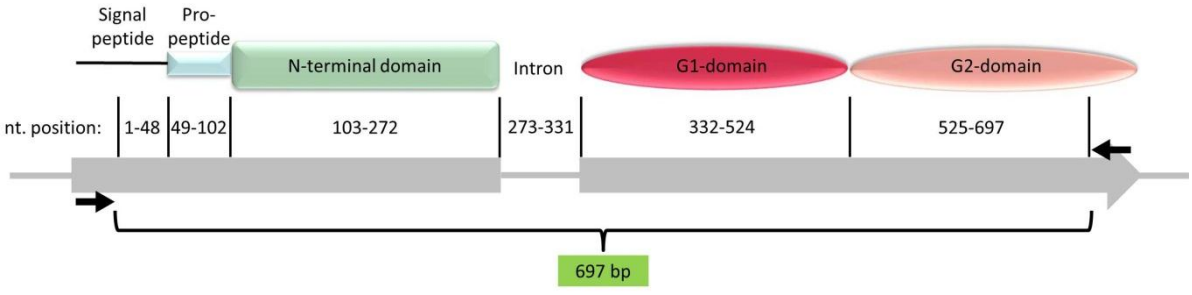


Figure 4.10: *Attacin A* protein domains according to *D. melanogaster* homologues and respective DNA regions on *D. paulistorum* genome, shown in grey. Black arrows represent PCR primers, the total PCR fragment size (without primers) is highlighted in green. "nt position" refers to the region in the alignment of PCR fragments coding for the corresponding protein domain.

A

AttA nt Position	048	129	131	189	234	249	267	280	286	292	308	318	324	446	449	455	479	536	575	584	605	629	662
Type	**S**	**S**	**R**	**R**	**S**	**S**	**S**	**N**	**N**	**N**	**N**	**N**	**N**	**S**	**S**	**S**	**S**	**S**	**S**	**S**	**S**	**S**	**R**
A28	T	A	G	A	A	G	T	t	a	t	t	t	t	G	C	C	C	C	T	T	G	C	G
C2-I	T	G	G	A	T	G	C	t	c	c	a	t	t	G	T	C	C	C	T	C	A	C	G
C2-II	C	G	G	A	T	G	C	t	c	c	a	t	a	G	C	C	C	C	T	C	A	C	G
I1	T	A	G	A	T	C	C	t	–	t	t	c	t	G	C	C	C	C	G	T	G	G	G
MS	T	G	A	C	T	C	C	t	a	t	t	t	t	G	C	C	T	T	T	C	G	C	G
O11	T	G	A	C	T	C	C	c	a	t	t	t	t	G	C	C	C	C	T	C	G	C	G
SM	T	A	G	A	T	G	C	t	c	t	t	t	t	A	C	G	C	C	T	C	A	C	A
D. wls	T	A	A	A	T	C	C	t	a	t	t	t	c	G	C	C	C	C	T	C	A	C	G
D. equi	T	G	G	A	T	C	C	a	a	t	t	t	c	G	C	C	C	C	T	C	G	C	G
D. trop	C	G	G	A	T	A	C	t	a	a	t	t	c	G	C	C	C	C	T	T	A	C	G
D. cap	T	G	A	T	–	T	C	a	a	t	t	t	t	A	C	C	C	C	T	C	T	C	G
D. neb	T	G	A	T	–	T	C	–	–	t	t	t	t	A	C	C	C	C	T	C	T	C	G
D. salt	T	G	G	T	–	A	C	t	g	c	t	t	t	G	T	C	C	C	T	T	T	A	G

B

AttA nt Position	131	189	662
Type	**E**	**N**	**St**
A28	S	K	W
C2-1	S	K	W
C2-2	S	K	W
I1	S	K	W
MS	N	N	W
O11	N	N	W
SM	S	K	*
D. wls	N	K	W
D. equi	S	K	W
D. trop	S	K	W
D. cap	N	D	W
D. neb	N	D	W
D. salt	S	N	W

Table 4.2: Polymorphism table (**A**) and amino acid substitution table (**B**) for *Attacin A* gene determined in the six *D. paulistorum* semispecies: Amazonian (A28); Centroamerican with two potential paralogues (C2-I & C2-II); Interior (I1); Andean-Brazilian (MS); Orinocan (O11); and Transitional (SM). As outgroups, *D. willistoni* (*D. wls*), *D. equinoxialis* (*D. equi*), *D. tropicalis* (*D. trop*), *D. capricorni* (*D. cap*), *D. nebulosa* (*D. neb*) and *D. saltans* (*D. salt*) are shown. **A:** "nt Position" refers to nucleotide position in the alignment (of *D. paulistorum* semispecies without primers, without outgroups) in exonic (grey) and intronic (white) boxed regions. Type refers to noncoding (N), silent (S) or replacement (R) mutation; green filling represents changes from the majority consensus sequence and hyphen (–) refers to single nucleotide deletion. **B:** "nt Position" refers to nucleotide position in the alignment responsible for the amino acid substitution. Type refers to chemically equivalent (E) and non-equivalent (N) substitution. Yellow filled cell (St) refers to stop-codon (*) in aa 201 of *Attacin A* in the SM strain.

4.1.6.2 *Attacin C*

The 406 - 414 bp long fragment (without primers) of *Attacin C* covers 13 bp of the signal peptide, the entire propetide, N-terminal domain and intron (varying between 289 and 361 bp) as well as 51 bp of the G1 domain (**Fig. 4.11**). Within the six *D. paulistorum* semispecies, the fragment comprises four variable sites in the coding region and 14 in the intron (4.4% divergence), including eight indels in MS and O11 strain (**Table. 4.3 A**). Two SNPs in the coding region are non-synonymous, one in the N-terminal, another in the G1 domain (**Table. 4.3 B**).

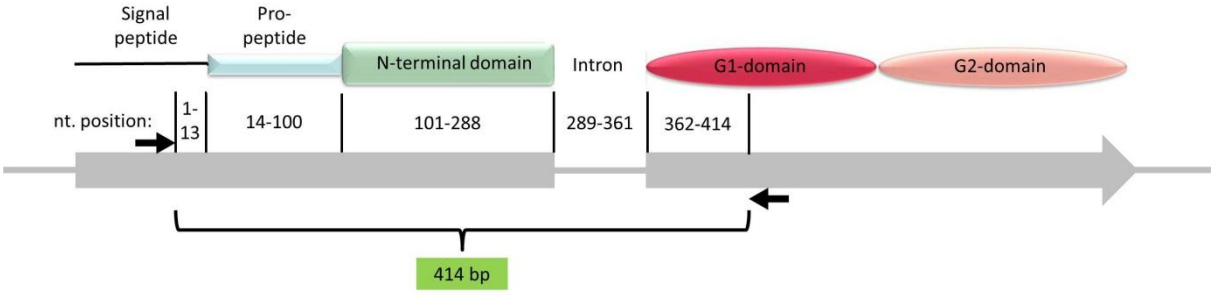


Figure 4.11: *Attacin C* protein domains according to *D. melanogaster* homologues and respective DNA regions on *D. paulistorum* genome, shown in grey. Black arrows represent PCR primers, the total PCR fragment size (without primers) is highlighted in green. "nt position" refers to the region in the alignment of PCR fragments coding for the corresponding protein domain.

A

AttC **nt Position**	**031**	**082**	**243**	**300**	**301**	**318**	**320**	**322**	**323**	**324**	**325**	**326**	**327**	**328**	**329**	**332**	**350**	**380**
Type	**S**	**S**	**R**	**N**	**N**	**N**	**N**	**N**	**N**	**N**	**N**	**N**	**N**	**N**	**N**	**N**	**N**	**R**
A28	C	G	A	t	g	a	a	g	g	t	c	a	t	a	a	t	c	T
C2	T	A	A	t	g	a	a	g	g	t	t	a	t	a	a	t	c	T
I1	C	G	A	t	g	a	a	g	g	t	t	a	t	a	a	t	c	T
MS	T	A	A	a	a	g	t	-	-	-	-	-	-	-	-	c	t	A
O11	T	A	T	a	a	g	t	-	-	-	-	-	-	-	-	c	t	A
SM	C	G	A	t	g	a	a	g	g	t	t	a	t	a	a	t	c	T
D. wls	C	G	A	a	a	g	t	-	-	-	-	-	-	-	-	c	t	C
D. equi			A	a	a	g	t	-	-	-	-	-	-	-	-	c	t	T
D. trop			A	a	a	g	t	-	-	-	-	-	-	-	-	c	t	T
D. cap			-	a	g	c	a	-	-	-	-	-	-	-	-	t	t	T
D. neb			-	-	-	c	a	-	-	-	-	-	-	-	-	t	t	T

B

AttC **nt Position**	**243**	**380**
Type	**N**	**E**
A28	Q	D
C2	Q	D
I1	Q	D
MS	Q	E
O11	L	E
SM	Q	D
D. equi	Q	D
D. wls	Q	D
D. trop	Q	D
D. cap	-	D
D. neb	-	D

Table 4.3: Polymorphism table (**A**) and amino acid substitution table (**B**) for *Attacin C* gene determined in the six *D. paulistorum* semispecies: Amazonian (A28); Centroamerican (C2); Interior (I1); Andean Brazilian (MS); Orinocan (O11); and Transitional (SM). As outgroups, *D. willistoni* (*D. wls*), *D. equinoxialis* (*D. equi*), *D. tropicalis* (*D. trop*), *D. capricorni* (*D. cap*) and *D. nebulosa* (*D. neb*) are shown. **A:** "nt Position" refers to nucleotide position in the alignment (of *D. paulistorum* semispecies without primers, without outgroups) in exonic (grey) and intronic (white) boxed regions. Type refers to noncoding (N), silent (S) or replacement (R) mutation; green filling represents changes from the majority consensus sequence and hyphen (–) refers to single nucleotide deletion. For *D. equi., D. trop., D. cap.* and *D. neb*. only the smaller fragment of 242 bp was examined (see Fig. 4.1 B). **B:** "nt Position" refers to nucleotide position in the alignment responsible for the amino acid substitution. Type refers to chemically equivalent (E) and non-equivalent (N) substitution.

4.1.6.3 *Attacin D*

In *D. melanogaster* and *D. willistoni Attacin D* is the only of the examined AMPs lacking a signal- and propetide (**Fig. 4.11**). It may therefore be either non-functional, or it could have an intracellular function (Hedengren, 2000). In the *D. paulistorum* species complex, 44 variable sites where found in the *Attacin D* PCR fragments, varying in size without primers from 555 to 566 bp (7.7 – 7.9% divergence). Interestingly, two different types were found in four out of six strains (A28-I & -II, C2-I & -II, I1-I & -II, SM-I & -II; 0.9 – 3.7 divergence between the paralogues). As all strains are highly inbred isofemale lines, those types cannot represent different alleles of a population, but are more likely paralogues at different loci originated by duplications. Those potential paralogues differ in at least four positions in coding regions with at least one non-synonymous substitution (**Table 4.4 A**). The divergence in the intronic regions is for some strains even higher (each deletion was counted): 14 positions in the intron distinguish C2-I from C2-II. The highest divergence was observed between supposed paralogues of the I1 strain and the SM strain (I1-I & -II; SM-I & -II) with four synonymous and four non-synonymous substitutions in the coding region and 12 different positions in the intron. Remarkably, the sequence types of I1 and SM strains (I1-I & SM-I; I1-II & SM-II) are identical. In total, all *Attacin D* sequences of *D. paulistorum* semispecies show 20 variable positions in coding regions (exon) and 24 in non-coding regions (intron, see **Table 4.4 A**). With five non-synonymous sites out of seven variable positions, the N-terminal domain (nt positions 1-142) shows the most divergent peptide, followed by the G1 domain (nt positions 209-388) with four out of six. The most conserved protein region of *D. paulistorum Attacin D* is the G2 (nt positions 389-566) domain with only one non-synonymous sites out of seven variable positions. In total, ten non-synonymous substitutions were observed in *D. paulistorum*, six of which lead to a chemically non-equivalent amino acid change in the translation (**Table 4.4 B**).

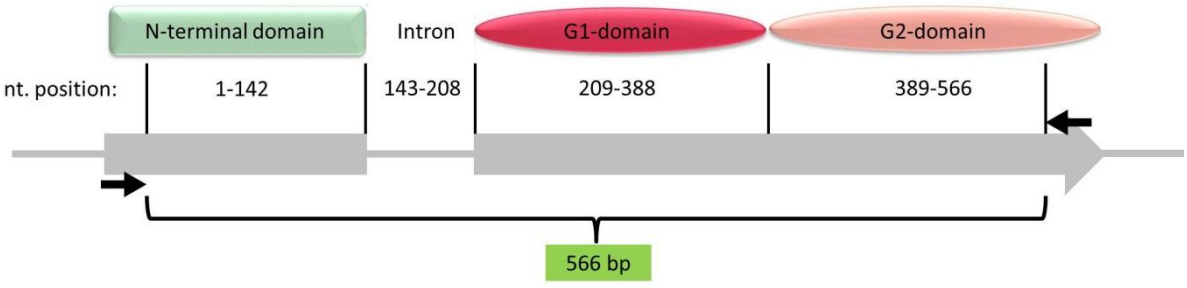


Figure 4.11: *Attacin D* protein domains according to *D. melanogaster* homologues and respective DNA regions on *D. paulistorum* genome, shown in grey. Black arrows represent PCR primers, the total PCR fragment size (without primers) is highlighted in green. “nt position” refers to the region in the alignment of PCR fragments coding for the corresponding protein domain.

A

AttD nt Position	022	030	059	065	076	100	134	155	161	170	171	172	173	174	175	176	177	178	179	180	181	182	183	184	185	186	188	189	192	193	203	257	269	301	309	354	364	397	457	478	493	507	541	562
Type	S	R	R	R	S	R	R	N	N	N	N	N	N	N	N	N	N	N	N	N	N	N	N	N	N	N	N	N	N	N	N	R	R	S	R	R	S	S	S	S	S	R	S	S
A28-I	C	G	G	G	A	A	G	c	a	a	a	a	c	g	t	t	t	g	c	a	c	t	c	a	t	a	g	c	t	a	g	G	T	A	C	A	A	T	A	G	C	G	T	T
A28-II	C	G	G	G	A	A	G	c	a	a	a	a	c	g	t	t	t	g	c	a	t	t	c	a	t	a	g	c	t	a	g	G	G	A	C	C	A	T	A	G	T	G	T	C
C2-I	C	A	G	G	A	A	A	c	t	a	a	a	–	–	t	t	t	–	–	–	t	a	c	a	c	c	t	t	a	a	g	G	G	A	C	A	A	T	T	T	C	G	T	C
C2-II	C	A	G	G	A	A	G	t	a	a	a	a	c	g	t	t	t	g	c	a	t	t	c	a	t	a	g	c	t	a	t	G	G	T	C	A	A	T	A	G	T	G	T	C
I1-I	T	A	G	A	A	A	G	c	t	a	a	a	c	g	t	t	t	g	c	a	t	t	c	a	t	a	g	c	t	a	g	A	G	A	C	A	A	C	A	G	T	G	T	C
I1-II	C	A	T	C	C	A	G	c	a	–	–	–	–	–	–	–	–	–	–	–	t	t	c	a	t	a	g	c	t	a	t	G	G	A	C	A	A	T	A	G	T	C	G	C
MS	C	A	G	G	A	T	G	c	a	a	a	a	c	g	–	–	–	–	–	–	–	–	–	–	t	a	g	c	t	a	t	G	G	A	C	A	G	T	T	T	C	G	T	T
O11	C	A	G	G	A	A	G	c	a	a	a	a	c	g	–	–	–	–	–	–	–	–	–	–	t	a	g	c	t	a	t	G	G	A	G	A	A	T	T	T	T	G	T	C
SM-I	T	A	G	A	A	A	G	c	t	a	a	a	c	g	t	t	t	g	c	a	t	t	c	a	t	a	g	c	t	a	g	A	G	A	C	A	A	C	A	G	T	G	T	C
SM-II	C	A	T	C	C	A	G	c	a	–	–	–	–	–	–	–	–	–	–	–	t	t	c	a	t	a	g	c	t	a	t	G	G	A	C	A	A	T	A	G	T	C	G	C
D. wls	C	A	G	G	A	A	G	c	a	a	a	a	c	g	t	a	t	c	c	a	t	t	c	a	t	a	g	t	a	c	g	G	T	A	C	A	T	T	T	A	C	G	T	T
D. equi	C	A	G	G	A	A	G	c	g	a	a	t	c	g	t	t	t	g	t	a	t	t	c	a	t	a	g	t	t	a	g	G	G	A	C	A	T	T	T	T	C	G	T	T
D. trop	C	A	G	G	A	A	G	c	a	a	a	c	c	t	g	t	t	c	c	a	t	t	c	a	t	a	g	t	t	a	c	G	G	A	C	A	T	T	T	T	C	G	T	C

B

AttD nt Position	030	059	065	100	134	257	269	309	354	507
Type	E	N	E	N	E	N	E	N	N	N
A28-I	S	A	V	K	V	G	L	S	E	G
A28-II	S	A	V	K	V	G	V	S	A	G
C2-I	N	A	V	K	M	G	V	S	E	G
C2-II	N	A	V	K	V	G	V	S	E	G
I1-I	N	A	I	K	V	S	V	S	E	G
I1-II	N	S	L	K	V	G	V	S	E	A
MS	N	A	V	N	V	G	V	S	E	G
O11	N	A	V	K	V	G	V	C	E	G
SM-I	N	A	I	K	V	S	V	S	E	G
SM-II	N	S	L	K	V	G	V	S	E	A
D. wls	N	A	V	K	V	G	L	S	E	G
D. equi	N	A	V	K	V	G	V	S	E	G
D. trop	N	A	V	K	V	A	V	S	E	G

Table 4.4: Polymorphism table (**A**) and amino acid substitution table (**B**) for *Attacin D* gene determined in the six *D. paulistorum* semispecies: Amazonian with two potential paralogues (A28-I & A28-II); Centroamerican with two potential paralogues (C2-I & C2-II); Interior with two potential paralogues (I1-I & I1-II); Andean-Brazilian (MS); Orinocan (O11); and Transitional with two potential paralogues (SM-I & SM-II). As outgroups, *D. willistoni* (*D. wls*), *D. equinoxialis* (*D. equi*) and *D. tropicalis* (*D. trop*) are shown. **A:** "nt Position" refers to nucleotide position in the alignment (of *D. paulistorum* semispecies without primers, without outgroups) in exonic (grey) and intronic (white) boxed regions. Type refers to noncoding (N), silent (S) or replacement (R) mutation; green filling represents changes from the majority consensus sequence and hyphan (–) refers to single nucleotide deletion. **B:** "nt Position" refers to nucleotide position in the alignment responsible for the amino acid substitution. Type refers to chemically equivalent (E) and non-equivalent (N) substitution.

4.1.6.4 *Diptericin*

The *Diptericin* protein of *D. melanogaster* and *D. willistoni* lacks both the N-terminal and the G1- domain, it consists only of a signal- and propetide and a G2 domain (**Fig. 4.12**). The *Diptericin* PCR fragment covers 281 bp (without primers), which represents almost the whole gene. Within the *D. paulistorum* semispecies, *Diptericin* has a total of 16 variable sites (5.7% divergence), 12 in the G2 domain, two in the signal- and propeptide, respectively (**Table 4.5 A**). One site in the signal peptide and one in the G2 domain are non-synonymous, but chemically equivalent (**Table 4.5 B**).

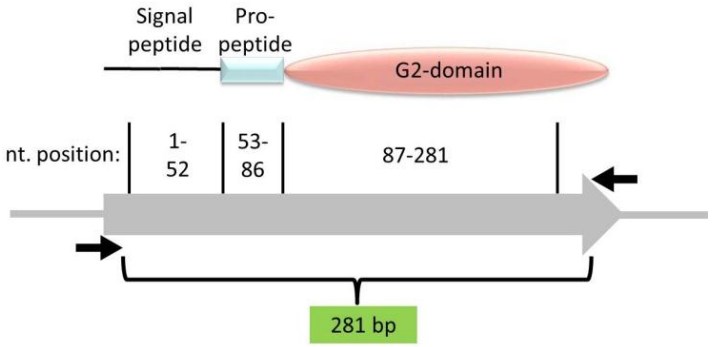


Figure 4.12: *Diptericin* protein domains according to *D. melanogaster* homologues and respective DNA regions on *D. paulistorum* genome, shown in grey. Black arrows represent PCR primers, the total PCR fragment size (without primers) is highlighted in green. "nt position" refers to the region in the alignment of PCR fragments coding for the corresponding protein domain.

A

Dipt **nt Position**	**023**	**039**	**075**	**081**	**096**	**102**	**105**	**120**	**128**	**165**	**186**	**210**	**252**	**261**	**267**	**273**
Type	**R**	**S**	**S**	**S**	**S**	**S**	**S**	**S**	**R**	**S**	**S**	**S**	**S**	**S**	**S**	**S**
A28	T	C	A	A	G	A	C	T	T	C	G	T	T	C	C	C
C2	C	C	G	T	G	T	C	C	A	C	G	T	T	C	C	C
I1	C	C	A	A	A	T	C	C	T	T	A	G	C	T	T	C
MS	C	T	G	T	G	T	T	C	T	C	A	C	T	C	C	T
O11	C	T	G	T	G	T	T	C	T	C	A	C	T	C	C	T
SM	C	C	A	A	G	T	C	C	T	T	A	G	C	T	T	C
D. wls	C	C	A	T	G	T	C	C	T	C	G	T	T	T	C	C
D. equi	C	C	A	T	G	T	C	C	T	C	A	T	T	C	C	C
D. cap	C	C	A	T	G	T	C	C	T	C	G	T	T	C	C	C

B

Dipt **nt Position**	**023**	**128**
Type	**E**	**E**
A28	V	F
C2	A	Y
I1	A	F
MS	A	F
O11	A	F
SM	A	F
D. wls	A	F
D. equi	A	F
D. cap	A	F

Table 4.5: Polymorphism table (**A**) and amino acid substitution table (**B**) for *Diptericin* gene determined in the six *D. paulistorum* semispecies: Amazonian (A28); Centroamerican (C2); Interior (I1); Andean-Brazilian (MS); Orinocan (O11); and Transitional (SM). As outgroups, *D. willistoni* (*D. wls*), *D. equinoxialis* (*D. equi*) and *D. capricorni* (*D. cap*) are shown. **A:** "nt Position" refers to nucleotide position in the alignment (of *D. paulistorum* semispecies without primers, without outgroups). Type refers to silent (S) or replacement (R) mutation; green filling represents changes from the majority consensus sequence. **B:** "nt Position" refers to nucleotide position in the alignment responsible for the amino acid substitution. Type refers to chemically equivalent (E) substitution.

4.1.6.5 *Diptericin B*

Similar to *Diptericin, Diptericin B* protein of *D. melanogaster* and *D. willistoni* is lacking the N-terminal and G1 domain. A 70 - 74 bp long Intron is situated in the coding region for the propetide (**Fig.4.12**). The PCR fragment for *Diptericin B* covers 257 - 261 bp (without primers), including 19 bp of the signal peptide, the entire propeptide and intron as well as 94 bp of the G2 domain. The six *D. paulistorum* semispecies show five variable sites in the coding region and nine in the intron (5.4% divergence), four of them represent deletions (**Table 4.6 A**). Three different sequence types were found in the Centroamercian strain C2 (C2-I, C2-II, C2-III, 3% divergence within those paralogues), the Andean-Brazilian strain showed two potential paralogues (0.3% divergence). One SNP in the signal peptide (nt. position 17) and one in Propetide (nt. position 139) lead to a change in the amino acid sequence, both are chemically non-equivalent (**Table 4.6 B**).

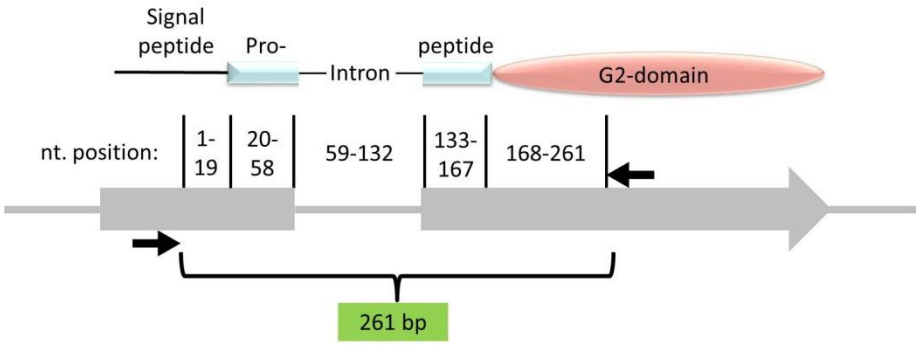


Figure 4.12: *Diptericin B* protein domains according to *D. melanogaster* homologues and respective DNA regions in *D. paulistorum* genome, shown in grey. Black arrows represent PCR primers, the total PCR fragment size (without primers) is highlighted in green. “nt position” refers to the region in the alignment of PCR fragments coding for the corresponding protein domain.

A

Dipt B nt Position	017	052	069	074	075	092	115	116	117	118	119	139	165	189
Type	**R**	**S**	**N**	**N**	**N**	**N**	**N**	**N**	**N**	**N**	**N**	**R**	**S**	**S**
A28	C	C	a	c	g	t	g	t	-	-	-	T	G	A
C2-I	C	C	a	c	g	t	t	t	-	-	-	C	C	A
C2-II	T	C	t	c	a	t	t	c	t	t	g	C	C	A
C2-III	C	T	c	c	g	t	t	-	-	-	-	C	C	A
I1	C	C	a	c	g	g	c	t	-	-	-	C	C	T
MS-I	C	C	a	c	g	t	t	t	t	-	-	C	C	T
MS-II	C	C	a	t	g	t	t	t	t	-	-	C	C	T
O11	C	C	a	c	g	t	t	t	t	-	-	C	C	T
SM	C	C	a	c	g	g	t	t	-	-	-	C	C	T
D. wls	C	C	g	a	g	t	t	t	-	-	a	C	C	A
D. equi	C	C	a	a	g	t	t	-	-	-	-	C	C	A
D. trop	C	C	g	t	g	g	t	-	-	-	-	C	C	A
D. cap	C	C	g	a	g	t	t	c	-	-	-	C	T	A
D. neb	C	C	g	a	g	t	t	c	-	-	-	C	T	A

B

Dipt B nt Position	017	139
Type	**N**	**N**
A28	P	S
C2-I	P	P
C2-II	S	P
C2-III	P	P
I1	P	P
MS-I	P	P
MS-II	P	P
O11	P	P
SM	P	P
D. wls	P	P
D. equi	P	P
D. trop	P	P
D. cap	P	P
D. neb	P	P

Table 4.6: Polymorphism table (**A**) and amino acid substitution table (**B**) for *Diptericin B* gene determined in the six *D. paulistorum* semispecies: Amazonian (A28); Centroamerican with three potential paralogues (C2-I; C2-II & C2-III); Interior (I1); Andean-Brazilian with two potential paralogues (MS-I & MS-II); Orinocan (O11); and Transitional (SM). As outgroups, *D. willistoni* (*D. wls*), *D. equinoxialis* (*D. equi*) *D. tropicalis* (*D. trop*), *D. capricorni* (*D. cap*) and *D. nebulosa* (*D. neb*) are shown. **A:** "nt Position" refers to nucleotide position in the alignment (of *D. paulistorum* semispecies without primers, without outgroups) in exonic (grey) and intronic (white) boxed regions. Type refers to noncoding (N), silent (S) or replacement (R) mutation; green filling represents changes from the majority consensus sequence and hyphen (–) refers to single nucleotide deletion. **B:** "nt Position" refers to nucleotide position in the alignment responsible for the amino acid substitution. Type refers to chemically non-equivalent (N) substitution.

4.1.6.6 *Drosocin*

Drosocin is with 171 bp the smallest of the examined AMP genes. The protein structure comprises a signal peptide, propeptide and a C-terminal domain, which is homologue to the N-terminal domain of the other AMPs (**Fig.4.13**). Within five strains of *D. paulistorum* semispecies (A28, C2, I1, MS, SM), the 121 bp long *Drosocin* PCR fragment (without primer) has five variable sites, all of them synonymous (2.9% divergence, **Table 4.7**). Orinocan strain (O11) showed no PCR product with *Drosocin* primers.

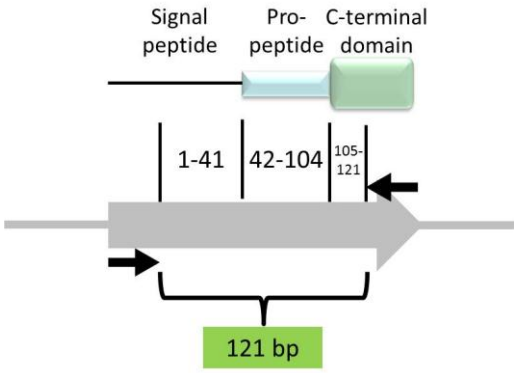


Figure 4.13: *Diptericin* protein domains according to *D. melanogaster* homologues and respective DNA regions on *D. paulistorum* genome, shown in grey. Black arrows represent PCR primers, the total PCR fragment size (without primers) is highlighted in green. “nt position” refers to the region in the alignment of PCR fragments coding for the corresponding protein domain.

Dro **nt Position**	**011**	**020**	**025**	**053**	**110**
Type	**S**	**S**	**S**	**S**	**S**
A28	C	T	T	A	G
C2	C	C	T	G	C
I1	T	T	T	A	C
MS	C	T	A	A	G
O11	-	-	-	-	-
SM	T	T	T	A	C
D. wls	C	T	T	A	C
D. equi	C	T	T	A	C
D. trop	C	T	T	A	C

Table 4.7: Polymorphism table for *Drosocin* gene determined in the six *D. paulistorum* semispecies: Amazonian (A28); Centroamerican (C2); Interior (I1); Andean-Brazilian (MS); Orinocan (O11); and Transitional (SM). As outgroups, *D. willistoni* (*D. wls*), *D. equinoxialis* (*D. equi*) and *D. tropicalis* (*D. trop*) are shown. “nt Position” refers to nucleotide position in the alignment (of *D. paulistorum* semispecies without primers, without outgroups). Type refers to silent (S) mutation; green filling represents changes from the majority consensus sequence. Hyphen (-) with yellow background highlights the missing PCR signal for Orinocan semispecies.

A summary of all variable sites detected in the respective domains of the six examined AMPs is given in **Figure 4.14**. *Attacin D* showed the highest level of total nucleotide divergence and the highest proportion of non-synonymous substitutions in the N-terminal and G1 domain. It represents therefore a prime candidate for a marker gene of the species complex of *D. paulistorum* and also for further studies of divergent adaptation to different microbial challenges.

4.1.7 Variable sites in AMP sequences of *D. paulistorum*

	n strains	Fragment length (bp)	Signal Peptide	Pro-peptide	N-termnial Domain	G1 Domain	G2 Domain	Intron	Total
Attacin A	6	696 - 697	1/0 (48 bp)	0/0 (54 bp)	6/2 (171 bp)	5/0 (192 bp)	5/* (173 bp)	6 (58 - 59 bp)	17/2 * (+ 6)
Attacin C	6	406 - 414	0/0 (13 bp)	2/0 (87 bp)	1/1 (188 bp)	1/1 (53 bp)	-	14 (65 - 73 bp)	4/2 (+ 14)
Attacin D	6	555 - 566	-	-	7/5 (142 bp)	6/4 (180 bp)	7/1 (178 bp)	24 (55 - 66 bp)	20/10 (+24)
Diptericin	6	281	2/1 (52 bp)	2/0 (34 bp)	-	-	12/1 (195 bp)	-	16/2
Diptericin B	6	257 - 261	1/1 (19 bp)	3/1 (74 bp)	-	-	1/0 (94 bp)	9 (70 - 74 bp)	5/2 (+ 9)
Drosocin	5	121	3/0 41 bp	1/0 (63 bp)	1/0 (17 bp)	-	-	-	5/0

Figure 4.14: Summary of variable sites within AMPs in *D. paulistorum* semispecies. “n strains” = number of strains among which the variable sites were determined. “Fragment length” = the length of the PCR product of the respective genes (without primers). The variable sites (first digit) in relation to the non-synonymous substitutions (second digit) are displayed for each functional section of the protein (signal peptide, propeptide, N-terminal domain, G1 domain, G2 domain). In brackets, the length of the section in the PCR product is shown. “Total” gives the sum of all variable sites in the coding regions of the PCR product (first digit) in relation to the sum of non-synonymous substitutions (second digit). In brackets, the variable sites of the intron are added. The cells highlighted in yellow indicate the most variable domains of *Attacin D.* Asterisk (*) in G2 domain of *Attacin A* represents stop codon in SM strain.

4.1.8 Southern blot Analysis of *Attacin D*

Multiple sequence types were observed for some *D. paulistorum* strains in *Attacin A* (C2-I, C2-II) and *Diptericin B* (C2-I, C2-II, C2-III, MS-I, MS-II). The highest number of potential paralogues, however, was found in *Attacin D* (A28-I, A28-II, C2-I, C2-II, I1-I, I1-II, SM-I, SM-II, see **Table 4.4 A**). In order to verify these observations as true paralogues, genomic Southern blots were performed with P^{32}-labelled *Attacin D* probes on DNA of *D. paulistorum* semispecies.

Cloned *Attacin D* fragments from O11 strain were used to prepare probes for labelling. As none of the *Attacin D* sequences harbours a restriction site for *Hin*dIII, the genomic *D. paulistorum* DNA used to prepare the blot was previously digested with this restriction enzyme. The paralogous sequences should therefore be coded on different restriction fragments, probably varying in size (restriction fragment length polymorphism, RFLP). As a reference, *Hin*dIII-digested DNA of λ-phages was loaded. As λ-DNA does not hybridize with the labelled probe, the bands on the membrane were marked manually with a pencil under UV light. Two independent membranes were generated and successfully utilized for Southern hybridizations. Results from these experiments are presented below.

On the first membrane (**Fig. 4.15 A**), *Hin*dIII-digested DNA of the six *D. paulistorum* semispecies was hybridized with *Attacin D* probe. A28, MS and O11 showed a strong signal at 5,400 bp, for C2, I1 and SM a strong band appeared at 2,000 bp. The signal for C2 is faint compared to the others, presumably due to a smaller amount of DNA loaded on the gel used for this blot. All strains show a second, faint signal at 12,000 bp.

The second membrane (**Fig. 4.15 B**) was established with male and female DNA of the strains A28, C2 and O11. Female DNA of *D. willistoni* was used as a control group, as its sequenced genome only shows one Attacin D gene. Consistent with the first membrane were the signals at 2,000 and 12,000 bp for both sexes of the C2 probe and the 5,400 bp long fragment for female and male A28, as well as the signal at 12,000 bp in female A28. For O11, no bands were visible, probably because the DNA concentration was not sufficient for hybridization. Inconsistent with the results of the first membrane, however, were signals at 3,500 bp for both female and male A28 and a signal at 6,200 bp for male A28, which could not be detected on the first blot. Also, the DNA of male A28 does not show the expected 12,000 bp fragment. For C2 strain, the same signal at 3,500 bp was detected in both sexes, although at a much lower intensity. These findings suggest a possible third paralogue in A28 and C2, as well as a sex-specific RFLP pattern for A28. *D. willistoni* has a clear band at 3,600 bp and a weak signal at 12,000 bp. The low intensity of the 12,000 bp signal could be due to a lower transfer efficiency of high-molecular-weight DNA fragments. However, the occurrence of the second band at 3,600 bp in *D. willistoni* was unexpected, as the GenBank data for the genomic region harbouring the *Attacin D* (GK12085) gene

predicts only a single fragment of 11,835 bp in length for *Hin*dIII digestion. Hence, the 3,600 bp fragment cannot be explained with the published *D. willistoni* sequence data, presumably due to the incompletely sequenced genome. Alternatively, the sequenced *D. willistoni* strain could have lost the second paralogue, or it developed *de novo* in the strain used for this thesis. Based on our data, we conclude that the 12,000 bp fragments, also present in all *D. paulistorum* semispecies, however, most likely comprise the ancestral *Attacin D* paralogue.

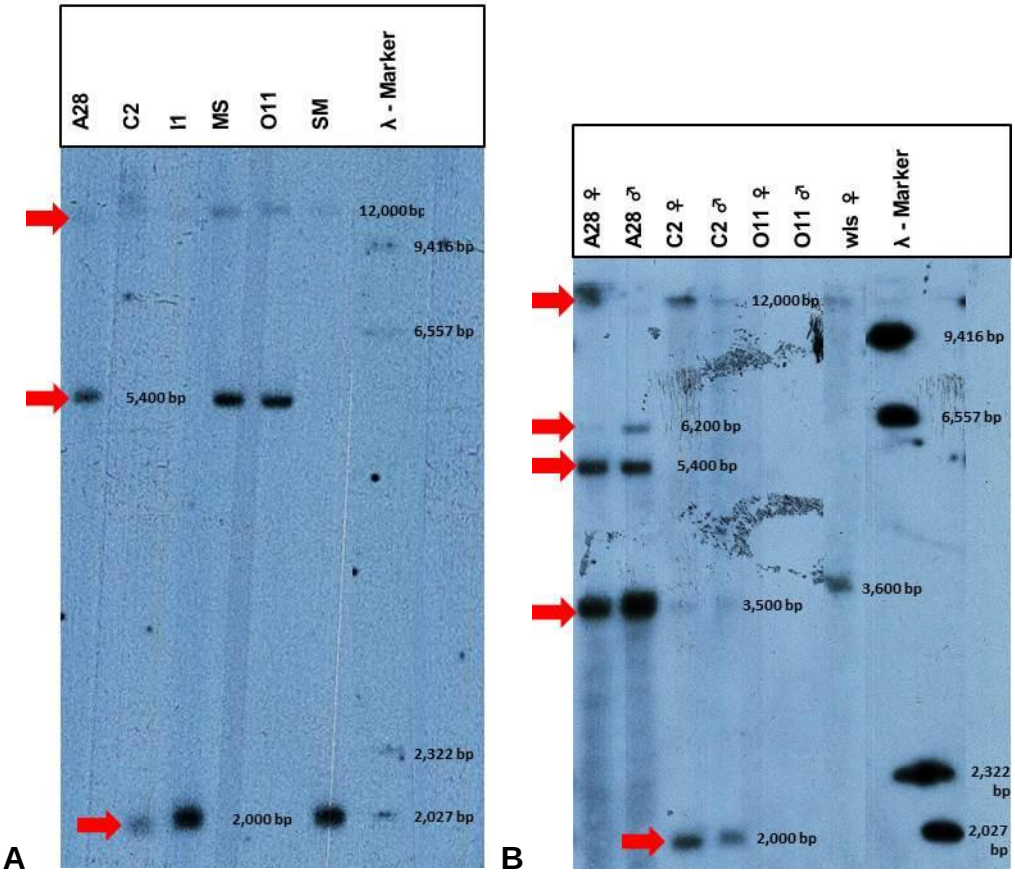


Fig. 4.15: Southern blot of *Hin*dIII-digested *D. paulistorum* DNA hybridized with P^{32}-labelled *Attacin D* probe. *Hin*dIII-digested λ-phage DNA was used as size marker, radioactive signals derive from pencil marks on the membrane. The images were captured on X-ray-film with three-day exposure. **A:** female DNA of the six *D. paulistorum* semispecies, multifly extraction (following Junakovic, 2004). **B:** male and female DNA of A28, C2 and O11 semispecies and female DNA *D. willistoni.* Red arrows indicate the observed bands.

4.2 Mitochondrial Sequences of *Drosophila paulistorum*

Additionally to nuclear AMP and immune related genes, mitochondrial sequences of *D. paulistorum* were examined. Due to the high mutation rate of mitochondrial DNA (Brown, 1979) it is a promising target to detect SNPs in closely related taxa, such as the six *D. paulistorum* semispecies. Primer targeting a 602 bp long fragment of the protein coding mitochondrial gene *NADH dehydrogenase subunit 4* (*ND4*) were designed by D. Schneider using the mitochondrial genome of *Drosophila willistoni* as reference. *ND4* was chosen for its high mutation rate and location in maximum distance to previously examined mitochondrial gene *cytochrome c oxidase 1* (*CO1*). **Figure 4.16** shows a schematic map of the mitochondrial genome of *D. willistoni.*

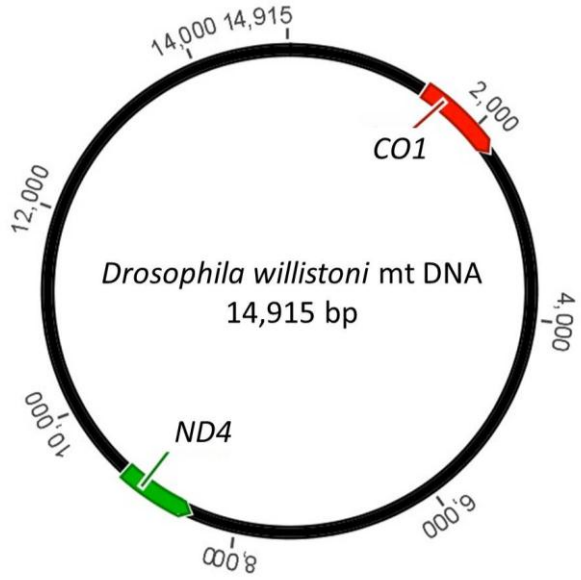


Figure 4.16: Scheme of *D. willistoni* mitochondrial genome, established with Geneious program software. *Cytochrome c oxidase 1* (*CO1*) gene is indicated in red, *NADH dehydrogenase subunit 4* (*ND4*) gene is indicated in green.

4.2.1 Cloning and sequencing of *NADH dehydrogenase subunit 4*

ND4 sequences of *D. paulistorum* semispecies and *D. equinoxialis* were analysed by cloning and sequencing. First attempts to sequence cloned *ND4* PCR-fragments failed or provided only incomplete reads, the number of obtained complete reads varies therefore. For I1 strain, none of the cloned sequences provided useful reads. Direct sequencing was performed subsequently to confirm the data obtained from cloned fragments. Simmilar to *Attacin D* fragments, direct sequencing was also performed with DNA templates provided by D. Schneider from the classic six *D.*

paulistorum semispecies in order to rule out strain contaminations. The number of complete reads used for *ND4* sequence analysis is shown in **Table 4.8**.

ND4	**cloning**	**Direct sequencing**				**Total**
		WD		DS		
		fwd	rev	fwd	rev	
A28	4/1	3	2	2	2	13/10
C2	2/2	2	2	2	2	10/10
I1	-	2	2	2	2	8/8
MS	3/1	2	2	2	2	11/9
O11	10/3	3	2	2	2	19/12
SM	1/1	2	2	2	2	9/9
D. equi.	-	1	1	-	-	2/2

Table 4.8: Number of complete *ND4* reads obtained from cloning (first digit indicates the number of reads, each from a different clone, the second digit indicates the number of PCRs performed for cloning) and direct sequencing. DNA extracted by W. Däuble (WD) and D. Schneider (DS) is indicated for direct sequencing. Fwd indicates direct sequencing reads obtained with *ND4* forward primer, rev stand for *ND4* reverse primer. Hyphen indicates that no sequencing was performed.

In addition to Sanger reads shown above, Illumina sequencing on embryonic DNA of A28, C2, I1, MS, and O11, performed in collaboration with Dr. Lisa Klasson from Uppsala University, Sweden, was used to examine mitochondrial genes.

4.2.2 *ND4* sequence analysis

Sanger reads of the 602 bp long fragment of *ND4* (without primers), established with adult DNA, show 31 SNPs within the *D. paulistorum* semispecies group (5.1% divergence, **Table 4.9**). All sequences obtained in this study derived from DNA of adult flies. For MS strain, a second mitotype was found in the direct sequencing reads, differing in seven nt positions from the cloned sequences (1.1% divergence). This mitotype was named “MS minor”, whereas the cloned mitotype was named “MS major”, which refers to the intensity of the peaks in the direct sequencing chromatograms. While Illumina reads, established with DNA of early embryos, show SNP patterns consistent with Sanger reads of A28, C2, I1 and MS major, the MS minor sequence obtained with Illumina differs only in two positions from MS major (0.3 % divergence) and in six positions from MS minor obtained with Sanger sequencing (1% divergence). Intriguingly, the Illumina “MS minor” version is identical with the *ND4* sequence obtained for C2 strain. Sanger reads obtained for the O11 strain are consistent with the “O11 major” Illumina reads. With the latter method,

however, a second mitotype ("O11 minor") was found at a lower frequency, differing in 17 nt positions from O11 major (2.8% divergence).

Three non-synonymous mutations were detected within *D. paulistorum* species complex in nt position 153 (leucine → valine), 215 (threonine → methionine) and 309 (isoleucine → valine). A DNA-based neighbour-joining tree, built with the Jukes-Cantor genetic distance model and *D. willistoni* as outgroup, shows the genetic distance of *ND4* within *D. paulistorum* and *D. equinoxialis* (**Figure 4.17**). Low bootstrap values received for MS minor and O11 (Illumina) minor sequences did not allow to resolve their topology with a threshold of 50%.

ND4 **nt Position**	**016**	**031**	**040**	**049**	**064**	**067**	**082**	**091**	**153**	**154**	**157**	**215**	**253**	**256**	**280**	**286**	**298**	**300**	**309**	**310**	**319**	**337**	**343**	**349**	**373**	**415**	**448**	**472**	**487**	**562**	**589**
Type	**S**	**S**	**S**	**S**	**S**	**S**	**S**	**S**	**R**	**S**	**S**	**R**	**S**	**S**	**S**	**S**	**S**	**S**	**R**	**S**	**S**	**S**	**S**	**S**	**S**	**S**	**S**	**S**	**S**	**S**	**S**
A28	C	A	T	C	A	T	T	C	A	C	C	G	A	T	T	C	G	G	C	C	C	T	T	T	C	G	G	C	T	C	C
A28 (Illumina)	C	A	T	C	A	T	T	C	A	C	C	G	A	T	T	C	G	G	C	C	C	T	T	T	C	G	G	C	T	C	C
C2	C	T	A	T	G	T	A	T	A	T	T	G	A	C	A	T	T	A	T	T	C	C	C	C	T	A	G	T	T	T	T
C2 (Illumina)	C	T	A	T	G	T	A	T	A	T	T	G	A	C	A	T	T	A	T	T	C	C	C	C	T	A	G	T	T	T	T
I1	C	A	T	C	A	T	T	C	A	C	C	G	G	T	T	C	A	G	T	T	C	T	T	T	C	A	G	C	C	C	C
I1 (Illumina)	C	A	T	C	A	T	T	C	A	C	C	G	G	T	T	C	A	G	T	T	C	T	T	T	C	A	G	C	C	C	C
MS major	T	T	A	T	G	T	A	T	A	T	T	A	A	C	A	T	T	A	T	T	C	C	C	C	T	A	A	T	T	T	T
MS major (Illumina)	T	T	A	T	G	T	A	T	A	T	T	A	A	C	A	T	T	A	T	T	C	C	C	C	T	A	A	T	T	T	T
MS minor	T	T	A	T	A	C	A	T	A	T	T	G	A	T	A	T	T	A	T	T	C	C	T	C	C	A	G	T	T	T	T
MS minor (Illumina)	C	T	A	T	G	T	A	T	A	T	T	G	A	C	A	T	T	A	T	T	C	C	C	C	T	A	G	T	T	T	T
O11	C	A	T	C	A	T	T	C	A	C	C	A	G	T	T	C	A	G	T	T	T	T	T	T	C	A	G	C	C	C	C
O11 major (Illumina)	C	A	T	C	A	T	T	C	A	C	C	A	G	T	T	C	A	G	T	T	T	T	T	T	C	A	G	C	C	C	C
O11 minor (Illumina)	T	T	A	T	G	T	A	T	C	T	T	A	A	C	T	C	A	G	T	T	T	C	C	C	T	A	A	T	T	C	C
SM	C	A	T	C	A	T	T	C	A	C	C	G	G	T	T	C	A	G	T	T	C	T	T	T	C	A	G	C	C	C	C
D. equi	C	T	A	T	G	C	A	T	A	C	T	G	A	T	A	T	T	A	T	T	T	T	T	C	T	A	A	T	C	C	T
D. wls	C	T	T	T	G	T	A	T	A	T	T	G	A	T	C	T	T	A	T	A	C	T	T	C	C	A	C	T	T	T	T

Table 4.9: Polymorphism table for *ND4* gene determined in the six *D. paulistorum* semispecies: Amazonian (A28); Centroamerican (C2); Interior (I1); Andean-Brazilian (MS); Orinocan (O11); and Transitional (SM). As outgroups, *D. willistoni* (*D. wls*) and *D. equinoxialis* (*D. equi*) are shown. Illumina sequences are indicated in brackets, minor and major refers to peak height in the chromatogram (Sanger sequencing) or abundance (Illumina). "nt Position" refers to nucleotide position in the alignment (of *D. paulistorum* semispecies without primers, without outgroups). Type refers to silent (S) or replacement (R) mutation; green filling represents changes from the majority consensus.

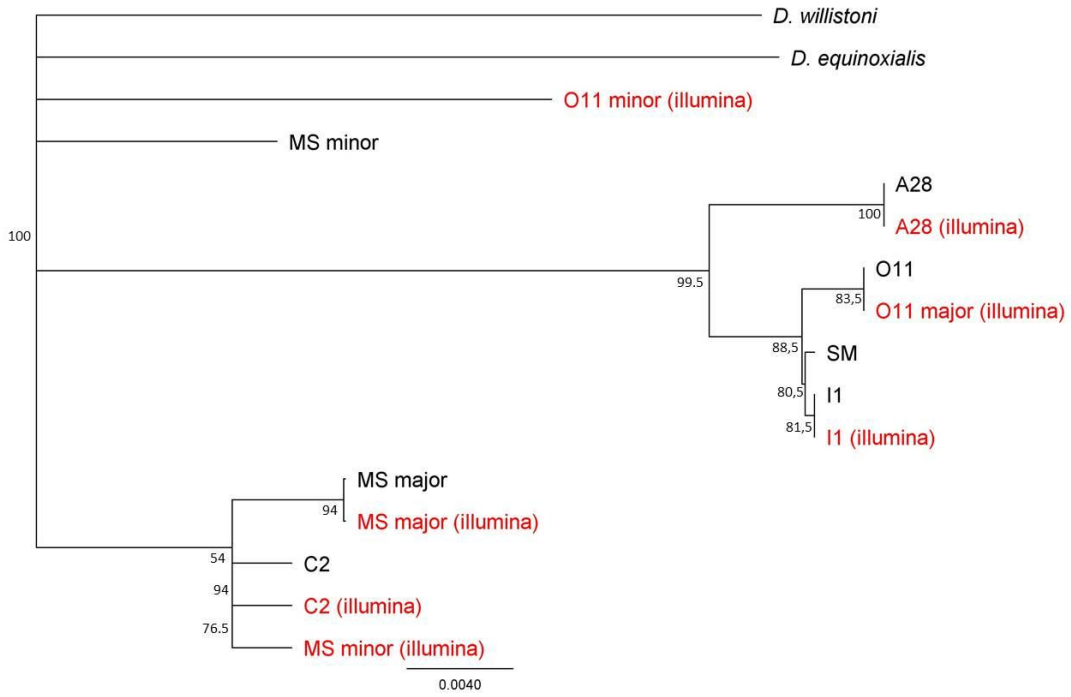


Figure 4.17: Neighbour – joining tree of *ND4* Sanger and Illumina (typed in red) sequences of *D. paulistorum* strains, *D. equinoxialis* and *D. willistoni.* Illumina sequences are typed red, Sanger sequences in black. Geneious software was used to build the tree with Jukes-Cantor genetic distance model. The tree was resampled with Bootstrap method and 200 replicates (bootstrap values indicated at the nodes of the tree), the threshold was at 50%. *D. willistoni* was used as outgroup

4.2.3 Southern blot analysis of *ND4*

4.2.3.1 Hybridisation of *Psi*I digested DNA with radioactively labelled *ND4*

To verify the existence of multiple mitotypes observed in the sequence data set and to establish strain-specific markers, we performed restriction fragment length polymorphism (RFLP) analysis via Southern Hybridization with radioactively P^{32}-labelled *ND4* fragments. Illumina sequencing of embryonic mitochondrial DNA of *D. paulistorum* strains A28, C2, I1, MS and O11 allowed *in silico* predictions of expected restriction fragment lengths. Using this data, we were able to differentiate the *Psi*I RFLP pattern of major and minor O11 mitotypes, as well as of A28 and I1. However, MS and C2 could not be distinguished with this method. For SM strain, no Illumina data was available. Restriction sites for *Psi*I endonuclease in the up- and downstream flanking region of the *ND4* gene and the resulting restriction fragments for the examined strains are shown in **Fig. 4.18**. Two sites (highlighted in red) at positions 525 and 1618 were selectively diagnostic for A28, I1, O11 major and O11 minor mitotype.

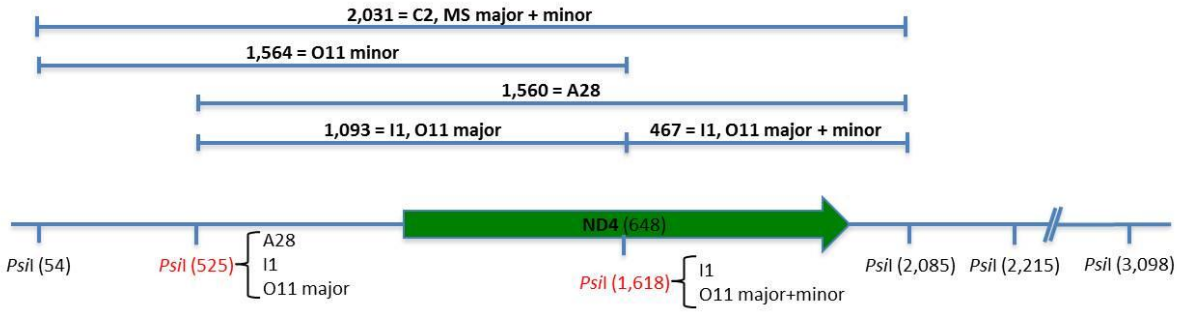


Fig. 4.18: Schematic illustration of the 648 bp long *D. paulistorum ND4* PCR fragment on the mitochondrial genome. *Psi*I restriction sites are indicated with their respective position, starting 1300 bp upstream of the *ND4* fragment. Diagnostic sites are marked in red, the strains which show those sites (according to Illumina sequence data) are indicated next to it. The resulting RFLPs are shown above, their respective size (in nucleotides) and the corresponding strains are indicated.

The RFLP pattern for *Psi*I digestion predicted with Illumina sequence data allowed to perform an *in silico* electrophoresis using Geneious software (**Fig. 4.19 A**). Those predicted RFLPs were tested with two independent Southern blots. DNA extracted from 10 females of each *D. paulistorum* semispecies (multifly extraction after Junakovic, 2004) was digested with *Psi*I and electrophoresed for each blot. After the transferring of the DNA sample sets onto nylon membranes, they were hybridized with P^{32}-labelled *ND4* fragments (**Fig. 4.19 B, C**). In order to compare quality and yield of different extraction methods, DNA from 10 females and 15 males of O11 strain was extracted using the Puregene extraction kit and processed in the same way as the other DNA sets (**Fig. 4.19 D**).

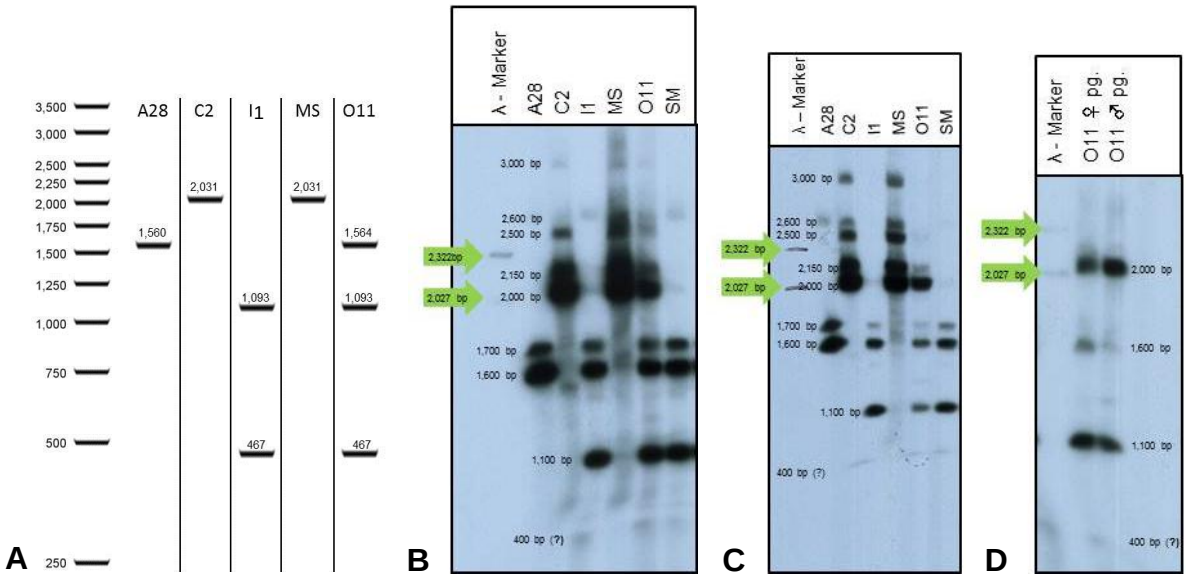


Fig. 4.19 A: Virtual electrophoresis of *Psi*I-digested fragments of mtDNA coding for *ND4* based on Illumina sequence data. **B, C:** Southern blot of *Psi*I-digested DNA of *D. paulistorum* semispecies (female, multifly extraction after Junakovic, 2004), hybridized with P^{32}-labelled *ND4* probe. **D:** Southern blot of *Psi*I-digested DNA of O11 semispecies extracted with Puregene kit (=pg., female ♀ and male ♂ DNA). **B-D:** Green arrows indicate 2,027 and 2,322 bp fragments of HindIII-digested λ-phage DNA (size marker), radioactive signals derive from pencil marks on the membrane. X-ray-film, **B, D:** 19 h of exposure, **C:** 13 h of exposure

Contradictory to the RFLPs calculated with Illumina sequences from embryonic DNA, Southern blots of adult DNA showed more bands than expected. Two fragments of A28 strain hybridized with the labelled probe at 1,600 and 1,700 bp, respectively. C2 and MS showed multiple signals at 2,000, 2,150, 2,500, 2,600 and 3,000 bp. For I1 and SM strains, three bands were detected at 1,100, 1,600 and 1,700 bp. O11 strain showed for the multifly extraction method after Junakovich (2004) even five bands at 1,100, 1,600, 1,700 2,000 and 2,150 bp. With the Puregene extraction, this number could be reduced to two strong bands at 1,100 and 2,000 bp and a faint band at 1,600 bp. A fourth, very weak band is visible at the bottom of the membrane (marked with 400 bp); this band also appears for I1 and SM strains. As the agarose gels for Southern blotting all showed an unusual negative smile (**Fig. 4.20**), the fragments in this region might not have been separated properly. Therefore, it cannot be ruled out that this band is the predicted 467 bp fragment. **Table 4.10** summarizes the expected and observed bands.

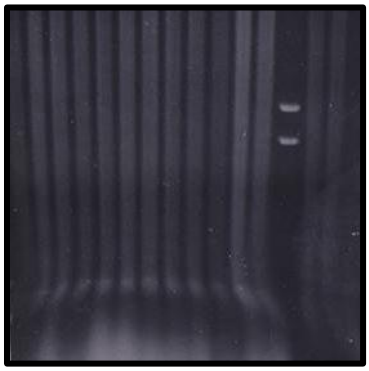


Fig. 4.20: Gel electrophoresis of *Psi*I-digested *D. paulistorum* DNA. **Fig. 4.19** (**B, D**) shows the corresponding *ND4* - hybridized Southern blot. The unusual negative smile shape of the DNA at the running front might contain and mask the predicted 467 bp long fragment.

ND4	Expected Bands Illumina (bp)	Observed Bands (bp)
A28	1,560	**1,600**, 1,700
C2	2,031	**2,000**, 2,150; 2,500, 2,600, 3,000
I1	467, 1,093	**400 (?)**, **1,100**, 1,600, 1,700
MS	major + minor: 2,031	**2,000**, 2,150, 2,500, 2,600, 3,000
O11	major: 467, 1,093 minor: 467, 1,564	**400 (?)**, **1,100**, **1,600**, 1,700, 2,000, 2,150
SM	-	400 (?), 1,100, 1,600, 1,700

Table 4.10: Summary of the expected and observed restriction fragments of *ND4* region of *D. paulistorum* semispecies, digested with *Psi*I. Expected bands were determined with Illumina sequence data. Observed bands correlating with the expected ones are typed in bold. “major” and “minor” refers to multiple mitotypes found by Illumina sequencing. No Illumina data was available for SM strain.

A possible explanation for the unexpected observed bands is an incomplete digestion of *Psi*I or additional mitotypes which lost the restriction sites and could not be detected with Illumina sequencing of embryonic DNA, as most of the observed fragments can be correlated with the loss of one or more restriction sites (see **Fig. 4.18**): For A28, the absence or incomplete digestion of the restriction site at position 2,085 leads to a fragment of 1,690 bp. The strains C2 and MS produce a fragment of 2,161 bp without the site at position 2,085. If, additionally, the site at position 2,215 is lost or not digested, it would lead to a fragment of 3,044 bp. The observed bands of 2,500 and 2,600 bp in C2 and MS, however cannot be explained with incomplete digestion of the sequenced mitotypes, but under the assumption of an additional mitotype with the restriction site at position 525. If the *Psi*I restriction site in position 1,618 of I1 strain is lost or not digested, a fragment of 1,560 bp would be the result. If additionally the site at position 2,215 is not cut, this would lead, similar to A28 strain, to a fragment of 1690 bp. For O11 major mitotype, a fragment of 1,560 bp would emerge without the restriction site at position 1,618, if the site at position 525 is ignored too, this would lead to a 2,031 bp long fragment. By leaving out a third restriction site at position 2,085, a fragment of 2,161 bp would be the result. The loss of the latter two restriction sites lead to the same fragments in O11 minor mitotype.

Intriguingly, O11 DNA extracted with the Puregene kit (**Fig.4.19 D**) did not show the band at ~1,700 and ~2,150 bp, suggesting that the restriction digestion was more efficient in those samples. Nevertheless, more fragments than the two expected major and minor mitotypes were detected. In order to test if the assumed incomplete digestion can be reduced with the Puregene DNA extraction kit, the experiment was

repeated with three *D. paulistorum* strains (A28, C2, O11) and *D. willistoni* as control (**Fig. 4.21**).

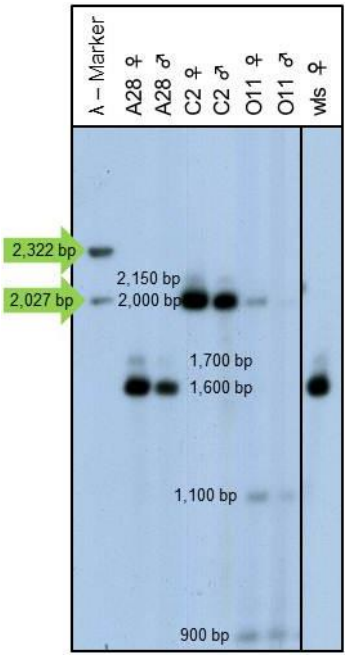


Fig. 4.21: Southern blot of *Psi*I-digested female and male DNA of *D. paulistorum strains* Amazonian (A28), Centroamerican (C2), Orinocan (O11) and *D. willistoni* (wls). Multifly DNA extraction with Puregene kit, hybridized with P^{32}-labelled *ND4* probe. *Hin*dIII-digested λ-phage DNA was used as size marker, radioactive signals derive from pencil marks on the membrane. X-ray-film, 40 hrs exposed.

In fact, the number of bands on the blot in **Fig. 4.21** could be reduced: For A28 down to a single intense band of 1,500 bp and a faint one of 1,600 bp; for C2 down to one strong signal at the height of 2,000 bp and a weak signal at 2,150 bp. O11 overall showed faint signals, but the number was reduced from eight down to three: one at 2,000 bp, one at 1,100 bp at a third one near the running front. As shown in **Fig. 4.20**, we also observed unusual band shapes at the bottom of the gel, which may have resulted from inefficient separation. Therefore, the third signal may represent the 467 bp fragment, although the calculated length is 900 bp.

While for A28 and C2, the digestion seems to be complete with this extraction method, O11 still shows the band at 2,000 bp which could be the result of three uncut restriction sites for O11 major and two for O11 minor mitotype, respectively. As described above, either due to incomplete digestion or an additional mitotype which is missing the restriction sites.

The 1,560 bp long fragment specific for O11 minor mitotype is missing in the Southern blot analysis shown in **Fig. 4.21**. Hence, the minor mitotype detected with Illumina sequence data of embryonic O11 DNA could not be confirmed. The band at ~1,600 bp shown in **Fig. 19 B-D**, however, could be assigned to O11 minor mitotype, although it could also be the result of incomplete digestion of O11 major mitotype. A28 can be distinguished from the other strains with *Psi*I RFLP pattern of *ND4*, while I1 and SM show the same fragments. MS mitotypes cannot be distinguished with this approach, since both show the same fragment size as C2 strain.

4.2.3.1 Hybridisation of *Sau*3AI digested DNA with DIG labelled *ND4*

RFLP analysis was also performed with Southern blots of *Sau*3AI digested DNA of *D. paulistorum* semispecies, hybridised with digoxigenin (DIG) labelled *ND4* fragments. For this experiment, Illumina sequence data of embryonic DNA predicted signals at 993 bp for A28, I1 and O11 major mitotype, while C2, MS major and minor mitotypes and O11 minor mitotype showed an additional restriction site (position 1,692), generating two fragments of 483 and 510 bp (**Fig. 4.22**). As described earlier, no Illumina data was available for SM strain.

Sanger sequence data of adult DNA, however, did not show the additional restriction site in the MS minor mitotype, predicting three signals of 483, 510 and 993 bp for the two mitotypes of this strain. Furthermore, only one mitotype for O11 strain, identical with O11 major mitotype found with Illumina method, was detected with Sanger sequencing, predicting a single fragment of 993 bp.

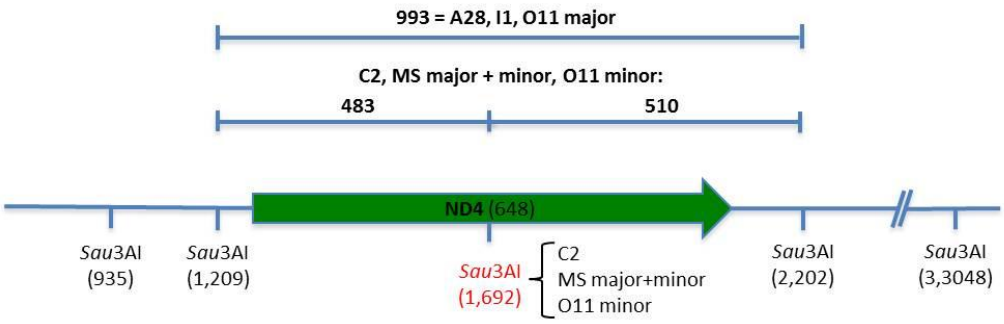


Fig. 4.22 Schematic illustration of the 648 bp long *D. paulistorum ND4* PCR fragment on the mitochondrial genome. *Sau*3AI restriction sites are indicated with their respective position, starting 1300 bp upstream of the *ND4* fragment. Diagnostic sites are marked in red, the strains which show those sites (according to Illumina sequence data) are indicated next to it. The resulting RFLPs are shown above, their respective size (in nucleotides) and the corresponding strains are indicated.

In silico electrophoresis established with Geneious software for *Sau*3A1 digested *ND4* fragments based on Illumina sequence data is shown in **Fig. 23 A**. Southern blot hybridized with DIG labelled ND4 fragments is shown in **Fig. 23 B**. The observed fragments for MS and O11 correspond with the restriction sites found with Sanger sequencing and are conflicting with Illumina sequence data.

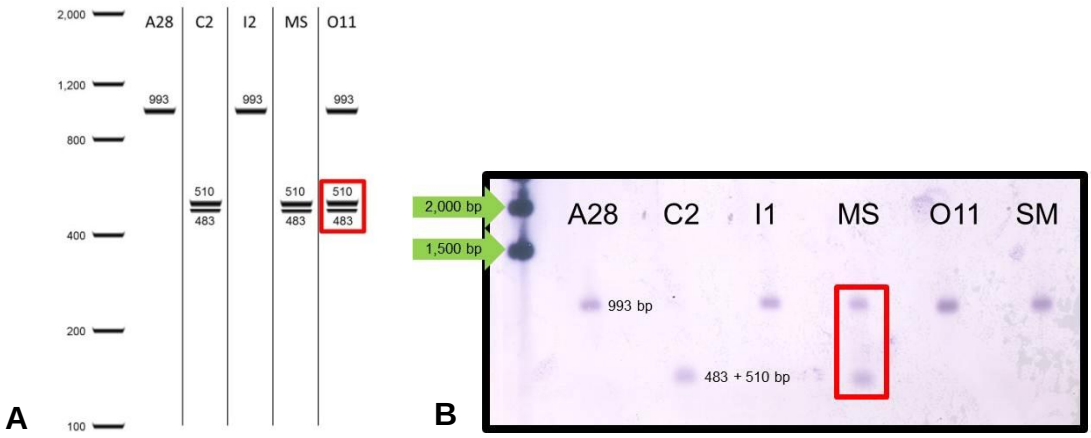


Fig. 4.23 A: Virtual electrophoresis of *Sau*3AI-digested fragments of mtDNA coding for *ND4* based on Illumina sequence data. **B:** Southern blot of *Sau*3AI-digested DNA of *D. paulistorum* semispecies (female, multifly extraction, Puregene DNA extraction kit), hybridized with DIG-labelled *ND4* probe. Green arrows indicate 1,500 and 2,000 bp fragments of Generuler 1kb DNA ladder. Red boxes indicate conflicting results.

4.2.4 Southern blot analysis of *CO1*

Additional to the *ND4* analysis, Southern blots were also performed with P^{32}-labelled fragments of the gene encoding for cytochrome c oxidase subunit 1 (*CO1*). Prediction of *Psi*I RFLPs based on Illumina sequencing for the strains A28, C2, I1, MS and O11 are shown in **Fig. 4.24**. For SM strain, no Illumina data was available. Size Polymorphisms downstream of *CO1* are responsible for the varying fragment sizes of A28, I1, O11 minor and major. One diagnostic site at position 3,394 bp for MS and C2 was detected.

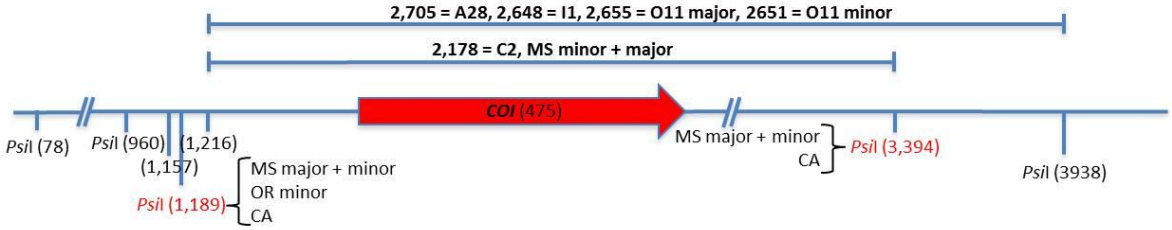


Fig. 4.24: Schematic illustration of the 475 bp long *D. paulistorum CO1* fragment on the mitochondrial genome. *Psi*I restriction sites are indicated with their respective position, starting 1,696 bp upstream of the *CO1* fragment. Diagnostic sites are marked in red, the strains which show those sites (according to Illumina sequence data) are indicated next to it. The resulting RFLPs are shown above; their respective size (in nucleotides) and the corresponding strains are indicated.

In silico electrophoresis of *CO1* fragments derived from embryos established with Geneious software is shown in **Fig. 4.25 A**. Membranes used for *ND4* hybridization (**Fig. 4.19 B, D**) were stripped and rehybridized with P^{32}-labelled *CO1* fragments (**Fig. 4.25 B, C**). A summary of the observed and expected bands is given in **Table 4.11.**

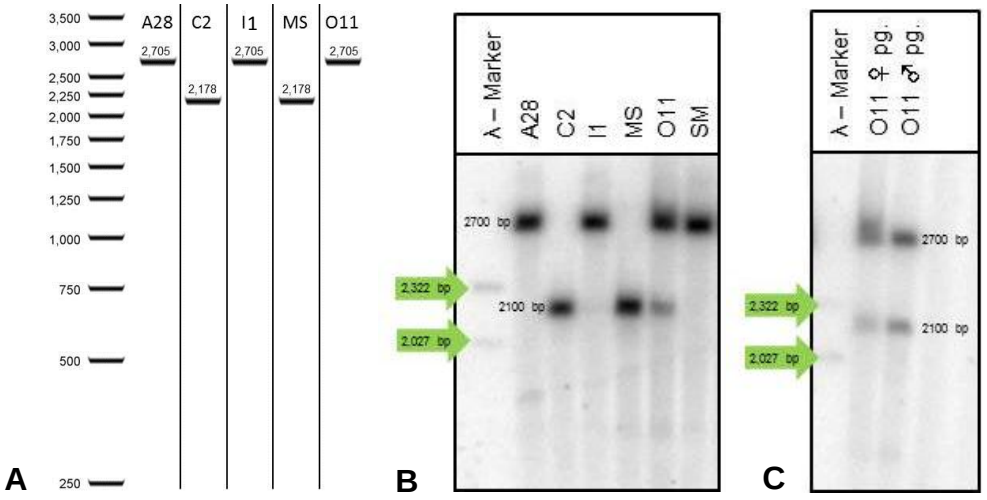


Fig. 4.25 A: Virtual electrophoresis of *Psi*I-digested fragments of mtDNA coding for *CO1* based on Illumina sequence data. **B:** Southern blot of *Psi*I-digested DNA of *D. paulistorum* semispecies (female, multifly extraction after Junakovic, 2004), hybridized with P^{32}-labelled *CO1* probe. **C:** Southern blot of *Psi*I-digested DNA of O11 semispecies extracted with Puregene kit (pg: female ♀ and male ♂ DNA). **B, C:** Green arrows indicate 2,027 and 2,322 bp fragments of *Hin*dIII-digested λ-phage DNA (size marker), radioactive signals derive from pencil marks on the membrane. Phospoimager, 35 min of exposure.

CO1	**Expected Bands Illumina (bp)**	**Observed Bands (bp)**
A28	2,722	**2,700**
C2	2,178	**2,100**
I1	2,722	**2,700**
MS	major + minor: 2,178	**2,100**
O11	major + minor: 2,722	2,100, **2,700**
SM	-	2,700

Table 4.11: Summary of the observed and expected restriction fragments of *CO1* region, digested with *Psi*I. Expected bands were determined on basis of Illumina sequence data. Observed bands correlating with the expected ones are typed in bold. “major” and “minor” refers to multiple mitotypes found with Illumina sequencing. The 2100 bp fragment of O11 is the only observed fragment differing from the prediction.

The obtained signals for *CO1* RFLP analysis correlate with the predicted fragment length for all strains except for O11. Here, additionally to the expected band at 2,700 bp, a signal at 2,100 bp was detected, in a similar height as band for C2 and MS. This signal cannot be explained with the Illumina sequence data obtained for O11 strain. *CO1* RFLP pattern is neither suitable for distinguishing minor from major MS and minor from major O11 mitotypes, nor as marker system to distinguish the *D. paulistorum* semispecies.

4.3 Diagnostic Single Nucleotide Polymorphisms in *Drosophila paulistorum* semispecies

Using the obtained sequence data shown above (see section **4.1.6, 4.2.2**), nuclear AMP genes plus the mitochondrial *ND4* gene were examined for diagnostic Single Nucleotide Polymorphisms (SNPs) for each *D. paulistorum* semispecies (**Table 4.12**). The positions of the diagnostic SNPs refer to the PCR fragment without primers and differ in some cases from the position in the alignment used for the polymorphism tables due to deletions/insertions in the Intron.

	AttA	*AttC*	*AttD*	*Dipt*	*DiptB*	*Dro*	*ND4*
A28	**234** T → A **267** C → T	-	**030** A → G **I: 181** T → C **I: 269** G → T **II: 354** A → C	**23** C → T **102** C → T **120** C → G	**115** T → G **136** C → T **162** C → G	-	**309** T → C **310** T → C **415** A → g
C2	**292** T → C **308** T → A **I: 449** C → T **II: 48** T → C **II: 324** T → A	-	**I: 134** G → A **I: 177** T → A **I: 180** T → C **I: 181** A → C **I: 183** G → T **I: 184** C → T **I: 187** T → A **II: 155** C → T **II: 301** A → T	**128** T → A	**II: 17** C → T **II: 69** A → T **II: 75** G → A **II: 116** T → C **II: 119** A → G **III: 52** C → T **III: 69** A → C	**20** T → C **53** A → G	-
I1	**317** T → C **574** T → G **628** C → G	-	**I: 22** C → T **I: 65** G → A **I: 257** G → A **I: 397** T → C **II: 59** G → T **II: 65** G → C **II: 76** A → C **II: 496** G → C **II: 530** T → G	**96** G → A	**115** T→ C	-	-
MS	**479** C → T **536** C → T	-	**100** A → T **354** A → G	-	**II: 74** C → T	**25** T → A	m. s.: **67** T → C
O11	**280** T → C	**243** A → T	**299** C → G	-	-	-	**319** C → T m. i.: **153** A → C
SM	**446** G → A **455** C → G **662** G → A	-	SM-I = I1-I SM-II = I1-II	-	-	-	-

Table 4.12: Summary of all diagnostic SNPs found in examined nuclear (*AttA, AttC, AttD, Dipt, DiptB, Dro*) and mitochondrial (*ND4*) genes of *D. paulistorum* semispecies. The indicated SNPs are only present in the associated semispecies. Position of the SNP in the PCR fragment without primers is indicated first, followed by the nucleotide change. Roman numerals represent SNPs only present in one potential paralogue, “m.s.” and “m.i.” indicate SNPs only present in the minor mitotype found with Sanger (m.s.) or Illumina (m.i.) sequencing. *AttA, AttC, AttD = Attacin A, -C, -D; Dipt, DiptB = Diptericin, -B; Dro = Drosocin*.

4.4 Applying diagnostic SNPs for *D. paulistorum* strain typing

Newly collected *D. paulistorum* isofemale lines (see **table 3.2** for complete strain list) deriving from individuals collected in South Brazil (JB, MLC, MTS, PoA1, PoA10, RP), mid- and Mideast Brazil (Ama, Bra, NEG-DED, TRI-POU, SAL-13, SET-RIA, ZOO1) and mutant strains from Honduras (prune, white, yellow) were examined to determine their relationship to the described semispecies.

4.4.1 Sequence Analysis of newly collected Brazilian lines

In order to perform strain typing with the newly collected *D. paulistorum* lines, nuclear genes *Attacin C* and *Attacin D* and the mitochondrial gene *ND4* were amplified by PCR. *Attacin C* PCR was only performed with the strains from south Brazil. Cloning of the fragments was also performed with this reduced set of strains, the whole set was used for direct sequencing of *Attacin D* and *ND4*. For SAL-13, no complete reads were obtained (**Table 4.13**). The complete sequences were examined for SNPs as described earlier for the six *D. paulistorum* semispecies.

	AttC	***AttD***			***ND4***		
	cloning	cloning	direct seq.		cloning	direct seq.	
			fwd	rev		fwd	rev
Ama	-	-	2	2	-	2	2
Bra	-	-	2	2	-	2	2
JB	3 / 1	3 / 1	2	2	3 / 2	2	2
MLC	3 / 1	3 / 1	2	2	-	2	2
MTS	3 / 1	3 / 1	2	2	2 / 1	2	2
NEG-DED	-	-	2	2	-	-	-
PoA1	3 / 1	3 / 1	2	2	-	2	2
PoA10	3 / 1	3 / 1	2	2	5 / 2	2	2
RP	3 / 1	3 / 1	4	4	1 / 1	4	3
SET-RIA	-	-	2	2	-	2	2
SAL-13	-	-	-	-	-	-	-
TRI-POU	-	-	2	2	-	2	2
ZOO1	-	-	2	2	-	2	2
prune	3 / 1	3 / 1	2	2	1 / 1	-	-
white	2 / 1	3 / 1	2	2	3 / 1	-	-
yellow	3 / 1	3 / 1	2	2	1 / 1	-	-

Table 4.13: Number of complete reads obtained from sequencing of *Attacin C – D* and *ND4* PCR products of 11 newly collected Brazilian (JB. MLC, MTS, RP, PoA1, PoA10, Ama, Bra, SET-RIA, TRI-POU, ZOO1) and three mutant strains (prune, white, yellow). For cloned sequences, the first digit represents the number of complete reads, the second digit indicates the number of PCRs performed to obtain the sequenced clones.

4.4.1.1 *Attacin C*

With DNA of six strains deriving from south Brazil (JB, MLC, MTS, PoA1, PoA10, RP) and three mutant strains (prune, white, yellow), *Attacin C* fragments from a single PCR for each strain were cloned and sequenced (**Table 4.13**). Within those strains, only one new SNP in PoA10 was detected not present in any of the six semispecies (**Table 4.14 A**). This mutation does not change the amino acid sequence (**Table 4.14 B**).

A

AttC **nt Position**	**031**	**082**	**243**	**286**	**300**	**301**	**318**	**320**	**322**	**323**	**324**	**325**	**326**	**327**	**328**	**329**	**332**	**350**	**380**
Type	**S**	**S**	**R**	**S**	**N**	**N**	**N**	**N**	**N**	**N**	**N**	**N**	**N**	**N**	**N**	**N**	**N**	**N**	**R**
A28	C	G	A	T	t	g	a	a	g	g	t	c	a	t	a	a	t	c	T
I1	C	G	A	T	t	g	a	a	g	g	t	t	a	t	a	a	t	c	T
SM	C	G	A	T	t	g	a	a	g	g	t	t	a	t	a	a	t	c	T
MLC	C	G	A	T	t	g	a	a	g	g	t	c	a	t	a	a	t	c	T
white	C	G	A	T	t	g	a	a	g	g	t	t	a	t	a	a	t	c	T
C2	T	A	A	T	t	g	a	a	g	g	t	t	a	t	a	a	t	c	T
MS	T	A	A	T	a	a	g	t	-	-	-	-	-	-	-	-	c	t	A
JB	T	A	A	T	a	a	g	t	-	-	-	-	-	-	-	-	c	t	A
PoA1	T	A	A	T	a	a	g	t	-	-	-	-	-	-	-	-	c	t	A
prune	T	A	A	T	a	a	g	t	-	-	-	-	-	-	-	-	c	t	A
yellow	T	A	A	T	a	a	g	t	-	-	-	-	-	-	-	-	c	t	A
PoA10	T	A	A	C	a	a	g	t	-	-	-	-	-	-	-	-	c	t	A
O11	T	A	T	T	a	a	g	t	-	-	-	-	-	-	-	-	c	t	A
MTS	T	A	T	T	a	a	g	t	-	-	-	-	-	-	-	-	c	t	A
RP	T	A	T	T	a	a	g	t	-	-	-	-	-	-	-	-	c	t	A
D. wls	C	G	A	T	a	a	g	t	-	-	-	-	-	-	-	-	c	t	C

B

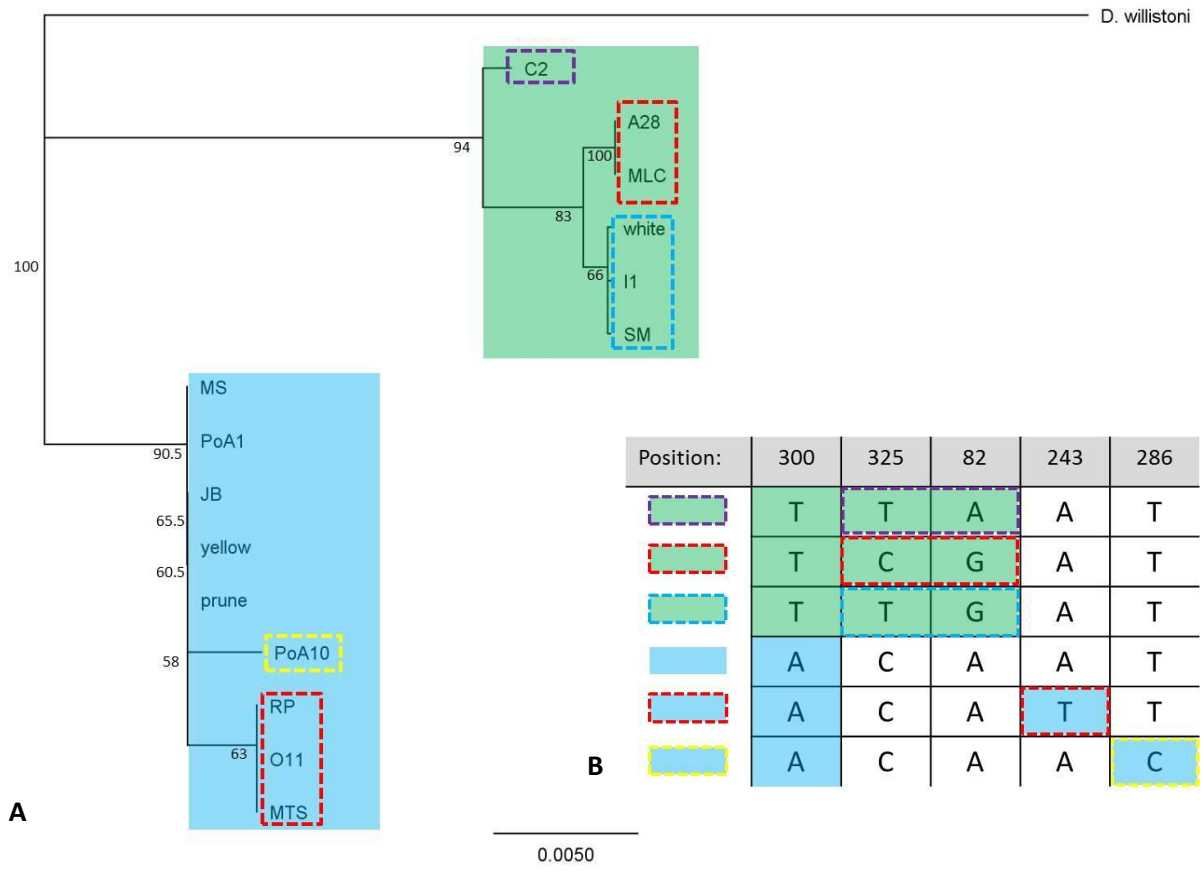
AttC **nt Position**	**243**	**380**
Type	**N**	**E**
A28	Q	D
I1	Q	D
SM	Q	D
MLC	Q	D
white	Q	D
C2	Q	D
MS	Q	E
JB	Q	E
PoA1	Q	E
prune	Q	E
yellow	Q	E
PoA10	Q	E
O11	L	E
MTS	L	E
RP	L	E
D. wls	Q	D

Table 4.14: Polymorphism table (**A**) and amino acid substitution table (**B**) for *Attacin C* gene determined in newly collected *D. paulistorum* strains and mutant strains, grouped with *Attacin C* sequences of described semispecies. As outgroup, *D. willistoni* (*D. wls*) is shown. **A:** "nt Position" refers to nucleotide position in the alignment (of *D. paulistorum* semispecies without primers) in exonic (grey) and intronic (white) boxed regions. Type refers to non-coding (N), silent (S) or replacement (R) mutation; green filling represents changes from the majority consensus sequence and hyphen (–) refers to single nucleotide deletion. **B:** "nt Position" refers to nucleotide position in the alignment responsible for the amino acid substitution. Type refers to chemically equivalent (E) and non-equivalent (N) substitution.

Single diagnostic SNPs for the classic semispecies could not be found in *Attacin C*. A combination of multiple sites, however, enables the identification of the Amazonian (A28), Centroamerican (C2), Andean-Brazilian (MS) and Orinocan (O11) strain. In contrast, Interior (I1) and Transitional (SM) could not be distinguished with this marker. (**Fig. 4.26**)

Of the examined Brazilian strains, PoA1 and JB are identical to MS in *Attacin C*, this extends to the mutant strains prune and yellow. MTS and RP show sequences similar to O11, whereas MLC groups with A28. The white mutant is in *Attacin C* identical to I1 and SM. PoA10 has a unique SNP in *Attacin C*, although it shows a high degree of homology with MS and O11 (99.6 and 99.8%, respectively; **Fig. 26 A**).

A single site at position 300 can be used to distinguish between the MS/O11 and the A28/C2/I1/SM cluster. Inclusion of two more SNPs at 325 and 82 allows identifying A28, I1/SM and C2 – like sequences. The Orinocan semispecies (O11) has a diagnostic site at position 243 and the unique SNP of PoA10 is found at position 286 (**Fig. 26 B**). The indicated positions refer to the nucleotide positions in the PCR fragments without primers and may differ in some cases from the position in the alignment (used for polymorphism tables) due to the varying length of the intron.



Position:	300	325	82	243	286
	T	T	A	A	T
	T	C	G	A	T
	T	T	G	A	T
	A	C	A	A	T
	A	C	A	T	T
	A	C	A	A	C

Fig. 4.26: ***Attacin C*** - marker system for *D. paulistorum strains.* **A:** Neighbour-Joining tree with *D. willistoni* as outgroup, calculated with Jukes-Cantor genetic distance model in Geneious programme The tree was resampled with Bootstrap method and 200 replicates (bootstrap values indicated at the nodes of the tree), the threshold was at 50%. A28 group is highlighted in green, MS group in blue. Coloured dashes show subgroups. **B:** Table of SNPs defining the groups highlighted in the respective colour/dashes. Position refers to the nucleotide position in the sequence without primer.

4.4.1.2 *Attacin D*

Attacin D sequences of all newly collected and mutant strains were examined. Nine out of 16 strains showed two different sequence types (Ama, Bra, MLC, MTS, NEG-DED, PoA1, RP, SAL-13, and SET-RIA) with a divergence of 0.7 – 3.4% to each other, respectively (2.08% average divergence). The sequences of SAL-13 and the second sequence type of RP (RP-II) were incomplete and hence excluded from further analysis. NEG-DED was treated as outgroup, as its sequence showed high homology to *D. willistoni.* In the aligned coding region of all complete *Attacin D* sequences of newly collected *D. paulistorum* strains, 18 out of 38 variable sites were not found in the described semispecies, eight of those represent a change in the amino acid sequence of the protein. In the intron, only one out of 24 variable sites could not be found in one of the classic six lines (nt 198 of PoA1-II, **Table 4.15 A**). Six of the non-synonymous mutations in the new lines lead to a chemically non-equivalent amino acid substitution (**Table 4.15 B**).

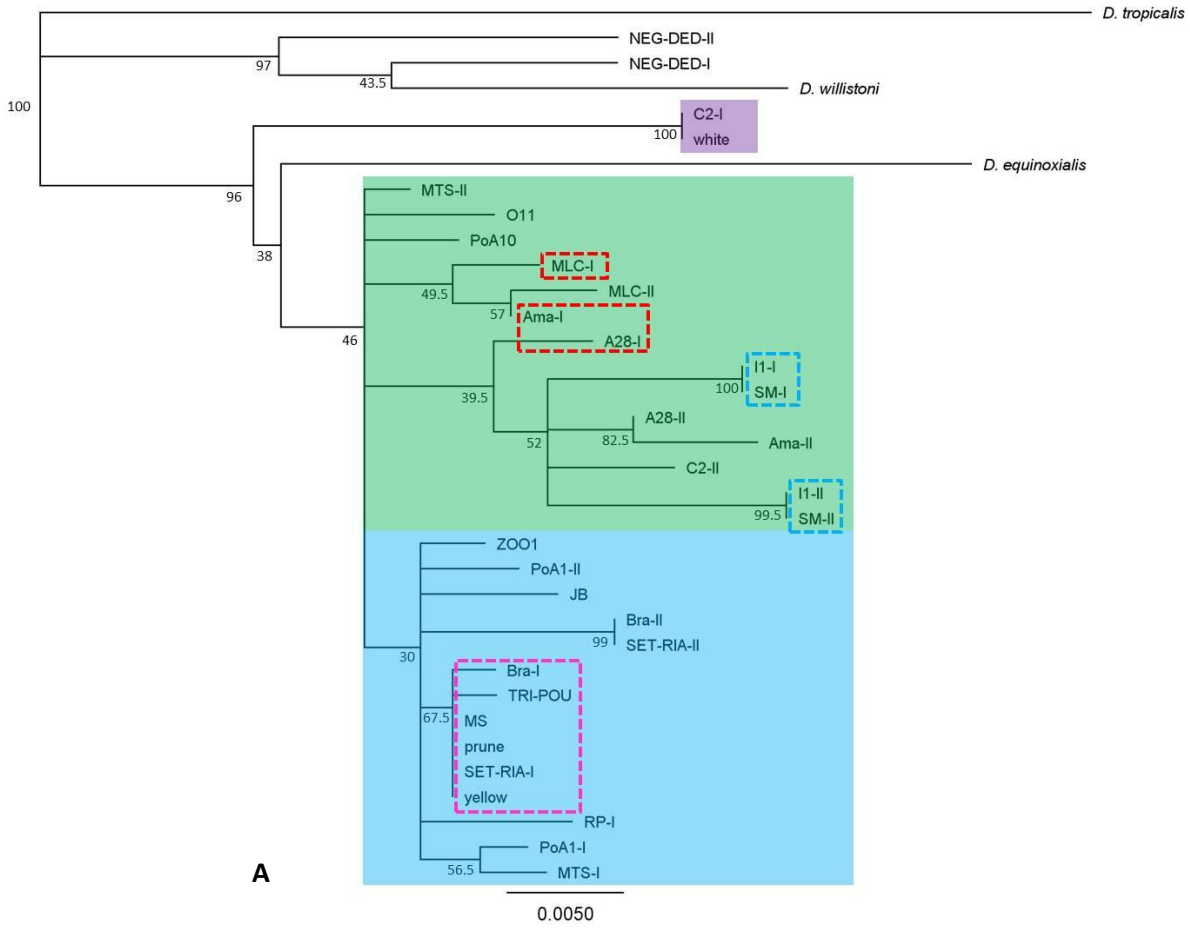
Fig. 4.27 A shows that the *Attacin D* sequences allow the identification of some of the newly collected strains: Prune and yellow mutant strains as well as one sequence version of SET-RIA are identical to the Andean-Brazilian strain MS. TRI-POU and one *Attacin D* version of the Brasilia strain (Bra) also show a high homology with MS. The white mutant sequence is identical to one version of the Centroamerican semispecies. Both sequence types of the Amazonica strain (Ama-I & -II) group closely with the two types of the Amazonian semispecies (A28-I & -II), the two types found in the MLC strain are closely related to A28-I. The two *Attacin D* types of I1 and SM are identical, respectively. The established marker system shown in **Fig. 4.27 B** allows an identification of MS-like *Attacin D* sequences based on a guanine at position 364 (pink dashed box). A diagnostic site for the A28-I sequence, shared with MLC-I and Ama-I, was found at position 181, showing cytosine instead of thymine (red dashed box). At position 65, both *Attacin D* sequences of I1 and SM can be distinguished from other semispecies by an adenine or cytosine instead of a thymine. At position 100, a big cluster comprising the *Attacin D* sequence of the Andean-Brazilian semispecies and the majority of the newly collected strains (JB, Bra, MTS-I, PoA1, prune, SET-RIA, TRI-POU, RP-I, yellow, ZOO1) is characterized by a thymine instead of an adenine. The second big cluster comprising the *Attacin D* sequences of the other classic semispecies (except for C2-I) and the new lines Ama, MLC and MTS-II can be identified by two diagnostic sites at position 100 (adenine) and 134 (guanine). C2-I and white mutant sequence has an adenine at position 134 as a diagnostic SNP. As mentioned earlier for *Attacin C*, the positions indicated in **Fig. 4.27 B** refer to the nt position in the PCR fragment without primer and may differ from the position in the alignment used for the polymorphism table.

AttD nt Position	004	022	030	036	059	065	076	079	100	130	134	155	161	170	171	172	173	174	175	176	177	178	179	180	181	182	183	184	185	186	188	189	192	193	203	274	279	299	269	271	274	279	299	301	309	349	354	364	382	397	398	404	407	430	445	457	472	478	493	502	507	541	562
Type	S	S	R	R	R	R	S	S	R	S	R	N	N	N	N	N	N	N	N	N	N	N	N	N	N	N	N	N	N	N	N	N	N	N	S	R	R	R	S	R	R	R	S	S	S	S	S	S	S	R	S	S	R	R	R	S	S	S	S	S	S	R	S	S
A28-I	C	C	G	G	G	G	A	A	A	G	G	c	a	A	a	a	c	g	t	t	t	g	c	a	c	t	c	a	t	a	g	c	t	a	g	G	A	G	T	G	G	A	G	A	C	A	A	A	A	T	G	G	G	G	T	A	C	G	C	G	G	T	T
MLC-II	T	C	A	G	G	G	A	C	A	A	G	c	a	A	a	a	c	g	t	t	t	g	c	a	c	t	c	a	t	a	g	c	t	a	g	G	A	G	T	G	G	A	G	A	C	A	A	A	A	T	G	G	G	G	T	T	C	T	C	A	G	T	T
AMA-I	T	C	A	G	G	G	A	A	A	G	G	c	a	A	a	a	c	g	t	t	t	g	c	a	c	t	c	a	t	a	g	c	t	a	g	G	A	G	T	G	G	A	G	A	C	A	A	A	A	T	G	G	G	G	T	T	C	T	C	G	G	T	T
MLC-I	T	C	A	G	G	G	A	A	A	G	G	c	a	A	a	a	c	g	t	t	t	g	c	a	t	t	c	a	t	a	g	c	t	a	g	G	A	G	G	G	G	A	G	A	C	A	A	A	A	T	G	G	G	G	T	T	C	T	C	A	G	T	T
A28-II	C	C	G	G	G	G	A	A	A	G	G	c	a	A	a	a	c	g	t	t	t	g	c	a	t	t	c	a	t	a	g	c	t	a	g	G	A	G	G	G	G	A	G	A	C	A	C	A	A	T	G	G	G	G	T	A	C	G	T	G	G	T	C
AMA-II	C	C	G	G	G	G	A	C	A	A	G	c	a	A	a	a	c	g	t	t	t	g	c	a	t	t	c	a	t	a	g	c	t	a	g	G	A	G	G	G	G	A	G	A	C	A	C	A	A	T	G	G	G	G	T	A	C	G	T	A	G	T	C
I1-I	C	T	A	G	G	A	A	A	A	G	G	c	t	A	a	a	c	g	t	t	t	g	c	a	t	t	c	a	t	a	g	c	t	a	g	G	A	G	G	G	G	A	G	A	C	A	A	A	A	A	C	G	G	G	G	T	A	C	G	T	G	G	T	C
SM-I	C	T	A	G	G	A	A	A	A	G	G	c	t	a	a	a	c	g	t	t	t	g	c	a	t	t	c	a	t	a	g	c	t	a	g	G	A	G	G	G	G	A	G	A	C	A	A	A	A	A	C	G	G	G	G	T	A	C	G	T	G	G	T	C
I1-II	C	C	A	G	T	C	C	A	A	G	G	G	c	a	–	–	–	–	–	–	–	–	–	–	–	t	t	c	a	t	a	g	c	t	a	t	G	A	G	G	G	G	A	G	A	C	A	A	A	A	A	T	G	G	G	G	T	A	C	G	T	G	C	G	C
SM-II	C	C	A	G	T	C	C	A	A	G	G	G	c	a	–	–	–	–	–	–	–	–	–	–	–	t	t	c	a	t	a	g	c	t	a	t	G	A	G	G	G	G	A	G	A	C	A	A	A	A	A	T	G	G	G	G	T	A	C	G	T	G	C	G	C
C2-II	C	C	A	G	G	G	G	A	A	A	G	G	G	t	a	a	a	a	c	g	t	t	t	g	c	a	t	t	c	a	t	a	g	c	t	a	t	G	A	G	G	G	G	A	G	T	C	A	A	A	A	A	T	G	G	G	G	T	A	C	G	T	G	G	T	C
ZOO1	C	C	A	G	G	G	G	A	A	T	G	G	G	c	a	a	a	a	c	g	t	t	t	g	c	a	t	t	c	a	t	a	g	c	t	a	g	G	A	G	G	G	G	A	G	A	C	A	A	A	A	A	T	A	G	G	G	T	T	C	T	C	G	G	T	C
PoA1-I	C	C	A	G	G	G	G	A	A	T	G	G	G	c	a	a	a	a	c	g	t	t	t	g	c	a	t	t	c	a	t	a	g	c	t	a	g	A	A	G	G	G	G	A	G	A	C	G	A	A	A	T	G	G	G	G	A	T	C	T	C	G	G	T	T
MTS-I	C	C	A	G	G	G	G	A	A	T	G	G	G	c	a	a	a	a	c	g	t	t	t	g	c	a	t	t	c	a	t	a	g	c	t	a	g	G	A	A	G	G	G	G	A	A	A	C	G	A	A	A	T	G	G	G	G	A	T	C	T	C	G	G	T	T
RP-I	C	C	A	G	G	G	G	A	A	T	G	G	G	c	a	a	a	a	c	g	t	t	t	g	c	a	t	t	c	a	t	a	g	c	t	a	g	G	A	G	G	G	G	A	G	A	C	A	A	T	T	C	G	T	G	G	A	T	C	T	C	G	G	T	T
MTS-II	C	C	A	G	G	G	G	A	A	A	G	G	G	c	a	A	a	a	c	g	–	–	–	–	–	–	–	–	–	t	a	g	c	t	a	t	G	A	G	G	G	G	A	G	A	C	A	A	A	A	A	T	G	G	G	G	T	T	C	T	C	G	G	T	T
O11	C	C	A	G	G	G	G	A	A	A	G	G	G	c	a	a	a	a	c	g	–	–	–	–	–	–	–	–	–	t	a	g	c	t	a	t	G	A	G	G	G	G	A	G	A	G	A	A	A	A	A	T	G	G	G	G	T	T	C	T	T	G	G	T	C
PoA10	C	C	A	G	G	G	G	A	A	A	G	G	G	c	a	a	a	a	c	g	–	–	–	–	–	–	–	–	–	t	a	g	c	t	a	t	T	A	G	G	G	T	A	G	A	C	A	A	A	A	A	T	G	G	G	G	T	T	C	T	C	G	G	T	C
PoA1-II	C	C	A	G	G	G	G	A	A	T	G	G	G	c	a	a	a	a	c	g	–	–	–	–	–	–	–	–	–	t	a	g	c	t	g	t	G	G	G	G	G	G	G	G	A	C	A	A	A	A	A	T	G	G	G	G	T	T	C	T	C	G	G	T	C
JB	C	C	A	G	G	G	G	A	A	T	G	G	G	c	a	a	a	a	c	g	–	–	–	–	–	–	–	–	–	t	a	g	c	t	a	t	G	A	G	G	T	G	A	G	A	C	A	A	A	A	A	T	G	G	G	G	T	T	C	T	C	G	G	T	C
BRA-II	C	C	A	A	G	G	A	A	T	G	G	G	c	a	a	a	a	a	c	g	–	–	–	–	–	–	–	–	–	t	a	g	c	t	a	t	G	A	G	G	G	G	A	G	A	C	A	A	A	A	A	T	G	G	C	A	A	T	T	T	C	G	G	T	C
SET-R.-II	C	C	A	A	G	G	A	A	T	G	G	G	c	a	a	a	a	a	c	g	–	–	–	–	–	–	–	–	–	t	a	g	c	t	a	t	G	A	G	G	G	G	A	G	A	C	A	A	A	A	A	T	G	G	A	A	A	T	T	T	C	G	G	T	C
BRA-I	C	C	A	G	G	G	A	A	T	G	G	G	c	a	a	a	a	a	c	g	–	–	–	–	–	–	–	–	–	t	a	g	c	t	a	t	G	A	G	G	C	G	A	G	A	C	A	A	G	A	T	G	G	G	G	G	T	T	C	T	C	G	G	T	T
TRI-P.	C	C	A	G	G	G	A	A	T	G	G	G	c	a	a	a	a	a	c	g	–	–	–	–	–	–	–	–	–	t	a	g	c	t	a	t	G	A	G	G	G	G	A	G	A	C	A	A	G	A	T	G	G	G	G	G	T	T	C	T	C	G	G	T	T
MS	C	C	A	G	G	G	A	A	T	G	G	G	c	a	a	a	a	a	c	g	–	–	–	–	–	–	–	–	–	t	a	g	c	t	a	t	G	A	G	G	G	G	A	G	A	C	A	A	G	A	T	G	G	G	G	G	T	T	C	T	C	G	G	T	T
prune	C	C	A	G	G	G	A	A	T	G	G	G	c	a	a	a	a	a	c	g	–	–	–	–	–	–	–	–	–	t	a	g	c	t	a	t	G	A	G	G	G	G	A	G	A	C	A	A	G	A	T	G	G	G	G	G	T	T	C	T	C	G	G	T	T
SET-R.-I	C	C	A	G	G	G	A	A	T	G	G	G	c	a	a	a	a	a	c	g	–	–	–	–	–	–	–	–	–	t	a	g	c	t	a	t	G	A	G	G	G	G	A	G	A	C	A	A	G	A	T	G	G	G	G	G	T	T	C	T	C	G	G	T	T
yellow	C	C	A	G	G	G	A	A	T	G	G	G	c	a	a	a	a	a	c	g	–	–	–	–	–	–	–	–	–	t	a	g	c	t	a	t	G	A	G	G	G	G	A	G	A	C	A	A	G	A	T	G	G	G	G	G	T	T	C	T	C	G	G	T	T
C2-I	C	C	A	G	G	G	A	A	A	G	A	c	t	a	a	a	–	–	t	t	t	–	–	–	t	a	c	a	c	c	t	t	a	a	g	G	A	G	G	G	G	A	G	A	C	A	A	A	A	T	G	G	G	G	T	T	C	T	C	G	G	T	C
white	C	C	A	G	G	G	A	A	A	G	A	c	t	a	a	a	t	t	t	t	–	–	–	–	a	c	a	c	c	t	t	a	a	g	G	A	G	G	G	G	A	G	A	C	A	A	A	A	A	T	G	G	G	G	T	T	C	T	C	G	G	T	C
NEG-DED-I	C	C	A	G	G	G	A	A	A	G	G	c	a	a	a	a	c	g	t	a	t	c	c	a	t	t	c	a	t	a	g	t	a	c	g	G	A	G	G	G	G	A	G	A	C	A	A	C	A	T	G	G	G	G	A	T	C	C	C	G	G	T	T
NEG-DED-II	C	C	A	G	G	G	A	A	A	G	A	c	a	a	a	a	c	g	t	a	t	c	c	a	t	t	c	a	t	a	g	t	a	c	g	G	A	G	G	G	G	A	G	A	C	A	A	T	A	T	G	G	G	G	A	T	C	C	C	G	G	T	T
D. equi	C	C	A	G	G	G	A	A	A	G	G	c	g	a	a	t	c	g	t	t	t	g	t	a	t	t	c	a	t	a	g	t	t	a	g	G	A	G	G	G	G	A	G	A	C	A	A	T	A	T	G	G	G	G	T	T	T	T	C	G	G	T	T
D. wls	C	C	A	G	G	G	A	A	A	G	G	c	a	a	a	a	c	g	t	a	t	c	c	a	t	t	c	a	t	a	g	t	a	c	g	G	A	G	T	G	G	A	G	A	C	A	A	T	A	T	G	G	G	G	A	T	C	A	C	G	G	T	T
D. trop	T	C	A	G	G	G	A	A	A	G	G	c	a	a	a	a	c	c	t	g	t	t	c	c	a	t	t	c	a	t	a	g	t	t	a	c	G	A	G	G	G	G	A	G	A	C	A	A	T	G	T	G	G	G	G	A	T	C	T	C	G	G	T	C

Table 4.15 A: Polymorphism table for *Attacin D* gene determined in newly collected *D. paulistorum* strains and mutant strains, grouped with *Attacin D* sequences of described semispecies. As outgroups, NEG-DED-I & -II, *D. equinoxialis* (*D. equi*), *D. willistoni* (*D. wls*) and *D. tropicalis* (*D. trop*) are shown. “nt Position” refers to nucleotide position in the alignment (of *D. paulistorum* semispecies without primers) in exonic (grey) and intronic (white) boxed regions. Type refers to noncoding (N), silent (S) or replacement (R) mutation; green filling represents changes from the majority consensus sequence and hyphen (–) refers to single nucleotide deletion.

AttD nt Position	030	036	059	065	100	134	257	266	269	279	299	309	354	382	398	404	407	507
Type	**E**	**N**	**N**	**E**	**N**	**E**	**N**	**N**	**E**	**E**	**N**	**N**	**N**	**N**	**N**	**N**	**E**	**N**
A28-I	S	G	A	V	K	V	G	T	L	N	G	S	E	Q	A	A	V	G
MLC-II	N	G	A	V	K	V	G	T	V	N	G	S	E	Q	A	A	V	G
Ama-I	N	G	A	V	K	V	G	T	V	N	G	S	E	Q	A	A	V	G
MLC-I	N	G	A	V	K	V	G	T	L	N	G	S	E	Q	A	A	V	G
A28-II	S	G	A	V	K	V	G	T	V	N	G	S	A	Q	A	A	V	G
Ama-II	S	G	A	V	K	V	G	T	L	N	G	S	A	Q	A	A	V	G
MTS-II	N	G	A	V	K	V	G	T	V	N	G	S	E	Q	A	A	V	G
I1-I	N	G	A	I	K	V	S	T	V	N	G	S	E	Q	A	A	V	G
SM-I	N	G	A	I	K	V	S	T	V	N	G	S	E	Q	A	A	V	G
I1-II	N	G	S	L	K	V	G	T	V	N	G	S	E	Q	A	A	V	A
SM-II	N	G	S	L	K	V	G	T	V	N	G	S	E	Q	A	A	V	A
C2-II	N	G	A	V	K	V	G	T	V	N	G	S	E	Q	A	A	V	G
O11	N	G	A	V	K	V	G	T	V	N	G	C	E	Q	A	A	V	G
PoA10	N	G	A	V	K	V	G	T	V	N	G	S	E	Q	A	A	V	G
PoA1-II	N	G	A	V	N	V	G	T	V	S	G	S	E	Q	A	A	V	G
JB	N	G	A	V	N	V	S	T	V	N	G	S	E	Q	A	A	V	G
Bra-II	N	E	A	V	N	V	G	T	V	N	G	S	E	Q	A	A	M	G
SET-RIA-II	N	E	A	V	N	V	G	T	V	N	G	S	E	Q	A	A	M	G
ZOO1	N	G	A	V	N	V	G	T	V	N	G	S	E	Q	T	A	V	G
PoA1-I	N	G	A	V	N	V	G	T	V	N	G	S	E	Q	A	A	V	G
MTS-I	N	G	A	V	N	V	G	T	V	N	R	S	E	Q	A	A	V	G
RP-I	N	G	A	V	N	V	G	T	V	N	G	S	E	H	A	S	V	G
Bra-I	N	G	A	V	N	V	G	T	V	N	G	S	E	Q	A	A	V	G
TRI-POU	N	G	A	V	N	V	G	A	V	N	G	S	E	Q	A	A	V	G
MS	N	G	A	V	N	V	G	T	V	N	G	S	E	Q	A	A	V	G
prune	N	G	A	V	N	V	G	T	V	N	G	S	E	Q	A	A	V	G
SET-RIA-I	N	G	A	V	N	V	G	T	V	N	G	S	E	Q	A	A	V	G
yellow	N	G	A	V	N	V	G	T	V	N	G	S	E	Q	A	A	V	G
C2-I	N	G	A	V	K	M	G	T	V	N	G	S	E	Q	A	A	V	G
white	N	G	A	V	K	M	G	T	V	N	G	S	E	Q	A	A	V	G
NEG-DED-I	N	G	A	V	K	V	G	T	V	N	G	S	E	Q	A	A	V	G
NEG-DED-II	N	G	A	V	K	M	G	T	V	N	G	S	E	Q	A	A	V	G
D. equi	N	G	A	V	K	V	G	T	V	N	G	S	E	Q	A	A	V	G
D. wls	N	G	A	V	K	V	G	T	L	N	G	S	E	Q	A	A	V	G
D. trop	N	G	A	V	K	V	A	T	V	N	G	S	E	Q	A	A	V	G

Table 4.15 B: Amino acid substitution table for *Attacin D* gene determined in newly collected *D. paulistorum* strains and mutant strains, grouped with *Attacin D* sequences of described semispecies. As outgroups, *D. equinoxialis* (*D. equi*), *D. willistoni* (*D. wls*) and *D. tropicalis* (*D. trop*) are shown. “nt Position” refers to nucleotide position in the alignment responsible for the amino acid substitution. Type refers to chemically equivalent (E) and non-equivalent (N) substitution.



B

Position:	100	134	65	181	364
	A	A	G	T	A/T
	A	G	G	T	A
	A	G	G	C	A
	A	G	A/C	T	A
	T	G	G	T	A/T
	T	G	G	T	G

Fig. 4.27 ***Attacin D*** - marker system for *D. paulistorum strains.* **A:** Neighbour-Joining tree with *D. tropicalis* as outgroup, calculated with Jukes-Cantor genetic distance model in Geneious programme. The tree was resampled with Bootstrap method and 200 replicates (bootstrap values indicated at the nodes of the tree), the threshold was at 30%. The diagnostic sites position 100 and 134 allows to distinguish between three main groups (highlighted in blue, green and purple), Coloured dashes show subgroups. **B:** Table of SNPs defining the groups highlighted in the respective colour/dashes. Position refers to the nucleotide position in the sequence without primer.

4.4.1.3 *ND4*

Diagnostic SNP typing of 602 bp long *ND4* sequence was performed with the complete set of newly collected and mutant strains except for SAL-13 and NEG-DED. As shown before in **Table 4.13**, the data used for the strains JB, MTS, PoA10, RP derived of cloned *ND4* PCR fragments and direct sequencing. For mutant strains prune, white and yellow, only cloned sequences were used while for the other strains, only direct sequencing reads were available. For RP, additional Illumina sequencing data, provided by Lisa Klasson, Uppsala University, Sweden, was included.

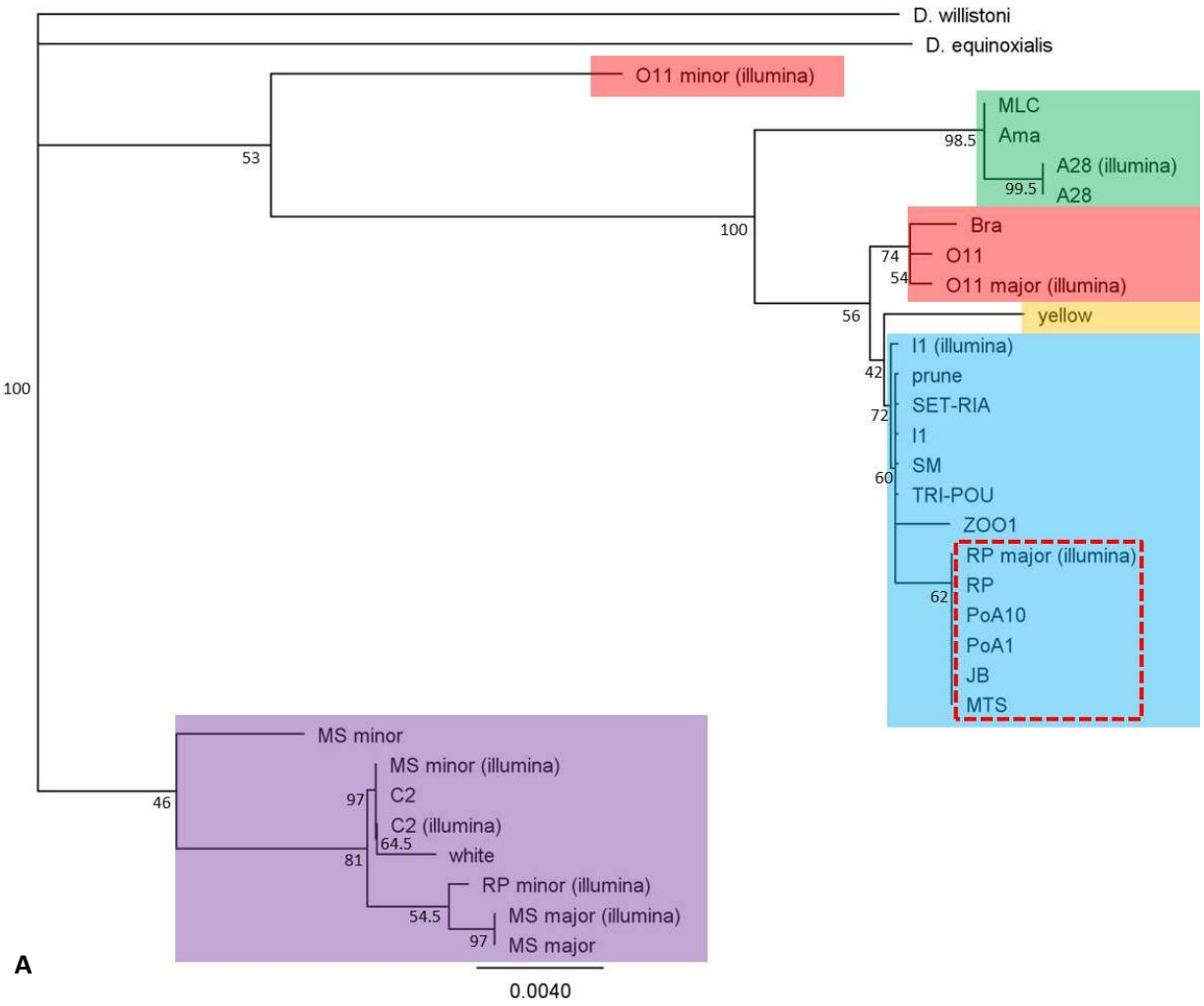
ND4 **nt Position**	**016**	**031**	**040**	**049**	**064**	**067**	**082**	**091**	**153**	**154**	**157**	**175**	**215**	**253**	**256**	**280**	**286**	**298**	**300**	**309**	**310**	**319**	**337**	**343**	**349**	**373**	**415**	**430**	**448**	**459**	**472**	**487**	**552**	**562**	**571**	**588**	**589**
Type	**S**	**S**	**S**	**S**	**S**	**S**	**S**	**S**	**R**	**S**	**S**	**S**	**R**	**S**	**S**	**S**	**S**	**S**	**S**	**R**	**S**	**S**	**S**	**S**	**S**	**S**	**S**	**S**	**S**	**R**	**S**	**S**	**R**	**S**	**S**	**S**	**S**
A28	C	A	T	C	A	T	T	C	A	C	C	A	G	A	T	T	C	G	G	C	C	C	T	T	T	C	G	A	G	T	C	T	T	C	C	A	C
A28 Illu.	C	A	T	C	A	T	T	C	A	C	C	A	G	A	T	T	C	G	G	C	C	C	T	T	T	C	G	A	G	T	C	T	T	C	C	A	C
MLC	C	A	T	C	A	T	T	C	A	C	C	A	G	A	T	T	C	G	G	T	C	C	T	T	T	C	G	A	G	T	C	T	T	C	C	A	C
Ama	C	A	T	C	A	T	T	C	A	C	C	A	G	A	T	T	C	G	G	T	C	C	T	T	T	C	G	A	G	T	C	T	T	C	C	A	C
C2	C	T	A	T	G	T	A	T	A	T	T	A	G	A	C	A	T	T	A	T	T	C	C	C	C	T	A	A	G	T	T	T	T	T	C	A	T
C2 Illu.	C	T	A	T	G	T	A	T	A	T	T	A	G	A	C	A	T	T	A	T	T	C	C	C	C	T	A	A	G	T	T	T	T	T	C	A	T
MS min. Illu.	C	T	A	T	G	T	A	T	A	T	T	A	G	A	C	A	T	T	A	T	T	C	C	C	C	T	A	A	G	T	T	T	T	T	C	A	T
white	C	T	A	T	G	T	A	T	A	T	T	G	G	A	C	A	T	T	A	T	T	C	C	C	C	T	A	A	G	T	T	T	T	T	C	A	T
RP min. Illu.	C	T	A	T	G	T	A	T	C	T	T	A	A	A	C	A	T	T	A	T	T	C	C	C	C	T	A	A	G	T	T	T	T	T	C	A	T
MS maj.	T	T	A	T	G	T	A	T	A	T	T	A	A	A	C	A	T	T	A	T	T	C	C	C	C	T	A	A	A	T	T	T	T	T	C	A	T
MS maj. Illu.	T	T	A	T	G	T	A	T	A	T	T	A	A	A	C	A	T	T	A	T	T	C	C	C	C	T	A	A	A	T	T	T	T	T	C	A	T
O11 min. Illu.	T	T	A	T	G	T	A	T	C	T	T	A	A	A	C	T	C	A	G	T	T	T	C	C	C	T	A	A	A	T	T	T	T	C	C	A	C
MS min.	T	T	A	T	A	C	A	T	A	T	T	A	G	A	T	A	T	T	A	T	T	C	C	T	C	C	A	A	G	T	T	T	T	T	C	A	T
I1	C	A	T	C	A	T	T	C	A	C	C	A	G	G	T	T	C	A	G	T	T	C	T	T	T	C	A	A	G	T	C	C	T	C	C	A	C
I1 Illu.	C	A	T	C	A	T	T	C	A	C	C	A	G	G	T	T	C	A	G	T	T	C	T	T	T	C	A	A	G	T	C	C	T	C	C	A	C
SM	C	A	T	C	A	T	T	C	A	C	C	A	G	G	T	T	C	A	G	T	T	C	T	T	T	C	A	A	G	T	C	C	T	C	C	A	C
SET-RIA	C	A	T	C	A	T	T	C	A	C	C	A	G	G	T	T	C	A	G	T	T	C	T	T	T	C	A	A	G	T	C	C	T	C	C	A	C
TRI-POU	C	A	T	C	A	T	T	C	A	C	C	A	G	G	T	T	C	A	G	T	T	C	T	T	T	C	A	A	G	T	C	C	T	C	C	A	C
prune	C	A	T	C	A	T	T	C	A	C	C	A	G	G	T	T	C	A	G	T	T	C	T	T	T	C	A	A	G	T	C	C	T	C	C	A	C
O11	C	A	T	C	A	T	T	C	A	C	C	A	A	G	T	T	C	A	G	T	T	T	T	T	T	C	A	A	G	T	C	C	T	C	C	A	C
O11 maj. Illu.	C	A	T	C	A	T	T	C	A	C	C	A	A	G	T	T	C	A	G	T	T	T	T	T	T	C	A	A	G	T	C	C	T	C	C	A	C
Bra	C	A	T	C	A	T	T	C	A	T	C	A	A	G	T	T	C	A	G	T	T	T	T	T	T	C	A	A	G	T	C	C	T	C	C	A	C
JB	C	A	T	C	A	T	T	C	A	C	C	A	G	G	T	T	C	A	G	T	T	C	T	T	T	C	A	T	G	T	C	C	T	C	C	A	C
MTS	C	A	T	C	A	T	T	C	A	C	C	A	G	G	T	T	C	A	G	T	T	C	T	T	T	C	A	T	G	T	C	C	T	C	C	A	C
PoA1	C	A	T	C	A	T	T	C	A	C	C	A	G	G	T	T	C	A	G	T	T	C	T	T	T	C	A	T	G	T	C	C	T	C	C	A	C
PoA10	C	A	T	C	A	T	T	C	A	C	C	A	G	G	T	T	C	A	G	T	T	C	T	T	T	C	A	T	G	T	C	C	T	C	C	A	C
RP	C	A	T	C	A	T	T	C	A	C	C	A	G	G	T	T	C	A	G	T	T	C	T	T	T	C	A	T	G	T	C	C	T	C	C	A	C
RP maj. Illu.	C	A	T	C	A	T	T	C	A	C	C	A	G	G	T	T	C	A	G	T	T	C	T	T	T	C	A	T	G	T	C	C	T	C	C	A	C
ZOO1	C	A	T	C	A	T	T	C	A	C	C	A	G	G	T	T	C	A	G	T	T	C	T	T	T	C	A	A	G	T	C	C	C	C	C	A	C
yellow	C	A	T	C	A	T	T	C	A	C	C	A	G	G	T	T	C	A	G	T	T	C	T	T	T	C	A	A	G	C	T	C	T	C	T	G	C
D. equi	C	T	A	T	G	C	A	T	A	C	T	A	G	A	T	A	T	T	A	T	T	T	T	T	C	T	A	A	A	T	T	C	T	C	T	A	T
D. wls	C	T	T	T	G	T	A	T	A	T	T	A	G	A	T	C	T	T	A	T	A	C	T	T	C	C	A	A	C	T	T	T	T	T	T	A	T

Table 4.16 A: Polymorphism table for mitochondrial *ND4* gene determined in newly collected *D. paulistorum* strains and mutant strains, grouped with *ND4* sequences of described semispecies. As outgroups, *D. equinoxialis* (*D. equi*) and *D. willistoni* (*D. wls*) are shown. “nt Position” refers to nucleotide position in the alignment. Type refers to silent (S) or replacement (R) mutation; green filling represents changes from the majority consensus sequence. “maj.” and “min.” indicate ND4 sequences of the potential major and minor mitotype detected with sanger sequencing, “maj. Illu.” and ”min. Illu.” represent potential major and minor mitotype detected with Illumina sequencing.

Table 4.16 A shows that the SNP in position 430 detected in the South Brazilian strains JB, MTS, PoA1, PoA10 and RP (Sanger sequencing and major mitotype of Illumina sequencing) could not be found in any *ND4* sequence of the described semispecies. The same is true for the SNP found in ZOO1 in position 552 as well as the SNPs of the yellow mutant strain in position 459, 571 and 588. SNP 450, diagnostic for the South Brazilian strains, is a silent mutation; in contrast, the mutations found in the yellow mutant (nt position 459, aa position 48) and in ZOO1 strain (nt position 552, aa position 17) lead to amino acid substitutions. In the case of the yellow mutant strain, this substitution is chemically non-equivalent (**Table 4.16 B**). A second mitochondrial haplotype was found less abundant for the RP strain by Illumina sequencing (RP min. Illu.), which could not be confirmed by Sanger reads. This mitotype could be restricted only to embryonic stages, while in adult probes the mayor mitotype becomes predominant.

ND4 **nt Position**	**153**	**215**	**309**	**459**	**552**
Type	**E**	**N**	**E**	**E**	**E**
A28	L	T	V	M	I
A28 Illu.	L	T	V	M	I
MLC	L	T	I	M	I
Ama	L	T	I	M	I
C2	L	T	I	M	I
C2 Illu.	L	T	I	M	I
MS min. Illu.	L	T	I	M	I
white	L	T	I	M	I
RP min. Illu.	V	M	I	M	I
MS maj.	L	M	I	M	I
MS maj. Illu.	L	M	I	M	I
O11 min. Illu.	V	M	I	M	I
MS min.	L	T	I	M	I
I1	L	T	I	M	I
I1 Illu.	L	T	I	M	I
SM	L	T	I	M	I
SET-RIA	L	T	I	M	I
TRI-POU	L	T	I	M	I
prune	L	T	I	M	I
O11	L	M	I	M	I
O11 maj. Illu.	L	M	I	M	I
Bra	L	M	I	M	I
JB	L	T	I	M	I
MTS	L	T	I	M	I
PoA1	L	T	I	M	I
PoA10	L	T	I	M	I
RP	L	T	I	M	I
RP maj. Illu	L	T	I	M	I
ZOO1	L	T	I	M	V
yellow	L	T	I	V	I
D. equi	L	T	I	M	I
D. wls	L	T	I	M	I

Table 4.16 B: Amino acid substitution table for *ND4* gene determined in newly collected *D. paulistorum* strains and mutant strains, grouped with *ND4* sequences of described semispecies. As outgroups, *D. equinoxialis* (*D. equi*) and *D. willistoni* (*D. wls*) are shown. “nt Position” refers to nucleotide position in the alignment responsible for the amino acid substitution. Type refers to chemically equivalent (E) and non-equivalent (N) substitution. “maj.” and “min.” indicate *ND4* sequences of the potential major and minor mitotype detected with sanger sequencing, “maj. Illu.” and ”min. Illu.” represent potential major and minor mitotype detected with Illumina sequencing.



B

Position:	280	310	253	319	430	459
	A	T	A	C	A	T
	T	C	A	C	A	T
	T	T	G	C	A	T
	T	T	G	C	T	T
	T	T	G	T	A	T
	T	T	G	C	A	C

Fig. 4.28: ***ND4*** - marker system for *D. paulistorum strains.* **A:** Neighbour-Joining tree with *D. willistoni* as outgroup, calculated with Jukes-Cantor genetic distance model in Geneious programme. The tree was resampled with Bootstrap method and 200 replicates (bootstrap values indicated at the nodes of the tree), the threshold was at 30%. Strains grouping with A28 are highlighted in green, the Brazilian cluster in blue, strains grouping with C2-I/MS/white in purple. O11 and Bra can be identified with SNP in position 319 and are highlighted in red. Yellow mutant strain is highlighted in yellow, coloured dashes represent subgroups. **B:** Table of SNPs defining the groups highlighted in the respective colour/dashes. Position refers to the nucleotide position in the sequence without primer. Illumina sequences identical to Sanger reads (A28, C2, I1, MS major, RP major, and O11 major) are not shown.

In **Fig. 4.28,** the marker system established with *ND4* sequences is shown. Most of the newly collected Brazilian strains group closely to the Interior (I1) and Transitional (SM) semispecies. However, due to the low bootstrap support, they could also group with OR. They are highlighted with a blue box and can be identified with two markers in position 253 (guanine) and 319 (cytosine). TRI-POU, SET-RIA and the prune mutant strain are identical to I1/SM while South Brazilian strains JB, MTS, PoA1, PoA10 and RP (red dashed box) share a common diagnostic SNP in position 430 (thymine). Centroamerican (C2), Andean-Brazilian (MS), white mutant strain and the Illumina minor sequence of the Brazilian strain RP form a cluster with a common diagnostic SNP in position 280 (adenine). Ama and MLC strain group closely to A28 (green box) and can be identified in position 310 (cytosine). Sanger and major Illumina sequence of O11 strain show high homology to Bra strain, together with the minor Illumina sequence of O11 they can be typed with the SNP in position 319 (thymine, red box). The yellow mutant strain has a unique site in position 459 (cytosine, yellow box).

4.4.2 Southern blot analysis of newly collected lines

In order to confirm the multiple sequence types found for *Attacin D* (**Fig. 4.15**) and *ND4* (**Fig. 4.19**) for the classic set of semispecies, Southern blot analysis was performed with the newly collected Brazilian strains.

The *Attacin D* hybridization of *Hin*dIII-digested DNA (**Fig.4.29**) showed for all Brazilian strains a band at 5,400 bp, at the same height as the strains of the described semispecies Amazonian (A28), Andean-Brazilian (MS) and Orinocan (O11). The second, very faint signal at 12,000 bp observed in the six semispecies was also detected in all new strains except for JB, were the second signal appeared at 13,500 bp. Additionally, a third signal at 2,500 bp was detected in this strain. MTS strain also showed a third signal at 6,600 bp, PoA1 a third at 4,000 bp and a fourth at 11,000 bp.

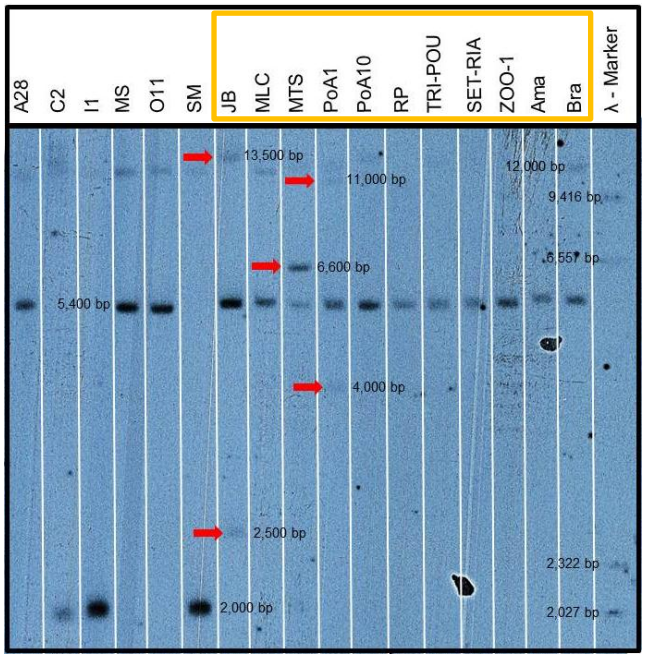


Fig. 4.29: Southern blot of *Hin*dIII-digested *D. paulistorum* DNA (female, multifly extraction following Junakovic, 2004) hybridized with P^{32}-labelled *Attacin D* probe. Newly collected Brazilian strains (indicated by yellow rectangle) were loaded next to the described semispecies. White lines serve as aid lines to assign corresponding strains and signals. *Hin*dIII-digested λ-phage DNA was used as size marker; radioactive signals derive from pencil marks on the membrane. The images were captured on X-ray-film with three-day exposure. Red arrows indicate bands only observed in newly collected strains.

Southern blot analysis of mitochondrial sequences was performed with *Psi*I – digested DNA, hybridized with P^{32}-labelled probes of *ND4* and *CO* (**Fig. 4.30 A, B**). For *CO1* blot, a smaller set of strains was available.

With *ND4* probes (**Fig. 4.30 A**), MLC and Ama strain showed the same RFLP pattern as the Amazonian semispecies (A28) at 1,600 and 1,700 bp. As shown earlier for A28 (**Fig. 4.21**), the 1,700 bp band is the result of incomplete digestion. The RFLP patterns of the other Brazilian strains are the same as for the Interior (I1) and transitional (SM) semispecies with two diagnostic bands at 1,100 and 1,600 bp, a faint band at 1,700 and an obscure band, potentially representing the calculated 461 bp fragment (see **Fig. 4.20 A**).

CO1 RFLP patterns are identical for A28, I1 and SM and hence also all for examined Brazilian strains with a single band at 2,700 bp (**Fig. 4.20 B**).

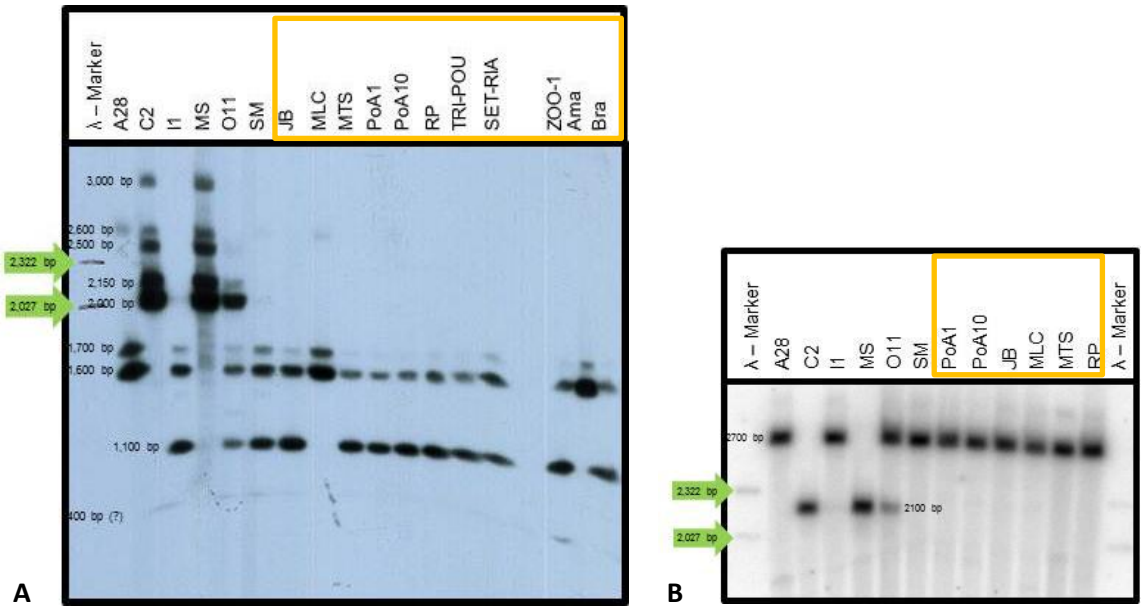


Fig. 4.30: Southern blot of *Psi*I-digested DNA of *D. paulistorum* semispecies and newly collected Brazilian strains (indicated by yellow rectangle). DNA of female flies (multifly extraction after Junakovic, 2004) was **A:** hybridized with P^{32}-labelled *ND4* probe, **B:** hybridized with P^{32}-labelled *CO1* probe. Green arrows indicate 2,027 and 2,322 bp fragment of *Hin*dIII-digested λ-phage DNA (size marker); radioactive signals derive from pencil marks on the membrane. **A:** X-ray-film, 19 hrs of exposure, **B:** Phospoimager, 35 min of exposure.

4.5 Evolutionary pressure on *Attacin D* domains of *D. paulistorum* strains

The obtained *Attacin D* sequences derived from *D. paulistorum* strains give the opportunity to test this gene for signs of purifying, diversifying or neutral selection. As *D. paulistorum* semispecies harbour different strains of *Wolbachia* endosymbionts (Miller et al., 2010), diversifying selection of effector proteins of the host immunity, which interact with *Wolbachia* is to be expected.

In order to test *Attacin D* sequences of *D. paulistorum* strains for regions of high amino acid diversity, a sliding window analysis was performed, with *D. paulistorum strains* and distantly related species *D. equinoxialis, D. tropicalis and D. willistoni* (**Fig. 4.31 A, B**). Compared to the other species, *D. paulistorum* showed elevated divergence in the middle of the N-terminal and G1 domain and especially in the N-terminal region of the G2 domain.

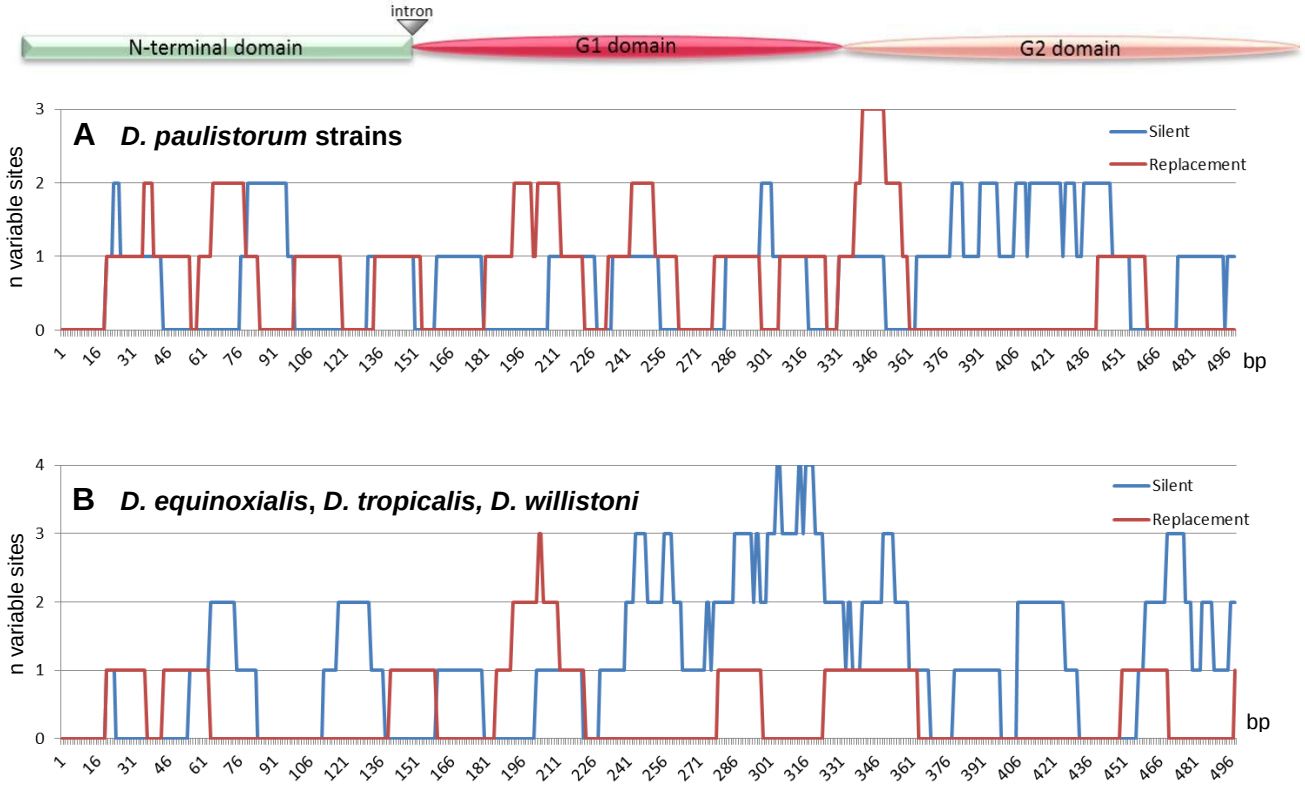


Figure 4.31: Sliding window analysis of variable sites in the coding region of ***Attacin D*** PCR fragment (500 bp) among 20 ***D. paulistorum*** strains (**A**) and within more distantly related species ***D. equinoxialis*, *D. willistoni* and *D. tropicalis*** (**B**). Numbers of silent (blue) or replacement (red) sites in a window of 20 nucleotides are displayed at the last (20th) position of the window. X-axis shows the nucleotide position, Y-axis shows the number of variable sites. The schematic on top of the figure illustrates the corresponding domain region.

To test if the diversity of *Attacin D* observed in *D. paulistorum* is the result of evolutionary pressure, the K_a/K_s values of each domain were determined. DNAsp v.

5.10 was used to calculate the ratio of the number of non-synonymous substitutions per non-synonymous site (K_a) to the number of synonymous substitutions per synonymous site (K_s) in the *Attacin D* fragment. We calculated the K_a/K_s ratios for each *D. paulistorum* strain and compared it to *D. willistoni*. *D. equinoxialis, D. tropicalis* and NEG-DED were used as outgroups. The K_a/K_s ratio was determined separately for each domain. In theory, a value greater than one indicates diversifying selection, less than one purifying selection, whereas a value of one suggests neutral selection (Li et al., 1985). For all domains, the K_a/K_s values are clearly smaller than one (**Table 4.17**). The average K_a/K_s value in *D. paulistorum* strains for the N-terminal domain is 0.18, for the G1 domain 0.07 and for the G2 domain 0.06. For Sal-13 (both paralogues) and RP-II no comparable K_s and K_a values could be calculated for the N-terminal and G2 domain, as they lack 41 - 43 bp at each flank.

	N-terminal domain			G1 domain			G2 domain		
	Ks	Ka	Ka/Ks	K_s	K_a	K_a/K_s	K_s	K_a	K_a/K_s
A28-I	0.092	0.019	0.21	0.1313	0	0	0.2758	0.0151	0.05
MLC-I	0.1602	0.0095	0.06	0.1313	0	0	0.2432	0.0151	0.06
Ama-I	0.1253	0.0095	0.08	0.1313	0	0	0.2119	0.0151	0.07
MLC-II	0.1602	0.0095	0.06	0.1308	0.0073	0.06	0.2432	0.0151	0.06
A28-II	0.092	0.019	0.21	0.1296	0.0146	0.11	0.2758	0.0151	0.05
Ama-II	0.1602	0.019	0.12	0.1296	0.0146	0.11	0.099	0.0151	0.15
RP-II	-	-	-	0.1308	0.0073	0.06	-	-	-
MTS-II	0.092	0.019	0.21	0.1606	0.0146	0.09	0.2432	0.0151	0.06
I1-I	0.1259	0.019	0.15	0.1319	0.0146	0.11	0.3099	0.0151	0.05
SM-I	0.1259	0.019	0.15	0.1319	0.0146	0.11	0.3099	0.0151	0.05
SM-II	0.1246	0.0288	0.23	0.1308	0.0073	0.06	0.3099	0.0228	0.07
I1-II	0.1246	0.0288	0.23	0.1308	0.0073	0.06	0.3099	0.0228	0.07
C2-II	0.1253	0.0095	0.08	0.1599	0.0073	0.05	0.2758	0.0151	0.05
O11	0.092	0.0095	0.10	0.1316	0.0146	0.11	0.2119	0.0151	0.07
PoA10	0.092	0.0095	0.10	0.1599	0.0073	0.05	0.1819	0.0151	0.08
PoA1-II	0.092	0.019	0.21	0,1308	0.0146	0.11	0.1819	0.0151	0.08
JB	0.092	0.019	0.21	0.1765	0.0182	0.10	0.1819	0.0151	0.08
Bra-II	0.0929	0.0287	0.31	0.1308	0.0073	0.06	0.2148	0.0227	0.11
SET-RIA-II	0.0929	0.0287	0.31	0.1308	0.0073	0.06	0.2148	0.0227	0.11
ZOO1	0.092	0.019	0.21	0.1308	0.0073	0.06	0.1819	0.0228	0.13
PoA1-I	0.092	0.019	0.21	0.1906	0.0073	0.04	0.1819	0.0151	0.08
MTS-I	0.092	0.0095	0.10	0.1308	0.0073	0.06	0.2119	0.0151	0.07
RP-I	0.092	0.019	0.21	0.1027	0.0146	0.14	0.2119	0.0228	0.11
Bra-I	0.092	0.019	0.21	0.1452	0.0109	0.08	1.2119	0.0151	0.01
TRI-POU	0.092	0.019	0.21	0.1308	0.0146	0.11	0.2119	0.0151	0.07
MS	0.092	0.019	0.21	0.1308	0.0073	0.06	0.2779	0.0151	0.05
prune	0.092	0.019	0.21	0.1308	0.0073	0.06	0.2119	0.0151	0.07
SET-RIA-I	0.092	0.019	0.21	0.1308	0.0073	0.06	0.2119	0.0151	0.07
yellow	0.092	0.019	0.21	0.1308	0.0073	0.06	0.2119	0.0151	0.07
C2-I	0.0934	0.019	0.20	0.1308	0.0073	0.06	0.1819	0.0151	0.08
white	0.0934	0.019	0.20	0.1308	0.0073	0.06	0.1819	0.0151	0.08
D. equinoxialis	0.1253	0.019	0.15	0.2197	0.0146	0.07	0.2454	0.0227	0.09
NEG-DED-I	0.0594	0.0387	0.65	0.1308	0.0073	0.06	0.1819	0.0228	0.13
NEG-DED-II	0.0609	0.0286	0.47	0.1902	0.0073	0.04	0.1819	0.0228	0.13
SAL13-I	-	-	-	0.1308	0.022	0.10	-	-	-
SAL13-II	-	-	-	0.2218	0.022	0.06	-	-	-
D. tropicalis	0.092	0.019	0.21	0.2865	0.0221	0.08	0.1819	0.0228	0.13

Table 4.17: K_s, K_a values and **K_a/K_s** ratio for *Attacin D* domains (**N-terminal, G1, G2**) of all *D. paulistorum* strains (red filling) and outgroups (blue filling) versus *D. willistoni.*

4.6 Phylogenetic relationships derived from mitochondrial and nuclear marker in *D. paulistorum*

Mitochondrial and nuclear phylogeny in *D. paulistorum* species complex showed in previous studies intriguingly different topologies (Gleason et al., 1998, Robe et al., 2010). In order to compare the mitochondrial *ND4* data with the obtained nuclear sequences, a concatemeric nuclear tree was built (**Fig. 4.32 A**). For this tree, the sequences for the AMPs described above (except *Drosocin*) and the immune receptor gene *SrC1* plus the sequences of the genes for *Desaturase F-η* and *Desaturase F-ζ* (provided by D. Schneider) came to use. The sequences were concatenated in the following order: *Attacin A – Attacin C – Attacin D – Diptericin – Diptericin B - SrC1 – Desaturase F-η - Desaturase F-ζ*. For strains with multiple types of some genes, the presumably more ancestral first paralogue was chosen. The resulting concatemer had a length of 4068 bp in the consensus sequence.

In this concatenated nuclear tree, the grouping of C2 and O11 with the other semispecies is intriguingly different compared to the tree of mitochondrial *ND4* sequences (**Fig. 4.32 B**). Considering the nuclear sequences, O11 strain groups closely to MS, and the C2 strain groups in close proximity to I1/SM. The mitochondrial *ND4* sequence of O11 however groups closely to I1/SM/A28 while the *ND4* sequence of C2 shows high homology to MS.

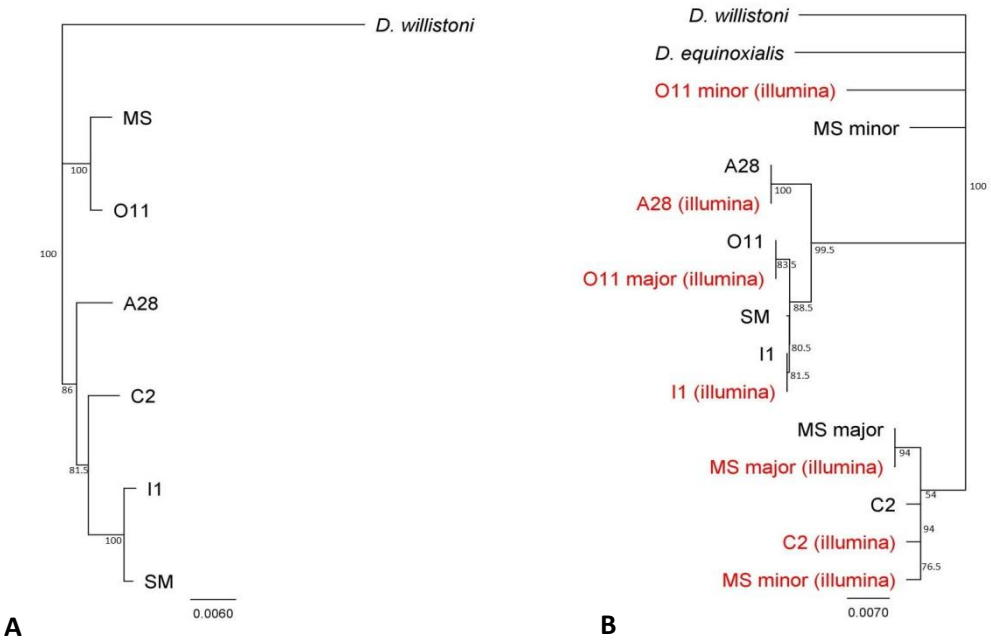


Fig. 4.32 A: Concatemeric Neighbour-Joining tree of sequences of *Attacin A – Attacin C – Attacin D – Diptericin – Dipericin B – SrC1 – Desaturase F-η - Desaturase F-ζ*. **B:** Neighbour-Joining tree of mitochondrial *ND4* Sanger and Illumina (typed in red) sequences of *D. paulistorum* strains, *D. equinoxialis* and *D. willistoni.* Illumina sequences are typed red, Sanger sequences in black. **A + B:** Bootstrap values are indicated at the nodes of the tree. Trees were calculated using the Geneious tree builder with Jukes-Cantor genetic distance model, Neighbour-Joining tree building method and *D. willistoni* as outgroup. The trees were resampled with Bootstrap method and 200 replicates, the threshold was at 50%.

4.7 Paternal transmission of mitochondria in *D. paulistorum*

Nigro and Prout (1990) showed with *D. simulans* strains that *Wolbachia* is associated with mitochondrial haplotypes. In hybrids of *D. paulistorum semispecies* (A28xO11^{R} and C2xA28^{R}), paternal transmission of *Wolbachia* has been observed recently by W. Miller and D. Schneider (unpublished data). To test if the paternal transmission of *Wolbachia* in these hybrids is accompanied by paternal heritage of mitochondria, DNA of A28xO11^{R} and C2xA28^{R} hybrids (F28 - F29) was examined with previously established *Psi*I RFLP patterns of mitochondrial genes *ND4* and *CO1* in Southern blots.

4.7.1 *ND4* RFLP pattern in *D. paulistorum* hybrids

We performed Southern blots with *Psi*I-digested male and female DNA of A28, C2, O11 and A28xO11^{R} and C2xA28^{R} hybrids (F28 – F29). As controls, female DNA of *Wolbachia*-uninfected *D. melanogaster* white mutant (w^{1118}) and *Wolbachia*-infected *D. equinoxialis* and *D. willistoni* (wls) were used. The membranes were hybridized with radioactively-labelled *ND4* fragments (**Fig. 4.33** and **Fig. 4.34**). DNA for those Southern blots was extracted following Junakovic (2004), the multiple bands appear as observed earlier, most likely because of incomplete digest.

Surprisingly, the pattern of the paternal O11 strain, i.e., two characteristic bands at 2,000 and 2,100 bp, (indicated by red box in **Fig. 4.33**) is undoubtedly observed in A28xOR^{R} hybrids. As the restriction fragment lengths of A28 also appear in O11, it cannot be excluded that the hybrids carry both, A28 and O11 mitochondrial haplotypes.

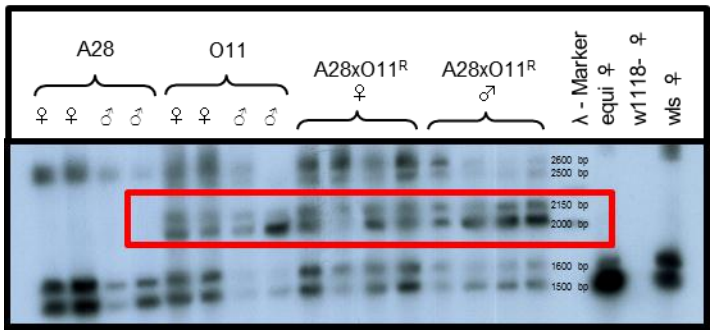


Fig. 4.33: Southern blot of *Psi*I-digested *D. paulistorum* DNA from multifly extraction after Junakovic (2004), hybridized with radioactively-labelled ***ND4*** probe. DNA of male (♂) and female (♀) A28 and O11 individuals plus A28xO11^{R} hybrids was used. As controls, equally treated DNA of female *D equinoxialis (equi)*, *D. melanogaster* white mutant (w^{1118}) and *D. willistoni* was loaded. *Hin*dIII-digested λ-phage DNA was used as size marker. Red box indicates the paternal signature. X-ray film, 21 hrs of exposure.

Figure 4.34 shows the RFLP pattern of C2 and A28 compared to C2xA28^{R} hybrids. Here, the paternal pattern of A28 does not include the fragment heights of C2 (2,000 and 2,150 bp), so this experiment clearly shows the absence of maternal mitochondrial signal in the hybrids, and solely the paternal signature (1,500 and 1,600 bp) instead.

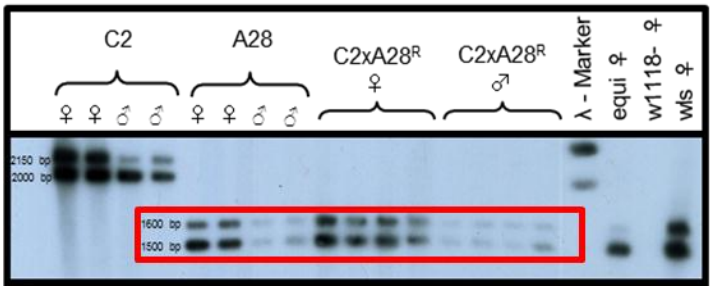


Fig. 4.34: Southern blot of *Psi*I-digested *D. paulistorum* DNA from multifly extraction after Junakovic (2004), hybridized with radioactively-labelled ***ND4*** probe. DNA of males (♂) and females (♀) of C2 and A28 individuals plus C2xA28^{R} hybrids was used. As controls, equally treated DNA of female *D equinoxialis (equi)*, *D. melanogaster* white mutant (w^{1118}) and *D. willistoni* was loaded. *Hin*dIII-digested λ-phage DNA was used as size marker; radioactive signals derive from pencil marks on the membrane. Red box indicates the paternal signature. X-ray film, 22 hrs of exposure.

4.7.2 *CO1* RFLP pattern in *D. paulistorum* hybrids

In order to confirm the results obtained through *ND4* hybridization, the membranes were stripped and rehybridized with radioactively labelled *CO1* probe. The paternal signature of O11 (2,100 bp) was observed in A28xO11^{R} hybrids (**Fig 4.35**). In C2xA28^{R} hybrids, the paternal A28 signature (2,700 bp) could be determined (**Fig. 4.36)**.

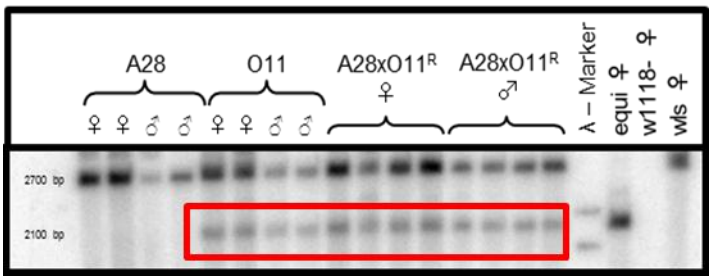


Fig. 4.35: Southern blot of *Psi*I-digested *D. paulistorum* DNA from multifly extraction after Junakovic (2004), hybridized with radioactively labelled ***CO1*** probe. DNA of male (♂) and female (♀) of A28 and O11 individuals plus A28xO11^{R} hybrids was used. As controls, equally treated DNA of female *D equinoxialis (equi)*, *D. melanogaster* white mutant (w^{1118-}) and *D. willistoni* was loaded. *Hin*dIII-digested λ-phage DNA was used as size marker. Red box indicates the paternal signature. Phosphoimager, 50 min.

Fig. 4.35 shows that the restriction fragments of A28 (2,700 bp) appear again also in O11. Hence, although the pattern of the paternal O11 strain, i.e. the band at 2,100 bp, is undoubtedly present in A28xO11^{R} hybrids, once more it cannot be excluded that the A28xO11^{R} hybrids carry both, A28 and O11 mitochondrial haplotypes (similar to *ND4*, **Fig. 4.33**). C2xA28^{R} hybrids (**Fig. 4.36**) clearly show the presence of paternal A28 signature (2,700bp) while the maternal C2 pattern (2,100 bp) is completely missing.

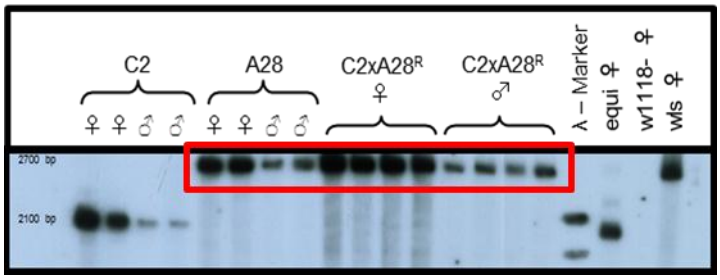


Fig. 4.36: Southern blot of *Psi*I-digested *D. paulistorum* DNA from multifly extraction after Junakovic (2004), hybridized with radioactively labelled ***CO1*** probe. DNA of males (♂) and females (♀) of C2 and A28 individuals plus and C2xA28^{R} hybrids was used. As controls, equally treated DNA of female *D equinoxialis (equi)*, *D. melanogaster* white mutant (w^{1118-}) and *D. willistoni* was loaded. *Hin*dIII-digested λ-phage DNA was used as size marker; radioactive signals derive from pencil marks on the membrane. Red box indicates the paternal signature. X-ray film, 48 hrs of exposure.

5. Discussion

5.1 Informative genes in *D. paulistorum*

5.1.1 Nuclear genes

Most sets of primers designed for *D. willistoni* homologues of nine immune-related genes (six AMPs, three immunoreceptors) provided robust PCR signals of expected size in *D. paulistorum* strains. However, the first PCR fragment of *Hemomucin* could not be detected in any neotropical sample, whereas the PCR fragment of *Drosocin* was missing exclusively in the Orinocan semispecies. The absence of the *Drosocin* signal could be a result of accumulated SNPs in one of the primer-site-binding region, gene deletion or a large insertion in the O11 orthologue. It is, however, an interesting coincidence that the Orinocan semispecies is the only semispecies tolerating high densities of *Wolbachia* (Miller et al. 2010), a fact leaving room for speculation, as the loss of this gene could be promoting a higher abundance of the symbiont. However, other Brazilian strains such as RP and PoA1, that also harbour *Wolbachia* at high titer levels express *Drosocin* properly (D. Schneider, unpublished). The negative results obtained from *Hemomucin* specific PCR on *D. paulistorum* can most likely be explained by a mismatching *forward* primer, as none of the examined strains showed a PCR product, whereas the PCR for the second *Hemomucin* fragment provided a signal with the complementary reverse primer of the first fragment.

The six sequenced AMP genes *Attacin A*, *-C*, *-D*, *Diptericin*, *-B* and *Drosocin* showed elevated rates of nucleotide divergence within the six semispecies (2.4% for *Attacin A* and 7.9% for *Attacin D*), which is consistent with observations made earlier by Lazzaro and Clark (2001) for *Attacins* in different *D. melanogaster* strains. However, within a *D. paulistorum* semispecies, multiple sequence types were found for *Attacin A*, *Attacin D* and *Diptericin B*, potentially deriving from paralogous gene conversions, as the highly inbred lines used for this thesis are free of gene polymorphisms. The finding of multiple paralogues of *Attacins* in *D. melanogaster* (Lazzaro and Clark, 2001) supports this conclusion. The abundance among *D. paulistorum* semispecies and the divergence of those paralogues, however, are varying: *Attacin A* gene shows two paralogues only in C2 (0.4% divergence) while for *Diptericin B,* three paralogues were found in C2 (3% divergence) and two in MS (0.3% divergence) strain. Intriguingly, four of the six examined strains (A28, C2, I1 and SM) showed a second paralogue of *Attacin D,* with a divergence ranging from 0.9 to 3.7%. Those paralogues were confirmed by direct sequencing.

This remarkably high abundance of paralogues of *Attacin D* was further tested by hybridization of Southern blots of *Hin*dIII digested *D. paulistorum* DNA with radioactively labelled *Attacin D* fragments. Two signals could be detected in all samples, proving, under the assumption of complete digestion, two distinct gene loci for *Attacin D* sequences not only for the four strains that showed multiple sequence

types, but for all six *D. paulistorum* strains. Further sequencing might be necessary to detect the second paralogue also in MS and OR strain. The signal at 12,000 bp, present in all *D. paulistorum* strains, was also detected in *D. willistoni* and could hence be considered as ancestral paralogue. The second signal of C2, I1 and SM strain has a length of 2,000 bp, A28, MS and O11 showed a second band at 5,400 bp.

For A28 and C2 strain, however, conflicting results were detected with a second Southern blot, as a third band appeared at 3,500 bp in both sexes of those strains. For C2, however, this signal was very faint, while in the A28 samples it represented the most intense band. Beyond that, male A28 DNA showed a RFLP signal of 6,200 bp, while the band at 12,000 bp was missing in this sample. This sex-specific pattern of A28 samples was undetectable in the first blot as only female DNA was used. The strong signals at 3,500 bp for A28 in the second blot, however, remain obscure. The second blot also showed for *D. willistoni* a signal at 3,600 bp which could not be explained with data of the published sequenced genome, either due to the incomplete shotgun-sequencing or due to a recent duplication in the *D. willstoni* reference strain used for this thesis.

5.1.2 Mitochondrial genes

Sequencing of cloned PCR fragments of mitochondrial *ND4* gene of adult *D. paulistorum* semispecies as well as direct sequencing of the *ND4* PCR products were performed in this thesis using Sanger sequencing. Illumina sequencing of embryonic mitochondrial DNA, performed in parallel by Lisa Klasson, University of Uppsala, Sweden, provided additional sequence data for all strains except SM, including up- and downstream regions. *ND4* gene showed a nucleotide divergence between all semispecies of *D. paulistorum* of 5.1%, in accordance with previous studies on *Drosophila* species, showing that *NADH dehydrogenase* subunits are commonly fast evolving genes (Garesse, 1988; Moriyama and Powell, 1997).

For MS strain, two mitotypes were detected by the presence of consistent double peaks in direct sequencing chromatograms, of which only the more abundant major mitotype (showing higher peaks) was also detected in cloned sequences. Hence, it cannot be ruled out that the direct sequencing reads provided signals for more than two mitotypes, as all aberrant SNPs were assigned to the minor mitotype. The minor mitotype detected with Illumina sequencing, however, did not appear in any of our Sanger sequencing reads. While the Sanger mitotypes of MS strain differ from each other in 1.1% of their nucleotides, Illumina mitotypes only differ in 0.3% of their nucleotides. In addition, a minor mitotype of O11 strain, differing in 2.8% of its nucleotides from the major mitotype was detected by Illumina sequencing and could not be confirmed by any Sanger read.

To verify those mitochondrial haplotypes, Southern blots with *Psi*I digested DNA of *D. paulistorum* semispecies where hybridized with P^{32}-labelled *ND4* and *CO1*

fragments. The first Southern blots for *ND4* showed an unexpected high number of bands. Although the restriction fragments predicted with Illumina reads could be detected for all semispecies (except for the smallest 467 bp fragments, presumably due to negative smile of the DNA on the agarose gel), between one to four additional bands appeared in all tested samples. The length of those fragments correlates with undigested restriction sites, incompletely digested DNA is therefore the most likely explanation for those findings. DNA extraction with the Puregene kit indeed showed a reduction to a single strong band for A28 and C2 strain, supporting this hypothesis. Intriguingly, with this DNA extraction method, the O11 sample did not show the fragment specific for the minor mitotype (Illumina sequencing) at 1,564 bp, in contrast to the other Southern blots, a band at 2,000 bp was detected instead. Presuming that the Puregene DNA extraction was completely digested, this signal most likely represents a mitotype that was not detected either by Sanger or Illumina sequencing. For I1, MS and SM, no Southern blotting of DNA extracted with the Puregene kit was performed, hence, it cannot be ruled out that the additional signals derive from mitotypes lacking *Psi*I restriction sites.

The different mitotypes of the MS strain detected with Sanger or Illumina sequencing could not be confirmed with *Psi*I RFLPs of *ND4*, as they share the same restriction sites. In the Sanger sequences of MS, however, *Sau*3A1 restriction enzyme cuts within the *ND4* sequence of the major mitotype, while the minor mitotype is lacking this restriction site. The resulting fragments of 483 and 510 bp for the major mitotype and 993 for the minor mitotype were detected with DIG-labelled *ND4* probes on Southern blots of*Sau3*A1 digested DNA. Illumina data, on the other hand, predict for both mitotypes fragments of 483 and 510 bp. The O11 minor mitotype, detected solely with Illumina sequences, is also lacking the *Sau*3A1 restriction site within the *ND4* fragment but could not be observed with the DIG labelled probe. Hybridization of Southern blots of *Psi*I digested DNA with P^{32}-labelled *CO1* probe could not solve those conflicts either, as they showed a second signal for O11, at the same height as the fragment of C2 and MS strain, which was not predicted from Illumina sequences (Sanger sequence data was not available for the regions up and downstream of *CO1*, where the restriction sites are located).

In summary, the two mitotypes detected for MS with Sanger sequencing were confirmed with *Sau*3A1 RFLP pattern, while the mitotypes detected with Illumina sequencing for MS and O11 could not be confirmed. Importantly, as Illumina sequencing was only performed with embryonic DNA, whereas Sanger sequences exclusively derive from adult DNA, a possible explanation for those findings are the differences in frequencies of mitotypes in distinct developmental stages. Presumably, such so called “mitochondrial bottlenecks” occur during early embryogenesis (Bergstrom & Pritchard 1998), and can promote a sweep of mitotypes. However, this implies a heteroplasy in the zygote, which is easily conceivable with regard to studies on *D. simulans* (Satta et al., 1988; Matsuura et al., 1991; Dean et al., 2003).

5.1.3 Incongruence between nuclear and mitochondrial phylogenies in the *D. paulistorum* species complex

The phylogenies of *D. paulistorum* semispecies, published by Gleason et al. (1998) and Robe et al. (2010), were mostly consistent at nuclear level: The Amazonian semispecies differentiated first, followed by the separation of two lineages; one comprising Andean-Brazilian and Orinocan semispecies, the other clustering of Transitional, Interior and Centroamerican semispecies. Gleason and colleagues (1998) described in this clade the Centroamerican semispecies as more distantly related to Transitional and Interior semispecies, whereas Robe et al. (2010) found that the Transitional semispecies branches off first, while Centroamerican and Interior represent sister taxa. The concatemeric tree of nuclear AMP, immunoreceptor and *Desaturase* genes built for this thesis, shows a different branching compared to those studies. In its first node, the clade of Andean-Brazilian and Orinocan semispecies separates from a larger cluster, in which Amazonian semispecies branches off early, followed by Centroamerican semispecies and eventually showing Interior and Transitional semispecies as sister taxa. Since previous studies exclusively focused on housekeeping genes (*adh, amd, hb, per*), and since this study examined mainly immune-related genes, which are not constitutively expressed, the deviating clustering of Amazonian semispecies could be explained with a divergent evolution of those types of genes.

Mitochondrial data of Gleason et al (1998) and Robe et al. (2010) provided a very conflicting phylogeny with a low bootstrap support. The phylogeny established with *ND4* mitochondrial gene in this study provided in total an improved resolution, but could not confirm one of the previous studies. Gleason et al. (1998) showed a clustering of Andean-Brazilian, Orinocan and Transitional semispecies with Centroamerican and Interior in the same clade and Amazonian as distantly related semispecies. Robe et al. (2010), however, integrated Amazonian, Andean-Brazilian, Interior and Orinocan semispecies in a first and Transitional and Centroamerican semispecies in a second clade. The present study groups Amazonian, Interior, Orinocan and Transitional semispecies together, while Andean-Brazilian and Centroamerican semispecies form a more distinct clade. The minor mitotypes of MS found with Sanger and O11 found with Illumina sequencing could not be integrated in the phylogeny due to insufficient bootstrap values. The finding of multiple mitotypes in this species complex and the observed paternal transmission of mitochondria could serve as an explanation for those very different phylogenies found in the respective studies.

5.2 Marker system for strain typing of *D. paulistorum* semispecies and newly collected lines

In order to classify newly collected and yet unclassified mutant *D. paulistorum* strains, sequence data obtained for this thesis was used to find a minimal set of diagnostic SNPs. *Attacin C* and *Attacin D* were selected to create an improved nuclear marker system, *ND4* data was applied for classification of mitochondria. Six strains of South Brazil (JB, MLC, MTS, PoA1, PoA10, RP), six strains recently collected in mid- and Mideast Brazil (Ama, Bra, NEG-DED, TRI-POU, SET-RIA, ZOO1) and three mutant strains from Honduras (prune, white, yellow), all kept as isofemale lines, were analysed.

5.2.1 *Attacin C*

Attacin C, one of two AMPs without multiple paralogues, was used as marker gene because of its higher number of informative SNPs compared to *Diptericin*, the second AMP without paralogues*.* Although *Attacin C* sequence showed only a single diagnostic SNP (243:T) for Orinocan semispecies, a combination of two to five SNPs allowed the classification of Amazonian (82:G, 325:C) Andean-Brazilian (82:A, 243:A, 286:T, 300:A, 325:C) and Centroamerican (82:A, 325:T) semispecies. Interior and Transitional were identical for *Attacin C* (82:G, 325:T) and could therefore not be distinguished. From the unclassified *D. paulistorum* lines, *Attacin C* sequence was only available for the South Brazilian and mutant strains. JB, PoA1 and the prune and yellow mutant strain showed identical *Attacin C* sequences as Andean-Brazilian semispecies, while *Attacin C* of RP and MTS was identical to the sequence of Orinocan semispecies. MLC could be identified as Amazonian semispecies, while the white mutant had the same *Attacin C* sequence as Interior and Transitional semispecies, which group very closely to Centroamerican semispecies. Although highly homologous to Andean-Brazilian semispecies, PoA10 showed a unique SNP in position 286, with cytosine instead of thymine.

5.2.2 *Attacin D*

The high number of diagnostic SNPs within the six semispecies made *Attacin D* a prime candidate for a strain typing marker gene of *D. willistoni* group species. As described previously for *Attacin C,* Interior and Transitional semispecies showed identical sequences for both of their paralogues and hence could not be distinguished. The other four semispecies, however, could be differentiated. A complete set of unclassified strains from South and Mideast Brazil, as well as mutant strains were tested for their SNPs in *Attacin D.* However, the finding of multiple paralogues of this gene in nine of the strains and a generally high diversity made it quite complicated to apply *AttD* as a reliable marker system. Many of the diagnostic SNPs for the six semispecies also appeared in strains with an otherwise very

divergent sequence. For some topologies, the genetic information was not sufficient to create a tree with a higher bootstrap value threshold than 30%.

Nevertheless, some of the sequences could be assigned clearly to a defined semispecies group: *Attacin D* of the white mutant strain was identical to the first paralogue of Centroamerican semispecies, which is consistent with *Attacin C* data. The second paralogue of the C2 strain might be undetected in the white mutant, or the mutant branched off before the second paralogue emerged. This mutant strain, however, can be assigned to the Centroamerican semispecies, both by an adenine in position 134 as diagnostic SNP.

Attacin D sequence of prune and yellow mutant strain and one paralogue of SET-RIA were identical to Andean-Brazilian strain. Other sequences of unclassified strains showed a high homology to this semispecies: TRI-POU and the first paralogue of Bra both have a 99.8% identical sequence, the second paralogues of Bra and SET-RIA have an identity of 99%. On *Attacin D* level, those strains could therefore be classified as Andean-Brazilian and show a diagnostic SNP in position 364 (guanine instead of adenine or thymine). Bra and SET-RIA, however, show a recent duplication of this gene. For yellow and prune mutant strain, this classification is supported by *Attacin C* data.

The two mitotypes of MLC are in 98.8% of their *Attacin D* sequence identical to the first paralogue of Amazonian strain A28. The two paralogues of Ama show 99.3 and 99.5% identity to the two corresponding A28 paralogues. This finding suggests that MLC is very closely related to the A28 strain, but branched off before the second A28 paralogue emerged. Instead, a recent duplication in MLC created a second paralogue, which has acquired only very few SNPs compared to the first. The Ama strain, however, has for both of the A28 paralogues a corresponding counterpart of high identity. On *Attacin D* level, MLC and Ama strain can therefore be considered as members of the Amazonian semispecies, and can be identified with a cytosine in position 181. *Attacin C* data of MLC supports this finding.

The strains JB, MTS, PoA1, RP and ZOO1 could not be assigned unambiguously to distinct semispecies with *Attacin D* sequence. The second mitotype of MTS groups closely to Orinocan semispecies, a finding which is consistent with *Attacin C* data, although the first paralogue of MTS groups in closer proximity to Andean-Brazilian semispecies. PoA10 also clusters with Orinocan semispecies, while JB, PoA1, RP and ZOO1 group with Andean-Brazilian strain. Although *Attacin C* data supports this finding for JB and PoA1, the grouping of RP and the second mitotype of MTS differs for this gene. The bootstrap values for the grouping of those strains on *Attacin D* level are, however, too low to draw conclusions about their relationships to the described semispecies.

Southern blot analysis of *Attacin D* with DNA of South and Mideast Brazilian strains did not help to determine their affiliation to a semispecies, as only two patterns were observed within the six semispecies. Most of the examined strains showed the same

pattern as Amazonian, Andean-Brazilian and Orinocan semispecies. For JB, MTS and PoA1, additional signals were detected which could not be correlated with any other strain. They could, however, indicate additional undetected paralogues.

5.2.3 *ND4*

ND4 sequence of MLC and Ama showed in accordance with nuclear data very high identity (99.8%) to Amazonian *ND4*. As diagnostic site for this semispecies, cytosine in position 310 was determined. *ND4* sequence of white mutant strain showed, corresponding with *Attacin C* and *–D* data, a very high identity to Centroamerican semispecies. Adenine in position 280 can be used as diagnostic marker for white mutant strain in *ND4*. Many unclassified strains, however, revealed a very different grouping at mitochondrial level, compared to the nuclear genes described above. For the Bra strain and Orinocan major mitotype, 99.8% identity was observed on *ND4* level. A thymine in position 319 can be used as diagnostic SNP. Bra grouping with Andean-Brazilian semispecies in *Attacin D* contradicts this finding on mitochondrial level. For RP strain, the less frequent mitotype detected with Illumina sequencing grouped with Andean Brazilian semispecies, while the more abundant *ND4* sequence showed high identity to Interior and Transitional Semispecies. The same is true for *ND4* sequence of South Brazilian strains JB, MTS, PoA1 and PoA10, as they are identical to the RP major mitotype. *ND4* sequence of ZOO1 strain also groups in close proximity to Interior and Transitional. TRI-POU, SET-RIA and prune *ND4* sequences are identical with those semispecies.

Southern blots with P^{32}-labelled *ND4* fragments supported those results, the pattern of Interior and Transitional was also detected in all Brazilian strains except Ama and MLC strain, which showed the same patterns as Amazonian semispecies. The finding of Transitional/Interior-like mitochondrial sequences in all Brazilian strains (except Ama and MLC) is surprising, as nuclear data and geographic distribution suggest a closer relationship to Andean-Brazilian semispecies. Considering the minor RP mitotype, which groups close to Andean-Brazilian semispecies, however, it would appear plausible that the *ND4* sequences detected for the Brazilian strains only represent the most abundant mitotype, while minor mitotypes remain undetected. This would assume a mitochondrial introgression from Interior/Transitional to Andean Brazilian semispecies, which has spread all over Brazil within the last 60 years since the collection of the MS strain. To examine *Wolbachia* infections in those strains and compare them to Transitional/Interior infections could shed light on this question. If the bacteria are assumed to hitchhike, they should provide a similar phylogeny as mitochondria.

Table 5.1 summarizes the strain typing established with nuclear and mitochondrial sequence data as well as Southern blot data for *ND4*. Results obtained with *Attacin D* Southern blots are not included in this table, as they did not provide enough information for strain typing.

Unclassified strains	*Attacin C*	*AttacinD*	*ND4*	
	DNA sequence	DNA sequence	DNA sequence	Southern blot
Ama	-	AM	AM	AM
Bra	-	AB	OR	IN/TR
JB	AB	?	IN/TR	IN/TR
MLC	AM	AM	AM	AM
MTS	OR	?	IN/TR	IN/TR
PoA1	AB	?	IN/TR (AB)	IN/TR
PoA10	AB	?	IN/TR	IN/TR
RP	OR	OR	IN/TR	IN/TR
SET-RIA	-	AB	IN/TR	IN/TR
TRI-POU	-	AB	IN/TR	IN/TR
ZOO-1	-	?	IN/TR	IN/TR
Prune mutant	AB	AB	IN/TR	-
White mutant	IN/TR/CA	CA	CA	-
Yellow mutant	AB	AB	?	-

Table 5.1: Summary of the strain typing results. Abbreviations for described semispecies: AB = Andean Brazilian, AM = Amazonian, CA = Centroamerican, IN = Interior, OR = Orinocan, TR = Transitional. Sequences that did not allow unambiguous grouping of the corresponding strain into a distinct *D. paulistorum* semispecies are indicated with a question mark, hyphen represents missing data. Strains with consistent nuclear and mitochondrial data (Ama, MLC, white mutant) are highlighted in yellow.

5.3 Evolutionary pressure on *Attacin D*

The ratio of the number of non-synonymous substitution per non-synonymous site to the number of synonymous substitutions per synonymous site (K_a/K_s) can be used as indicator of selective pressure acting on a protein-coding gene (Miyata and Yasunaga, 1980). The coding region of *Attacin D* of all *D. paulistorum* strains used for this thesis was examined for traits of selective pressure. As *Attacin D* proteins are the only AMP lacking their characteristic signal- and propeptide domains and thus probably acting intracellularly (Hedengren, 2000), they are prime candidates for possible coevolutionary interactions with intracellular, obligate mutualistic *Wolbachia.* If direct interactions with the symbiont are assumed, *Attacin D* should be subjected to positive selection in terms of an “arms race” scenario (Dawkins and Krebs, 1979), as it has to coadapt to the fast evolving bacteria.

K_a/K_s ratio of all three protein domains of the *Attacin D* gene showed signs of negative instead of positive selection, although in every domain of the protein regions

with high density of replacing mutations were found. The occurrence and maintenance of multiple divergent paralogues for a majority of the examined strains, however, can be interpreted as an adaptation to a divergent set of symbionts. Those findings are in accordance with previous studies of Lazzaro and Clark (2003), suggesting that the arms race model does not accurately describe the evolutionary dynamics of AMPs in *D. melanogaster*, but the high frequency of paralogue polymorphisms and the linkage disequilibrium of AMP genes still suggest positive selection.

5.4 Mitochondrial transmission in *D. paulistorum* hybrids

Paternal leakage of mitochondria has been considered as the most plausible explanation for the emergence of heteroplasy, that were observed in crossing experiments of *D. simulans* strains with distinct mitochondrial haplotypes (Wolff et al., 2012). The hitchhiking of *Wolbachia* with mitochondria is known since the early 1990s (Nigro and Prout, 1990; Turelli and Hoffmann, 1991). As paternal transmission has been observed recently in *D. paulistorum* hybrids of Amazonian and Orinocan semispecies (AMxOR^{R}, Miller and Schneider, unpublished), the question whether or not also paternal leakage of associated mitochondria occurred was adressed in this work.

Not only AMxOR^{R} hybrids were examined, but also hybrids of Centroamerican and Amazonian semispecies (CAxAM^{R}) were included in the experiments. Southern blots with P^{32}-labelled *ND4* and *CO1* fragments were performed on *Psi*I-digested DNA of both sexes of the maternal, paternal and hybrid strains. Unexpectedly, the presence of paternal signals could be confirmed with both genes beyond all doubt in both hybrid types. For AMxOR^{R} hybrids, the paternal signal was not suitable to exclude the presence of maternal DNA; maternal signals were only missing the distinctly paternal bands. Therefore, it could not be excluded that AMxOR^{R} hybrids still harbour both, paternal and maternal mitotypes. Further experiments are needed to find restriction enzymes or hybridization fragments that provide suitable RFLPs for exclusion of maternal DNA. The CAxAM^{R} hybrids, however, did not show any traces of former maternal mitotypes, hence, it can be assumed that they are in a state of paternal homoplasy.

This finding of paternal mitochondrial transmission supports the evidence for heteroplasy found in several strains, *i.e.*, in newly collected as well as in the “classic” semispecies. These data show that this species complex in *statu nascendi* is able to maintain other than maternal mitotypes over several generations, and even can replace former maternal mitotypes completely with the paternally inherited. The challenge for further studies will be to determine the “mitochondrial bottleneck”, i.e., at what hybrid generation the sweep of the mitotypes occurs. Furthermore, the finding of heteroplasy and paternal mitochondrial leakage has to be included in the establishment of a *D. paulistorum* phylogeny. The concept of six semispecies, as it

was introduced in the 1950´s, might be outdated. Large scale next generation whole genome analysis with numerous newly collected samples from different geographic locations might answer the question on how this species complex might have evolved in the past 60 years, and on how their *Wolbachia* endosymbionts and associated mitochondria are triggering incipient speciation under laboratory conditions and in nature.

6. References

Åsling B. et al. (1995) Identification of early genes in the Drosophila immune response by PCR-based differential display: the *Attacin A* gene and the evolution of *Attacin*-like proteins. Insect Biochem. Mol. Biol. **25**:511–518

Ayala F.J., Tracey M.L. (1973) Genetic differentiation and reproductive isolation between two subspecies of *Drosophila willistoni*. J. Hered. **64**:120–124

Ayala F.J. et al. (1974) Genetic differentiation during the speciation process in *Drosophila.* Evolution **28**:576–592

Bandi C. et al. (1998) Phylogeny of *Wolbachia* in filarial nematodes. Proc. Biol. Sci. **265:**2407-2413

Bates D., Ehrman L., Perelle I. (1977) Intersemispecific hybridity and egg hatchability in *D. paulistorum*. Dros. Inf. Serv. **52**: 94

Bergstrom C.T., Pritchard J. (1998) Germline bottlenecks and the evolutionary maintenance of mitochondrial genomes. Genetics **149**:2135–2146

Binder A. (2001) β2 Adrenergic Receptor Genetics and its Role in Human Hypertension, http://www.uni-graz.at/~binder/science/thesis/start_htmlsu46.html

Birky C.W. (2001) The inheritance of genes in mitochondria and chloroplasts: laws, mechanisms and models. Ann. Rev. of Gen. **35**:125–148

Bouchon D. et al. (1998) Evidence for widespread *Wolbachia* infection in isopod crustaceans: molecular identification and host feminization. Proc. R. Soc. Lond. B **265**:1081-1090

Boutros M. et al. (2002) Sequential Activation of Signalling Pathways during Innate Immune Responses in *Drosophila*. Devel. Cell **3**:711–722

Burla H. et al. (1949) The *willistoni* group of sibling species of *Drosophila*. Evolution **3**:300–314

Breton S. et al. (2007) The unusual system of doubly uniparental inheritance of mtDNA: isn´t one enough? Trends in Genetics **23:**465-474

Brown W.M., et al. (1979) Rapid evolution of animal mitochondrial DNA. Proc. Natl. Acad. Sci. USA **76:** 1967–1971

Dawkins R., Krebs J. R. (1979) Arms races between and within species. Proc. R. Soc. Lond. B Biol. Sci. **205**:489–511

Daniels S., Ehrman L. (1974) Embryonic pole cells and mycoplasma like symbionts in *Drosophila paulistorum*. J. Invertebr. Pathol. **24**: 14–19

Dean M.D. et al. (2003) Influence of two Wolbachia strains on population structure of East African Drosophila simulans. Genetics **165**: 1959-1969

DeLuca S.Z., O'Farrell P.H. (2012) Barriers to male transmission of mitochondrial DNA in sperm development. Dev. Cell. **22:**660–668

Dobzhansky T., Ehrman L., Pavlovsky O. (1957) *Drosophila insularis*, a new sibling species of the *willistoni* group. Univ. Texas Publ. **5721**:39–47

Dobzhansky T., Spassky B. (1959) *Drosophila paulistorum*, a cluster of species in statu nascendi. Proc. Natl. Acad. Sci. U S A **45**:419–428

Dobzhansky T., Pavlovsky O. (1971) Experimentally created incipient species of Drosophila. Nature **230:**289–292

Drosophila 12 Genomes Consortium (2007) Evolution of genes and genomes on the *Drosophila* phylogeny. Nature **450**:203-218

Ehrman L. (1962) The transitional races of *Drosophila paulistorum*: a study of hybrid sterility. Proc. Natl. Acad. Sci. U S A. **48**:157-159

Ehrman L. (1968) Antibiotics and infections hybrid sterility in *Drosophila paulistorum*. Mol. Gen. Genet. **103:**218–222

Ehrman L., Kernaghan, R.P. (1972) Infectious heredity in *Drosophila paulistorum*. Ciba Foundation Symposium: Pathogenic Mycoplasmas. Amsterdam: Assoc. Scientific Publishers. pp 227–250

Ferrandon D. et al. (1998) A drosomycin-GFP reporter transgene reveals a local immune response in *Drosophila* that is not dependent on the toll pathway. EMBO J. **17**:1217-1227

Garcia A.C.L. et al. (2006) Identification of the sibling species of the *Drosophila willistoni* subgroup through the electrophoretical mobility of acid phosphatase-1. J. of Zool. Syst. and Evol. Res. **44**:212-216

Garesse R. (1988). *Drosophila melanogaster* mitochondrial DNA: Gene organization and evolutionary considerations. Genetics **118:**649–663

Gleason J.M. et al. (1998) A Molecular Phylogeny of the *Drosophila willistoni* Group: Conflicts Between Species concepts? Evolution **52**:1093-1103

Gyllenstein U. et al. (1991) Paternal Inheritance of mitochondrial DNA in mice. Nature **352:**255-257

Hedengren M. et al. (2000) Expression and Evolution of the *Drosophila Attacin/Diptericin* Gene Family. Biochem. and Biophys. Research Communications **279:**514-581

Hertig M., Wolbach S.B. (1924) Studies on *Rickettsia*-like microorganisms in insects. Journal of Medical Research **44**: 329-74

Hurst, G.D.D. et al. (1999) Male-killing *Wolbachia* in two species of insect. Proc. R. Soc. Lond. B **266:**735-740

Imler J.L., Bulet P. (2005) Antimicrobial peptides in *Drosophila*: structures, activities and gene regulation. Chem. Immunol. Allergy **86**:1–21

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

Jones S.L., et al. (1999) Phagocytosis. Fundamental Immunology, 4th edn. Paul, W.E. (ed). Philadelphia: Lippincott-Raven Publishers p. 997–1020

Junakovich N. (2004) Southern Blotting analysis of individual *Drosophila* flies. Methods Mol. Biol. **260**:41-57

Kastritsis C.D., Dobzhansky T. (1966) *Drosophila pavlovskiana*, a race or a species? The Am. Midl. Nat. **78**:244–247

Kernaghan R.P., Ehrman L. (1970) Antimycoplasmal antibiotics and hybrid sterility in *Drosophila paulistorum*. Science **169:**63–64.

Kondo R. et al. (1992) Further observation of paternal transmission of *Drosophila* mitochondrial DNA by PCR selective amplification method. Genet. Res., Camb. **59:**81-84

Kvist L. et al. (2003) Paternal leakage of mitochondrial DNA in the great tit (*Parus major*). Mol. Biol. Evol. **20:**243-247

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

Lazzaro B.P., Clark A.G. (2001) Evidence for Recurrent Paralogous Gene Conversion and Exceptional Allelic Divergence in the *Attacin* Genes of *Drosophila melanogaster.* Genetics **159**:659-671

Lazzaro B.P., Clark A.G. (2003) Molecular Population Genetics of Inducible Antibacterial Peptide Genes in *Drosophila melanogaster.* Mol. Biol. Evol. **20**:914-923

Lazzaro B.P (2008) Natural selection on the *Drosophila* antimicrobial immune system. Curr. Opin. Microbiol. **11**:284-289

Li W.H., Wu C.I. and Luo C.C. (1985) A new method for estimating synonymous and nonsynonymous rates of nucleotide substitution considering the relative likelihood of nucleotide and codon changes. Mol Biol Evol **2**:150-174

Luo C., Zheng L. (2000) Independent evolution of Toll and related genes in insects and mammals. Immunogenetics **51**:92–98

Luo C. et al. (2001) Tehao functions in the Toll pathway in *Drosophila melanogaster.* Possible roles in development and innate immunity. Insect Mol. Biol. **10**:457-464

Miyata T., Yasunaga T. (1980) Molecular evolution of mRNA: a method for estimating evolutionary rates of synonymous and amino acid substitutions from homologous nucleotide sequences and its applications. J. Mol. Evol., 16: 23–36

Matsuura E.T. (1991) Mitochondrial DNA heteroplasmy maintained in natural populations of *Drosophila simulans* in Reunion. Genet. Res. **57:**123–126

Mayr E. (1970) Populations, species and evolution: an abridgment of animal species and Evolution, p. 287. Harvard University Press,Cambridge.

Miller J.S. et al. (1999) Eicosanoids mediate nodulation reactions to bacterial infections in adults of the cricket, *Gryllus assimilis*. J. Insect Physiol. **45**: 75-83

Miller W.J., Ehrman, L., Schneider, D. (2010) Infectious Speciation Revisited: Impact of Symbiont-Depletion on Female Fitness and Mating Behaviour of *Drosophila paulistorum*. PLoS Pathog. **6**: e1001214

Moriyama E.N., Powell J.R. (1997). Synonymous substitution rates in *Drosophila*: Mitochondrial versus nuclear genes. J. Mol. Evol. **45**: 378–391

Nappi A.J. et al. (1995) Superoxide anion generation in Drosophila during melanotic encapsulation of parasites. Eur. J. Cell Biol. **68**:450-456

Nappi A.J. et al. (2004) Signalling pathways implicated in the cellular innate immune responses of Drosophila. Inv. Surv. J. **1:**5-33

Nigro L. and Prout T. (1990) Is there selection on RFLP differences in mitochondrial DNA?. Genetics **125**:551–555

Pearson A. et al (1995) Expression cloning of dSR-CI, a class C macrophage-specific scavenger receptor from *Drosophila melanogaster*. Proc. Natl. Acad. Sci. USA **92**:4056-4060

Perez-Salas S. et al. (1970) The interior semispecies of *Drosophila paulistorum*. Evolution **24**:519–527

Powell J.R. (1997) Progress and Prospects in Evolutionary Biology: The *Drosophila* Model. Oxford University Press, Inc. New York BIOSIS ID: 1102964

Raychoudhury R. et al. (2011) Modes of acquisition of *Wolbachia*: Horizontal transfer, hybrid introgression and codivergence in the *Nasonia* species complex. Evolution **63:**165-183

Rigaud T. et al. (1991). Feminizing endocytobiosis in the terrestrial crustacean *Armadillidium vulgare* Latr. (Isopoda): recent acquisitions. Endocyt. Cell Res. **7:**259–273

Robe L.J. et al. (2010) Taxonomic boundaries, phylogenetic relationships and biogeography of the *Drosophila willistoni* subgroup (Diptera: Drosophilidae). Genetica **138:**601-607

Satta Y.N. et al. (1988) Dubious maternal inheritance of mitochondrial DNA in *D. simulans* and evolution of *D. mauritiana.* Genet. Res. **52**:1–6

Spassky B. (1957) Morphological differences between sibling species of *Drosophila*. Univ. Texas Publ. **5721**:48–61

Spassky B. et al. (1971) Geography of the sibling species related to Drosophila willistoni, and of the semispecies of the Drosophila paulistorum complex. Evolution **25**:129–143

Stouthamer R. et al. (1990) Antibiotics cause parthenogenetic Trichogramma (Hymenoptera/Trichogrammatidae) to revert to sex. Proc. Natl Acad. Sci. USA **87:**2424–2427

Sutovsky P. et al. (2000) Ubiquitinated sperm mitochondria, selective proteolysis, and the regulation of mitochondrial inheritance in mammalian embryos. Biol Reprod **63:**582–590

Schlenke T.A., Begun D.J. (2004) Linkage Disequilibrium and Recent Selection at Three Immunity Receptor Loci in *Drosophila simulans.* Genetics **169:**2013-2022

Theopold U. et al. (1996) Helix pomatia Lectin, an Inducer of *Drosophila* Immune Response, Binds to Hemomucin, a Novel Surface Mucin. J. of Biol. Chem. **271**:12708-12715

Townsend J.L. (1954) Cryptic subspeciation in *Drosophila* belonging to the subgenus *Sophophora*. Am. Nat. **88**:339–351

Turelli M. and Hoffmann A.A. (1991) Rapid spread of an inherited incompatibility factor in California *Drosophila*. Nature **353**:440–442

Werren JH et al. (2008) *Wolbachia* - master manipulators of invertebrate biology. Nat. Rev. Microbiol. **6**:741-51

Williamson D.L., Ehrman L., Kernaghan R.P. (1971) Induction of sterility in *Drosophila paulistorum*: effect of cytoplasmic factors. Proc. Natl. Acad. Sci. U.S.A. **68**:2158–2160

Wolff J.N. et al. (2012) Paternal transmission of mitochondrial DNA as an integral part of mitochondrial inheritance in metapopulations of *Drosophila simulans.* Heredity 1-6 (Epub ahead of print)

Yen J.H., Barr A.R. (1971) New Hypothesis of the Cause of Cytoplasmic Incompatibility in *Culex pipiens.* L. Nature **232**:657 - 658

Zasloff M. (2002) Antimicrobial peptides of multicellular organisms. Nature **415**:389-395

7. *Curriculum vitae*

Wolfgang Däuble
Stolberggasse 10/21
1050 Vienna, Austria

Mobile: 0043 650 932 02 39
Email: wolfgang.daeuble@gmail.com

Date of birth:	14.09.1983
Place of birth:	Karlsruhe, Germany
Citizenship:	Austrian/German
Civilian Service:	accomplished 2002-2003

Education

2011-2012	**Diploma thesis:** '*Drosophila paulistorum* trinity: Strain typing, immunity, mitochondrial transmission' Wolfgang Miller Group, Centre ofAnatomy and Cellbiology, Medical University of Vienna, Austria
Since 2003	Student of Molecular Biology, University of Vienna, Austria
1996 – 2002	Highschool: Friedrich Ebert Gymnasium, Bonn, Germany
1993 – 1996	Highschool: Schola Europaea, Brussels, Belgium
1990 – 1993	Elementary school: GGS-Gothenschule, Bonn, Germany
1989 – 1990	Elementary school: Collegio Humboldt, Caracas, Venezuela

Practical training and conference contributions

September 2011	Poster presentation EMBO 2011, Vienna, Austria: "Friend or foe? Insect-directed strategies for differentiating between microbial pathogens and beneficial mutualists".

June 2011	Oral presentation 6th Summer Conference of Center of Anatomy and Cell Biology, Medical University of Vienna, Austria
April 2011	COST (European Cooperation in Science and Technologie) training in Bioinformatics, Uppsala, Sweden
February - March & July – August 2007	Practical Training Novartis Institutes for Biomedical Research, Department Innovative Screening Technologies, Vienna, Austria, group Claudia-Nicole Meissner

Additional

Internship in November 2012 at the science department of the Austrian Broadcast Cooperation (Radio). Since then working as a freelance journalist on science topics.

Member of the Symphonic Choir of the University of Vienna, Austria, baritone

Member of the Symphonic Orchestra of the University of Vienna, Austria, cellist

Founding member of the Voice Club choir of the University of Vienna and the Molecular Biology Orchestra, Vienna, Austria

Part-time member of the Cantus Novus choir, Archdiozese Vienna and the Ensemble Coinonia, Vienna, Austria