A Theoretical Model on the Role of Lateral Gene Transfer in the Evolution of Endosymbiotic Genomes

D I S S E R T A T I O N

zur Erlangung des akademischen Grades

Doctor Rerum Naturalium
im Fach Biologie

eingereicht an der
Mathematisch-Naturwissenschaftlichen Fakultät I
Humboldt-Universität zu Berlin

von
Dipl. Biol. Víctor Hugo Anaya Muñoz , M.Sc.
geboren am 03.06.1977 in Mexiko Stadt, Mexiko

Präsident der Humboldt-Universität zu Berlin:
Prof. Dr. Jan-Hendrik Olbertz

Dekan der Mathematisch-Naturwissenschaftlichen Fakultät I:
Prof. Dr. Andreas Herrmann

Gutachter:
1. Prof. Dr. Peter Hammerstein
2. Prof. Dr. Hans-Jörg Rheinberger
3. Prof. Dr. Daniel Piñero Dalmau

eingereicht am: 28. April 2010
Tag der mündlichen Prüfung: 25.Oktober 2010

Abstract

Lateral gene transfer has played a key role in the evolution of living beings. This process was first acknowledged in (1978) by Schwartz and Dayhoff but considered a relatively infrequent eccentricity and ignored. Later on, as DNA and protein sequences accumulated and more refined phylogenies were reconstructed, the contribution of lateral (or horizontal) gene transfer to the evolutionary history of living organisms gained relevance. Besides, gene transfer is known to occur not only between independent organisms but also, and more frequently between endosymbionts including eukaryotic organelles. I developed a theoretical model to study the lateral gene transfer process between cell organelles (but extendible to other endosymbionts) and the cell nucleus. The model explores the role of the lack of recombination in the organelles (Muller's ratchet) as well as deviations from Muller's ratchet in the form of non-symmetrical homologous recombination in relation with the gene transfer process. Also, nuclear incompatibilities resulting from the inclusion of a transferred gene, and cyto-nuclear incompatibilities between the mutant endosymbiotic genomes and the modified nuclear genome are investigated. The results obtained show that under certain circumstances the existence recombination or its non-existence produce the same results, and that deviations from symmetry in the recombination process might have important effects on the frequency of different alleles. It is also clear that there is a strong relation between nuclear and endosymbiotic genomes, and that the evolutionary fate of one largely depends on the forces affecting the other. When nuclear and cyto-nuclear incompatibilities are introduced in the model, the results show that lateral gene transfer-induced incompatibilities could potentially play a role in the speciation process similar to the one produced by mitochondria in the *Nasonia* species.

Zusammenfassung

Laterale Gentransfer wurde zuerst von Schwartz und Dayhoff (1978) entdeckt, die es aber als eine Exzentrizität werteten und als solche ignorierten. Später, als mehrere DNS- und Eiweißsequenzen sequenziert und raffiniertere Phylogenien rekonstruiert wurden, hat die Rolle an Relevanz gewonnen, die der laterale (oder horizontale) Gentransfer in der evolutionären Geschichte von lebendigen Organismen gespielt hat. Außerdem existiert auch zwischen Endosymbionten und Zellkernen statt. Ich habe ein theoretisches Modell entwickelt, das den lateralen Gentransfer zwischen Endosymbionten und dem Zellkern repräsentiert. Das Modell erforscht die Bedeutung des Fehlens von Rekombination in den Organellen (Muller's Ratchet) sowie Abweichungen von Muller's Ratchet in Form der *non-symmetrical homologous recombination* in Gentransfermechanismen. Ich habe zum einen Zellkern-Inkompatibilitäten, die aus der Übertragung eines Gens resultieren, und zum anderen Zyto- und Zellkern-Inkompatibilitäten zwischen den mutierten endosymbiotischen Genomen und dem modifizierten Zellenkern untersucht. Die Ergebnisse zeigen, dass unter bestimmten Bedingungen die Existenz oder Nicht-Existenz von Rekombination die gleiche Wirkung haben können. Es zeigte sich auch, dass Rekombination, wenn sie vorkommt und wenn sie nicht symmetrisch ist, starke Auswirkungen auf die Allelenfrequenz einer Population haben kann. Es wurde auch klar, dass es eine starke Beziehung zwischen dem Zellkern und endosymbiotischen Genomen gibt, und dass das evolutionäre Schicksal des einen größtenteils von den evolutionären Kräften abhängig ist, die das andere beeinflussen. Wenn man Zellkern- und Cyto-Zellkerninkompatibilitäten in das Modell einführt, dann zeigen die Ergebnisse, dass die Inkompatibilitäten, die der laterale Gentransfer produziert hat, möglicherweise eine ähnliche Rolle im Speziationsmechanismus spielen könnten wie die Inkompatibilitäten zwischen Mitochondrien und Zellkernen in verschiedenen *Nasonia*-Arten.

Contents

1. Overview

Lateral gene transfer refers to the transmission of genetic information between organisms that are not related by descent and that could even belong to different species, families, or reigns. A particular case of lateral gene transfer that has had profound influence in the constitution of the nuclear genomes of eukaryotes and that of their endosymbionts.

The present work explores lateral gene transfer form several perspectives. It presents a historical reconstruction and analysis of the way in which lateral gene transfer has been studied; from the description of the "transforming principle" in *Pneumococcus*, until the recognition of its evolutionary relevance as a source of genetic novelty and the shadows of doubt it casted on the possibility of reconstructing a Universal Tree of Life.

As many evolutionary relevant phenomena lateral gene transfer has been studied from the theoretical point of view. Population genetics models have specially focussed on the effects of endosymbiotic gene transfer in the conformation of the endosymbiotic genome. Using experimental approaches it has been possible to asses the feasibility of successfully transfer and activate a gene from an organelle into the nuclear genome and measure the transfer frequency.

Organellar genomes, on the other hand, are characterized by an extreme reduction their genomes, it is assumed that endosymbiotic gene transfer has played an important role during the reduction process, but it has also been proposed that Muller's ratchet (the accumulation of deleterious mutations due the lack of recombination in asexual organisms) might be responsible of the genomic reduction in the modern organelles and other cellular endosymbionts .

The present work analyses from the population genetics perspective how endosymbiotic gene transfer in the context of Muller's ratchet and other recombination scenarios in the organelle can affect the fixation of deleterious mutations in the organellar genomes and transferred genes can invade the nuclear genome at a population level. The results presented here support the idea that recombination, specially in the form or gene transformation are important to create heteroplasmic scenarios or induce the fixation of deleterious or mildly deleterious mutations in the organellar population under certain nuclear conditions.

2. Introduction

2.1. Lateral gene transfer

Lateral gene transfer is the process of transmission of genetic information between organisms that belong to different, sometimes very distant taxa. It constitutes one of the mechanisms by which genetic novelty may be obtained in prokaryotic and eukaryotic organisms; it could also be responsible for important evolutionary novelties in non-directly related branches of the so-called tree-of-life.

Bacterial transformation is one of the clearest and best understood examples of lateral gene transfer. It was first described by Fred Griffith in 1928 while he was working on the diversity of Pneumococcus types in pneumonia patients, and he noticed the transition between the *smooth* (S) and *rough* (R) colonial types (cited in Brock, 1990). The physicochemical nature of the process driving bacterial transformation was not entirely clear at the time but it was quickly acknowledged as medically and genetically relevant. It was until the work of Avery, Macleod, and McCarty (1944) on the "transforming principle" that the role of DNAas carrier of the genetic information was clear.

The evolutionary and phylogenetic importance of gene transfer was first addressed by Schwartz and Dayhoff (1978) though it was quickly dismissed as an irrelevant phenomenon; lately as gene sequences started to accumulate and the "purity" of genomes became more dubious, Doolittle (1998, 1999b,a) proposed that Lateral Gene Transfer could be an important process that may account for the current conformation of the genomes and which, in the most extreme case, may actually preclude us from reconstructing a "Universal Tree of Life". This is, of course, not undebated and not all the parties involved in the discussion agreed on the importance of lateral gene transfer or on its evolutionary long term effects[1].

The great amount of genomic information that began to accumulate during the decades of the 1980s and 1990s, made it possible to find an increasing number of examples of transferred genes either individually or by group even between different kingdoms. There are however, several difficulties to positively detect (O'Malley and Boucher, 2005; Suárez-Díaz and Anaya-Muñoz, 2008). Lateral gene transfer can be safely assumed if a gene shows a lower similarity with an ortholog in a closely related organism than with a probable homolog from an organism in a distant taxon, producing (usually) unexpected tree topologies; another signature of lateral gene transfer can be detected when gene syntheny is conserved between distant lines (Koonin et al., 2002, 2001; Wolf et al., 2002).

[1] Several authors have discussed this issue, see for example: Koonin et al., 2002; Kurland, 2005; O'Malley and Boucher, 2005, Suárez-Díaz and Anaya-Muñoz, 2008, Section 3.6, and in Spanish Suárez-Díaz and Anaya-Muñoz, 2009, Section 2.5

In any case, the conservation of a functional transferred gene requires that it provides some kind of selective advantage, otherwise it would behave like a neutral sequence that has not been able to invade the population (Martin and Herrmann, 1998; Berg and Kurland, 2002; Dyall et al., 2004; Huang et al., 2005). In that sense, there are three main kinds of transference that can be potentially maintained: the acquisition of a novel gene not present in the lineage, the acquisition of a gene paralogous of one already present in the lineage and, the displacement of a gene by an orthologous gene from a distant group (Koonin et al., 2002, 2001).

The first strong (genomic) indication of the importance of gene transfer in the evolution of organisms came from the multi-factorial analysis of the codon usage frequencies in the (at the time) existing fragments of the *Escherichia coli*'s genome. The study revealed that around 15% of the sequences presented a significant deviation from the general codon usage pattern (Medigue et al., 1991; Koonin et al., 2002). Some other bacteria include also a significant amount of genes that were probably object of gene transfer (see Table 2.1), among them it is worth mentioning *Rickettsia prowazekii*, an α-proteobacteria related to those that probably gave origin to mitochondria (Andersson et al., 1998; Müller and Martin, 1999), and *Synechocystis* sp. a cyanobacteria that is related to those that could have originated chloroplasts (Gray, 1999). Some other examples are akin to the evolution of antibiotic resistance, pathogenicity and some metabolic pathways; in particular the hypertermophilic bacteria seem to have scavenged more archaeal genes than other bacteria.

Examples of laterally transferred genes detected in eukaryotes are still not as many as those documented for prokaryotes; however a view over some paradigmatic cases suffices to give a general idea of the importance that this phenomenon had in the evolution of those organisms. Koonin et al. (2001, 2002) comment on the approximately 185 sequences from the genome of *Caenorhabditis elegans* with relatively strong similarity (higher that with other eukaryotes) to bacterial genes. Also, in the 2001 draft of the human genome, a careful examination of around 223 gene products yielded 113 sequences that are common in bacteria and could potentiay represent cases of lateral gene transfer.

A genome-wide analysis of Bdelloids, a parthenogenic class of rotifers, found traces of massive gene transfer of bacterial, fungal and plant origins (Gradysehev et al., 2008). A specially intriguing case is represented by the sea slug *Elysia chlorotica*. These organisms are able to abduct the chloroplasts of its usual prey, the secondary alga *Vaucheria litorea*; the sequestered chloroplasts remain active for the 10 months of their life span providing the slug with carbon and energy. The photosyntetic activity of the chloroplasts continues without the presence of the algal nucleus, because the genome of *Elysia chlorotica* has been able to integrate the genes that are necessary to keep the chloroplast functioning however, they are not able to reproduce in the slug and it needs to get new chloroplasts in every generation early in development to assure its maturation (Rumpho et al., 2008).

There are also examples of gene transfer in fungi; Morris et al. (2009) reported evidence or multiple gene fusions and lateral gene transfer events in oomycetes genomes, in particular the regulatory networks of these plant pathogens are built from domains from different origins that include metazoans, bacteria, plants and other fungi.

Table 2.1.: Candidate horizontal transfers between bacteria, archaea, and eukaryote (Modified from Koonin et al., 2001)

Species	**Referece taxon**	**Acquired genes (number and %)**
Aeropyrum pernix	Archaea	47 (2.5)
Methanobacterium thermoautotrophicum	Archaea	100 (4.2)
Methanococcus jannaschii	Archaea	39 (2.3)
Pyrococcus horikoshii	Archaea	39 (1.9)
Pyrococcus abyssi	Archaea	39 (2.2)
Thermoplasma acidophilum	Archaea	54 (3.7)
Halobacterium sp	Archaea	174 (7.2)
Aquifex aeolicus	Bacteria	45 (3.0)
Thermotoga maritima	Bacteria	53 (2.9)
Deinococcus radiodurans	Bacteria	45 (1.5)
Bacillus subtilis	Bacteria	28 (0.7)
Bacillus halodurans	Bacteria	40 (1.0)
Mycobacterium tuberculosis	Bacteria	62 (1.7)
Escherichia coli	Bacteria	13 (0.3)
Haemophilus influenzae	Bacteria	3 (0.2)
Rickettsia prowazekii	Bacteria	7 (0.8)
Pseudomonas aeruginosa	Bacteria	39 (0.7)
Neisseria meningitidis	Bacteria	5 (0.2)
Vibrio cholerae	Bacteria	16 (0.4)
Xylella fastidiosa	Bacteria	8 (0.3)
Treponema pallidum	Bacteria	4 (0.4)
Borrelia burgdorferi	Bacteria	6 (0.7)
Synechocystis PCC6803	Bacteria	115 (3.6)
Chlamydophyla pneumoniae	Bacteria	9 (0.9)
Mycoplasma pneumoniae	Bacteria	1 (0.1)
Ureaplasma urealyticum	Bacteria	1 (0.2)
Helicobacter pylori	Bacteria	3 (0.2)
Campylobacter jejuni	Bacteria	4 (0.2)

2.1.1. Endosymbiotic gene transfer

Several authors agree that one of the major sources of transferred genes in eukaryotes could be intracellular symbionts (Doolittle, 1998; Selosse et al., 2001; Henze et al., 2002; Martin, 2003; Dyall et al., 2004). The α-proteobateria *Wolbachia*[2] is a maternally inherited endosymbiont that is known to infect proximally 60% of the insect species (Hilgenboecker et al., 2008), as well as some nematode species. It is also known for its ability to manipulate the reproductive mechanisms of its host causing nuclear incompatibilities, male killing, parthenogenesis or feminization of males. Hotopp Dunning et al. (2007) reported that besides several *Wolbachia* genetic insertions in fruit flies, wasps, and nematodes previously reported, they found the presence of 44 out of 45 *Wolbachia* genes, in the nuclear genome of the fly *Drosophila ananassae* (Hawaii) representing almost the whole genome of the bacterium (see Table 2.2).

Among endosymbints a special case is that of mitochondria, plastids and organelles

[2]A relative of *Rickettsia prowazekii*, and therefore related to mitochondria.

derived from the former. The evolution of organellar genomes has been characterized by an extensive genome reduction that in some times has lead to the (almost) complete deletion of the whole genome, as show the cases of hydrogenosomes and mitosomes. This genomic erosion has produced an important compartmentalization of the biochemical activities in the eukaryotic cell.

Table 2.2.: Summary of *Wolbachia* sequences and evidence for lateral gene transfer in public databases (Modified from Hotopp et al., 2007)

Organism	**Total traces screened**	***Wolbachia*traces**
Brugia malayi (filarial nematode)	1,260,214	22524
Culex pipiens quinquefasciatus (mosquito)	7,380,430	21304
Drosophila ananassae (fruit fly)	3,878,537	38605
Drosophila sechellia (fruit fly)	1,203,722	1
Drosophila simulans (fruit fly)	2,321,958	7473
Nasonia giraulti (wasp)	540,102	2
Nasonia longicornis (wasp)	447,736	1
Nasonia vitripennis (wasp)	3,360,694	30

The origin of organelles is strictly related to the endosymbiotic theory. Broadly, in its different versions and reformulations it implies the incorporation of a bacteria into a different organism that was either a proto-eukaryote without organelles (except nucleus), or another bacteria (or a combination of bacteria) (Sagan, 1967; Margulis, 1981; Gray et al., 1999; Henze et al., 2002; Emelyanov, 2007). The idea of endosymbiotic relations as the origin for eukaryotic cells traces back to the end of the XIXth Century by Schimper (for chloroplasts) and Altman (for mitochondria) and got extended during the first decade of the XXth Century by Meresckovsky(Allen et al., 2007; Henze et al., 2002; Emelyanov, 2007). Today the endosymbiotic theory is widely accepted as explanation for the origins of chloroplasts, mitochondria and other organelles. However, it seems that not all organelles a were product of endosymbiotic processes. The hybridization between cell types has also played a significant role. Apparently the cell nucleus originated as the combination of an Archeobacteria and an Eubacterial ancestor from which around 75% of the nuclear genes seem to derive (Gray et al., 1999; Allen et al., 2007).

Doolittle (1998) proposed a very simple idea that explains some sources of transferred genes from proto-organelles and other endosymbionts to the nucleus of the proto-eukaryote cell. The process (called "Gene Transfer Ratchet") contemplates endocytic organisms that "hunted" bacteria, digested them and released their remains in the cytoplasm. The released content would contain also fragmented genetic material of the digested bacteria which, in turn, was freely available for the genome to take. This process could then be repeated several times, just because the eating habits of the proto-eukaryote put it in such disposition. The durability of the transfers depended on selective advantages and conservation of the imported gene. If a particular transferred sequence was lost in the nuclear genome it could be reintroduced some other time, if it was lost in the proto-organelle it could have been compensated by the nuclear version.

Two main evolutionary trends can be distinguished when organellar genomes are stud-

ied[3]: genomes that preserve clear evidence of their eubacterial origin, and genomes that have strongly diverged from it. The first pattern is characterized by the preservation of an (almost) complete set of tRNAS, bacterial-like genomic structure (i.e. very few or no introns) and a not so drastic reduction of the genome. In contrast, the second pattern consists of a very drastic gene number reduction (as in *Chlamydomonas* or *Plasmodium*), divergence in ribosomal DNA and rRNA structures and accelerated nucleotide substitution rates. It is, however, common to both trends that the organelle's genome size is reduced and that it becomes dependent on the cell nucleus for various tasks.

Genome sequencing projects have revealed that gene transfer between mitochondria and cell nucleus is an ongoing process. One of the evidences that the authors provide is the presence of several nuclear sequences of mitochondrial origin (known as *numts* pronounced "new-mights"), that are, in the case of humans, related to genetic diseases (Hazkani-Covo et al., 2010). However, this is not a process limited to mitochondria, examples of similar processes occurring in plastids are also found both in nature[4] and in the laboratory.

2.1.2. Experimental Gene Transfer

Experimental gene transfers and gene activation from either mitochondria or chloroplasts to the cell nucleus have been achieved under strong selective conditions. These experiments showed that under certain conditions gene transfer is not an infrequent phenomenon and they also show the important adaptive role that gene transfer could play (or has played), when radical environmental changes are met.

Thorsness and Fox (1990) measured the escape of mitochondrial DNA in *Saccharomyces cerevisiae* by use of an independently replicating plasmid (2μ) that carried selection markers for both mitochondrial and nuclear maintenance. The authors do not actually measure the transfer of one gene and its activation, but that of the whole plasmid[5] which happens at a frequency of $\sim 2 * 10^{-5}$ events per cell per generation; however, this frequency can vary depending on factors such as the genetic background or the temperature of culture[6].

Direct measurement of the transference of a functional gene from an organelle to the nucleus was achieved by Huang et al. (2003) and Stegemann et al. (2003). In both cases the plastidic genome was transformed to include selectable (antibiotic) markers for the chloroplast and nucleus. The frequency of gene transfer was then quantified in pollen grains (Huang et al., 2003) or leaf-cells (Stegemann et al., 2003). For the first case the frequency calculated was 1 in $16'000$ pollen grains whereas for the second case the transfer frequency was estimated to be approximately 1 of $5'000'000$ events per cell per generation, which the authors consider "remarkably similar" to the frequency obtained by Thorsness and Fox (1990).

[3] This distinction is, however, not entirely clear-cut.

[4] Martin et al. (see for example 1998), Bock and Timmis (2008) and Huang et al. (2005)

[5] This would be more similar to the transfer of a complete micro chromosome.

[6] And probably is more related to the integrity of the mitochondria and nuclear membranes than to the genetic stability of both mitochondria and the nucleus.

Interestingly, Huang et al. (2003) consider that even though there is a high frequency of gene transfer and given that it seems that the tobacco genome is not in continuous expansion; there should be some kind of equilibrium between integration and deletion. On the other hand, Stegemann et al. (2003) argue that: *a)* gene transfer is an ongoing mechanism for nuclear genome evolution by means of frequent acquisition of organellar DNA sequences and, *b)* that there is an important contribution of promiscuous DNA insertions to intra-specific genetic variation in multicellular eukaryotes. Non of the authors of any of those two papers actually acknowledges that they are working on particularly strong selective conditions and that this factor could also account for the high gene-transfer frequency reported.

All the previous examples explore the transfer of a gene that is ready to be functional in the nucleus. However, It is usually the case that an organellar gene is not active in the cell nucleus and thus it requires certain changes to become functional. Stegemann and Bock (2006) experimentally reconstructed the transfer and activation of a chloroplastic gene in the nucleus. They used the same experimental set-up used in the Stegemann et al. (2003) work but instead of selecting for the nuclear transfer of the nuclear marker, they selected for the transfer and activation of the chloroplasts marker which is usually inactive in the cell nucleus. They report that the frequency at which transfer and activation of the monitored genes under their experimental conditions is $\sim 3 * 10^{-8}$, however this number might change depending on the chromosomal location of the insertion.

On a different experimental approach, Sorek et al. (2007) tried to determine the experimental barriers to horizontal gene transfer between non-eukaryotes using genome-wide techniques. The authors argue that the cloning phase of a genome sequencing project can be viewed as multiple transfer events to the host (*Escherichia coli*) with an extra-chromosomal plasmid. According to them in every genome-sequencing project, there are, for most of the cases, certain DNA fragments that are quite difficult to clone and sequence, according to the authors those fragments correspond to genes belonging to gene families whose presence in multicopy could be "toxic" to the host. Their main point is that there are universal barriers to the genes that can be transferred between organisms, and that those barriers are constituted mainly by gene toxicity.

As has can be seen, lateral gene transfer should not be consider as seldomly occurring phenomenon, but as one that has taken place, and probably still does, at a relatively high rate during the evolutionary history of life. It is probably one of the most influential evolutionary forces based on very basic bio-molecular mechanisms that are now very well understood; however, the details of the population dynamics that could take a molecular mechanism to become an evolutionary force has not been as deeply studied. I will be one of the objectives of this work to investigate some of those details in the context of endosymbiotic gene transfer.

3. A genomic example of lateral gene transfer: The *Nasonia* genome[1]

3.1. Sequencing the *Nasonia* genome

Eukaryotic examples of lateral gene transfer are not as common as prokaryotic ones. There are a number of reasons that have been invoked to explain this; they include the amount of prokaryotic genomes sequences compared the eukaryotic genomes so far accumulated, the difficulty of positively identifying transferred sequences; the frequency at which the process may have occurred or still occures, and others. Nevertheless, as already mentioned, Hotopp et al. (2007) have presented evidence of massive lateral gene transfer from the endosymbiont *Wolbachia* to the nuclear genome of its hosts, among them the parasitoid jewel wasp *Nasonia*.

Recently Werren et al. (2010) published the genome of three *Nasonia* species[2]. These small hymenopterans (2-3 mm long) are specially well suited for laboratory work and genetical experiments, their life cycle is short (14-15 days depending on the temperature), genetic screening is simple due to the haplo-diploid sex determination system, pupae and larval stages can be stored, and several genetic markers already exist thanks to almost 50 years of genetic research in these organisms, in particular on *Nasonia vitripennis* (Pultz and Leaf, 2003; Werren et al., 2004).

Nasonia vitripennis, *N. giraulti*, and *N. longicornis* are usually reproductively isolated by the action of the intracellular endosymbiont *Wolbachia*, however if the bacteria is removed using antibiotics the three species are able to interbreed. Interspecies interbreed provides in this case with the opportunity to conduct fine-scale genetic mapping and positional cloning of any gene of interest, specially those involved in species differentiation (Werren et al., 2004, 2010).

Sequencing techniques have become increasingly fast since the beginning of the Human Genome Project[3] in 1990. However, the basic sequencing methodology remains pretty similar to the one reported by Fleischmann et al. (1995) for the *Haemophilus influenzae* Rd genome, which in turn is based on the method developed by Sanger et al. (1982) to obtain the nucleotide sequence of the λ-bacteriophage.

The procedure is as follows; a huge number of randomly created fragments of genomic DNA are cloned into bacterial vectors, the inserts of those vectors would be individually

[1] This work was done as part of the *Nasonia* Genome Project and published in Werren, J.H., Richards, S., Desjardins, C.A.,... **Anaya, V.H.**, ... et al.(2010), Functional and Evolutionary Insights from the Genomes of Three Parasitoid *Nasonia* Species *Science* 327:343-348.

[2] *Nasonia vitripennis*, *N. giraulti*, and *N. longicornis*.

[3] And derived projects as the *Escherichia coli*, yeast, or *Drosophila* genome projects.

sequenced and all the sequences are assembled (as a puzzle) using specialized software without actually needing a physical map of the genome. The process, called Whole Genome Shotgun (WGS), requires a big number of clones[4] to assure good coverage of the genome and high sequence's quality.

The main technical difficulty of the WGS strategy arise form the management of an enormous amount of sequence and sequence-related information, and from the assembly process where important computational resources are needed to determine the order of thousands of overlapping "reads"[5] and define where the sequencing gaps are located. (Sanger et al., 1982; Fleischmann et al., 1995; Pop et al., 2002). Once there is a general picture of the genome the annotation process starts, computational resources are also used to predict coding and regulatory regions as well as other genetic features.

The *Wolbachia*-free genome form *N. vitripennis* was sequenced using a WGS strategy with a six-fold sequence coverage; the draft sequence was then compared for completeness with 16 finished Bacterial Artificial Chromosomes (BAC), and 18′000 Expressed Sequence Tags (EST), an error rate of $5.9 * 10^{-4}$ was estimated. Two highly inbreed lines of *N. giraulti* and *N. longicornis* were sampled at single-fold sequence coverage (Sanger) and 12-fold 45-base pair Illumina genome coverage, and then aligned to the *N. vitripennis* to generate the respective assemblies.

However, the "purity" of the genome is generally a concern in every sequenciation project. Depending on the organism there can be organellar or other endosymbiotic genomes that can be detected in the nuclear sequence which may derive either from lateral gene transfer events or from technical artifacts like genetic contamination.

Since one of the objectives of the *Nasonia* Genome Project was to study the extent of lateral gene transfer in the genome, it was necessary to differentiate the true examples of gene transfer from those that were simple contaminants. In the *Nasonia* genome this resulted not too complicated. lateral gene transfer from *Wolbachia* was already reported (Hotopp et al., 2007), besides, since the wasp was cured form its bacterial endosymbiont, in principle all *Wolbachia*-related sequences were most likely located in the nuclear genome. Another source of confirmation came from the deep coverage of the sequence.

The published draft of the *Nasonia* genome presents, among other interesting features on the genome structure of the organism and its evolutionary and biological consequences, several examples of laterally transferred genes from *Wolbachia* to the nuclear genome. In fact, some of the examples include a number of Pox-Virus related sequences that were transferred originally into the *Wolbachia* genome an from there into the wasp's genome where it diversified. The sequences of the three species included in this study, showed that the *Wolbachia* transferred sequences are species-specific. A number of other sequences, grouped in 319 scaffolds were also found in the genome, most of them apparently belong to a possible bacterial commensal of the wasp belonging to the genus *Proteus* Werren et al. (2010, specially Table S55)

[4] A single DNA-base has a probability of $P_o = e^{-m}$ of not being sequenced, where m is the sequence coverage, thus, the higher the coverage, the lower the probability that a base remains unsequenced (Fleischmann et al., 1995).

[5] Every sequenced fragment.

3.2. Involvement in the *Nasonia* genome project

During my participation in the Nasonia Genome Project I got involved in the detection of sequences of bacterial origin that needed to be removed from the genome's draft because they did not constituted examples of lateral gene transfer but of DNA contamination. I also collaborated with the identification of indel-markers that could be used to design hybridation oligonucleotides that latter could be used for genetic mapping of genes in the three species of Nasonia, specially in interspecies breeding experiments.

3.2.1. Detection of bacterial sequences from the *Nasonia* genome

To detect bacteria-originated sequences in the *N. vitripennis* genome I made BLAST comparisons of the annotated genome sequences from *Yersinia pestis* (NC_009381) and *Photorhabdus luminescens* (BX470251) against the draft of the *N. vitripennis* genome; then used self-written Perl scripts to parse the BLAST results and create two lists of hits (one per bacteria species) containing those with high identity levels (above 85%) and full fragment coverage.

In the case of *Y. pestis*-originated sequences, I located 87 hits; seven of them related to the 5S-RNA gene, and the rest related to alanine, methionine, phenylalanine, glycine and lysine t-RNA's. All of the hits had matches with different scaffolds from the *N. vitripennis* genome and some times more than one hit per scaffold was found.

When I performed the comparison between the *N. vitripennis* draft and the *P. luminescens* genome it was possible to find 145 sequences of probable bacterial origin. As in the *Y. pestis* case, the BLAST search found hits against the 5S-RNA gene and t-RNA's for alanine, phenylalanine, lysine (most of them), methionine, histidine and glycine, also multiple hits for each sequence were located in different scaffolds and occasionally several times in the same scaffold.

3.2.2. Indel data-base curation

Regions where deletions or insertions (indels) are good candidates for genetic markers since they can be either species-specific and can be used to pin-point a particular region in its native genome and the be re-localized in another species' genome if it is transferred by recombination.

The mapping strategy followed required that the candidate indel sequences were unique to any of the species, longer than 50 base-pairs (bp) and no tandem repeats should be present within the sequence to avoid the risk of self complementarity and secondary structures during the hybridization experiments. To find good candidates that covered all the pre-requisites several Perl scripts were prepared.

A list with around 13′000 indels from the versions 0.5 of the *N. vitripennis* and *N. giraulti* genomes was provided by the Baylor Center where the genome's sequencing and assembly was being performed. The list contained, besides other information, the indels and the sequence of the 100bp upstream and downstream that surround the indel. The first step consisted in cleaning the list from entries that were duplicated, had sequencing

problems, were classified as possible transposons or contained masked regions due to sequence simplicity. This curation process reduced the original list to around 7100 sequences that were then sorted by species giving 1259 sequences for *N. giraulti* and 2374 for *N. vitripennis* after the entries shorter than 50bp were removed.

To obtain the *N. vitripennis* candidates I looked for inserted sequences with a percentage of similarity of more than 96% and a complete coverage within *N. vitripennis* genome of the indel plus 100bp of context according to BLAST comparisons (Altschul et al., 1990). The sequences obtained were then compared with the *N. giraulti* genome where the selection criteria were to have high similarity (96%) and whole coverage for the flanking region that ideally would be a continuous 200 bp track. In case there was an indel present in *N. giraulti* it should have a low similarity with the *N. vitripennis* sequence (less than 50%).

The *N. giraulti* candidate sequences were obtained by the inverted mechanism. The sequences of *N. giraulti* should have a low similarity with any of the *N. vitripennis* genome or in the best scenario not being present at all. However, it was also necessary that the 100bp region up and downstream from the indel had a high similarity with a preferably continuous 200bp segment from *N. vitripennis* . The candidate sequences thus obtained were checked for tandem repeats using STRING (Parisi et al., 2003). By the time this process was finished the version 1.0 the *N. vitripennis* was ready; the candidate sequences were once again "blasted" against the new genomic sequences and those fulfilling the mentioned above criteria where submitted to Dr. John Werren in whose laboratory the hybridization oligonucleotides would be designed.

4. Endosymbiotic gene transfer, Muller's ratchet and the reduction of organellar genomes.

4.1. Eroding the organellar genome

Genome reduction has been one of the main characteristics of the evolutionary history of organelles and other obligate endosymbionts. It has been proposed that lateral gene transfer might have been, and probably continues to be, one of the driving forces leading to genome reduction. In this section I propose that lateral gene transfer plays an important mechanistic role in reduction process and explore its effects along with those of Muller's ratchet. I also claim that both mechanisms could have strong significance in the conformation of the eukaryotic genome.

4.1.1. Muller's ratchet

The origin and advantages of sexual reproduction has been the subject of long and everlasting debate. Since the 1930's (and even before, in Weissman's idea of *apomixis*) classic works on the advantages of sexual over asexual reproduction have tried to explain under which circumstances sexual reproduction is advantageous. It was soon recognized that the two main features of sexual reproduction are chromosome segregation and recombination; of them we can only asses recombination with an evolutionary value because it allows the testing of all kinds of combinations, which could eventually create some advantageous alternatives. Without chromosome recombination and sexual reproduction, all the favourable mutations could only compete against each other. This means that, if recombination is occurring, mutations need not to accumulate serially, but that they can happen in parallel and can be combined within various lines of descent (Fisher, 1930; Muller, 1932, 1958).

The main disadvantage of asexual reproduction can be better described on Muller's own words:

> Under conditions where only stability of type is needed, a non-recombining population does not actually degenerate as a result of an excess of mutation over selection, after the usual equilibrium between these pressures is reached. However, a kind of irreversible mechanism exists in the non-recombining species (unlike the recombining ones) that prevents selection, even if intensified, from reducing the mutational loads below the lightest that were in existence when intensified selection started, whereas, contrariwise 'drift' and

> what might be called 'selective noise' must allow occasional slips of the lightest load in the direction of increased weight.
>
> [...]If we disregard advantageous mutations, including those of reverse and suppressor types, and concentrate our attention in the effects of selection, we find that an asexual population incorporates a kind of ratchet mechanism, such that it can never get to contain, in any of its lines, a load of mutations smaller than that already existing in its present least-loaded lines. However, the latter lines can (in some of their branches, at any rate) become more heavily leaded by mutation (Muller, 1964).

The previously described mechanism known as Muller's ratchet (Felsenstein, 1974) holds a strong theoretical and biological meaning. It has extended from the discussions on the origins of sexual reproduction to practically any field where asexual reproduction is taking place. It has been studied from the theoretical (Felsenstein, 1974; Haigh, 1978; Kondrashov, 1994; Gordo et al., 2002; Fontanari et al., 2003), bioinformatics and experimental (Lynch, 1996; Moran, 1996; Schön and Martens, 2003; Hoekstra, 2005) perspectives. It has been shown that it is not only theoretically feasible but also that it can occur in asexually growing entities such as bacteria or DNA-containing organelles in eukaryotes[1].

The theoretical approaches have studied Muller's ratchet mainly to understand how quick and under which conditions an asexual population might be led to extinction. Scenarios considering finite and infinite populations, different mutational rates and selection values, as well as epistatic effects have been put forward. Among the results produced by these works it is worth mentioning that the ratchet operates even if the assumption of not allowing back mutations is removed [2](Haigh, 1978) However, under certain circumstances the Ratchet can be slowed down due to synergistic epistasis (Kondrashov, 1994). Population size seems to play also an important role, given that according to Fontanari et al. (2003) the effects of Muller's ratchet become stronger if population bottlenecks are present. In any case, Muller's ratchet is a very sensitive process and its outcome can vary according to the parameters used; it is very complicated to draw any definitive conclusion on the "true" effect of this process due to our lack of knowledge on the real values of selection and mutation[3] (Gordo et al., 2002).

Since DNA-bearing organelles are contained within sexually reproducing organisms, they are often inherited from only one parent and some times exhibit close to zero or zero recombination rates[4], which makes them effectively asexual, theory predicts that

[1] But see Roughgarden (2009) for a different view of the whole interpretation of the role of sex in evolution

[2] Back-mutation and other "positive" mutations are considered in the original formulation of the Muller's ratchet principle, however, it is not fundamental to it because it deals with the lack of recombination and its long time effects

[3] Several mutation rates have been measured, Viruses and Bacteria, it has been estimated on 1 in 10^{-8} cell divisions. For Maize, Fly (*Drosophila*), and Humans it is between 1 mutation in about 10^{-6} to 10^{-5} gametes produced, in Mice it is around 1 mutation in 10^{-5} to 10^{-4} gametes Klug and Cummings (1999)

[4] This however, may be questioned, there are examples of symmetrical and non-symmetrical recombina-

they should be under the influence of Muller's ratchet. If Muller's ratchet is indeed operating, a higher fixation rate of deleterious alleles would be expected, which also coincides with the advance of the ratchet (Charlesworth and Charlesworth, 1997). Wolfe et al. (1987) calculated the rates of synonymous substitutions in mitochondrial, chloroplast, and nuclear DNAs of plants and compared them with the nuclear and mitochondrial substitution rates of animals; they found that both mitochondrial and chloroplast genomes have higher substitution rates than the nuclear genomes do, suggesting that Muller's ratchet is acting on the organelles.

Based on this idea Lynch (1996) determined and compared the nucleotide substitution rate of mitochondrial and nuclear t-RNA's from several animal groups finding that indeed, mitochondrial substitution rates were between 5 and 25 fold higher than the nuclear counterparts. Moran (1996) achieved similar findings on the relative rates of substitution of the 16S RNA on five clades of endosymbiotic bacteria as compared with those of their free-living relatives.

There are, however, ways to escape from the effects of Muller's ratchet. The ostracods family Darwinulidae and the rotifers of the class Bdelloidea represent very interesting examples of ways to avoid the deleterious effects of Muller's ratchet. The Darwinulidae family is believed to lack sexual reproduction and reproduce parthenogenically. One of its representatives *Darwinula stevensoni*, showed very little genetic divergence; the opposite of what was expected from an asexually reproducing organism. Effective DNA-reparation mechanisms and particular ecological niches are invoked as explanations for this situation. In any case, there had been some doubts cast upon the asexuality of this group but the debate on this issue is still an ongoing one (Schön and Martens, 2003; Check Hayden, 2008; Martens and Schön, 2008). On the other hand, bdelloids apparently have been reproducing asexually for several million years, yet there are no traces of vestigial male reproductive structures, and eggs are produced from primary oocytes after two mitotic divisions without chromosome pairing or reduction.

The bdeloid Muller's ratchet scape mechanism could be a consequence of their ability to endure extreme desiccation periods. Membrane disruption as well as DNA fragmentation and later repair may occur during dehydration and recovery; this in in turn would facilitate the transmission of DNA fragments between different individuals. It is still not clear if the replacement of homologous segments is taking place; but if it were the case, it would represent a sexual-like genetic exchange (Gradysehev et al., 2008). Besides that, according to a publication by Ekelund and Rønn (2008) bdelloids are apparently *degenerate tetraploids*[5] and that may be in some way related to their mechanism to hinder Muller's ratchet.

---

tion in organelles from fungi (Van Winkle-Swift and Birky, 1978), plants (both under experimental and field conditions)(Lonsdale et al., 1988; Saville et al., 1998) and mammals (Thyagarajan et al., 1996)

[5]Tetraploid degeneracy is not restricted to this group; also the *Saccharomyces sensu stricto* and this characteristic group and the subphylum Vertebrata are degenerate tetraploids which followed very different evolutionary patterns.

4.1.2. Outlook of previous work on similar problems

Population genetics of DNA-carrying organelles has been the subject of various studies aiming to understand the effects of different evolutionary forces on their genomes. One of the first works on the population dynamics of organelles was published by Birky, Maruyama, and Fuerst (1983). They propose an infinite allele model, taking into consideration the effects of mutation and random drift on selectively neutral alleles; to represent the uniparental inheritance and independent segregation of organelles they redefine the effective population size (N_e) from the classical models of population genetics as organelle effective population size ($1/(2N_e)$). Their results show that the effective population size of the organelles' genes is of about one-fourth of the nuclear genes in the same population. This increases the gene fixation rate and decreases gene diversity.

Albert et al. (1996) developed a three-level selective process (intermolecular, intermitochondrial and intercellular) to model the conservation of the mitochondrial information. The authors used computer simulations of a recombination-selection cycle where genomes of different sizes and information contents were created as a result of the replication advantage that smaller genomes had. However, when multi-level selection was contemplated sub-optimal states (from the individual genome perspective) where privileged, as would be expected when a genomic conflict (in the terms proposed by Cosmides and Tooby, 1981) is occurring. One of their main findings is that there is a functional equivalence between a single mitochondrial genome containing the whole genetic information, and a number of partial genomes that taken together contain the same information as the single molecule.

O'Fallon (2007) proposes that the increased substitution rates and accumulation of mutations characteristic of intracellular symbionts derive emerge from the lost of their ability to be horizontally transmitted; conflicting levels of selection and the population structure that is imposed upon them by the host. Putting those factors altogether, the author tries to explain the evolutionary paths followed by organelles and other symbionts.

Considering acquisition and loss of genes in bacteria, Berg and Kurland (2002) developed a birth-and-death model that describes the dynamics of gene diffusion in microbial populations. The authors propose that the role of lateral gene transfer as a source of novel genes has been overrated. According to them the probability of fixation of an alien gene that is functionally equivalent to from the endogenous one is $1/(2N)$, which allows them to conclude that the alien genes that are fixed in global populations are very likely to provide new functions to the organism.

Novozhilov et al. (2005) explores gene transfer between bacteria from a theoretical perspective. They postulate a deterministic model of two classes of organisms growing. The fist one presents a determined sequence (i.e is *infected* by the sequence) and the other one receives the transferred sequence (i.e is *not infected*). Gene transfer is considered to occur at a constant rate, the same as mutation. Also a stochastic model of the process is proposed; it assumes the same basic conditions and proceeds using a birth and death process. They conclude that transferred sequences need not to be strongly beneficial, as it is generally considered, and that given the properly long amount of time a sequence (specially if fixed) could become beneficial, or even indispensable.

Yamauchi (2005) proposed a model in which he explores the role of intracellular competition in the process of gene transfer between mitochondria and the nucleus. He proposes a scenario in which an essential mitochondrial gene is transferred to the nucleus and lost in the organelle. Apparently, the intensity of intracellular competition is determinant for the success of the transfer process; if competition between cells is too intense there is little chance that the sequence would be fixed in the nucleus. This results in an increase of the population that has recently lost the mitochondrial version (due to strong intra genomic selection) of the gene but still has had no chance to obtain a functional nuclear copy of it through sexual reproduction. The work of (Yamauchi, 2005) is centered on the role of transfer rates and competition, no explicit role for Muller's ratchet is assumed.

4.2. The Model

Lateral gene transfer and Muller's ratchet are recognized as processes that have had an important role in the evolution of organellar genomes; both phenomena have been broadly studied from a number of different perspectives, but, to my knowledge, there have been no efforts to investigate their joint effects on the genomic evolution of DNA-containing organelles and other endosymbionts.

In this section I present a deterministic model to study the evolutionary fate of a gene that undergoes endosymbiotic gene transfer between a mitochondria-like organelle and the nuclear genome. The model is divided in two parts, one representing the dynamics on the organellar genome and the other representing those of the nucleus. The organellar portion of the model aims to study how endosymbiotic gene transfer may influence organellar genome reduction under Muller's ratchet and other recombination scenarios. The nuclear part of the model explores the fixation in the nuclear genome of a transferred gene and how cyto-nuclear and nuclear incompatibilities caused by the newly acquired gene can affect this process; finally by implementing both parts of the model together I intend to stablish a link between all those molecular processes and evolutionary forces at the population level.

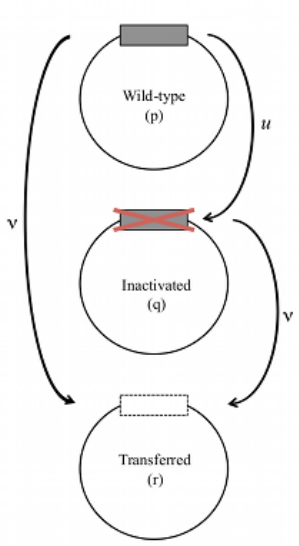

Figure 4.1.: **Alleles considered in the organellar model.** Three alleles are considered in the organellar model: Wild-type allele (with frequency p), inactivated allele due to mutation (with frequency q), and transferred allele (with frequency r). The latter is represented by an empty space where the wild-type or the inactivated alleles used to be located before the transfer event. Inactivation (u) and transfer (ν) occur at constant rates. The inactivated and transferred alleles have the same functional defect which is reflected on their fitness value w_{or}; however, since the genomes containing the wild-type and the inactivated alleles are larger that those where a gene was transferred, the size difference provides the latter with a replicative advantage (s).

In this sense, if endosymbiotic gene transfer and Muller's ratchet are acting on the organellar genomes, the former could represent a mechanism that would allow to "rescue" the information contained in the organellar genome by sending it to the nuclear genome, while at the same time probably reducing the selective pressure on the organelle's genome making it easier for mutations to accumulate and fixate in the organelle's population. However, if recombination is taking place, specially in the form of gene transformation, it could be possible that organellar gene reduction might not take place or that polymorphisms could be located.

4.2.1. Organelle's genome model

I am assuming that the organellar genomes in a population can be considered as an almost panmictic population in which gene transfer (ν) and gene inactivation (u) occur at a constant rate. As proposed by the experimental results of Stegemann et al. (2003) and Stegemann and Bock (2006) the back-mutation rate and gene transfer from the nuclear genome to the organelle is extremely rare and not considered in the model.

Let p, q and r represent the frequencies of the three possible states of one genetic locus: wild type, inactivated or transferred (see figure 4.1). These frequencies vary depending on the effects of mutation (u), gene transfer (ν) and their relative fitness value (w_{or}). Organelle's genome fitness, w_{or}, is defined as $1 - \sigma_{or}$ where σ_{or} is the cost of having an inactivated allele or a transferred one. The value of σ_{or} depends on the alleles present in the nucleus where one locus where two alleles (A and a) are being considered. The allele A stands for a wild type nuclear allele where the organellar gene can be located and activated (represented by a). The values of w_{or} are assigned according to Table 4.1 and differ from one generation to the next one depending on the frequency of a in the nucleus. Genome-replication advantage is assumed for those organelles that have lost a gene, because if one organelle that lost a gene it is shorter than the the wild-type or inactivated ones and is able to replicate faster. In the model, this replicatory advantage is represented by the factor s and takes values between 1 and 0.

Table 4.1.: Organellar fitness

	Organelle		
Nuclear genotype	***wild-type***	***inactivated***	***transferred***
AB|AB	1	0	0
AB|Ab	1	0	0
AB|aB	1	$1 - \sigma_{or}$	$1 - \sigma_{or}$
AB|ab	1	$1 - \sigma_{or}$	$1 - \sigma_{or}$
Ab|Ab	1	0	0
Ab|aB	1	$1 - \sigma_{or}$	$1 - \sigma_{or}$
Ab|ab	1	$1 - \sigma_{or}$	$1 - \sigma_{or}$
aB|aB	0	0	0
aB|ab	0	0	0
ab|ab	1	1	1

The fitness value constitutes a link between the organelle's genome and the nuclear genome and establishes them as a functional unit. It represents the idea that completeness of function and information can be achieved not only when the organelle's genome is intact but also as a product of heteroplasmy and/or gene transfer from the organelle to a different cell compartment. Both heteroplasmy and gene transfer have been investigated to a certain detail by Albert et al. (1996) and Yamauchi (2005) respectively.

As mentioned before, organelles are likely to be under the influence of Muller's Ratchet. By definition this process occurs when no recombination is taking place[6]. How ever, it would be interesting to investigate how does recombination might influence the the evolutionary dynamics of the organelle's genomes.

For that purpose I will assume, as usually done, that the organelles are haploid entities and with very low levels of heteroplasmy; all the former would imply that recombination would be effectively equivalent to the exchange of only the portions involved in the process, while the rest of the DNA molecule remains without changes making it unnecessary to follow the rest of the loci in the genome (see figure 4.2). Let $\beta_{i,j}$ describe the proportion of i-alleles that are exchanged for j-alleles when their respective chromosomes get in contact; the encounters' incidence between the chromosomes carrying the different alleles depends on the frequencies of each of them. This is formalized as $a_i a_j \beta_{i,j}$; the complementary process is represented by $a_j a_i \beta_{j,i}$, where $\beta_{i,j} + \beta_{j,i} = 1$. Finally, the changes in p, r and r in the next generation are calculated using equations (4.1) through (4.3).

$$p' = \frac{w_{or} p[1 - u - \nu + q(\beta_{q,p} - \beta_{p,q}) + r(\beta_{r,p} - \beta_{p,r})]}{\overline{w_{or}}} \tag{4.1}$$

$$q' = \frac{w_{or}[q(1 - \nu + p(\beta_{p,q} - \beta_{q,p}) + r(\beta_{r,q} - \beta_{q,r})) + pu]}{\overline{w_{or}}} \tag{4.2}$$

$$r' = \frac{w_{or} s[\nu(q + p) + r(1 + p(\beta_{p,r} - \beta_{r,p}) + q(\beta_{q,r} - \beta_{r,q}))]}{\overline{w_{or}}} \tag{4.3}$$

Where $\overline{w_{or}}$ is the sum of all nominators on that generation. Equations (4.1), (4.2) and (4.3) allow the inclusion of symmetrical homologous recombination and non-symmetrical homologous recombination. The former is represented when $\beta_{i,j} = \beta_{j,i}$, whereas the for later it suffices to satisfy $\beta_{i,j} \neq \beta_{j,i}$, keeping in mind that $\beta_{i,j} + \beta_{j,i} = 1$ for both cases.

[6]Recombination exists in two forms: homologous and non-homologous, however since Muller's ratchet was formulated in the context of recombination in relation to sexual reproduction, it only considers the homologous form. It is then important to mention that homologous recombination can occur two forms: symmetrical homologous recombination and non-symmetrical homologous recombination, the former refers the process where information is exchanged between both chromosomes. In the later case information is transferred from one chromosome to the other, as a result the chromosome that receives the information gets "transformed" into the other one.

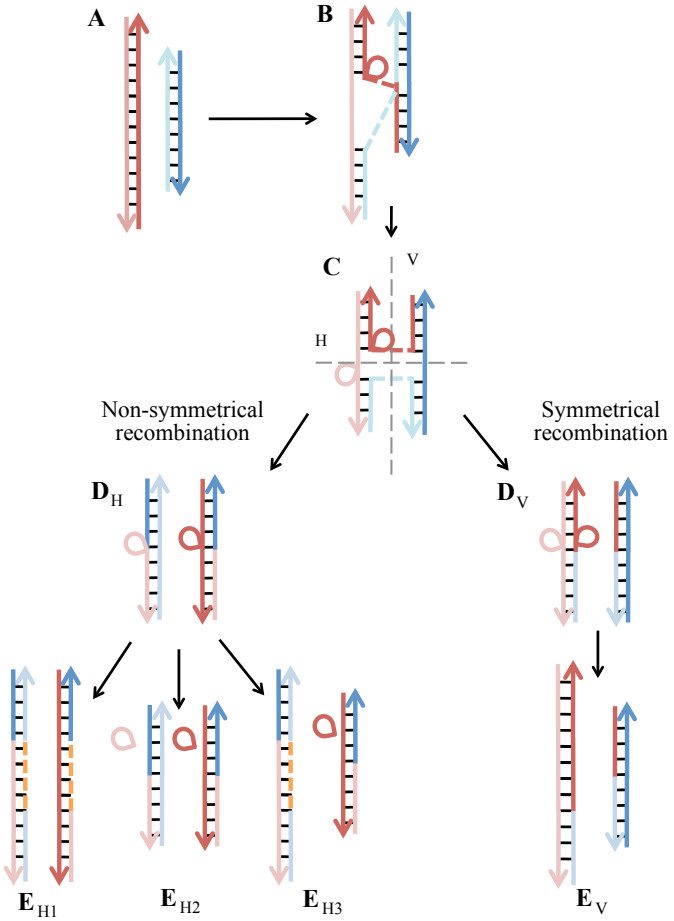

Figure 4.2.: **Symmetric recombination and the resolution of Holiday structure in haploid genomes.** A) Blue and red chromosomes are homologous and identical with the exception of a "missing" transferred gene in the blue one. B) Crossover between both chromosomes leads to the formation of a loop on the red strand due size difference. C) Formation of a Holiday structure and the two possible axes of resolution. D_A) Using axis A to solve the Holiday structure produces symmetrical homologous recombination; the products (E_A) are identical to those in panel A. D_B) Solutions resulting from axis B produce non-symmetrical homologous recombination. E_{B1}) The gene is transformed into the wild-type sequence; the "missing" segment is filled by the DNA reparation mechanisms (yellow dashed line). E_{B2}) Gene transformation in the direction of the short sequence, the loops resulting form the misalignment are excised and remain free DNA. E_{B3}) Double gene transformation, one string is completed while the other is shortened liberating a DNA segment.

4.2.2. Nuclear model

The nuclear model derives from the classical two-loci model with selection. In the present case I contemplate a scenario with two loci with two alleles each. The first loci is defined in a relatively loose way; it does not describe a particular "place" within the genome as it is usually done, but two possible states of the genome namely with or without a transferred and functional gene from the organelle. In this sense, I consider all the nuclear genome as a locus where a gene-insertion can occur, if no insertion has taken place I assume the locus as bearer of the wild-type allele A, on the other hand, if there has been a gene-insertion and activation allele A changes to allele a. The second locus is defined in the conventional way; it is a polymorphic locus with two alleles called B and b. Both alleles in this second locus are equally fitted and functionally equivalent if the first locus is homozygous for A however, B has different levels of incompatibility with a and it is completely incompatible if the first locus is homozygous for a (see table 4.2).

A second kind of incompatibility also represented in the form of a fitness cost, is caused by the interaction of organellar and nuclear genomes: a nucleus that is homozygous for A is completely incompatible with the inactivated and transferred alleles in the

organelle since there would be lacking a functional version of the gene. If the first locus is heterozygous it has a fitness reduction if confronted with the inactivated and transferred alleles. No added cost of double heterozygotes is assumed.

The nuclear cycle starts when a fraction of A is substituted by a at rate (μ)[7]. Let p_1 and q_1 represent the frequencies of A and a , similarly p_2 and q_2 represent those of B and b; the frequencies of p_1 and q_1 after gene transference are obtained using formulas (4.4) and (4.5); the frequencies of p_2 and q_2 remain unaltered during this phase and vary along the life cycle only as a product of the incompatibilities with loci A and a.

$$p_1' = p_1(1-\mu) \tag{4.4}$$
$$q_1' = q_1 + p_1\mu \tag{4.5}$$

During the second step of the cycle the organisms produce gametes. Four kinds are possible: $AB, Ab, aB,$ and ab; their respective frequencies are represented by x_i where $\sum_i x_i = 1$. The gamete frequencies are then calculated as shown in equations (4.6) to (4.9).

$$x_1 = p_1p_2 \tag{4.6}$$
$$x_2 = p_1q_2 \tag{4.7}$$
$$x_3 = p_2q_1 \tag{4.8}$$
$$x_4 = q_1q_2 \tag{4.9}$$

The third step of the process is the formation of embryos by the union of male and female gametes. Both sexes are assumed to be on the same proportions, but organelles are maternally inherited. As mentioned before the organelles and the new formed nuclei can suffer from incompatibilities depending on the combination of transferred genes, active and inactive alleles as well as the presence of alleles B or b; the fitness (w) and the costs $(1-\sigma_{nuc})$ of these incompatibilities are assigned according to Table: 4.2. The total absence of the wild-type alelle in the organelle or a in the nucleus from a single organism, is considered lethal; if no incompatibilities between a and B are assumed Table 4.2 is reduced to Table 4.3.

The frequencies $(E_{i,j,w})$ of the embryos generated are calculated using equations (4.10) where $w_{(i,j,k)}$ corresponds to a value from Table 4.2 or Table 4.3 that considers the paternal (i), maternal (j) and organellar (k) effects.

$$E_{i,j,w} = x_ix_jw_{(i,j,k)} \tag{4.10}$$

[7] μ represents the activation rate of a organelle-originated transferred gene; only those genes that successfully transferred and activated are considered in the model. It is then assumed that a fraction form the DNA-strings that are liberated from the organelles is lost.

Table 4.2.: Nuclear fitness with incompatibilities

♀(j) | Org. (k) \ ♂(i)		**AB**	**Ab**	**aB**	**ab**
AB	wild-type	0.0	0.0	0.0	0.0
	inactivated	0.0	0.0	$1-\sigma_{nuc}$	$1-\sigma_{nuc}$
	transferred	0.0	0.0	$1-\sigma_{nuc}$	$1-\sigma_{nuc}$
Ab	wild-type	1.0	1.0	1.0	1.0
	inactivated	0.0	0.0	$1-\sigma_{nuc}$	$1-\sigma_{nuc}$
	transferred	0.0	0.0	$1-\sigma_{nuc}$	$1-\sigma_{nuc}$
aB	wild-type	1.0	1.0	0.0	0.0
	inactivated	$1-\sigma_{nuc}$	$1-\sigma_{nuc}$	0.0	0.0
	transferred	$1-\sigma_{nuc}$	$1-\sigma_{nuc}$	0.0	0.0
ab	wild-type	1.0	1.0	0.0	1.0
	inactivated	$1-\sigma_{nuc}$	$1-\sigma_{nuc}$	0.0	1.0
	transferred	$1-\sigma_{nuc}$	$1-\sigma_{nuc}$	0.0	1.0

Table 4.3.: Nuclear fitness without incompatibilities

♀(j) | Org.(k) \ ♂(i)		**A**	**a**
A	wild-type	1.0	1.0
	inactivated	0.0	$1-\sigma_{nuc}$
	transferred	0.0	$1-\sigma_{nuc}$
a	wild-type	1.0	1.0
	inactivated	$1-\sigma_{nuc}$	1.0
	transferred	$1-\sigma_{nuc}$	1.0

The resulting embryos would in turn start a new reproductive cycle and generate their own recombinant gametes. Equations (4.11) to (4.14) are used to calculate the frequencies of the recombinant gametes which are represented by y_i in an analogous form to x_i.

$$y_1 = \frac{E_{1,1,k} + \frac{E_{1,2,k}+E_{1,3,k}}{2} + \frac{E_{1,4,k}+E_{2,3,k}}{4}}{\bar{w}} \tag{4.11}$$

$$y_2 = \frac{E_{2,2,k} + \frac{E_{1,2,k}+E_{2,4,k}}{2} + \frac{E_{1,4,k}+E_{2,3,k}}{4}}{\bar{w}} \tag{4.12}$$

$$y_3 = \frac{E_{3,3,k} + \frac{E_{1,3,k}+E_{3,4,k}}{2} + \frac{E_{1,4,k}+E_{2,3,k}}{4}}{\bar{w}} \tag{4.13}$$

$$y_4 = \frac{E_{4,4,k} + \frac{E_{2,4,k}+E_{3,4,k}}{2} + \frac{E_{1,4,k}+E_{2,3,k}}{4}}{\bar{w}} \tag{4.14}$$

Where $\bar{w} = \sum_i^j E_{i,j,w}$. The final frequencies of each allele can be calculated using formulas (4.15) to (4.18).

$$p_1' = y_1 + y_2 \tag{4.15}$$
$$q_1' = y_3 + y_4 \tag{4.16}$$
$$p_2' = y_1 + y_3 \tag{4.17}$$
$$q_2' = y_2 + y_4 \tag{4.18}$$

The models presented for the organelle's gene transfer process, and the nuclear inclusion of the transferred gene, can be coupled to represent the transference of genetic information between sub-cellular compartments. In this case it is necessary to consider the frequencies of the nuclear as well as the organellar alleles to determine the presence of any of the possible embryo-types and to define the activation rate as a product of the nuclear activation frequency and the frequency of the transferred allele (r) as suggested by the experimental results of Stegemann and Bock (2006).

4.3. Results

4.3.1. Organelle model

Due to the number of parameters in equations (4.1), (4.2) and (4.3) it is not possible to analytically determine equilibrium points, except the trivial ones in which either p, q or $r = 1$ and the other two $= 0$, or when the mutation and transfer rates $= 0$ and the frequencies are maintained as they are, independently of the recombination frequencies[8]. To look for other equilibrium points I implemented simulations using different parameter values. The simulation were run until the differences between the alleles frequencies in subsequent generations were smaller than $1 * 10^{-10}$. Allele's extinction was assumed if its frequency fell below $1 * 10^{-12}$.

Nevertheless, it is possible to asses the role of symmetrical recombination in the organelle's genomes evolution without the help of simulations. When recombination is modelled as done in equations (4.1), (4.2) and (4.3), one of the main assumptions done besides haploidy, is homoplasmy. If this is truly the case and recombination takes place in the form of symmetrical homologous recombination (meaning that $\beta_{i,j} = \beta_{j,i} = 0.5$), then this is exactly equivalent to the scenario where Muller's ratchet is operating ($\beta_{i,j} = \beta_{j,i} = 0$) and in both cases the equations reduce to those shown in (4.19), (4.20) and (4.21)[9]. To asses the role of non-symmetrical homologous recombination simulations where also run and the results are show below.

[8] if the recombination is symmetric

[9] see also figure 4.2, especially the part relating to symmetrical homologous recombination

$$p' = \frac{w_{or}p(1-u-\nu)}{\overline{w_{or}}} \tag{4.19}$$

$$q' = \frac{w_{or}[q(1-\nu)+pu]}{\overline{w_{or}}} \tag{4.20}$$

$$r' = \frac{w_{or}s[r+\nu(p+q)]}{\overline{w_{or}}} \tag{4.21}$$

Considering the reported values for organellar mutation and gene-transfer rates the results of all the simulations presented here were done using values of $u = 1 * 10^{-6}$ and $\nu = 1 * 10^{-5}$; results of the implementation of different values for those parameters are shown in appendix A.

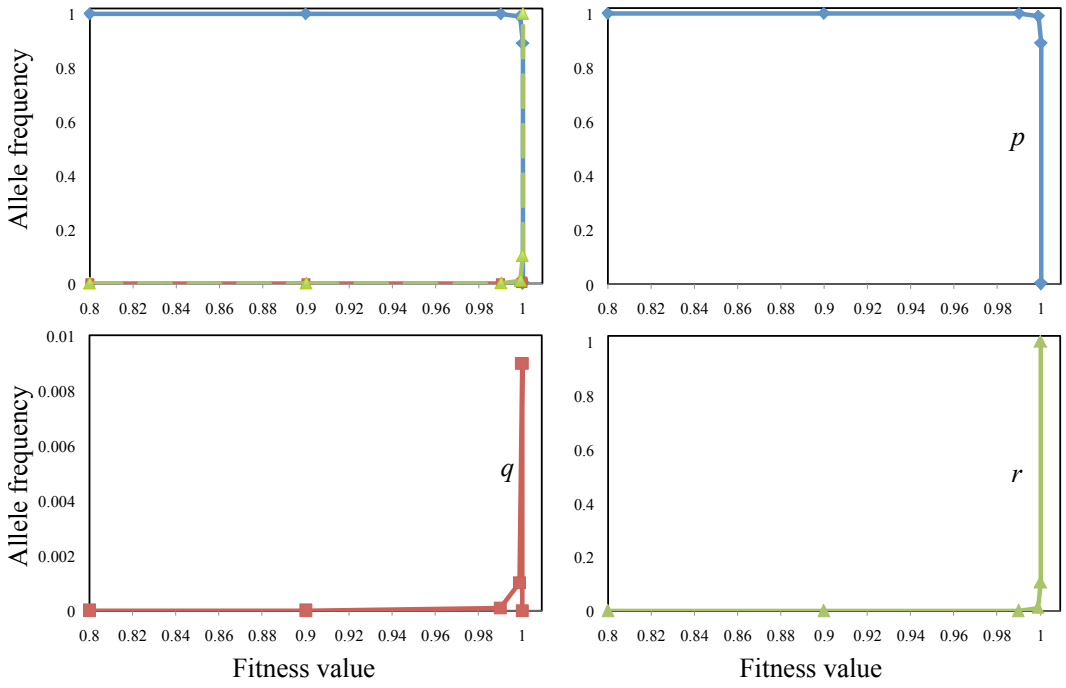

Figure 4.3.: **Organelle's alleles equilibrium frequencies under different fitness values for the inactivated and transferred alleles.**Panel A: Effects of the whole range of w_{or} used, p is presented in blue, q in red and r in green. Panels D, C and D depict a detailed view of the effects for p, q and r respectively. Note that polymorphisms are significant at w_{or} values between 0.99 and 1, with 1 taking the transferred allele to fixation. No recombination is taking place and there is no selective advantage for any allele.

To investigate the effects of fitness in the organelle's model, I obtained the equilibrium frequencies of the organelle alleles when different cost-values (σ_{or}) are affecting the mutants. There was no selection advantage for any of the alleles, and no recombination was occurring; and since the model was implemented by itself (i.e without the nuclear

section), no incompatibilities between the organelle and the cell nucleus were assumed. In all cases the wild-type allele had a fixed fitness value of 1.

As can be seen in figure 4.3 the only fitness values that allow mutant frequencies that are above the mutation and transfer ratios are those very close to 1, in particular between 0.99 and 1 that is, when the mutants are selectively neutral or almost neutral. Interestingly for the almost-neutral scenarios of σ_{or}, polymorphisms are observable (see figure 4.3 panels B, C and D); but when the fitness value for both mutant alleles equals 1, the only achievable equilibrium is the extinction of the wild-type and inactivated alleles and the fixation of the transferred one.

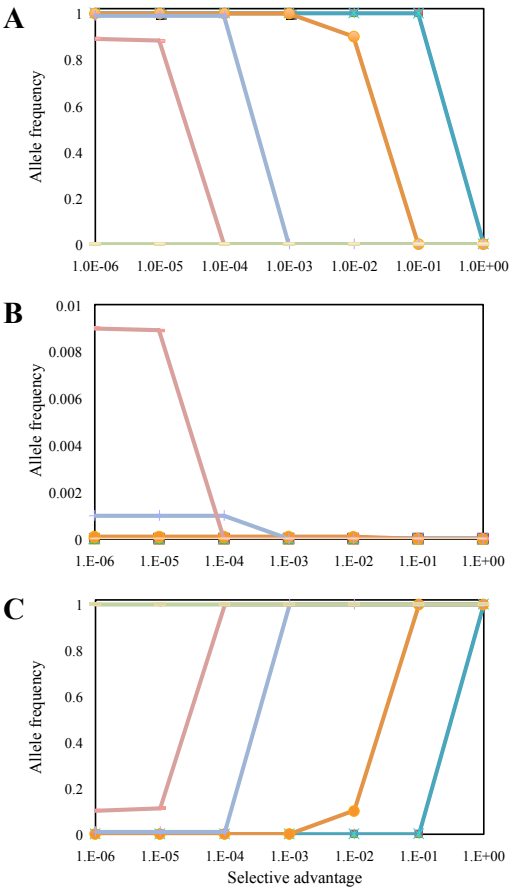

Figure 4.4.: **Equilibrium frequencies of the organelle's alleles as functions of fitness and selective advantage.** The effects of several fitness and s values and no recombination, in the equilibrium frequencies of the wild-type (A), transferred (B), and transferred r (C) alleles is explored. It can be seen that coexistence of the three alleles is only possible at certain combinations of w_{or} and s. The inactivated allele is present in the population at low frequencies (between the value of u and 0.009) only for values of w_{or} and relatively low values of S, otherwise either the wild-type or inactivated are the predominant alleles. Fixation of the transferred allele takes place when $w_{or} = 1$ for any value of s. Other values: the dark blue line stands for $0.5 \geq w_{org} \leq 0.9$, orange for 0.99, light blue for 0.999, light red for 0.9999 and green for 1.

In the model it is assumed that the transferred allele has a selective advantage because it is shorter than the wild-type and inactivated alleles and could replicate faster. To simulate that situation a range of s values between $1 * 10^{-6}$ and 1 where implemented together with w_{or} values between 0.5 and 1. Figure: 4.4 shows the results of these simulations. It can be seen that for all values of $w_{or} \leq 0.9$ the same results are obtained and are equivalent to those shown in figure 4.3, with the exception of $s = 1$ where the transferred allele gets fixed and the wild-type and inactivated alleles get lost (blue-green line in panels A and C). On the other hand, those values of w_{or} that already

allowed polymorphisms continue to do it as long as s is small enough, otherwise the transferred allele gets fixated. These results could suggest that the conditions simulated could be similar to those in which heteroplasmy is found in the organelles. A particularly interesting result is shown in panel B (light blue line); if $w_{or} = 0.999$ and $s < 1 * 10^{-4}$ the frequency of the inactivated allele (q) is higher than that of transferred one which remains at frequency close to ν. This result is important because it exemplifies a scenario where the transferred allele is not the predominant mutant even when selection is favouring it. As expected, if $w_{or} = 1$ the transferred allele gets fixed for any value of s.

One of the main objectives of this study was to study the role of recombination in the evolution of organellar genomes. Given that the action of symmetrical homologous recombination produces effectively the same results as Muller's ratchet (see above)[10], I decided to investigate the role of non-symmetrical homologous recombination. Four scenarios were simulated: gene conversion occurring with a high bias ($0.9 : 0.1$) either towards the "long" alleles (wild-type and inactivated) or towards the "short" one (transferred); and relatively slight deviations from symmetrical homologous recombination in both directions. As in the other cases a range of w_{or} and s values were simulated. For simplicity I assumed that $\beta_{1,2} = \beta_{2,1} = 0.5$; that is, gene conversion is only allowed between molecules of different sizes.

When non-symmetrical homologous recombination is strongly biased towards the wild-type and inactivated alleles for every value of s there is fixation of the wild-type allele at values close to 1 for every fitness value except $w_{or} = 1$ where the inactivated allele is fixed (see figure A.1, panels A and B). In the rest of the cases the inactivated allele become lost and the transferred allele was maintained at frequencies close to its emergence ratio ν in all circumstances (see figure A.1, panel C). In the opposite case, when recombination is strongly biased towards the transferred allele, the wild-type and inactivated alleles are lost except when $w_{or} = 0.5$ and $s \leq 0.1$ (see figure A.2). In the other cases studied the only equilibriums found were the extinction of the wild-type and inactivated alleles and the fixation of the transferred one.

If the deviation from symmetrical homologous recombination is reduced to a value of 0.05 and it is biased against the transferred allele, there is fixation of this allele at fitness values of 0.9 for $s \geq 1 * 10^{-2}$ (see figure A.3), fixation of the same allele at lower fitness values is only possible if $s > 0.1$. The value of the inactivated remains close to that of u except when $r = 1$. For comparison, in the case represented in figure 4.4, invasion by the transferred allele was only possible at higher s and fitness values. As expected, if $w_{or} = 1$ the fixation of the transferred allele occurs at any value of s. inactivated allele is able to be stable at a frequency close to 1 (see figure A.4, specially panel B). It is also noticeable that except for the cases when $s = 1$ and $w_{or} > 0.9$ where the transferred allele invades the population, all the other equilibria allow the coexistence of the three alleles, even if only at frequencies close to u and ν.

A general picture of the behaviour of the model can be drawn; invasion of the mutant kinds is possible under particular circumstances; the fixation of the transferred allele usually only takes place when the mutation cost (in term of fitness value) is small, that

[10]Lynch (2007) discusses also this point.

is, $\sigma_{or} \leq 0.01$, or at relatively high values of s. Counterintuitively the transferred allele does not benefits if recombination is biased towards it, actually this situation tends to favour the fixation of the inactivated one. On the other hand recombination working against the shorter variant does not imply that it will be eliminated, it does however, prevents it from growing at higher frequencies than its emergence one.

4.3.2. Nuclear model

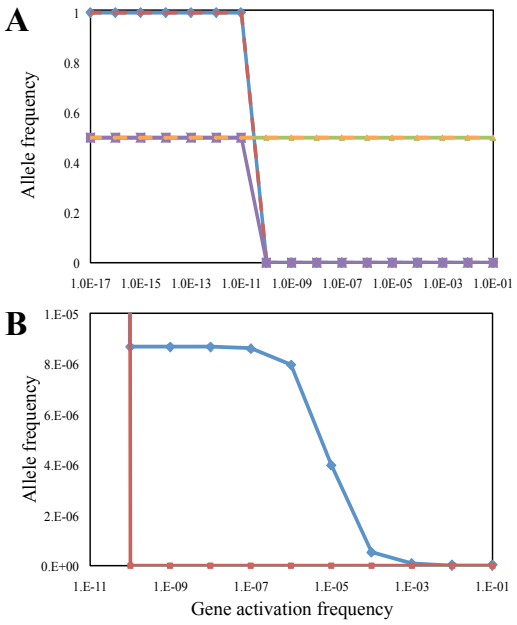

Figure 4.5.: **Nuclear allele equilibrium frequencies affected by cyto-nuclear and nuclear incompatibilities.** Panel A shows the effects of nuclear incompatibilities in the different alleles. For A there are no effects of nuclear incompatibilities (red line) or the lack of them (blue line), the fall in A frequency is due to the effects of the activation rate. In the case of the alleles B and b , their frequencies remain constant if there are no incompatibilities (orange and green lines), but if they are implemented (purple line), their frequencies vary according to that of A and a. In panel B it is possible to observe the effect of σ_{nuc}, when $\sigma_{nuc} = 0$ there is no elimination of A (blue line), except for high activation rates, but if $\sigma_{nuc} > 0$, then at an activation rate of $1 * 10^{-10}$ A gets lost.

As in the previous section, simulations implementing the nuclear equations described in section 4.2.2 were run until stability of the allele frequencies was reached. I incorporated the effects of incompatibilities between the newly transferred gene and the nuclear genes as well as incompatibilities between the nuclear genome and the organellar genome. I compared the latter with a scenario where no nuclear incompatibilities were assumed. In all cases gene activation rates were set at values between $1 * 10^{-18}$ and 0.1 and σ_{nuc} was set between 0 and 1.

As can be seen in the panel A of figure 4.5 equilibrium frequencies of A and a depend on the value of σ_{nuc} and μ. Nuclear incompatibilities only affect alleles B and b ; if incompatibilities are incorporated b gets fixated once a has also increased its frequency. If no incompatibilities are modelled then B and b remain unchanged for any value of μ (Panel A green and orange lines).

For values of $\sigma_{nuc} > 0$ there is no possibility of polymorphisms, independently of the σ_{nuc} value there is a frequencies' turnover point at an activation frequency of around $1 * 10^{-10}$ (see panel B).

4.3.3. Complete model

To find equilibrium points for the coupled organelle and nuclear models, simulations were run using a reduced repertoire of the parameters used in the previous sections. For all cases, the organelle mutation rate u was set to $1 * 10^{-6}$, organelle transfer rate $\nu = 1 * 10^{-11}$, the value of s was set to 0.1, nuclear activation rate (μ) was calculated every generation as a product of the frequency of r multiplied by a "turn-up" frequency that ranged from $1 * 10^{-8}$ to 0.1. The results presented also consider the effects of recombination and incompatibility scenarios.

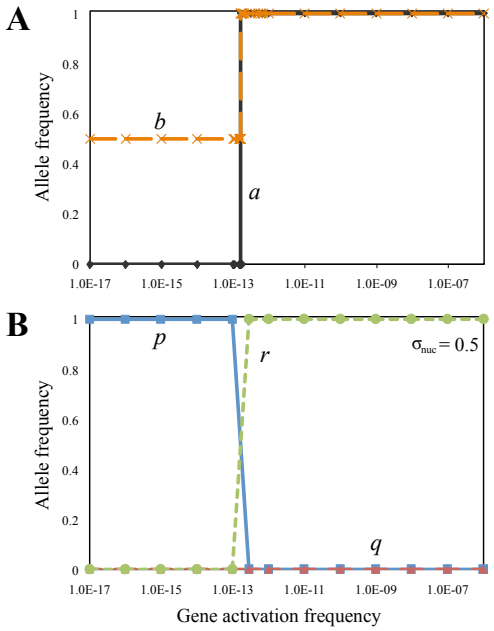

Figure 4.6.: **Equilibrium frequencies under the effects of Muller's ratchet and nuclear incompatibilities.** Panel A: equilibrium frequencies without incompatibilities and no recombination; Alleles a (red) and b (purple) got fixed while A and b are lost (not shown); the organellar alleles stay in their initial frequencies. Panel B: equilibrium frequencies with nuclear incompatibilities ($\sigma_{nuc} = 0.5$), The nuclear alleles behave as shown in panel A but the in the organelle the transferred allele goes to fixation (green), the wild-type (blue) allele is eliminated and the inactivated allele (red) remains at low frequency in the population.

The first result to be examined was the effects of Muller's ratchet (or symmetrical homologous recombination, according to the present model), when no incompatibilities are occurring neither at the nuclear nor at the cyto-nuclear levels[11]. As can be seen in the panel A of figure 4.6, there is a turning point around an activation rate of $1.66 * 10^{-13}$ similar to the one already detected by implementing the nuclear model by itself but displaced by two orders of magnitude. Also in this case a and b get fixated and A and B disappear from the population. At this point the organelle is still in its original allele conformation that is, the wild-type allele is very close to a frequency of 1 and the mutants' frequencies are close to u and ν respectively. The implementation of cyto-nuclear incompatibilities by themselves or together with nuclear incompatibilities does not affect the final result, but it introduces small shift in the allele's frequencies ($2.9 * 10^{-13}$). If only nuclear incompatibilities are implemented (see the panels B of figure 4.6), the frequencies of the nuclear alleles also get modified, producing the fixation of the transferred allele, the extinction of the wild-type and keeping the inactivated one

[11] except those that derive from the lack of a functional gene in both cellular compartments at the same time

around its emergence frequency (u). This implies that under the circumstances here modelled it is necessary to remove the selective pressure that the lack of the functional organellar gene imposes in the system, to allow the invasion of the mutant classes.

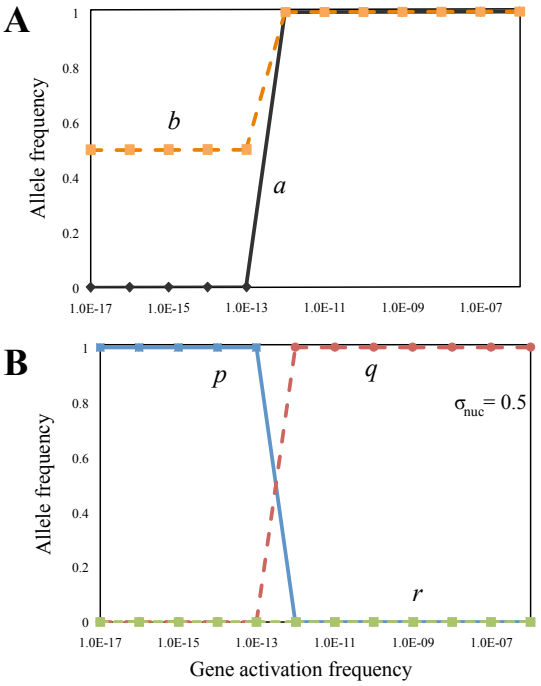

Figure 4.7.: **Equilibrium frequencies with recombination strongly biased towards the transferred allele** In the case that recombination is 9 : 1 biased towards the transferred allele the nuclear alleles (panel A) shows the same pattern independently of the incompatibilities implemented: The alleles A and B disappeared and a (black) and b (orange) invade the population starting at gene activation rates around $1 * 10^{-13}$. If nuclear incompatibilities are acting $\sigma_{nuc} = 0.5$ (Panel B), the nuclear changes are accompanied in the organelle by strong increase of the inactivated allele (green), decrease of the wild-type (blue), while the transferred allele (red) stays close to its original frequency.

To asses the effects of non-symmetrical homologous recombination in the organellar genome, four recombination scenarios were assayed: high and low bias towards the shorter mutant (the transferred allele) and high and low bias against it. When non-symmetrical homologous recombination is biased in a proportion of 9 : 1 towards the transferred allele and nuclear incompatibilities are implemented the nuclear alleles have the same dynamics (see panel A of figure 4.7) but an interesting change takes place in the organelle at the same gene activation frequency that generates the allele shift in the nucleus. As can be seen in the panel B of figure 4.7 the inactivated mutant increases its frequency in the population almost to 1 while that of wild-type allele descends to a level similar to u when the activation rate is around $1 * 10^{-13}$. The short mutant remains close to the value of μ at for all values of activation, this result goes in agreement with those shown in figure A.2. If cyto-nuclear incompatibilities alone or in addition to the nuclear ones are implemented, the results are practically identical to those depicted in the panel A of figure 4.7, but unlike the panel B of that figure, the frequencies of the alleles in the organelle remain unchanged.

In a scenario where the deviation from symmetrical homologous recombination only favours slightly the transferred allele ($\beta_{1,3} = \beta_{2,3} = 0.55$) the same basic results described in the image 4.7 are obtained, except for the changes shown in the panel B of figure 4.8: again for the nuclear alleles (Panel A), the implementation of incompatibilities has no effect; but in the organelle the effect of the nuclear incompatibilities ($\sigma_{nuc} = 0.5$)

produces the same observed increase in q and decrease of p, nevertheless allowing the coexistence of the transferred allele at values close to $1 * 10^{-3}$.

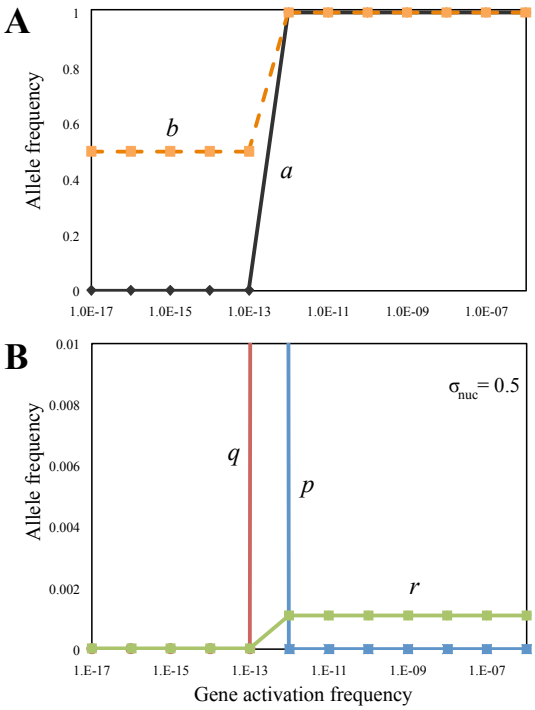

Figure 4.8.: **Gene conversion slightly biased towards the transferred allele.** The effects of a low deviation form symmetrical homologous recombination favouring the organellar the transferred allele allele ($\beta_{1,3} = \beta_{2,3} = 0.55$) are shown. In panel A are presented the allele's frequencies in the nucleus for all incompatibility scenarios; alleles A and B (not shown) become extinct at activation frequencies above $1 * 10^{-13}$ while a and b (black and orange) are fixed. Panel B shows the effects in the organelle when the nuclear incompatibilities ($\sigma_{nuc} = 0.5$) are implemented; the wild-type allele (blue) decreases, and the transferred allele is present at frequencies around $1 * 10^{-3}$.

Taken together the results presented in figures 4.7 and 4.8 it seems that at least for the parameters used, the transferred allele increments is frequency more when low recombination biases are implemented than when high ones are acting. The results also suggest that it is possible that more than one mutant allele can be maintained once a functional copy of the gene has been successfully transferred to the nucleus. It is also clear that the mutant allele that becomes more frequent is not the one that is benefited by the recombination bias or some selective advantage. It is again the case that the fixation of a and b is a necessary condition if any of the mutants is going to be fixed.

In the case that the bias in organellar recombination moves in the opposite direction and favours the long variants, instead of the short one, there are clear differences in the dynamics of the allele's frequencies, specially in the organelle. When non-symmetrical homologous recombination is biased in a proportion of $9 : 1$ against the transferred allele, the role of nuclear and cyto-nuclear incompatibilities become more relevant. As can be seen in the panel A of figure 4.9 there is only one stable-state for all the gene activation rates used and for nuclear incompatibilities alone or no incompatibilities at all: fixation in the nucleus of a and b and fixation of the transferred allele in the organelle. The rest of the variants are no longer present in the population. However, if only cyto-nuclear incompatibilities are operating (panel B), there is no change in the organellar allelic frequencies; the wild-type remains as the preponderant allele, the inactivated remains at its respective emergence frequencies and the transferred allele increases its frequency

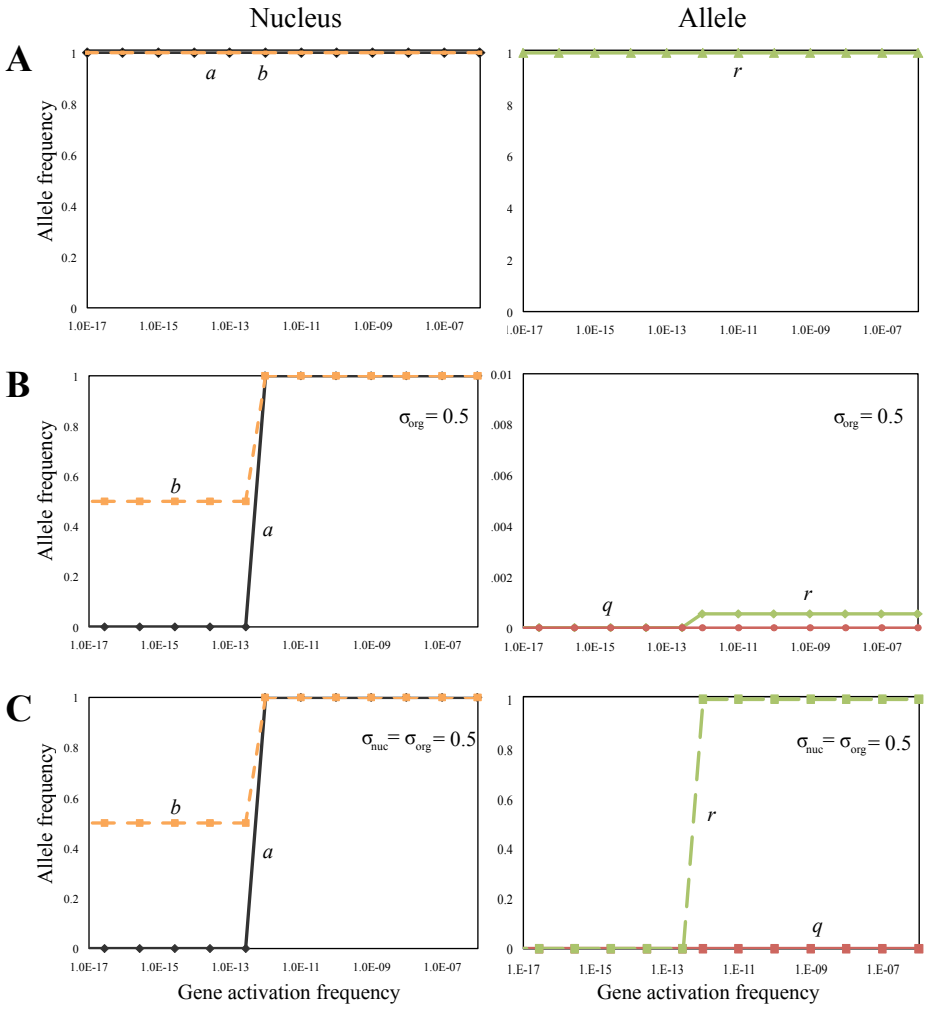

Figure 4.9.: **Equilibrium frequencies with recombination strongly biased against the transferred allele.** When a recombination bias of 9:1 is acting against the transferred allele and no incompatibilities or only the nuclear ones are implemented (panel A); there is only one equilibrium for all activation rates which is the fixation of a (black), b (orange) and the transferred allele (green). For the same recombination bias and cyto-nuclear incompatibilities (panel B) then there is fixation of a (black) and b (orange) in the nucleus and a increase of the transferred allele (green) up to around $5 * 10^{-4}$ in the organelle. When the same recombination bias is operating and both, nuclear and cyto-nuclear, incompatibilities are acting; the nuclear scenario remains as in B but the transferred allele (green) is fixed in and the inactivated allele (red) as well as the wild-type one are lost.

up to around $5 * 10^{-4}$. In the nucleus, a and b are fixed and their counterparts go to extinction when a critical activation rate is reached. The role of both cyto-nuclear and nuclear incompatibilities is interesting in this scenario. As can be seen in panel C they allow the fixation on transferred allele increases in the organelle when an activation rate of circa $1 * 10^{-13}$ is reached. The same happens in the nucleus for alleles a and b.

The last scenario that I studied was the effects of a slight bias from symmetrical

homologous recombination acting against transferred allele increases. In this case it is not possible to asses any effect to the presence or absence of nuclear or cyto-nuclear incompatibilities neither by themselves nor by their combined implementation. The effects on the nuclear alleles can be seen in figure A.6. As in most of the cases studied a and b increase in frequency reaching immediately a value of 1 and their counterparts are lost from the population once a critical activation rate is reached ($1 * 10^{-13}$). In the organelle the frequencies of the three alleles remain practically without any change as they were when all the simulations where started.

4.4. Discussion

In this work I have analysed the added effects of Muller's ratchet, symmetrical homologous recombination, non-symmetrical homologous recombination and lateral gene transfer on the evolutionary faith of an organellar gene located in the organelle's genome. I have also studied the effects of those processes in the fixation of that gene if it gets transferred to the nuclear genome where it can face genetic incompatibilities with the rest of the nuclear genome or with the organellar nuclear under certain circumstances.

The generally accepted biology of DNA-containing organelles include the following three assumptions a) organelles are (for most of the cases) haploid entities with more than one copy of the entire genome per organelle, b) all the genomes in one organelle are identical, and in a given population the level or variability is very low, c) organellar genomes are mostly clonal, i.e. they do not recombine. All these assumptions could support the idea that organellar genomes are candidates to be under the effect of Muller's ratchet (Birky, 1995, 2001; Ballard and Whitlock, 2004).

Homologous symmetric recombination by itself is not able to modify allelic frequencies because it does not create new alleles or eliminates others. It only modifies the genetic correlations in a population and, because some of those combinations might be better fitted than others, their allelic frequencies are affected by selection.

There is increasing evidence of the occurrence of recombination in organelles that might question the true role of Muller's ratchet in the organelles[12]. Reports of recombination, gene transformation, and solutions to Holiday structures (needed for homologous recombination) have been in the literature since the late 1970's but they have not reached enough resonance in the population genetics' community. The basic contradiction postulated by the traditional view of the organelle's biology and the role of Muller's ratchet is the following: If the three assumptions mentioned before hold, it means that in a given population all the organellar genomes are identical and new variants emerge by mutation, and those possible variants are mostly isolated due to maternal inheritance; thus, if at any given time recombination takes place between two molecules that are essentially identical, the end products of the process would have the same genetic cor-

[12] Saville et al. (1998) discuss a similar point, based on their sampling of mitochondrial DNA from natural populations of *Armillaria gallica* and the conclusion that recombination in the mitochondria of this group was the most parsimonious explanation for the mitochondrial molecular phylogenies that were reconstructing.

relations that they had at the beginning, the difference between both variants (if the mutation is selectively neutral), would only be their frequencies and it will not be affected by symmetrical homologous recombination. Therefore, the role of recombination (symmetrical homologous recombination) in general and Muller's ratchet in particular within this context is irrelevant unless one is willing to assume that genetic variability in the organelles might be higher. On the other hand; there are reports pointing out lower levels of homoplasty[13]; if this is the case, again, recombination has an important role to play.

An alternative can be found if recombination is not symmetrical. In that case recombination can, by itself, modify the allelic frequencies in a population and affect the correlations between different genes and gene combinations. Indeed, if one wants to provide recombination with a functional and evolutionary value under the conditions here presented, it is necessary then to consider not only symmetrical homologous recombination but also other variants of homologous recombination that could potentially change not only the correlations between alleles but also their frequencies. The fact that there is evidence of the resolution of Holliday junctions in mitochondria indicates that homologous recombination is taking place, and it has been suggested by Marechal and Brisson (2010) that in this case recombination could serve as a DNA repair mechanism. If recombination is indeed functioning as a genetic "fixing-tool" then it is actually more a case of non-symmetrical homologous recombination than symmetrical homologous recombination since the information of one of the DNA strands is modified to fit that of the other DNA-strand. Moreover, Van Winkle-Swift and Birky (1978) as well as Walsh (1992) have propose that gene conversion could be one of the mechanisms responsible for the low level of heteroplasmy usually found in organelles.

The results of the implementation of non-symmetrical homologous recombination confirm that this process is able to induce changes on the genetic composition of the organelles specially if there are incompatibilities at the nuclear or cyto-nuclear levels. The effects of non-symmetrical homologous recombination depend on how much it diverges from the strict symmetry and in which direction, but for some cases it produces the fixation of mutant alleles, even if the deviation is not favouring them. More over, as the results show, gene conversion does not have to substitute symmetrical homologous recombination, it only needs to occur at a certain frequency or every once in a while to induce changes in the allelic frequencies, it is therefore necessary to consider also the effects of these phenomena for the evolution of organellar genomes.

Taking into consideration the results obtained from the simulations and the body of information available in the literature, I propose that the role of Muller's ratchet needs to be redefined to include the possibility of its existence even under some recombinatory scenarios, in particular symmetrical homologous recombination under conditions of high homoplasmy

Special attention should be paid to the role of incompatibilities. Two types of incompatibilities were implemented: the cyto-nuclear one described the cost of a) losing one

[13] See for example Grztbowski (2000); Chat et al. (2002); Pearl et al. (2009); Magnacca and Brown (2010)

gene in the organelle and not transferring it to the nucleus or b) losing one gene in the organelle and having a transferred version of it in heterocygosis. These incompatibilities were practically removed once the transferred allele became fixed in the nucleus. In the case shown in panel B of figure 4.9 it can be seen that this parameter was responsible of the presence of polymorphisms. The second kind of incompatibility refers to conflicts between the newly transferred gene and its environment. This kind of incompatibility has more noticeable effects since it promotes (or hampers, depending on the situation) the fixation of mutant alleles in the organelle.

Under the conditions here modelled lateral gene transfer may promote the fixation of mutant alleles in the organelle, since it releases some of the selective pressure on the integrity of the organellar genome. Which mutant allele or alleles are going to be successful under this scenario depends on the direction and strength of the recombination's deviation and on the strength of the incompatibilities. On the other hand, the introduction of one new gene in the nuclear genome may also radically change the genetic conformation of a population if, again, there are incompatibilities between genes and their relative strength.

From the point of view of the genetic information, endosymbiotic gene transfer provides mechanism to maintain the genetic information relatively intact, at least up to a certain extent in which the function is at least partially conserved, specially in the case were this information is important for all the entities involved. In this sense the importance of the completeness of information, and a tendency to privilege the reduction of the organelle's genome has also been proposed by Albert et al. (1996); Yamauchi (2005) and O'Fallon (2007), however none of them establishes a relation between those elements.

Rand et al. (2004) presented a paper that sustained that a coevolutionary scenario could be observed between the nuclear genome of the modern eukaryotic cells and that of their organelles. This coevolutionary history derives from the interaction of slightly deleterious mutations in the organelles and compensatory mutations in the cellular nucleus. Oliveira et al. (2008) expand that idea and include a feedback-loop to the whole process in the form of genetic swaps in the organelle (in concrete in the mitochondrial genome) that would, in turn, reinforce the fixation of new slightly deleterious mutations.

A word on the role of population size and effective population size (usually depicted as N_e) in the evolution of organelles and their hosts[14]. This issue is not explicitly addressed in the model presented here but some ideas can be adventured. N_e is used to compensate for non-random reproductive opportunities, differences in gamete production, deviation from the 1:1 proportion in sex ratio, and so on, and represents the amount of he population that is actually transmitting its genetic information to the next generation. It is usually assumed that given haploidy and uniparental inheritance of organelles their N_e drops to about one-quarter of the nuclear one for a diploid population, how ever, Lynch et al. (2006) support the notion that it may not be the case.

As already mentioned before, the extreme reduction in genome numbers that takes place when organelles are transmitted in the mother line could be responsible for the

[14]see for example Lynch (2007), especially chapter 4.

elimination of most of the variation in the organelle genome. How ever, since at least apparently the sorting process of organelle genomes is random, there exist the chance that a non very frequent allele can increase in frequency and slowly get to fixation, especially if the nuclear conditions decreases the effect of selection on the organelle.

In the model presented here there is no reduction during the organelle inheritance, it is assumed that independently of the number of organelle genomes actually inherited to the next generation the proportions would keep as they where in the previous generation. However, I venture to say that, considering the effects that non-symmetrical homologous recombination can cause in the fixations of deleterious mutations, if a strong population reduction would be taking place it would probably increase the its fixation probability. On the other hand, for the nuclear genome the would be a stronger selective pressure to "adopt" the newly transferred allele but since it would be present (at least at the beginning) on low frequencies it could be the case that a successful invasion of the organelle population by the transferred allele does not implies a faster or even at all successful invasion of the nuclear population due to the biological constrains that the mutation could produce.

5. Conclusions

Lateral gene transfer has gained more attention in recent years. As genomic sequences of eukaryotes and prokaryotes accumulate, and new examples of ancient and recent events of gene transfer are confirmed, its importance as one of the sources of genetic novelty becomes more evident. It becomes also clear that even when vertical gene transmission is by far the most common form of genetic inheritance, there should be noticed that in special but not so eccentric cases, horizontal transmission is also a very important route of gene transmission.

In this dissertation I have presented a historical overview of how this process was identified, characterized, dismissed as evolutionarily relevant and later on reconsidered. As part of the latter, I have also presented a recent example of a group of organisms (three species of the *Nasonia* group) in which gene transfer was carefully characterized at the genomic level.

The model I presented is based on the assumption that endosymbitic gene transfer operates as a facilitator of the genome reduction process in the evolution of endosymbionts since it is constantly transporting genetic elements that can be fixed in the nuclear genome, releasing the selective pressure on the organelle. Another facilitator of the process is the existence of multiple genomes in one organelle which could attenuate the deleterious effects of inactivating mutations or the lost of genome segments; on the other hand, the presence of more than one organelle (or endosymbiont in general) in a single cell could also help to overcome the cost of having a defective one, making it possible to be maintained in the population at low frequencies until conditions that favour its expansion are present i.e the defective organelle finds itself sitting on a cell that al ready had included the organellar sequence in the nuclear genome.

Mutations can be fixed in the endosymbionts' population when the net functional defect caused by the mutation is close to zero. According to the results presented in chapter 4 selective neutrality or near-neutrality can be achieved by functional redundancy within the organelle or when the mutation has a very low cost. Functional redundancy can be obtained by the presence of multiple genomes in the organelles. They could able to compensate the functional role of at least one defective genome copy without compromising the integrity of the organelle's function. At other level, the presence of multiple organelles could buffer the damage of a non-functional organelle within the same cell.

However, the previous scenario is always transient, there is always the chance that the affected organellar genome or the affected complete organelle is eliminated from the population by a selective sweep, and in the best of the cases the mutant alleles are kept low frequencies. The best chance for such a mutation to invade the population is that functional redundancy could be achieved by a more permanent mechanism such as gene

transfer to the cell nucleus[1] and this new gene in the nucleus is fixed in the population.

If the fixation of a deleterious (or mildly deleterious) mutation is to occur following the route described before, then it is necessary that the gene (or genes) involved in the function are present and functional in the new location. In the case studied by the model of chapter 4 nuclear activation and population fixation of the tansferred gene needs to take place before the corresponding mutation[2] in the organelle can be fixed in the population.

When an organellar sequence is released in the cytoplasm and gets integrated (and activated) in the nuclear genome, it is likely that the integration, activation and expression of that transferred gene will interfere in some way with the regular function of the nuclear genome. The role of incompatibilities at the nuclear level and at the cytonuclear one have been widely explored in the framework of the speciation theory. The origin of nuclear incompatibilities is, for most of cases believed to be produced by the parallel evolution of two populations that re-establish reproductive contact after a period of separation. Cyto-nuclear incompatibilities have been usually depicted as the product of a conflict between cytoplasmatic endosymbionts (i.e. Wolbachia) and the nuclear genome. I propose that genes that have been transferred from the organelle to the nuclear genome can potentially create both kinds of incompatibilities.

I claim that not only slightly deleterious mutations in the organellar genomes can survive and invade the population but also highly deleterious mutations could survive and, eventually, invade the population. My model shows that if a certain level of heteroplasmy is allowed in the population, it could function as a sort of functional buffering mechanism also for highly deleterious mutations, even those that eliminate whole segments of DNA or inactivate gene function. In this context lateral gene transfer offers a mechanism that creates, and in certain way maintains, the conditions necessary for the compensatory mutations that are needed in the cell nucleus.

It has been proposed that Muller's ratchet could be responsible, at least in part, of the genome reduction, and in general of accumulation of mutations, in organelles and other endosymbionts because the lack of sexual reproduction (and in consequence of recombination between parental genomes) provokes the eventual elimination of the least mutation-loaded class in the population. Other assumptions on the biology of organelles assume a very low level of heteroplasticity, the almost clonality of their genomes and variability obtained mostly via mutation.

The results that I present here sustain the idea that, at least for the case of gene reduction, it is not a necessary condition that Muller's ratchet is operating, and that actually, recombination (specially if it is non-symetrical) may promote the fixation of deleterious or mildly deleterious mutations.

[1] This could also be achieved if other endosymbiont is able to produce a functionally equivalent role.
[2] Either the inactivation or the elimination of the sequence

Bibliography

B. Albert, B. Godelle, A. Atlan, R. De Paepe, and P.H. Guoyon. Dynamics of plant mitochondrial genome: Model of a three-level selection process. *Genetics*, 144:369–382, 1996.

C.A. Allen, M van der Giezen, , and J.F. Allen. Origin, function and transmition of mitochondria. In W. Martin and M. Müller, editors, *Origin of mitochondria and hydrogenosomes*, pages 39–56. Springer-Verlag, 2007.

S F Altschul, W Gish, W Miller, E W Myers, and D J Lipman. Basic local alignment search tool. *J Mol Biol*, 215(3):403–10, October 1990.

S.G. Andersson, A. Zomorodipour, J.O. Andersson, T Sicheritz-Pontën, U.C. Alsmark, R.M. Podowski, A.K. Näslund, A.S. Eriksson, H.H. Winkler, and C.G. Kurland. The genome sequence of *Rickettsia prowazekii* and the origin of mitochondria. *Nature*, 396: 133–140, 1998.

O T Avery, C M Macleod, and M McCarty. Studies on the chemical nature of the substance inducing transformation of pneumococcal types. *J. Exp. Med.*, 79:137–158, 1944.

J W O Ballard and M C Whitlock. The incomplete natural history of mitochondria. *Molecular Ecology*, 13:729–744, 2004.

O.G. Berg and C. G. Kurland. Evolution of microbial genomes: sequence acquisition and loss. *Mol. Biol. Evol.*, 19(12):2265–2276, 2002.

C W Jr Birky. Uniparental inheritance of mitochondrial and chloroplast genes: Mechanisms and evolution. *Proc. Natl. Acad. Sci.*, 92:11331–11338, 1995.

C W Jr Birky. The inheritance of genes in mitochondria and chloroplasts: laws, mechanisms, and models. *Annu. Rev. Genet.*, 35:128–148, 2001.

C W Jr Birky, T Maruyama, and P Fuerst. An approah to population and evolutionary genetic theory for genes in mitochondria and chloroplasts, and some results. *Genetics*, 103:513, 1983.

R. Bock and J.N. Timmis. Reconstructing evolution: gene transfer from plastids to the nucleus. *BioEssays*, 30:556–566, 2008.

T.D. Brock. *The emergence of bacterial genetics*, chapter 9, pages 213–241. Cold Spring Harbor Laboratory Press, 1990.

B. Charlesworth and D. Charlesworth. Rapid fixation of deleterious alleles can be caused by Muller's Ratchet. *Genet. Res. Camb.*, 70:63–73, 1997.

J. Chat, V. Decroocq, R. Decroocq, and R.J Petir. A case of chloroplast heteroplasmy in kiwifruit (*Actinidia deliciosa*) that is not transmitted during sexual reproduction. *The journal of Heredity*, 96:293–300, 2002.

E. Check Hayden. Scandal!, sex-starved and still surviving. *Nature*, 452:678–680, 2008.

L M Cosmides and J Tooby. Cytoplasmic inheritance and intragenomic conflict. *J. theor. Biol*, 89:83–129, 1981.

W. F. Doolittle. You are what you eat: A gene transfer ratchet could account for bacterial genes in eukaryotic nuclear genomes. *TIG*, 14(8):307–311, 1998.

W. F. Doolittle. Phylogenetic classification and the universal tree. *Science*, 284:2124–2126, 1999a.

W. F. Doolittle. Lateral genomics. *TIG*, 15(12):M5–M8, 1999b.

S D Dyall, M T Brown, and P J Johnson. Ancient invasions: From endosymbionts to organelles. *Science*, 304:253–257, 2004.

F. Ekelund and R. Rønn. If you don't need change, maybe you don't need sex. *Nature*, 453:587, 2008.

VV Emelyanov. Constantin Merezhkowsky and the endokaryotic hypotesis. In W. Martin and M. Müller, editors, *Origin of mitochondria and hydrogenosomes*, pages 201–237. Springer-Verlag, 2007.

J Felsenstein. The evolutionay advantage of recombination. *Genetics*, 78:737–756, 1974.

R A Fisher. *The Genetical Theory of Natural Selection*. Clarendon Press, Oxford, 1930.

RD Fleischmann, MD Adams, O White, RA Clayton, EF Kirkness, AR Kerlavage, CJ Bult, JF Tomb, BA Dougherty, JM Merrick, and et al. Whole-genome random sequencing and assembly of *Haemophilus influenzae* Rd. *Science*, 269(5223):496–512, 1995.

J F Fontanari, A Colato, and R S Howard. Mutation accumulation in growing asexual lineages. *Physical Review Letters*, 91(21):218101–1 – 218101–4, 2003.

I. Gordo, A. Navarro, and B Charlesworth. Muller's Ratchet and the pattern of variation at a neutral locus. *Genetics*, 161:835–848, 2002.

E.A. Gradysehev, M. Meselson, and I.R. Archipova. Massive horizontal gene transfer in bdelloid rotifers. *Science*, 320:1210–1213, 2008.

M.W. Gray. Evolution of organellar genomes. *Curr. Opin. Genet.*, 9:678–679, 1999.

M.W. Gray, G Burger, and B.F. Lang. Mitochondria evolution. *Science*, 283:1476–1481, 1999.

Tomasz Grztbowski. Extremely high levels of human mitochondrial DNA heteroplasty in single hair roots. *Electrophoresis*, (21):548–553, 2000.

J Haigh. The accumulation of deleterious genes in a population – Muller's Ratchet. *Theoretical Population Biology*, 14:251–267, 1978.

Einat Hazkani-Covo, Raymond M Zeller, and William Martin. Molecular poltergeists: mitochondrial dna copies (numts) in sequenced nuclear genomes. *PLoS Genet*, 6(2): e1000834, 2010. ISSN 1553-7404.

K. Henze, C. Schnarrenberger, and W. Martin. Endosymbiotic gene transfer: a special case of horizontal gene transfer germane to endosymbiosis, the origins of organelles and the origins of eukaryotes. In M. Syvanen and C. I. Kado, editors, *Horizontal Gene Transfer*, pages 351–361. Academic Press, 2002.

Kirsten Hilgenboecker, Peter Hammerstein, Peter Schlattmann, Arndt Telschow, and John H Werren. How many species are infected with wolbachia?–a statistical analysis of current data. *FEMS Microbiol Lett*, 281(2):215–20, April 2008. ISSN 0378-1097.

R. F. Hoekstra. Why sex is good. *Nature*, 434:571, 2005.

Julie C Dunning Hotopp, Michael E Clark, Deodoro C S G Oliveira, Jeremy M Foster, Peter Fischer, Mónica C Muñoz Torres, Jonathan D Giebel, Nikhil Kumar, Nadeeza Ishmael, Shiliang Wang, Jessica Ingram, Rahul V Nene, Jessica Shepard, Jeffrey Tomkins, Stephen Richards, David J Spiro, Elodie Ghedin, Barton E Slatko, Hervé Tettelin, and John H Werren. Widespread lateral gene transfer from intracellular bacteria to multicellular eukaryotes. *Science (New York, N.Y.)*, 317(5845):1753–6, September 2007. ISSN 1095-9203.

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

C.Y. Huang, M.A. Ayliffe, and J.N. Timmis. Direct measurement of the transfer rate of chloroplast DNA into the nucleus. *Nature*, 422:72–76, 2003.

C.Y. Huang, N. Grünheit, N. Ahmadinejad, J. J. Timmins, and W. Martin. Mutational decay and the age of chloroplast and mitochondrial genomes transferred recently to angiosperm nuclear chromosomes. *Plant Physiology*, 138:1723–1733, 2005.

William Klug and Michael Cummings. *Essentials of genetics*. Prentice Hall, 3rd edition, 1999. ISBN 0-13-080017-1.

A S Kondrashov. Muller's Ratchet under epistatic selection. *Genetics*, 136:1469–1473, 1994.

E. V. Koonin, K. S. Makarova, and L. Aravind. Horizontal gene transfer in prokaryotes: Quantification and classification. *Annu. Rev. Microbiol.*, 55:709–742, 2001.

E. V. Koonin, K. S. Makarova, Y. I. Wolf, and L. Aravind. Horizontal gene transfer and its role in the evolution of prokaryotes. In M Syvanen and C I Kado, editors, *Horizontal Gene Transfer*, pages 277–304. Academic Press, 2002.

C. G. Kurland. Paradigm lost. In J. Sapp, editor, *Microbial phylogeny and evolution Concepts and controversies*, pages 207–223. Oxford University Press, 2005.

D.M. Lonsdale, T. Brears, T.P. Hodge, S.E. Melville, and W.H. Rottmann. The plant mitochondrial genome: Homologous recombination as a mechanismos for generatin heterogeneity. *Phil. Trans. R. Soc. Lond B*, 319:149–163, 1988.

M. Lynch. Mutation accumulation in transfer RNAs: Molecular evidence for Muller's Ratchet in mitochondrial genomes. *Mol. Biol. Evol.*, 13(1):209–220, 1996.

M. Lynch. The origins of eukaryotic gene structure. *Mol. Biol. Evol.*, 23(2):450–468, 2006.

M Lynch, B Koskella, and S Schaack. Mutation pressure and the evolution of organelle genomic architecture. *Science*, 311:1727–1730, 2006.

Michael Lynch. *The origins of genome architecture*. Sinauer Associates, INC., 2007.

Karl Magnacca and Mark Brown. Mitochondrial heteroplasmy and dna barcoding in Hawaiian Hylaeus (nesoprosopis) bees (hymenoptera: Colletidae). *BMC Evolutionary Biology*, 10(1):174, 2010.

A. Marechal and N. Brisson. Recombination and the maintenance of plant organello genome stability. *New Phytologist*, 186:199–317, 2010.

L. Margulis. Symbiogenesis and symbionticism. In L. Margulis and R. Fester, editors, *Symbiosis as a source of evolutionary innovation*, pages 1–14. The MIT Press, 1981.

K Martens and I. Schön. Ancient asexuals: darwinulids not exposed. *Nature*, 453:587, 2008.

W. Martin. Gene transfer from organelles to the nucleus: Frequent and in big chunks. *Proc. Natl. Acad. Sci.*, 100(15):8612–8614, 2003.

W. Martin and R. G. Herrmann. Gene transfer from organelles to the nucleus: how much, what happens, and why? *Plant Physiol.*, 118:9–17, 1998.

W. Martin, B. Stoebe, V. Goremykin, S. Hansmann, M. Hasegawa, and K. V. Kowallik. Gene transfer to the nucleus and the evolution of chloroplasts. *Nature*, 393:162–165, 1998.

C. Medigue, T. Rouxel, A. Vigier, A. Hénaut, and A Danchin. Evidence for horizontal gene transfer in *Escherichia coli* speciation. *J. Mol. Biol.*, (222):851–856, 1991.

N. A. Moran. Accelerated evolution and Muller's Ratchet in endosymbiotic bacteria. *Proc. Natl. Acad. Sci.*, 93:2873–2878, 1996.

Paul Francis Morris, Laura Rose Schlosser, Katherine Diane Onasch, Tom Wittenschlaeger, Ryan Austin, and Nicholas Provart. Multiple horizontal gene transfer events and domain fusions have created novel regulatory and metabolic networks in the oomycete genome. *PLoS One*, 4(7):e6133, 2009.

H. J. Muller. Some genetic aspects of sex. *American Naturalist*, 64:118–138, 1932.

H. J. Muller. Evolution by mutation. *Bull. Am. Math. Soc.*, 64:137–160, 1958.

H. J. Muller. The relation of recombination to mutational advance. *Mutation Research*, 1:2–9, 1964.

M. Müller and W. Martin. The genome of *Rickettsia prowazekii* and some thoughts on the origin of mitochondria and hydrogenosomes. *BioEssays*, 21:377–381, 1999.

A. S. Novozhilov, G. P. Parev, and E. V. Koonin. Mathematical modeling of evolution of horizontally transferred genes. *Mol. Biol. Evol.*, 22(8):1721–1732, 2005.

B. O'Fallon. Population structure, levels of selection, and the evolution of intracellular symbionts. *Evolution*, 62(2):361–373, 2007.

Deodoro C S G Oliveira, Rhitoban Raychoudhury, Dennis V Lavrov, and John H Werren. Rapidly evolving mitochondrial genome and directional selection in mitochondrial genes in the parasitic wasp nasonia (hymenoptera: pteromalidae). *Molecular biology and evolution*, 25(10):2167–80, October 2008. ISSN 1537-1719. doi: 10.1093/molbev/msn159.

M. A. O'Malley and Y. Boucher. Paradigm change in evolutionary microbiology. *Studies in History and Philosophy of the Biological and Biomedical Sciences*, 36:183–208, 2005.

Valerio Parisi, Valeria De Fonzo, and Filippo Aluffi-Pentini. String: finding tandem repeats in dna sequences. *Bioinformatics*, 19(14):1733–8, September 2003.

Stephanie A. Pearl, Mark E. Welch, and David E. McCauley. Mitochondrial heteroplasmy and paternal leakage in natural populations of silene vulgaris, a gynodioecious plant. *Molecular Biology and Evolution*, 26(3):537–545, 2009.

M. Pop, S.L. Salzberg, and M. Shumway. Genome sequence assembly:algorithms and issues. *IEEE Computer*, 35(7):47–54, 2002.

Mary Anne Pultz and David S Leaf. The jewel wasp nasonia: querying the genome with haplo-diploid genetics. *Genesis*, 35(3):185–91, March 2003. ISSN 1526-954X.

David M Rand, Robert a Haney, and Adam J Fry. Cytonuclear coevolution: the genomics of cooperation. *Trends in ecology & evolution (Personal edition)*, 19(12):645–53, December 2004. ISSN 0169-5347. doi: 10.1016/j.tree.2004.10.003.

Joan Roughgarden. *The genial gene, deconstructing darwinian selfishness*. University of California Press, 1st edition, 2009.

Mary E Rumpho, Jared M Worful, Jungho Lee, Krishna Kannan, Mary S Tyler, Debashish Bhattacharya, Ahmed Moustafa, and James R Manhart. Horizontal gene transfer of the algal nuclear gene psbo to the photosynthetic sea slug elysia chlorotica. *Proc Natl Acad Sci U S A*, 105(46):17867–71, 2008.

L. Sagan. On the origins of mitosing cells. *J. Theor. Biol*, 14:225–274, 1967.

F. Sanger, A. R. Coulson, G. F. Hong, D. F. Hill, and G. B. Petersen. Nucleotide sequence of bacteriophage [lambda] dna. *Journal of Molecular Biology*, 162(4):729–773, 1982.

J.P Saville, Y. Kohli, and J.B. Anderson. mtDNA recombination in a natural population. *Proc. Natl. Acad. Sci.*, 95:1331–1335, 1998.

I Schön and K Martens. No slave to sex. *Proc. R. Soc. Lond B*, 270:827–833, 2003.

R. M. Schwartz and M. O. Dayhoff. Origins of prokaryotes, eukaryotes mitochondria and chloroplasts. *Science*, 199:395–403, 1978.

M. A. Selosse, B. Albert, and B. Godelle. Reducing the genome size of organelles favours gene transfer to the nucleus. *TREE*, 16(3):135–141, 2001.

R Sorek, Y Zhu, C Creevey, M P Francino, P Bork, and E M Rubin. Genome-wide experimental determination of the barriers to horizontal gene transfer. *Science*, 318: 1449–1452, 2007.

S. Stegemann and R. Bock. Experimental reconstruction of functional gene transfer from the tobacco plastid genome to the nucleus. *The plant cell*, 18:2869–2879, 2006.

S. Stegemann, S. Hartmann, S. Ruf, and R. Bock. High-frequency gene transfer from the chloroplast genome to the nucleus. *PNAS*, 100:8828–8833, 2003.

E. Suárez-Díaz and V.H. Anaya-Muñoz. History, objectivity, and the construction of molecular phylogenies. *Studies in History and Philosophy of Biological and Biomedical Sciences*, 39:451–468, 2008.

E. Suárez-Díaz and V.H. Anaya-Muñoz. La ansiedad metodológica como búsqueda de objetividad: la construcción de filogenias moleculares. *Ludus Vitalis*, 17(31):33–64, 2009.

P. E. Thorsness and T. D. Fox. Escape of DNA from mitochondria to the nucleus in *Saccharomyces cerevisiae*. *Nature*, 346:376–379, 1990.

B. Thyagarajan, R.A. Padua, and C. Campbell. Mammalian mitochondia possess homologous DNA recombination activity. *The Journal of Biological Chemistry*, 271(44): 27536–27543, 1996.

K.P. Van Winkle-Swift and C. W. Jr. Birky. The non-reciprocality of organelle gene recombination in *Clamydomonas reinhardtii* and *Saccharomyces cerevisiae*. *Molec. gen. Genet.*, 166:193–209, 1978.

J B Walsh. Intracellular selection, conversion bias and the expected substitution rate of organelle genes. *Genetics*, 130:939–946, 1992.

J. H. Werren, J. Gadau, L.W. Beukeboom, C. Desplan, J. Lynch, R. Rivers, S. Richards, and L. van de Zande. Proposal to sequence the *nasonia* genome. 2004.

John H. Werren, Stephen Richards, Christopher A. Desjardins, Oliver Niehuis, Jürgen Gadau, John K Colbourne, Leo W Beukeboom, Claude Desplan, Christine G. Elsik, Cornelis J. P. Grimmelikhuijzen, Paul Kitts, Jeremy Lynch, Terence Murphy, Deodoro C. S. G. Oliveira, Christopher D. Smith, Louis van de Zande, Kim C. Worley, Evgeny M. Zdobnov, Maarten Aerts, Stefan Albert, **Victor H. Anaya**, Juan M. Anzola, Angel R. Barchuk, Susanta K. Behura, Agata N. Bera, May R. Berenbaum, Rinaldo C. Bertossa, Márcia M G. Bitondi, Seth R. Bordenstein, Peer Bork, Erich Bornberg-Bauer, Marleen Brunain, Giuseppe Cazzamali, Lesley Chaboub, Joseph Chacko, Dean Chavez, Christopher P. Childers, Jeong-Hyeon Choi, Michael E. Clark, Charles Claudianos, Rochelle A. Clinton, Andrew G. Cree, Alexandre S. Cristino, Phat M. Dang, Alistair C. Darby, Dirk C. de Graaf, Bart Devreese, Huyen H. Dinh, Rachel Edwards, Navin Elango, Eran Elhaik, Olga Ermolaeva, Jay D. Evans, Sylvain Foret, Gerald R. Fowler, Daniel Gerlach, Joshua D. Gibson, Donald G. Gilbert, Dan Graur, Stefan Gründer, Darren E. Hagen, Yi Han, Frank Hauser, Dan Hultmark, Henry C. Hunter IV, Gregory D. D. Hurst, Shalini N. Jhangian, Huaiyang Jiang, Reed M. Johnson, Andrew K. Jones, Thomas Junier, Tatsuhiko Kadowaki, Albert Kamping, Yuri Kapustin, Bobak Kechavarzi, Jaebum Kim, Jay Kim, Boris Kiryutin, Tosca Koevoets, Christie L. Kovar, Evgenia V. Kriventseva, Robert Kucharskiand Heewook Lee, Sandra L. Lee, Kristin Lees, Lora R. Lewis, David W. Loehlinand John M. Logsdon Jr, Jacqueline A. Lopez, Ryan J. Lozado, Donna Maglott, Ryszard Maleszka, Anoop Mayampurath, Danielle J. Mazur, Marcella A. McClure, Andrew D. Moore, Margaret B. Morgan, Jean Muller, Monica C. Munoz-Torres, Donna M. Muzny, Lynne V. Nazareth, Susanne Neupertand Ngoc B. Nguyen, Francis M F. Nunes, John G. Oakeshott, Geoffrey O. Okwuonu, Bart A. Pannebakker, Vikas R. Pejaver, Zuogang Peng, Stephen C. Pratt, Reinhard Predel, Ling-Ling Pu, Hilary Ranson, Rhitoban Raychoudhury, Andreas Rechtsteiner, Justin T. Reese, Jeffrey G. Reid, Megan Riddle, Hugh M. Robertson, Jeanne Romero-Severson, Miriam Rosenberg, Timothy B. Sackton, David B. Sattelle, Helge Schlüns, Thomas Schmitt, Martina Schneider, Andreas Schüler, Andrew M. Schurko, David M. Shuker, Zilá L P. Simões, Saurabh Sinha, Zachary Smith, Victor Solovyev, Alexandre Souvorov, Andreas Springauf, Elisabeth Stafflinger, Deborah E. Stage, Mario Stanke, Yoshiaki Tanaka, Arndt Telschow, Carol

Trent, Selina Vattathil, Eveline C. Verhulst, Lumi Viljakainen, Kevin W. Wanner, Robert M. Waterhouse, James B. Whitfield, Timothy E. Wilkesand Michael Williamson, Judith H. Willis, Florian Wolschin, Stefan Wyder, Takuji Yamada, Soojin V. Yi, Courtney N. Zecher, Lan Zhang, and Richard A. Gibbs. Insights from three genomes of *Nasonia*: Insect natural enemies and emerging genetic model. *Science*, 327:343–348, 2010.

Y. I. Wolf, I. B. Rogozin, N. V. Grishin, and E. V. Koonin. Genome trees and the tree of life. *TRENDS in Genetics*, 18(9):472–479, 2002.

K. H. Wolfe, W. H. Li, and P. M. Sharp. Rates of nucleotide substitution vary greatly among plant mitochondrial, chloroplast, and nuclear DNAs. *Proc. Natl. Acad. Sci.*, 84:9054–9058, 1987.

A. Yamauchi. Rate of gene transfer from mitochondria to nucleus: Effects of cytoplasmic inheritance system and intensity of intracellular competition. *Genetics*, 171:1387–1396, 2005.

6. Appendix

A. Supplementary results for Chapter 4

In this appendix I am including some supplementary results that complement those presented in chapter 4. Following the same order used in that chapter, I will present and briefly discuss some results of the implementation of the organellar and nuclear portions of the model and finally results obtained from the implementation of the complete version

Organelle's model supplementary results

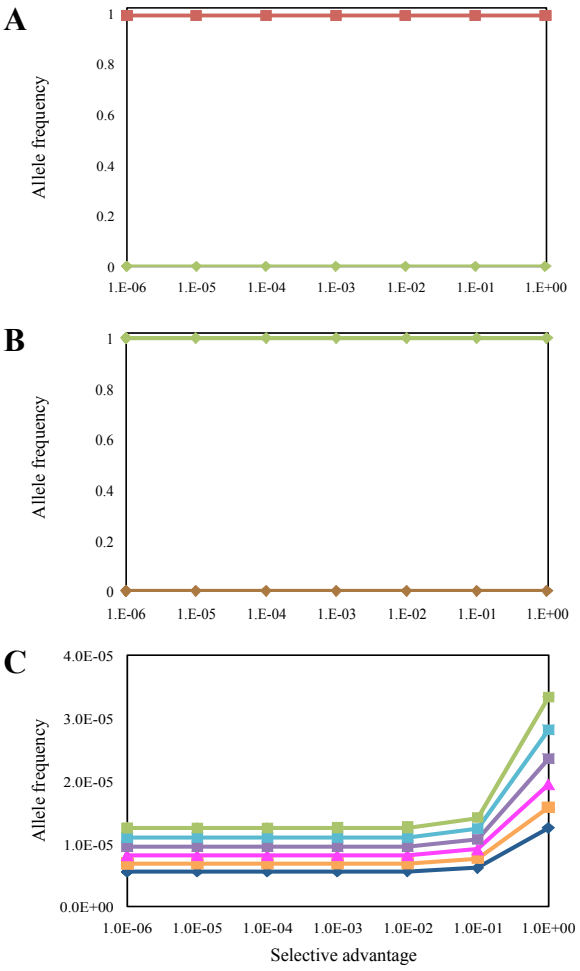

Figure A.1.: **Equilibrium frequencies if the organellar alleles under different values of w, s and non-symmetrical homologous recombination strongly biased against the transferred allele.** For all cases recombination was fixed at $\beta_{1,3} = \beta_{2,3} = 0.1$, $\beta_{3,1} = \beta_{3,2} = 0.9$. Panel A: Effects on p; the red line stands for fitness values of $0.5 \geq w_{or} \leq 0.9999$ whilst the green line stands for $w_{or} = 1$. Panel B: Effects on q; brown line stands for $0.5 \geq w_{or} \leq 0.99$ and the green line for $0.999 \geq w_{or} = 1$. Panel C: Effects on q, note that a different scale is being used. Fitness values: $w_{or} = 0.5$ in dark blue, $w_{or} = 0.6$ in orange, $w_{or} = 0.7$ in pink, $w_{or} = 0.8$ in purple, $w_{or} = 0.9$ in light blue, $w_{or} = 0.99$, $w_{or} = 0.999$, $w_{or} = 0.9999$ and $w_{or} = 1.0$ in green.

When recombination has a $9:1$ bias against the transferred allele (figure A.1), this

allele remains for all values at frequencies close to ν, and the inactivated allele has two equilibrium values: either it almost reaches a frequency of 1 (green line) or it disappears of the population. In all cases, except for $w_{or} = 1.0$, the wild-type allele stays close to 1.

In contrast to the previous case, as shown in figure A.2, the results of a $9:1$ bias towards the transferred are basically the opposite; the wild-type and inactivated alleles are lost from the population for all parameters except when $w_{or} = 0.5$ (blue line), and the transferred gene gets fixed for all other values.

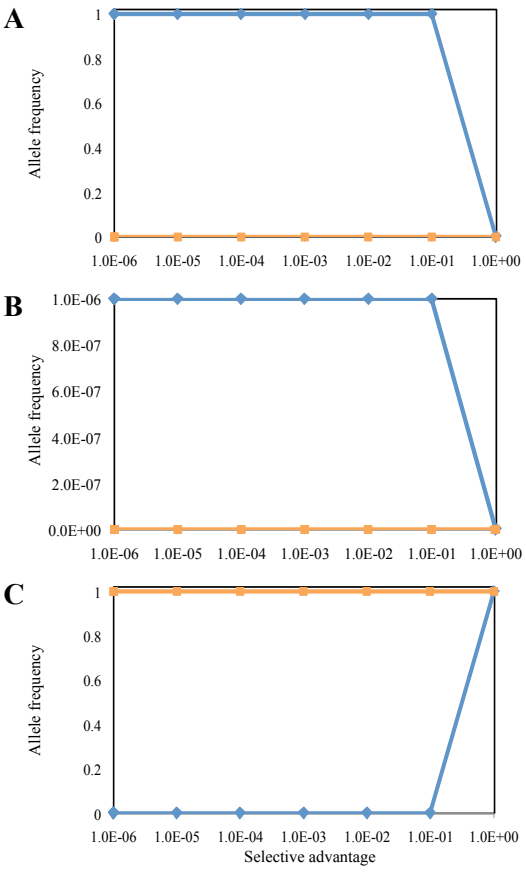

Figure A.2.: **Equilibrium frequencies of the organellar alleles for different values of w, s and non-symmetrical homologous recombination strongly biased towards the transferred allele.** Panel A: Effects on p . Panel B: Effects on q note that a different scale is being used. Panel C: Effects on r. For all cases recombination was fixed at 0.9 for $\beta_{1,3}$ and $\beta_{2,3}$, 0.1 for $\beta_{3,1}$ and $\beta_{3,2}$; fitness values are represented with a blue line for $w_{or} = 0.5$ and an orange line for $w_{or} \geq 0.6$

When non-symmetrical homologous recombination is acting against the allele the transferred allele it is possible to observe some interesting results; if the bias against the shorter variant is relatively mild ($\beta_{1,3} = \beta_{2,3} = 0.45$, the possibility of polymorphisms emerge, even if some of the alleles are present at low frequencies. As seen in figure A.3 the for various w_{or} values there are equilibria at which the inactivated allele is present (panel B). It can also be said that the transferred allele has a broader spectrum of parameters in which it can invade the population an fixate displacing the other two variants. In any case, if $w_{or} = 1$ fixation of the transferred allele takes place for all activation values

The effect of minor recombination bias towards the short variant is in some cases

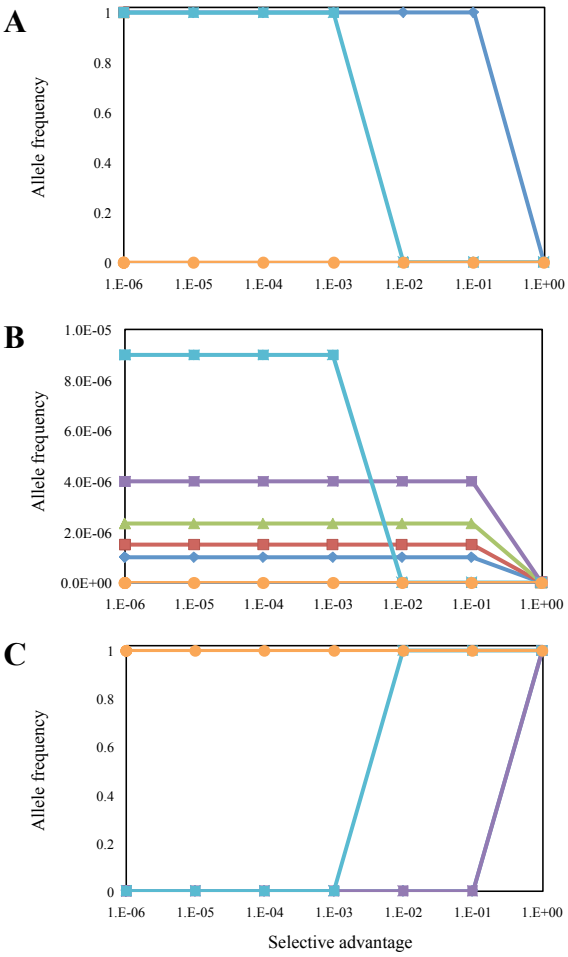

Figure A.3.: **Equilibrium frequencies of the organelle's alleles under the effects of w, s and non-symmetrical homologous recombination slightly biased against the transferred allele.** The effects on p, q and r are shown in panels A, B and C respectively. Recombination was fixed at $\beta_{1,3} = \beta_{2,3} = 0.45, \beta_{3,1} = \beta_{3,2} = 0.55$ for all cases; fitness values are presented in different colours:$w_{or} = 0.5$ in dark blue, $w_{or} = 0.6$ in red, $w_{or} = 0.7$ in green, $w_{or} = 0.8$ in purple, $w_{or} = 0.9$ in light blue, and $w_{or} = 1.0$ in orange.

unexpected, as can be seen in figure A.4, it is not the case that p increases as a result of the recombination bias but q does. If the inactivated allele does not has to pay any fitness cost or very low ones (panel B); then recombination is actually influencing its fixation in the population

Complete model, supplementary results

For completeness, simulations with different values of u and ν were run. For the nuclear part of the model only the results of allele A are shown (figure A.5 panel A), the results of B and b do not depend on the frequencies of migration or mutation and therefore not included. For the nuclear part, the equilibrium points of p, q and r are shown in panels B, C y D respectively. As expected, the equilibria values of the frequencies if the inactivated and transferred alleles increase when u or ν increase as well. For the wild-type allele there is no difference and it remains as the predominant allele for all cases.

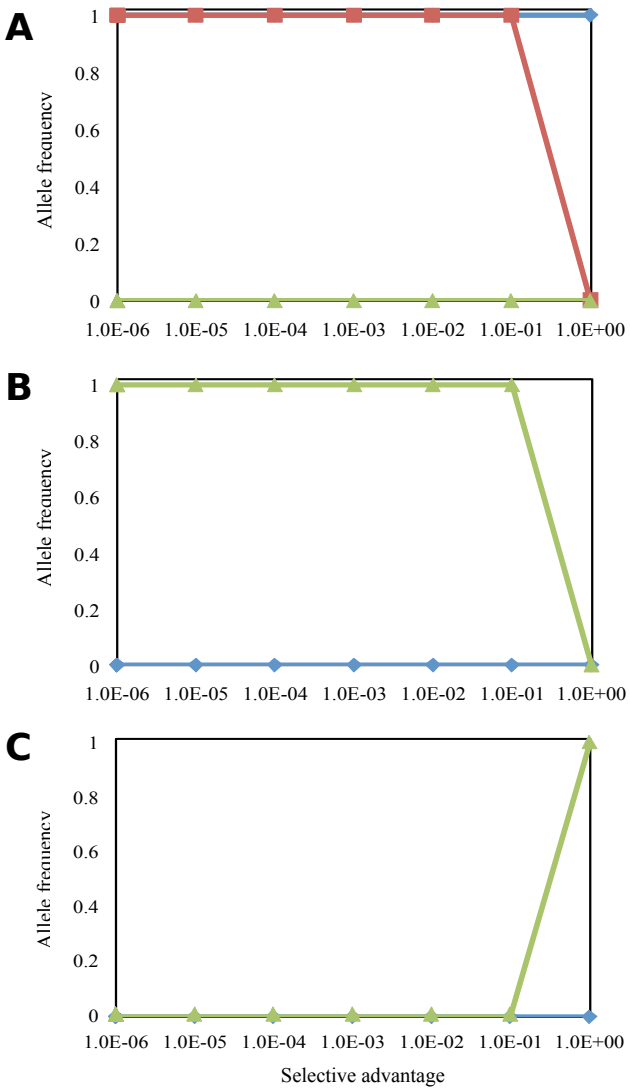

Figure A.4.: **Equilibrium frequencies of the organellar alleles under the effects of w, s and non-symmetrical homologous recombination slightly favouring the transferred effect.** Panel A: Effects on p the fitness values were set as $w_{or} = 0.5$ (blue line), $0.6 \geq w_{or} < 1$ (red line) and $w_{or} = 1$ (green line). Panel B: Effects on q, the fitness values were set as $0.5 \geq w_{or} \leq 0.999$ (blue line), $0.9999 \geq w_{or}$ (green line). Panel C: Effects on r the fitness values were set as $w_{or} = 0.5$ (blue line), $0.6 \geq w_{or} \leq 1$ (green line). For all cases the recombination values were set to $0.45 = \beta_{1,3} = \beta_{2,3}$ and $0.55 = \beta_{3,1} = \beta_{3,2}$.

Coming back to the effects of biased recombination towards the transferred allele (figures 4.7, 4.8, A.6 and A.7). For all scenarios studied, the allele that increments its frequency as a result of the recombination bias is the inactivated one. However, those differences deppend on the strength of the recombination bias: the more it favours the short variant the lower its equilibrium frequency stays. That is, for small deviations (0.05) from the strictly symmetrical homologous recombination, the transferred allele has chances to be present, but for strong deviations it practically disappears from the population. This result is interesting because it implies that gene conversion does not have to be the predominant form of recombination or take place very often to produce important effects.

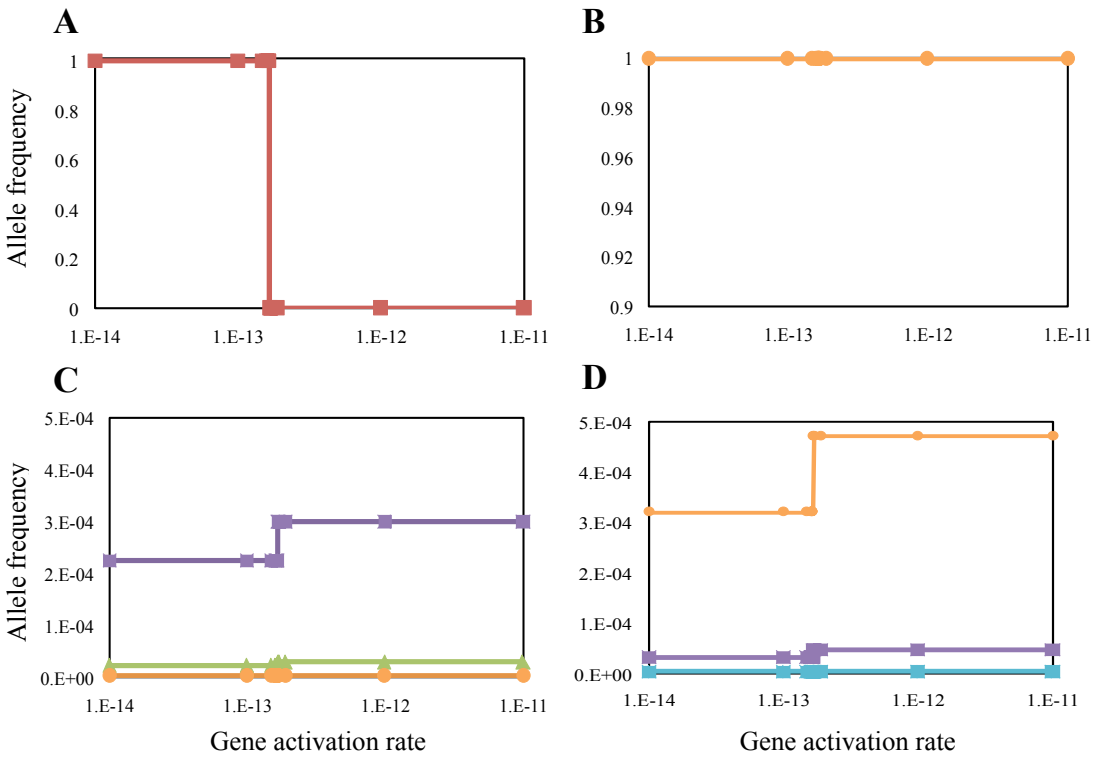

Figure A.5.: **Nuclear and organellar equilibria under different u and ν values.** The equilibria of allele A follows the same patter for all values used, with a critical activation rate of $1.66 * 10^{-13}$. Wild-type allele in the organelle (B) remains close to a frequency of 1 for all values. The equilibrium frequencies of the inactivated allele (C) increments as product of increasing u values. The same behaviour is seen for the case of the transferred allele (D) as product of increasing factors of ν. Mutation rates: $1 * 10^{-7}$ red line, $1 * 10^{-5}$ green line and $1 * 10^{-4}$ purple line. Transport rates: $1 * 10^{-6}$ light blue line, $1 * 10^{-4}$ orange line.

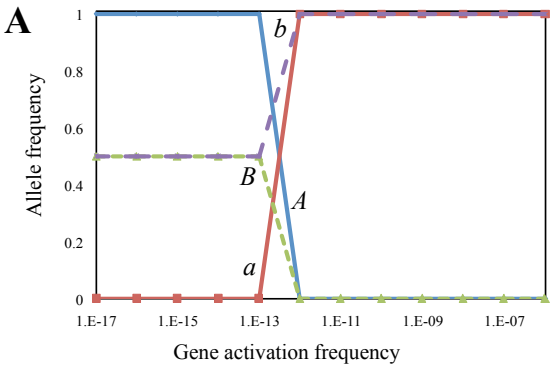

Figure A.6.: **Equilibrium frequencies with recombination slightly biased against the transferred allele.**If the recombination bias is operating against the transferred allele at a low ratio, the nuclear alleles maintain the same behaviour described in other cases: a and b are fixed when a critical activation rate is reached and their counterparts get extinct. In the organelle, there is coexistence of inactivated and transferred alleles, but the latter is present at lower frequencies.

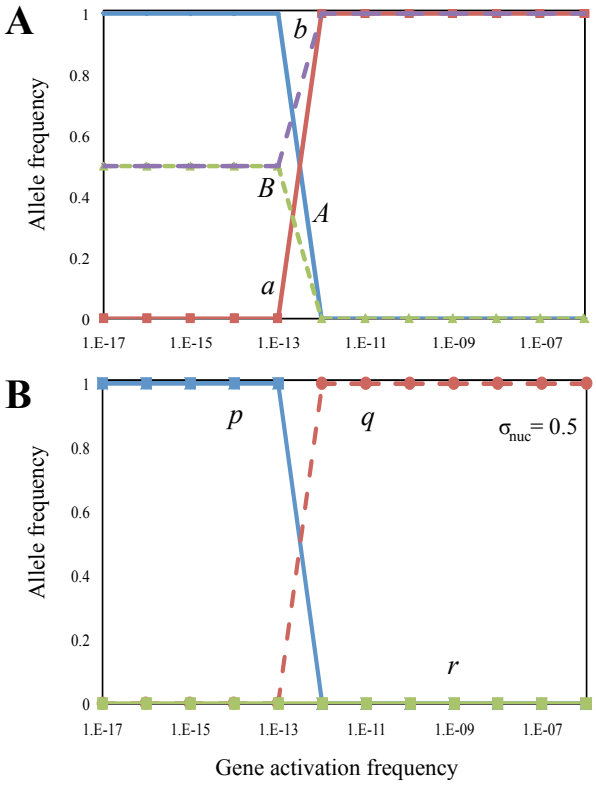

Figure A.7.: **Equilibrium frequencies with recombination slightly biased towards the transferred allele (complete).** When recombination is slightly biased towards r ($\beta_{1,3} = \beta_{2,3} = 0.55$) the general result obtained is quite similar to the one produced by strong deviation towards the transferred allele. In panel A are presented the allele's frequencies in the nucleus for all the incompatibility scenarios simulated; alleles A and B (red and blue respectively) become extinct when the activation frequencies are above $1 * 10^{-13}$ while a and b (purple and green) are fixed. Panel B shows effects of nuclear incompatibilities ($\sigma_{nuc} = 0.5$) in the organelle; q increases, p decreases and r is present at low values ($1 * 10^{-3}$).

List of Figures

List of Tables

Acknowledgements

The completion of this project would not be possible without the financial support of the *Deutsche Akademischer Austausch Dienst* (DAAD), and the *Sonderforschungsbereich* 618 *Theoretische Biologie: Robustheit, Modularität und evolutionäres Design lebender Systeme*.

I am especially grateful to Peter Hammerstein for accepting me in his research group and supporting and providing me guidance throughout this time. I would also like to thank Hans-Jöerg Rheinberger for giving the first impulse to this adventure and remained such a nice companion up to its end. I am also grateful to Daniel Piñero for his careful review of this manuscript and all the suggestions to improve it.

Finally, I am in debt with all the friends and family who always found a way to be present independently of the distance.

Selbständigkeitserklärung

Hiermit erkläre ich, die vorliegende Arbeit selbständig ohne fremde Hilfe verfaßt und nur die angegebene Literatur und Hilfsmittel verwendet zu haben.

Berlin, den 28.04.2011 Víctor Hugo Anaya Muñoz