Genome Evolution of Neurospora tetrasperma

YU SUN
Abstract


In this thesis work, I have used a comparative genomics approach to study a fungal model organism, *Neurospora tetrasperma*. My specific focus has been on genomic introgression, intron evolution, chromosomal structural rearrangements and codon usage. All of the studies are based on large-scale dataset generated by next-generation sequencing technology (NGS), combined with other techniques, such as Optical Mapping. In the introgression study, we detected large-scale introgression tracts in three *N. tetrasperma* lineages, and the introgression showed allele-specific and chromosomal-specific pattern. In the study of introns, we found indications of mRNA mediated intron loss and non-homologous end joining (NHEJ) mediated intron gains in *N. tetrasperma*. We found that selection is involved in shaping intron gains and losses, and associated with intron position, intron phase and GC content. In the study of chromosomal structural rearrangements, we found a lineage specific chromosomal inversion pattern in *N. tetrasperma*, which indicates that inversions are unlikely to associate with the origin of the suppressed recombination and the mating system transition in *N. tetrasperma*. The result suggests inversions are the consequences, rather than the causes, of suppressed recombination on the mating-type chromosome of *N. tetrasperma*. In the final study, analyses of codon usage indicated that the region of suppressed recombination in *N. tetrasperma* is subjected to genomic degeneration, and selection efficiency has been much reduced in this region.

**Keywords:** Neurospora tetrasperma, next-generation sequencing, introgression, intron, chromosomal inversion, codon usage

Yu Sun, Uppsala University, Department of Ecology and Genetics, Evolutionary Biology, Norbyväg 18 D, SE-752 36 Uppsala, Sweden.

© Yu Sun 2013

ISSN 1651-6214
urn:nbn:se:uu:diva-207330 (http://urn.kb.se/resolve?urn=urn:nbn:se:uu:diva-207330)
Dedicated to my family
Xi Fu
Chunmin Sun, Li Zhang
List of Papers

This thesis is based on the following papers, which are referred to in the text by their Roman numerals.


* These authors contributed equally to this work.

Reprints were made with permission from the respective publishers.
Paper not included in the thesis:


Abbreviations

DNA  deoxyribonucleic acid
RNA  ribonucleic acid
FGSC Fungal Genetic Stock Center
JGI DOE Joint Genome Institute
mat  mating type
NGS  next-generation sequencing
Introduction

The number of sequenced genomes has grown exponentially in the past decades. By using the enormous amount of available genomic data, a new research approach, comparative genomics, has drawn the attention of scientists, since it provides a new and unprecedented way to study the evolutionary history of organisms. In addition, the genomic research builds an information platform for many follow up laboratory experiments, such as finding the functional association between traits and genomic DNA, thus is critically important for the overall scientific progress.

Fungal species have proven to be excellent models to study comparative genomics due to short growth and generation times, relatively small genomic sizes compared to many other eukaryotic species, and ease of genome assembly and analyses due to low repeat content and so on. In this study, we have used *Neurospora tetrasperma*, a fungal species from Ascomycota phylum, to study the evolutionary history of the species and the cause and consequences of genomic changes. I hope this work contributes to the scientific community and in deepening the understanding of fungal genome evolution.
Comparative genomics

Advances in DNA genome sequencing progressed dramatically in the past few decades. The first DNA genome bacteriophage phi X174 was sequenced in 1977, with the total length of 5,375 base pair (bp) (Sanger, Air et al. 1977). In 1995, the first bacteria genome, *Haemophilus influenza*, was sequenced by The Institute for Genomic Research (TIGR) (Fleischmann, Adams et al. 1995), and one year later, the first sequenced eukaryotic genome, *Saccharomyces cerevisiae*, was released (Gofféau, Barrell et al. 1996). In 2001, two research teams published the first draft of human genomes (Lander, Linton et al. 2001; Venter, Adams et al. 2001). Ever since then, the number of sequenced and published genomes grow exponentially, and today thousands of genomes are deposited in various public databases (Lagesen, Ussery et al. 2010).

Comparative genomics is a newly arisen research field that uses computational approaches to compare the genomic DNA sequences of different species or strains. By this approach, huge amount of information can be generated to infer the evolutionary history and functional constraints of genomic characters (Hardison 2003). Numerous research topics, such as coding sequence or repetitive elements evolution, can be address within the comparative genomics framework. In my thesis work, I have mainly focused on the following aspects of comparative genomics of *N. tetrasperma*: introgression, chromosomal structural rearrangement, introns and codon usage, which are mostly related to the papers.

**Introgression and genome evolution**

Introgression, also known as introgressive hybridization, is a genetic process in which two species mate and produce a hybrid offspring, and the offspring repeatedly backcrosses to one of the parental species. The process results in gene flow from one species to another, and is an important evolutionary process shaping species diversification and adaptation (Rieseberg 1993; Baack and Rieseberg 2007). Introgression is the most characterized mechanism for gene flow between species in eukaryotes, but some other mechanisms may also contribute in the inter-specific gene transfer (Roper, Ellison et
Introgression is well-established in many plant systems, and estimated to occur in up to 25% of plant species (Mallet 2005). In the animal kingdom, introgression occurs much less frequently, but it is also recognized as an important factor driving genome evolution, speciation and adaptation (Feldman, Brodie et al. 2009; Kulathinal, Stevison et al. 2009; Green, Krause et al. 2010; Hayden, Pulcini et al. 2010).

In the fungal kingdom, introgression has been found in multiple species, e.g., species of Coniophora, Microbotryum, Aspergillus, and Stagnospora (Friesen, Stukenbrock et al. 2006; Kauserud, Svegarden et al. 2007; Le Gac, Hood et al. 2007; Slot and Rokas 2011). However, most of these studies are based on multi-locus sequence data, thus unable to reveal the actual size, abundance and distribution of introgression tracts. The growing number of genomic datasets from the fungal kingdom provides us a great opportunity to research on introgression process in fungal genomes.

Intron evolution

Introns are widely observed in nuclear genomes of eukaryotes (Berget, Moore et al. 1977; Chow, Gelinas et al. 1977; Evans, Fraser et al. 1977; Goldberg, Schwartz et al. 1977), but the functional role of introns and the factors affecting their turnover mechanism, rate, selection and population frequency are only beginning to be understood (Chorev and Carmel 2012; Rogozin, Carmel et al. 2012). For example, many intron loss and gain mechanisms have been proposed, such as reverse-transcribed mRNA mediated, non-homologous end joining mediated, intron transposition, transposon insertion and so on (Fink 1987; Li, Tucker et al. 2009; Yenerall and Zhou 2012), but the relative contribution of these mechanisms to the overall intron turnover pattern is largely unclear. In addition, only a few intron studies have been conducted on the population level, such as in Daphnia and Fusarium (Li, Tucker et al. 2009; Croll and McDonald 2012), thus how selection affects intron dynamics in eukaryotic genomes introns remains unclear. The growing availability of large-scale genomic datasets constitutes an important step towards addressing these fundamental issues in genome biology.

Chromosomal structural rearrangements

Chromosomal structural rearrangements include deletions, insertions, inversions and translocations. Chromosomal inversion involves the breakage of a segment of genomic sequences that reversely links to the original chromosome (Gardner and Sutherland 2004). Inversions are expected to play an important role for many evolutionary processes, such as speciation (Ayala
and Coluzzi 2005), by accumulation of reproductively incompatible genes (Noor, Grams et al. 2001), and the evolution of sex chromosomes (Lahn and Page 1999), by recombination suppression between the sex-determining locus and linkage for genes under sexual-antagonistic selection (Rice 1987; Charlesworth 1991). Taken together, inversions are expected to be adaptive and spread in the population when they prevent recombination from breaking up any combination of alleles with a fitness advantage, either in an ecological or sexual setting (Kirkpatrick 2010).

Synonymous codon usage in genome

Synonymous codons are not randomly used by many species. Natural selection increases the frequency of a subset of synonymous codons for more efficient and accurate translation, such as in *Caenorhabditis elegans*, *Drosophila* and *Arabidopsis* (Duret and Mouchiroud 1999; Duret 2000; Stoletzki and Eyre-Walker 2007). It is suggested that codon usage bias is under weak selection, based on the preferred/nonpreferred codon frequency comparison between *Drosophila melanogaster* and *Drosophila simulans* (Akashi 1994; Akashi 1995). Thus changes in effective population size ($N_e$) could strengthen or degenerate the codon usage biases in species. *Neurospora tetrasperma* is an emerging model organism for the study of early evolution of sex chromosomes (see section “Research in *Neurospora tetrasperma*”) (Gallegos, Jacobson et al. 2000; Menkis, Jacobson et al. 2008; Whittle and Johannesson 2011). A large suppressed recombination region is located in the middle of the mating-type chromosome (Menkis, Jacobson et al. 2008; Ellison, Stajich et al. 2011). The suppressed recombination could lead to reduced $N_e$ in the chromosomal region, and is likely to have impacts on codon usage biases.
Neurospora as a model system

Neurospora was first recognized as red bread mold, a contamination in French bakeries, as early as in the 19th century (Davis 2000). The first scientific paper systematically characterizing Neurospora as a genetic research model was published in 1927 (Shear 1927). In 1958, Beadle and Tatum were awarded the Nobel Prize in Physiology and Medicine based on their work in Neurospora and the “one-gene one-enzyme” hypothesis (Beadle and Tatum 1941). Since then, research in Neurospora, especially for *N. crassa*, has expanded significantly in the fields of genetics, enzyme regulation, metabolic pathways and circadian rhythms (Pittendrigh, Bruce et al. 1959; Jennings 1995; Davis 2000). In 2003, the first Neurospora genome (*N. crassa*) was fully sequenced by the Broad Institute (Galagan, Calvo et al. 2003), indicating that advanced sequencing technology had brought the old model organism into a new genomic era. In recent years, after two released genomes and a few phylogenetic and reproductive studies, Neurospora started to be used as a model system for the study of evolutionary biology (Dettman, Jacobson et al. 2003; Dettman, Jacobson et al. 2003; Ellison, Stajich et al. 2011).

Neurospora is a genus belonging to the Ascomycota. Species within the genus encompass three different mating systems: heterothallism (self-sterility), homothallism (self-fertility) and pseudohomothallism (self-fertility). For the heterothallic species, such as *N. crassa*, the ascospore contains either of the mating-type nuclei (*mat A* or *mat a*) (Figure 1), and mating between strains of two opposite mating-types are needed in order to complete the sexual cycle. For the homothallic species, such as *N. africana*, nuclei in the ascospores have no mating-type difference, and sexual reproduction can be done with themselves. Finally, for pseudohomothallic species, each ascospore contains nuclei of both mating-types, resulting in self-fertility (Figure 1) (Dodge 1927; Shear 1927; Raju 1994; Merino, Nelson et al. 1996). A Neurospora phylogenetic tree based on seven nuclear loci showed the homothallic, heterothallic and pseudohomothallic branches are mixed and nested with each other in the genus. It was suggested that the pattern is mainly shaped by homothallic species, which have evolved independently at least six times from a heterothallic ancestor during the evolution of Neurospora genus (Nygren, Strandberg et al. 2011). The hypothesis was supported by an investigation of the *mat*-locus structure of species with different mating systems (Gioti, Mushegian et al. 2012). Two pseudohomothallic species *N. tetras-
perma and *N. tetraspora* evolved independently in the evolutionary history: *N. tetrasperma* evolved from a heterothallic ancestor whereas *N. tetraspora* evolved from a homothallic ancestor (Nygren, Strandberg et al. 2011). Overall, Neurospora can be used as an excellent model to study the mechanism and consequences of reproductive mode evolution and mating-system transition.

**Figure 1.** Ascospore development for the heterothallic species *N. crassa* and the pseudohomothallic species *N. tetrasperma*. Red dots represent nuclei with *mat A* allele at the *mat*-locus, and blue dots represent nuclei with *mat a* allele in Neurospora. (Figure from Raju and Perkins 1994; modified by Jacobson, personal communication).

*N. crassa* is the most studied species in the genus, and has been widely used as an eukaryotic model organism in genetics, biochemistry and molecular studies (Davis 2000). The genome of *Neurospora crassa* was sequenced by the Broad Institute in 2003 (http://www.broadinstitute.org/annotation/genome/neurospora/MultiHome.html). It was the third sequenced fungal species after *Saccharomyces cerevisiae* and *S. pombe*. The *N. crassa* genome is ~ 41 million base pairs (Mbp), and a total of ~10,000 protein coding genes are distributed on the seven linear chromosomes (Galagan, Calvo et al. 2003). A striking feature of the *N. crassa* genome is that almost no sequence duplications can be found within the genome, due to the repeat-induced point mutation (RIP) mechanism (Selker, Cambareri et al. 1987;
Galagan, Calvo et al. 2003). RIP is a molecular process that accumulates GC to AT transition in genomes (Selker, Cambareri et al. 1987). It was first found in *N. crassa*, and then identified in many other fungal species, such as *Podospora anserina*, *Magnaporthe grisea* and *Leptosphaeria maculans* (Graia, Lespinet et al. 2001; Ikeda, Nakayashiki et al. 2002; Idnurm and Howlett 2003). In *N. crassa*, RIP causes efficient GC to AT mutations within the duplicated sequences, which have sequence length > 400 bp and sequence identity > 80% (Cambareri, Singer et al. 1991; Watters, Randall et al. 1999). One important consequences of RIP is that very few numbers of gene families can be found in the *N. crassa* genome, thus genome innovations from this approach is greatly restricted (Galagan, Calvo et al. 2003). On the other hand, the genome is compensated in terms of security and stability, by using RIP as an efficient mechanism for removing mobile DNA and selfish repetitive elements in the genome (Selker 1990).
Neurospora tetrasperma is a pseudohomothallic species that recently diverged from an ancestral heterothallic Neurospora ~ 1 million years ago, and a reproductive mode transition from heterothallism to pseudohomothallism accompanied the speciation event (Corcoran, Dettman et al. 2013). A large suppressed recombination region (~7.5 Mbp) is located in the center of the mating-type (mat) chromosome, with free recombining region in both chromosomal ends (pseudoautosomal regions) (Gallegos, Jacobson et al. 2000; Menkis, Jacobson et al. 2008). The suppressed recombination region is ~7.5 Mbp and covers ~80% of the chromosome, and is estimated to have evolved ~ 1 million years ago (Ellison, Stajich et al. 2011; Corcoran, Dettman et al. 2013). The suppressed recombination region is associated with mating-type segregation and correct nuclei packaging, and thus is crucial for the maintenance of the enforced self-fertile (pseudohomothallic) life style (Figure 1) (Merino, Nelson et al. 1996; Gallegos, Jacobson et al. 2000). Although recombination suppression between the mat locus and centromere is sufficient for mating-type segregation in N. tetrasperma, the suppressed recombination region extends over the majority of the mat chromosome. The large size of the suppressed recombination region, and other genomic features, such as codon usage degeneration and non-synonymous substitution accumulation, resembles the sex-chromosome model in animal and plants (Lahn and Page 1999; Gallegos, Jacobson et al. 2000; Nicolas, Marais et al. 2005).

Two factors are suggested to be associated with suppressed recombination: chromosomal structural rearrangements and genetic modifiers of recombination rate (Charlesworth, Charlesworth et al. 2005). By reciprocal introgression analysis between N. tetrasperma and N. crassa, Jacobson (2005) suggested both factors are involved in the suppressed recombination in N. tetrasperma. Menkis (2008) showed the evolution of suppressed recombination in N. tetrasperma involved two successive events, which is similar to the “evolutionary strata” found in human, mouse, chicken and plants (Lahn and Page 1999; Handley, Ceplitis et al. 2004; Sandstedt and Tucker 2004; Nicolas, Marais et al. 2005).

Neurospora tetrasperma FGSC 2508 mat A genome and FGSC 2509 mat a (referred as FGSC 2508 and 2509) genomes were sequenced by the Doe Joint Genome Institute (JGI) in 2011 (http://genome.jgi.doe.gov
FGSC 2508 and 2509 are different mating-type components that originated from the same heterokaryon. The genome size for *N. tetrasperma* is ~ 39 Mbp with ~ 10,000 – 11,000 genes. Two large inversions were found on the mating-type (*mat*) chromosome of *N. tetrasperma* 2508, which covers the majority of the suppressed recombination region on the chromosome (Ellison, Stajich et al. 2011). This study supports the model that the chromosomal rearrangements are strongly associated with, or even causing, the recombination suppression on the *mat* chromosome. Previous studies showed that the recombination suppression is crucial for correct nuclei package and the maintenance of self-fertilization for *N. tetrasperma* (Merino, Nelson et al. 1996; Gallegos, Jacobson et al. 2000; Ellison, Stajich et al. 2011). Thus, the structural rearrangements are suggested to be involved in the life style transition from heterothallism to pseudohomothallism for *N. tetrasperma*. In the *N. tetrasperma* genome study, Ellison (2011) also reported higher numbers of deleterious mutations in the suppressed recombination region of *N. tetrasperma* 2509 *mat a* than 2508 *mat A*, and suggested the pattern was caused by faster deleterious mutation accumulation rate in *N. tetrasperma* 2509. In addition, this study found genomic regions in *N. tetrasperma* also under RIP mutations, indicating the mechanism is active in the closely-related pseudohomothallic species (Ellison, Stajich et al. 2011).
Research aims

I Use Next-Generation Sequencing (NGS) data (Illumina paired-end reads) and comparative genomics approaches to address specific biological and evolutionary questions.

II Specify the size and abundance of introgression tracts in three *N. tetrasperma* lineages

III Identify introgression origin, direction, pattern and consequences in three *N. tetrasperma* lineages

IV Study rates, patterns and mechanisms of intron evolution in Neurospora

V Use population genomic data to study association between selection, genetic parameters and intron population frequency in *N. tetrasperma*

VI Use multiple library paired-end data and Optical Mapping techniques to construct a fine map for two *N. tetrasperma* lineages

VII Identify chromosomal structural variants (SVs) based on the fine map information for the two *N. tetrasperma* lineages

VIII Use bioinformatic approach based on abnormal read orientation and split read information to identify chromosomal SVs, and verify the SVs with fine map data

IX Study rates, patterns and mechanisms of intron evolution in Neurospora

X Study codon usage biases in *N. tetrasperma* and the impact of genomic degeneration on codon usage
Summaries of papers

Paper I – Introgression landscape in *Neurospora tetrasperma*

The significance of introgression as an evolutionary force shaping natural populations is well established in animal and plant systems (Rieseberg 1993; Mallet 2005; Kulathinal, Stevison et al. 2009). However, the abundance and size of introgression tracts, and to what degree interspecific gene flow is the result of adaptive processes, is largely unknown.

In the first study of the thesis, we presented medium coverage genomic data (Illumina 55bp pair-end reads, ~10X coverage for the genome) from three lineages of *N. tetrasperma*, and used comparative genomics to investigate the introgression landscape at the genomic level in this model species. We revealed one large introgression tract in each of the three phylogenetic lineages of *N. tetrasperma* (sizes of at least 5.6 Mbp, 5.2 Mbp and 4.1 Mbp, respectively). The introgression tract was located on the chromosome containing the locus conferring sexual identity, the mating-type (*mat*) chromosome. No introgression tract was found on autosomes or pseudoautosomal regions. The region of introgression was confined to the suppressed recombination region and was only found on one of the two *mat* chromosomes (*mat a*), which indicates an allele specific introgression pattern in *N. tetrasperma*. We used Bayesian concordance analyses to exclude incomplete lineage sorting as the cause for the observed pattern. From multilocus genealogy analyses with additional species of Neurospora, we showed that the introgressions are likely to originate from two closely related, freely recombining heterothallic species *N. hispaniola* and *N. crassa/N. perkinsii*. To detect the consequences of introgression, we investigated patterns of molecular evolution for the introgressed and non-introgressed region in *N. tetrasperma*. We found that introgression is correlated with a reduced level of molecular degeneration, and led to a shorter suppressed recombination time for the introgressed regions.

Overall, the chromosome specific (*mat*) and allele specific (*mat a*) introgression reported in this study comprise one of the largest introgression tracts reported to date from natural populations. Our data contradicts theoretical
predictions that introgression should be less likely occur on sex-determining chromosomes.

**Paper II - Intron evolution in Neurospora**

Intron evolution in eukaryotes is expected to be linked to many factors, such as life-history and lineage-specific factors, gene-specific factors and intron-specific factors (Jeffares, Mourier et al. 2006). However, the specific functional role of introns and the factors affecting their turnover are only beginning to be understood (Chorev and Carmel 2012).

In the second study of the thesis, we used comparative and population genomics approaches to study the rates, patterns, and causes of intron gain and loss in the fungal model genus Neurospora. Based on genome-wide analyses of *Neurospora crassa*, *N. tetrasperma*, and an outgroup, *N. discreta*, we identified a total of 9,495 intron positions from orthologous genes. Relative to other species of eukaryotes, we found a high intron loss rate (5.78 to 6.89 $\times 10^{-13}$ introns per site per year) and a moderate intron gain rate (7.53 to 13.76 $\times 10^{-10}$) in Neurospora. We reported that the 5’-bias, i.e., the relative number of introns located in the 5’ end of a gene, was statistically significant in both *N. crassa* and *N. tetrasperma*, and this pattern was mainly shaped by the 5’ biased intron gain/loss rate ratio. Specifically, the intron gain rate showed a 5’ biased pattern, and the intron loss rate showed an internal biased pattern. Our data indicated that intron losses are mediated by RT-mRNA and that intron gains are mainly mediated by the non-homologous end joining (NHEJ) system in Neurospora.

In addition to species-level analyses, we assessed the population frequency of derived (non-ancestral) intron mutations using 92 resequenced strains of *N. tetrasperma* (Illumina 90bp paired-end reads, 30X). We found that both derived intron gains and losses are skewed to higher frequency compared to neutral fourfold degeneration sites, suggesting positive selection is involved in shaping intron frequency in natural populations. Furthermore, the skewed intron frequency (of both gained and lost introns) was found in normally recombining regions, but not in regions of suppressed recombination in *N. tetrasperma*; this implies recombination facilitates more efficient intron fixation and removal. The frequency of introns in the population were associated with several genetic factors, such as intron position, intron phase and intron GC content, but not intron length. Furthermore, most gained introns likely originated before the populations diverged, suggesting they are not rapidly fixed nor eliminated at the intraspecific level.
Paper III – Large-scale chromosomal inversions in *Neurospora tetrasperma*

Chromosomal inversions are suggested to be an important factor facilitating major biological and evolutionary processes, such as speciation, adaptation and sex chromosome evolution. For *N. tetrasperma*, a large suppressed recombination region is located in the center of *mat* chromosome, and suggested to associate with the reproductive mode transition in the species. To test this hypothesis, we studied the chromosomal inversion landscape in *N. tetrasperma*.

We used next-generation sequencing (Illumina multiple pair-end library) and Optical Mapping data to construct genomic fine map for two *N. tetrasperma* lineages (Lineage 1 and Lineage 9). We found chromosomal inversions in *N. tetrasperma* based on comparative genomic analyses of the fine map data. In addition, we also predicted inversions with bioinformatics methods, and confirmed the prediction by genomic fine map data. Combined with public available *N. tetrasperma* Lineage 6 and outgroup *N. crassa*, we constructed the evolutionary history for large-scale chromosomal inversion events in *N. tetrasperma*. We found lineage-specific inversion patterns in *N. tetrasperma*, suggesting chromosomal inversion is unlikely to associate with the origin of the suppressed recombination and the sexual reproductive mode transition during the early evolution of the species. Furthermore, we infer that inversions are the consequences, rather than the causes of suppressed recombination on mating-type chromosome of *N. tetrasperma*.

Paper IV - Codon usage and sequence degeneration in *Neurospora tetrasperma*

Genomic degeneration, including mutation accumulation, gene silencing, gene loss and chromatin formation, is a key trait of sex chromosomes (e.g., Y chromosome in X/Y systems) (Bull 1983; Carvalho 2002; Skaletsky, Kuroda-Kawaguchi et al. 2003). At present, little is known about early stages of sex chromosome degeneration as most non-recombining sex chromosomes are so highly deteriorated that they retain few traces of the historical molecular events associated with their differentiation (Charlesworth and Charlesworth 2000; Charlesworth, Charlesworth et al. 2005).

*Neurospora tetrasperma* mating-type (*mat*) chromosomes have recently emerged as a model system for early sex chromosome evolution, as they contain a large and recently acquired region of suppressed recombination (>6.6Mbp and <4Mya, respectively) (Menkis, Jacobson et al. 2008). In the
final paper of the thesis, we assessed and compared preferred codon usage, which is one of the key traits associated with molecular-level adaptation (adapted for efficient and accurate translation), in *N. tetrasperma mat* chromosomes as a means to test for early signs of genomic degeneration associated with suppressed recombination. From the analysis of 290 newly identified genes in *N. tetrasperma mat a* and *mat A* strains and homologous genes in its close relative *N. crassa* (a total of 121,831 codon positions) we found marked evidence of degeneration in preferred codon usage in the region of suppressed recombination of the *N. tetrasperma mat* chromosomes. For example, 28.9% of genes located within stratum 2 (a more recently formed segment of suppressed recombination) and 63.2% of genes located in stratum 1 (an older segment) have different levels of preferred codons among alleles, while less than 3.3% of genes in the flanking pseudoautosomal regions show such variation. This time-dependent degeneration was shown to result directly from allele-specific switches from preferred to non-preferred codons.

Additional key findings include the fact that short genes undergo more extensive degeneration than longer genes and that each of the two *mat* chromosomes is subjected to deterioration in preferred codon usage i.e., there is no single “degenerative” chromosome (in contrast to highly evolved dimorphic (X/Y) systems). Comparative analysis of preferred codon usage, GC3, and GC content of introns, reveals that the degenerative changes in preferred codon usage in the region of suppressed recombination are best explained by reduced selection efficiency. The implications of these findings for our understanding of early sex chromosome evolution are discussed in the paper.
Conclusion and future perspectives

In this thesis, I have used comparative genomics approaches to study the genomic features of *N. tetrasperma* and how they are associated with evolutionary history and mating system transition.

In the first paper, I and my coworkers focused on the introgression landscape in *N. tetrasperma*. We found large-scale, chromosomal and allele specific introgression in *N. tetrasperma*. The introgression pattern, mixed with suppressed recombination and both sequence degeneration and invigoration patterns, makes Neurospora an interesting, but also complicated model to study the causes and consequences of introgression.

In the second paper, we zoomed in to focus on the gene level, to study intron evolution in Neurospora. We estimated the intron turnover rate in Neurospora, and found the intron gain and loss mechanism in the genus. We reported the association between selection and different genetic parameters, such as intron position, GC content and phase, and suggested intron frequency and fixation in population is affected by these parameters.

The third paper is about chromosomal structure rearrangements in *N. tetrasperma*. It is a technically challenging project that started in the beginning of my PhD, and has not finished yet. We used many techniques and approaches to get good quality genome sequencing, including 150bp, 500bp and 5kbp insert size paired-end Illumina data for hierarchical assembly and Optical Mapping data for the final placement of contigs. We found different structural variants in different *N. tetrasperma* lineages, which is contradictory to our previous predictions.

In the fourth paper, we focused on the coding-sequence in *N. tetrasperma*, to study the consequences of suppressed recombination. We found evidence for sequence degeneration in the suppressed recombination region, and showed that *N. tetrasperma* can be used as model organism to study the early evolution of sex chromosomes.

For Neurospora, further studies need to be done to answer many old or newly arisen questions, for example the structural rearrangements in natural population of *N. tetrasperma* and selection coefficient of the newly rear-
rangements. To solve these questions, high quality genomic sequencing for the natural population of *N. tetrasperma* is needed. Since sequencing with multiple insert libraries and Optical Mapping is expensive (>20,000 USD per strain), it is difficult to do the population fine map with the current techniques. Thus we are looking forward the third-generation sequencing technology to further reduce the sequencing price and generate longer reads facilitating genome assembly in the future.
Acknowledgements

First thanks to Sweden, Uppsala and Uppsala University, which give me a great opportunity to study in Sweden and get Master and Phd education in the past six years. Life in Sweden is great!

Great thanks to my supervisor Hanna Johannesson! She is always energetic, and helps me with lots of things, from broader perspective of Phd studies and researches, to the quite detailed grammar and spelling changes, no matter how busy she is. Thanks to my co-supervisor and master project supervisor Siv Andersson, who helped me to get into science and bioinformatics.

Thanks to Carrie-Ann Whittle. I worked several project with her, and get to learn how a scientist should work efficiently and productively. Thanks to Pádraic Corcoran, who helped a lot in my several project, and he always shows great interests to biology research and read tons of papers. Thanks for Jennifer Andersson, who helped editing and gave a lot of useful comments to my thesis. Thanks to all the current and previous group and lab-meeting members: Kristiina Nygren, Rebecka Strandberg, Niklas Samils, Anastasia Gioti, Alexander Mushegian, Ioana Brannstrom, Jesper Svedberg, Fan Han, Matilda Åslin, Anna Rosling, Bart Nieuwenhuis, Hector Urbina, Kerri Kluting and Haleh Hayatgheibi. Thanks for your collaboration or discussion with me!

Thanks to all the PIs in Evolutionary Biology: Hans Ellegren, Mattias Jacobsson, Tanja Slotte, Urban Friberg, Jochen Wolf, Anders Götherström, Simone Immler, Simon Whelan. Thanks for the contribution to make Evolutionary Biology a great place for Phd study, and provide enough fundings for Graduate School student.

Thanks to a lot of colleagues in Evolutionary Biology, who gives me a lot of good advises and discussion, including Sen Li, an excellent bioinformatician and computer hardware expert, Kimwoong Nam, who taught me a lot in bioinformatics, Johan Lindell, Holger Schielzeth, Per Sjödin and many others.
Thanks to colleagues in other department of EBC with their nice suggestion, including Biao Wang, Chen Hou, Pengjuan Zu, Björn Nystedt, Lionel Gay and many more.

Thanks for Lucie Gattepaille and Feifei Xu with their nice suggestion in teaching courses with me.

Thanks for the IT department, including Stefan Ås, Ben Gariod for their help in computer and network related things.

Thanks for Helge Ax:son Johanssons stiftelse and Nilsson-Ehle-donationerna from Kungl Fysiografiska Sällskapet i Lund and Graduate School from EBC with their grants and support to my Phd study.

Thanks to BGI with the nice collaboration project. Thanks Xuegang Wang, Junnian Liu, Zhiqin Lao, Li Li, Jiao Xue and many other people.

Thanks to my friends and Master program classmates in Uppsala: Di Wu, Yanling Cai, Jinzhi Hu, Shuang Gao, Xiaoyu Wang, Weilun Sun, Sungkyu Park, Li Li, Na Li, Jinghui Luo, Di Yu, Chuan Jin, Tao Cui, Fang Mao, Xiang Jiao, Yifei Wu, Shuangwei Hu, Chengxi Shi, Yang Sun, Yani Zhao, Lei Chen, Hua Huang, Bo Xu, Haisha Ma, Su-chen Li, Shangfeng Liu, Anqi Xiong and many more. We spend a great time together in Uppsala.

Thanks to my wife, Xi Fu, life is great with you!

Thanks to my parents, Chunmin Sun and Li Zhang. Thank you for your support and encourage!
References


A doctoral dissertation from the Faculty of Science and Technology, Uppsala University, is usually a summary of a number of papers. A few copies of the complete dissertation are kept at major Swedish research libraries, while the summary alone is distributed internationally through the series Digital Comprehensive Summaries of Uppsala Dissertations from the Faculty of Science and Technology.