Jury Member Report – Doctor of Philosophy thesis.

Name of Candidate: Aleksandra Bezmenova

PhD Program: Life Sciences

Title of Thesis: Evolutionary processes in hypervariable fungus Schizophyllum commune

Supervisor: Professor Georgii Bazykin

Co-supervisor: Professor Alexey Kondrashov, University of Michigan, USA

Name of the Reviewer: James B. Anderson

I confirm the absence of any conflict of interest	
	Date: 25-10-2021

The purpose of this report is to obtain an independent review from the members of PhD defense Jury before the thesis defense. The members of PhD defense Jury are asked to submit signed copy of the report at least 30 days prior the thesis defense. The Reviewers are asked to bring a copy of the completed report to the thesis defense and to discuss the contents of each report with each other before the thesis defense.

If the reviewers have any queries about the thesis which they wish to raise in advance, please contact the Chair of the Jury.

Reviewer's Report

Reviewers report should contain the following items:

- Brief evaluation of the thesis quality and overall structure of the dissertation.
- The relevance of the topic of dissertation work to its actual content
- The relevance of the methods used in the dissertation
- The scientific significance of the results obtained and their compliance with the international level and current state of the art
- The relevance of the obtained results to applications (if applicable)
- The quality of publications

The summary of issues to be addressed before/during the thesis defense

The Ph.D. thesis by Aleksandra Bezmenova explores aspects of mutation and recombination in the filamentous fungus *Schizophyllum commune*. The work is of high quality and is already represented by a beautiful publication in the high-end journal Molecular Biology and Evolution, of which the candidate is the first author. The stated topic of the thesis and the actual content are in close alignment. The methods used to address the research questions are appropriate and state-of-the-art. The thesis makes substantial contributions of broad interest in genetics and evolution and I am pleased to recommend that the candidate should defend the thesis formally. I thank the candidate and supervisors for inviting me to participate in the defense and I look forward to a lively discussion.

I now turn to a more detailed evaluation of each section of the thesis. I found the Introduction and literature review to be OK. The essential topics were covered, but not much more. I especially thought that the candidate could have gone deeper into the background on recombination (more on this below) and she could have been more specific and insightful about how her contributions fit into the broader picture. The writing in this section is a bit rough in places and I made a series of comments by means of strikethroughs, sticky notes, and insertions in the PDF file; of course, it is up to the candidate and supervisors about how to address each suggestion, or not. My impression was that these sections could use a close edit throughout.

In Chapter 3, things get exciting! I was familiar with the publication associated with this section and I'm glad to report that the writing here is excellent, having been through the full editorial process. This is an important study on mutation accumulation, with ample data to enable a good estimate of rate of mutation per base per cell generation and a high-res representation of the mutation spectrum.

I do have two points for consideration and possible discussion at the defense. In the thesis, the candidate implies that the haploid monkaryotic phase of S. commune is as prevalent in nature as the dikaryotic phase (thesis page 58, Chapter 4). This is not the case. Monokaryons tend to be fertilized rapidly by spores or hyphae and so the monokaryotic phase is extremely ephemeral in nature. (As an illustration, take a monokaryotic mycelium in a petri dish into the outdoors or indoors, remove the lid for 24 hours, and then sample the resident mycelium. It will by then have become dikaryotized by spores in the air!). Haploid monokaryons are extremely avid to take on fertilizing elements. This in no way impacts on the validity of the results in this chapter and it was gratifying that, in the next chapter, the candidate did look at dikaryons in the nature and found a somewhat comparable rate of mutation in the monokaryons in-vitro and the dikaryons in nature (1.24×10^7 vs. 2.45×10^7 mutations/nucleotide/m). The two-fold difference in rate actually might actually some sense. The monokaryons have one candidate nuclear type for mutation, while the dikaryon has two.

My other point concerns the idea that some filamentous organisms might show a decreasing rate of mutation the longer they grow vegetatively. I see no mechanistic reason why this should be the case under conditions of steady-state growth. And I interpret the Armillaria data (page 54) very differently than the candidate did. The more likely explanation for the data in Fig. 16 of the thesis is that the individual of Armillaria spread rapidly after its birth to fill up its present environment and then basically "ran in place" over the years, exploiting new food sources as they became available in the immediate locality. This means that isolates that were collected from points relatively close together in space may have almost as many cell divisions separating them from their common ancestor cell as isolates taken from points much further away from one another. In other words, the short distances drastically under-represent the actual number of cell divisions. The phylogeny of changes in (Fig. 2 in Anderson 2018) is consistent with this possibility – the terminal branches are all long compared to the internal, phylogenetically informative changes. No doubt, we can discuss this further at the defense.

Chapter 4 aimed to measure the mutation rates in naturally occurring dikaryons in nature and it largely succeeds in this, although the number of dikaryons sampled from multiple points was small; two dikaryons each with two fruitbodies identical for both nuclear types and one dikaryon with two fruitbodies sharing only one nuclear type. I suspect that this sample will need to be increased before publication. A better tactic going forward would be to first use the highly polymorphic mating types as indicators of individual identity. This is low-tech procedure, easy to do, and fast. The sequencing would then be applied only to known examples in which two (or more) fruitbodies sampled represent the growth of a single dikaryon.

Chapter 5. I am not sure why it is not possible to filter out the mutations happening during the growth of the dikaryon before fruitbody formation and then for the analysis to proceed with the mutations that were unique. Of course, it's best to minimize propagation of the dikaryon before fruitbody formation. But this growth phase is impossible to eliminate entirely. Even within the fruitbody, there are mitotic divisions of the dikaryon that precede basidium formation and meiosis.

Chapter 6. Nice design for using flanking regions to detect recombination in highly heterozygous and highly homozogous internal regions for a clean comparison! As appealing as these results are, and as much work that was required to get them, however, I suspect that it will probably be necessary to replicate this comparison in other chromosomal regions before publication. In setting the stage for this part, I also thought that there should be more background context on recombination with different levels of nucleotide sequence diverge. What about the yeast (Saccharomyces) example? With high levels of divergence, as in inter-species yeast hybrids, the mismatch repair machinery goes into overdrive and recombination is entirely blocked, leading to mis –segregation of entire chromosomes, resulting in widespread inviability of offspring. Indeed, this postzygotic mechanism is the main means of reproduction isolation between species. Might something similar be operating in S. commune in the highly divergent regions of the genome and less so in the regions with low divergence? It seems that the introduction should not ignore the history in Saccharomyces. Also, the historical use of flanking-marker regions in detecting recombination goes way back in genetics. Think Fogel and Hurst in the late sixties examining the link between gene conversion and crossing over using tools of flanking markers and tetrad analysis to reconstruct the whole meiosies. Although a technical challenge, having tetrads from basidia of S. commune would enable easy recovery of reciprocal recombination products and would detect non-reciprocal events (gene conversion), if they are occurring.

I also wonder about the underlying mechanism of the "homology search" in S. commune, the not-well-understood process by which weak synaptic interactions happen before any hybrid DNA is formed for example in repair of double-stranded breaks. Of course, we do not know about double-strand-break sites in S. commune and how their activity might be affected by levels of nucleotide-sequence divergence. Bottom line: the candidate could bring more nuance to the background on recombination than was presented. Certainly, any future publication on this topic should include an enhanced introduction and discussion. These ideas might also point out future paths if we truly want to understand how recombination is affected by the level of sequence divergence.

Conclusions Chapter. The end came rather abruptly. I've seen this many times in Ph.D. theses. The candidate has worked hard on experiments and writing and is maybe tired at this point. But please, dig a little deeper and give us more to go on for the future research in this area.

Sincerely yours,

James B. Anderson

Professor Emeritus

University of Toronto

X☐ X *I recommend that the candidate should defend the thesis by means of a formal thesis defense*

☐ *I recommend that the candidate should defend the thesis by means of a formal thesis defense only after appropriate changes would be introduced in candidate's thesis according to the recommendations of the present report*

☐ *The thesis is not acceptable and I recommend that the candidate be exempt from the formal thesis defense*

Skolkovo Institute of Science and Technology

EVOLUTIONARY PROCESSES IN HYPERVARIABLE
FUNGUS *SCHIZOPHYLLUM COMMUNE*

Doctoral Thesis

by

ALEKSANDRA BEZMENOVA

DOCTORAL PROGRAM IN LIFE SCIENCES

Supervisors
Professor Georgii Bazykin
Professor Alexey Kondrashov

Moscow - 2021

I hereby declare that the work presented in this thesis was carried out by myself at Skolkovo Institute of Science and Technology, Moscow, except where due acknowledgement is made, and has not been submitted for any other degree.

Candidate (Aleksandra Bezmenova)
Supervisor (Prof. Georgii Bazykin)

Abstract

The basidiomycete *Schizophyllum commune* has the highest level of genetic polymorphism known among living organisms. In previous studies, it was also found to have ~~a relatively high per generation mutation rate of $2x10^{-8}$~~. It was also shown that homologous recombination tends to occur in the more conservative parts of the *S. commune* genome, in particular inside exons, unlike other studied organisms. Here, we apply methods of comparative genomics to experimentally and naturally growing individuals and populations of *S. commune* to study the forces that shape its polymorphism. We ask what contributes to its hypervariability, and how hypervariability in turn modulates the effects of these processes. We focus on three basic population genetics forces: mutation, recombination and natural selection.

First, we used an experimental design that measures the rate of accumulation of somatic *de novo* mutations in a linearly growing ~~mononuclear~~ mycelium. We showed that *S. commune* accumulates mutations at a rate of $1.24x10^{-7}$ substitutions per nucleotide per meter of growth. In contrast to what has been observed in a number of species with extensive vegetative growth, this estimate does not decline in the course of propagation of a mycelium. As a result, even a moderate per cell division mutation rate in *S. commune* might translate into a very high per generation mutation rate when the number of cell divisions between consecutive meioses is large.

Second, we showed that the rate of accumulation of somatic mutations in dikaryotic mycelium in nature is similar to what we observe in monokaryotic mycelium under laboratory conditions; thus, we conclude that our mutation rate estimates are similar between stages of the life cycle and between in vitro systems and natural populations. Moreover, by comparing the genomes of fungi physically colocalized in the natural environment

(coinhabiting the same tree trunk), we confirm that they spread for large distances through vegetative growth, indicating that our estimates are relevant for understanding of the causes of real-life hypervariability.

Third, we directly showed that the homologous recombination rate is significantly higher when the genome segment is completely homozygous compared to when it is highly heterozygous.

Together, our findings showcase the power of whole-genome analysis of experimentally and naturally growing individuals for clarifying the causes and consequences of hypervariability in this system. They confirm the elevated mutation rate, and indicate that it contributes to the high polymorphism observed in *S. commune,* ~~shedding light on the causes of~~ extraordinary variability in this species. They also show that hypervariability suppresses recombination, confirming its role as a phenomenon that affects diverse facets of genome evolution.

Key words: mutation rate; somatic mutation rate; homologous recombination; *Schizophyllum commune*

Publications

1. Aleksandra V. Bezmenova, Elena A. Zvyagina, Anna V. Fedotova, Artem S. Kasianov, Tatiana V. Neretina, Aleksey A. Penin, Georgii A. Bazykin, and Alexey S. Kondrashov. Rapid accumulation of mutations in growing mycelia of a hypervariable fungus *Schizophyllum commune*. *Molecular Biology and Evolution*. 2020 Aug; 37(8):2279-2286.

2. Bezmenova AV, Bazykin GA, Kondrashov AS. Prevalence of loss-of-function alleles does not correlate with lifetime fecundity and other life-history traits in metazoans. *Biology Direct*. 2018 Mar 2; 13(1):4.

Table of Contents

List of Symbols, Abbreviations

BC - back cross

bp - base pair

CI - confidence interval

cM - centimorgan

CO - crossing-over

Gbp - gigabase pairs (10^9 nucleotides)

GwRR - genome-wide recombination rate

kb - kilobase (10^3 nucleotides)

MA line - mutation accumulation line

Mb - megabases (10^6 nucleotides)

NGS - next-generation sequencing

SAR - Stramenopila, Alveolata, Rhizaria

SNP - single nucleotide polymorphism

WGS - whole-genome sequencing

List of Figures

Fig. 13. Accumulation of mutations during the growth of the mycelium. Number of mutations that have reached 70% frequency in sequenced samples are shown. Replicas are displayed with different line types. (A) Narrow tubes. (B) Thick tubes.

Fig. 14. The relationship between the mutation accumulation rate and mycelium length. (A) Narrow tubes. (B) Thick tubes.

Fig. 15. Mutation accumulation rates in narrow and thick tubes (A) and for individual founding cultures (B).

Fig. 16. Relationship between the number of mutations (A) and mutation rate, and the distance between sequenced samples in Armillaria fungus. Obtained based on data from (Anderson et al. 2018).

Fig. 17. Distances between pairs of fruit bodies from the same tree trunk.

Fig. 18. The coverage distribution when mapping reads from one fruit body to genome assembly from another fruit body from the same trunk.

Fig. 19. Experimental layout. Colors represent two parental genotypes of the chromosome of interest.

Fig. **20.** Scheme of the primers used for determination of CO events in the central segment of the chromosome. Coordinates in the chromosome are not to scale.

Fig. 21. Scaffold genotypes of f1_13 and Z14 individuals. Scaffold of interest is circled.

Fig. 22. Primer names for the loci of interest in scaffold 3. Coordinates in the scaffold are not to scale.

List of Tables

Table A5. Distances between collected fruit bodies, Chapter 4.

Table A6. Assembly statistics for reference samples, Chapter 4.

Table A7. Statistics for scaffold assemblies for parents, F1, BC and F2 offsprings (Chapters 5 and 6).

Table A8. List of *de novo* mutations and mutational events in F1, BC and F2 crossings. Clustered mutations are shown in color.

Chapter 1. Introduction

1.1 Relevance and significance of the work

S. commune is a species with the highest known genetic polymorphism that can reach levels ~~higher than distances between some species~~ (Baranova et al. 2015). Such a high heterozygosity raises many questions and ~~gives a lot of~~ opportunities. Different evolutionary processes ~~are~~ to be studied to understand the potential sources of the ~~record-high~~ heterozygosity. In the meantime, the consequences of such polymorphism ~~are also an interesting subject to study~~. In general, *S. commune* i~~s a unique organism and all processes involving this fungus are of great interest for evolutionary biology.~~

Factors of Darwinian evolution are:

- mutation

- natural selection

- recombination

- genetic drift

- population structure

Mutational process and genetic drift are the ultimate factors responsible for the level of genetic polymorphism in a population: $H = 4N_e\mu$ (Kimura 1983), where H is the virtual heterozygosity (polymorphism) at neutral sites (the probability that two random individuals will have different alleles at a site), N_e is the effective population size that determines the random genetic drift, and μ is the spontaneous mutational rate. While the effective population size is extremely hard to estimate directly, the mutation rate can be relatively easily estimated using WGS methods.

S. commune is a ~~mycelial~~ fungus with ~~an~~ extensive vegetative growth (Niederpruem and Wessels 1969; Gooday 1995). Thus, the generational mutation rate in this fungus is likely to be determined by the per cell division somatic mutation rate. The mode at which somatic mutations are accumulated in the growing mycelium, and how these mutations are transferred to the subsequent generations may ~~in the end~~ largely determine the genetic polymorphism of this species.

While within-population natural selection does not affect the neutral heterozygosity, the within mycelium natural selection may affect the substitution ~~accumulation~~ rate inside the growing mycelium. Nearly neutral evolutionary theory states that slightly deleterious and slightly beneficial mutation ($|N_e s| \ll 1$) act as neutral and may be randomly fixed in a population due to genetic drift (Kimura 1983). Thus, in a mycelium with limited amount of hyphae more mutations will be fixed due to genetic drift, and thus the substitution accumulation rate will be higher than in the large mycelium.

To fully address the question of how mutational process operates in *S. commune*, in Chapters 3-5 we study and discuss the somatic and generational mutation rates, both under laboratory conditions and in nature. We also look at how the mycelium size and thus the natural selection within it affects the substitution accumulation rate.

Homologous recombination is the factor of evolution that probably does not affect the genetic polymorphism much, but which is itself most likely affected by the high level of heterozygosity. As long as ~~it~~ involves homologous chromosomes, it usually happens when the genetic ~~distance~~ between to exchanging parts of the genome is very small, usually less than 2% (Leffler et al. 2012). However, in *S. commune* this ~~distance~~ may reach 20%. It was previously shown that the level of heterozygosity may indeed affect the recombination rate (Waldman and Liskay 1988; Datta et al. 1997). In (Seplyarskiy et al. 2014), it was shown that

in *S. commune* the CO events tend to occur within more conserved regions, in particular exons. In Chapter 6, we directly study how the recombination rate may depend on the genetic distance between recombining regions of the genome.

1.2 Goals and objectives

We aim to study evolutionary processes involving hypervariable fungus *S. commune*. These ~~involves~~:

- studying of somatic mutational processes and natural selection inside growing mycelia of *S. commune*;
- studying of generational mutational process *in vitro* and *in vivo*;
- studying of the dependence of the homologous recombination rate on the level of chromosome heterozygosity in *S. commune.*

1.3 Study system

The following states of *S. commune* mycelia were used and analyzed in this work:

- dikaryon fruit bodies collected in nature;
- dikaryon fruit bodies obtained under laboratory conditions;
- monokaryon mycelia cultivated under laboratory conditions.

Whole genome and Sanger sequencing data ~~was~~ used in this work.

1.4 Implications of this work

Here, we show that evolutionary processes involving *S. commune* are indeed unique in some ways. In particular, we show the potential generational mutation rate may exceed almost all rates previously estimated for other species. Moreover, we show that the homologous recombination in *S. commune* is indeed suppressed by the high genetic distance between chromosomes, and may raise drastically given the more genetically close segments of a genome.

This shed light on how highly diverse species might appear and exist. Moreover, this knowledge might be of great help for further studies of evolutionary processes (such as natural selection within the population) that use *S. commune* as an object.

As a side observation, we show that the mutation rate estimations under laboratory conditions do not always fully represent what may happen in nature.

1.5 Personal contribution

All work in this thesis has been performed by myself, except for the explicitly listed below. In particular, I have performed all the cultivation work in Chapter 3, and all the bioinformatic analysis overall, except for the reference sequences assembling in Chapter 2.

Alexey Kondrashov gave the initial idea of the first two experiments of this work (Chapter 3-4), collected samples used in Chapter 4, and edited text of Chapter 3.

Vladimir Seplyarskyi gave the idea of the experimental layout in Chapter 6.

Elena Zvyagina conducted the cultivation part of the work in Chapters 5-6 and participated in the DNA extraction.

Tatiana Neretina participated in the DNA extraction and library preparation, and conducted all the Sanger sequencing in Chapter 6.

Alena Glagoleva participated in DNA extraction and library preparation.

Maria Logacheva, Alexey Penin and Anna Fedotova conducted all the whole-genome sequencing.

Artyom Kasianov conducted the *de novo* assembly of the reference genomes in Chapter 3.

Chapter 2. Review of the literature

2.1 *Schizophyllum commune* Fries

Introduction. *Schizophyllum commune* Fries is a widespread species of a basidiomycete genus Schizophyllum (Fig.1). *S. commune* is ~~a cosmopolitan species and occupies~~ all continents except for Antarctica (Cooke 1961). *S. commune* is a xilotrophe basidiomycete, however, it usually utilizes saprotrophic nutrition type, and is to be found on both dry and wet trunks. It can also attack living plants at damaged spots (Cooke 1961; Takemoto et al. 2010), causing white rot of wood (Takemoto et al. 2010). Unlike most of the homobasidiomycetes, *S. commune* can produce fruit bodies on artificial media, thus becoming one of the popular model basidiomycetes (Niederpruem and Wessels 1969).

Fig. 1. Fruit body of *S. commune* on a trunk (Palmer and Horton 2006).

Systematic classification. Systematic classification of *S. commune* (Anon):

- Opisthokonta
   - Fungi
      - Basidiomycota
        - Agaricomycotina
          - Agaricomycetes
            - Agaricomycetidae

- Agaricales
  - Schizophyllaceae
    - Schizophyllum
      - *Schizophyllum commune*

Initially Schizophyllum ~~genus~~ was attributed to the Agaricus ~~genus~~; however, they were eventually described as a separated genus (Watling and Sweeney 1974). Due to the widespread ~~previability~~ of this genus, its representatives were present in a huge number of collections under a huge number of species names. However, after the revision of approximately 4 000 collections most of the samples were attributed to the *S. commune* ~~species~~ (Cooke 1961).

Significance to the human population. More than 300 plant species are described as nutrition substrates of *S. commune*, the prevailing ones being deciduous trees, although coniferous trees ~~are also described.~~ The most prevalent substrate is *Pyrus malus*, with both wood and fruits being the target of the fungus. In tropics *S. commune* can be found on *Saccharum officinarum* (Cooke 1961).

S. commune can be described as a pest. It can damage agricultural crops - vegetables and berries, as well as fruit trees, especially at wounds. Moreover, *S. commune* can damage wooden buildings and structures. *S. commune* is one of the few fungi that can damage light balsa wood (Singer 1949; Cooke 1961).

Usually *S. commune* does not infect humans and animals; however, sporadic cases of *S. commune* infection in humans, dogs and some other animals are described. *S. commune* ~~was the patogene in~~ approximately half of basidiomycete infections in humans (Chowdhary et al. 2014). In most of the cases the fungus affected the respiratory tract: out of 114 described cases, 52% were classified as bronchopulmonary diseases, including allergic bronchial mycosis; 22% were classified as sinusitis (Chowdhary et al. 2013; Chowdhary et al. 2014).

21

Three cases of dog infection and one case of seal infection are described (Hanafusa et al. 2016).

Biology. Life cycle of *S. commune* is presented in Fig 2. It consists of two alternating stages - mono- and dikaryons. Monokaryons consist of mononuclear haploid cells, while dikaryons consist of cells that carry two haploid nuclei. Usually the mycelium is represented by a dykarion, which produces fruit bodies after receiving some external signals. Fruit bodies carry basidia cells, which produce haploid basidiospores as a result of karyogamy and meiosis. Basidiospores are spread and then give birth to monokaryon mycelia. Monokaryons can grow independently and occupy territory as well as dikaryons. When monokaryons with different mating types (see below) meet each other, they can form fusion and produce dykarion which will carry nuclei with genomes from both 'parental' monokaryons (Palmer and Horton 2006).
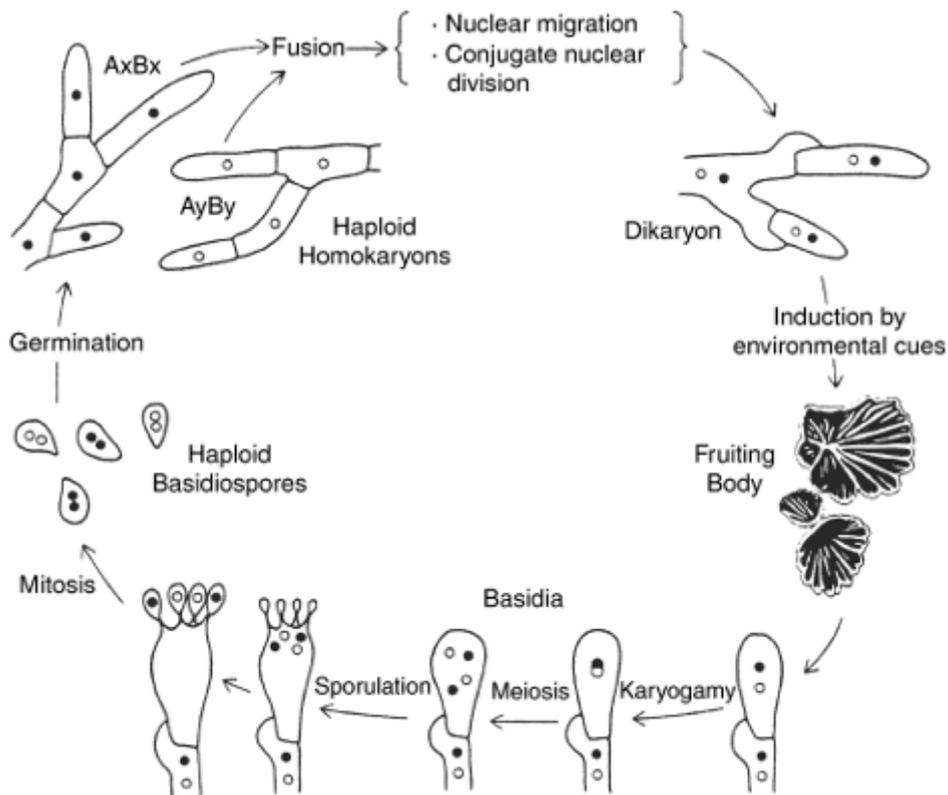
Fig. 2. Life cycle of *S. commune* (Palmer and Horton 2006).

Both mono- and dikaryons can be relatively easily cultivated *in vitro*. It is possible to cultivate isolated monokaryons that originated from a single spore. Such cultures are called monospore cultures, and when cultivated in isolation they show unlimited vegetative growth without producing fruit bodies (Niederpruem and Wessels 1969).

Mating types of *S. commune* are controlled by a common for basidiomycetes system of two loci - A and B (Raper 1996). Genes at locus A are responsible for the repression of the asexual spore formation, and control mating and nuclei division as well as cell formation in dikaryons. There are two closely located subloci in *S. commune genome* - Aα и Aβ. They carry functionally independent genes, however, the difference in either one of these subloci is enough for locus A to be considered different in terms of mating types. Genes located at locus B are responsible for the pheromones recognition pathways, and encode pheromones themselves and their receptors. As locus A, locus B consists of two subloci, and the difference in either one of these subloci is enough for locus B to be considered different in terms of mating types.

S. commune demonstrates an extremely high number of mating types - over 20 000 (Raper 1996). To compare, there are 25 types of locus B and 2 types of locus A in a model basidiomycete *Ustilago maydis,* which results in 50 different mating types (Puhalla 1970). In homobasidiomycetes that produce fruit bodies loci A and B are usually multiallelic, which results in a huge number of mating types; however, these numbers are usually less than that in *S. commune*. For example, 12 000 mating types are described for the basidiomycete *Coprinopsis cinerea*, which is already considered high (Raper 1996).

The formation of fruit bodies is possible only if both loci A and B are different in the interacting monokaryons. If both loci are the same, the resulting mycelium is indistinguishable from monokaryon. Same locus B and different locus A result in
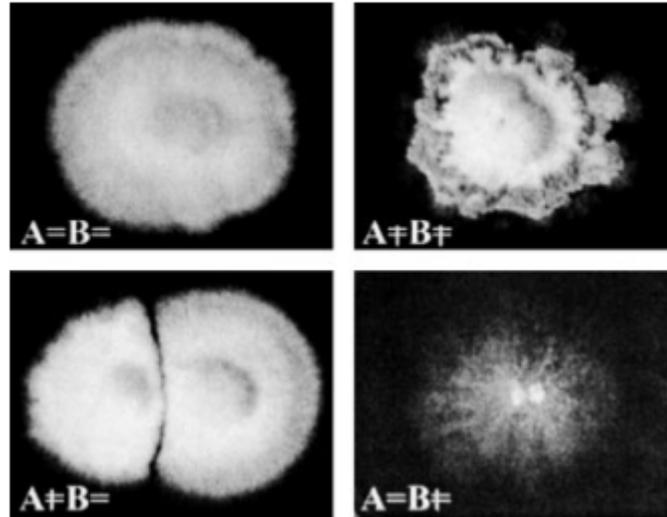
Fig. 3. Different phenotypes as a result of interaction of monokaryons with different mating compatibility:: A=B= - monokaryon phenotype; A≠B≠ - fertile dikaryon; A≠B= - non-interacting monokaryons; A=B≠ - "flat" phenotype (Kothe 1999).

non-interacting monokaryons. Same locus A and different locus B result in "flat" phenotype with hugely reduced aerial mycelium (Fig. 3) (Kothe 1999).

There are two types of *S. commune* hyphae - 3-5 and 1-2 um in diameter. Vegetative hyphae usually branch near the center of the cell. The cell length is usually in the 30-200 um range, with mean being 80 um (Essig 1922). Hyphae show apical growth, with rare intercalar growth observed in sporulating structures (Gooday 1995).

In (Clark and Anderson 2004) a long term cultivation of mono- and dikaryons of *S. commune* is described. In this work 12 monokaryon and 12 dikaryon cultures were cultivated on Petri dishes. Two protocols of mycelium transfer were used - in one case 2x2x2 mm fragments of the medium with the mycelium were transferred to new Petri dishes; in the second case, 2x9 mm fragments of the medium with the mycelium were transferred. This was done in an attempt to vary the effective population size of growing hyphae, and thus the effectiveness of

24

natural selection. During the 18 months of the experiment, the growth rate of mono- and dikaryons changed in a different manner (Fig. 4). At the beginning, monokaryons grew at a rate of ~8 mm/day, faster than dikaryons (~4 mm/day). However, eventually the growth rate of monokaryons did not change much, while some of the dikaryons showed an increase of the growth rate by several times (other dikaryons showed a change of growth similar to that for monokaryons) (Fig. 4).
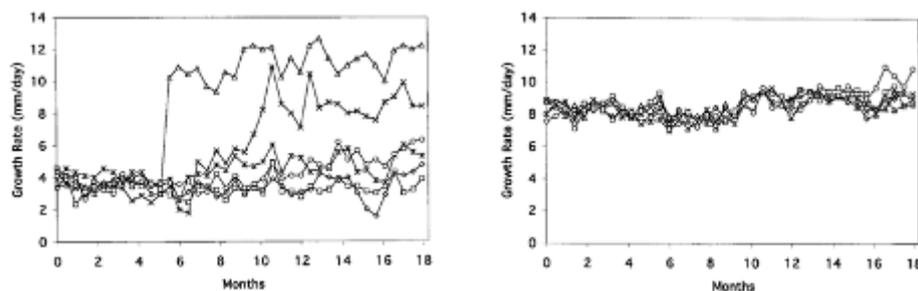
Fig. 4. Change of growth rates for dikaryons (left) and monokaryons (right) of *S. commune* with time (Clark and Anderson 2004).

Genome. *S. commune* genome was sequenced in 2010 (Ohm et al. 2010). ~~Its length appeared to be 38.5Mb.~~ Later the assembly ~~has~~ been slightly improved, and currently consists of 25 scaffolds, with 16 319 annotated genes (version v3.0) (JGI).

Genetic diversity. Within-population polymorphism, or virtual heterozygosity, is the probability of two random alleles from a population to be different. In (Leffler et al. 2012) it was shown that in most species the level of polymorphism is below 3%; out of 167 studied eukaryotic species only two showed heterozygosity above 5%. Among all studied groups of organisms the most polymorphic ones appeared to be arthropods (with mean polymorphism value at 1.25%), while the least polymorphic ones appeared to be chordates (with mean

polymorphism value at 0.26%). The least polymorphic species was *Lynx lynx* (0,01%). Human population has the polymorphic level at 0,12 – 0,15% (The 1000 Genomes Project Consortium 2015). For ~~the~~ long time the most polymorphic species was considered to be *Caenorhabditis brenneri* (16%) (Dey et al. 2013).

However, it was found that in *S. commune* the within-population genetic diversity can reach 20%, which is the current record among studied species (Baranova et al. 2015). After examining 24 individuals from Russian and USA populations, it was shown that the genetic polymorphism reaches 13% in Russian and 20% in USA populations. Moreover, it was shown that a relatively high proportion of SNPs are three- and four-allelic (17.7% and 2.5% respectively). Such a huge heterozygosity allows us to study evolutionary processes such as positive and negative natural selection and homologous recombination with resolution and precision previously achievable.

2.2 Mutational process and methods of mutation rates estimation

Mutational process is one of the key ~~aspects of evolution.~~ Spontaneous mutations are the source of genetic diversity which is necessary for adaptation. The rate at which mutations appear and fix, and how they affect fitness are important parameters of different models of evolution. Unlike many other parameters, such as effective population size, the rate of spontaneous mutagenesis can be measured, indirectly or directly. Before the NGS era, due to low probability of the emergence of *de novo* mutations, indirect methods were mostly used; however, after the development of the NGS methods and the appearance of whole-genome sequences, direct methods became also applicable (Kondrashov and Kondrashov 2010).

Indirect methods of mutation rate estimation include:

- screening of phenotypically visible mutations
- estimation of fitness changes with the course of mutation accumulation
- estimation of the rate of evolution in neutral sites using between species divergence data
- estimation of the mutation rate using within-population polymorphism data

Direct methods include:
- direct study of the mutations accumulated during a large number of generations (mutation accumulation (MA) lines)
- triad studies (whole genome sequencing of two parents and a child)

Screening of phenotypically visible mutations. A simple way to estimate mutation rate is to screen mutations that have phenotypically visible effects. To do so, one has to know the target, mutations in which lead to a distinguishable phenotype, and the length of the genome sequence, mutations in which lead only to that particular phenotype. Before the NGS era the application of this method was largely limited by the lack of knowledge about functional loci. This method was applied to humans (Sommer 1995; Kondrashov 2003), Drosophila (Yang et al. 2001) and some other species; however, it was never popular for studying eukaryotic mutation rates. The only broad study is described in (Yang et al. 2001). This method often leads to the underestimation of the mutation rate, because i) not all nonsense mutations lead to the desired phenotype; ii) mutants can be missed during screening. A large problem of this method is a huge amount of work required and a small number of mutations detected. For example, in (Yang et al. 2001) after screening of 900 000 flies only 16 mutations in 8 loci were found.

Estimation of the rate of evolution in neutral sites. In neutral sites the rate of evolution is equal to the mutation rate (Kimura 1983). This fact was used in a number of studies to

estimate the mutation rate using the between species divergence data (Crow 1993; Drake et al. 1998; Nachman and Crowell 2000; Scally et al. 2012). However, this method has several disadvantages. First, it is hard to prove that a site is indeed neutral. Usually synonymous sites are used as neutral; however, it is not always correct. To solve this problem the attempts to use orthologous pseudogenes between human and chimpanzee was made (Nachman and Crowell 2000). Second, usually the precise data about the divergence times and regeneration lengths is lacking, while this data is necessary for estimation of the number of generations separating two sequences. Third, recurrent mutations within a single site between distant species may affect the results.

In (Nachman and Crowell 2000) the generational mutation rate in humans was estimated at $2.5 \cdot 10^{-8}$ substitutions/nucleotide/generation, which is approximately twice as high as the estimations obtained later using direct methods.

Estimation of the mutation rate using within-population polymorphism data. Virtual heterozygosity in a diploid population that is at an equilibrium state is proportional to the effective population size and the mutation rate: $H = 4N_e\mu$. Thus, given the value of the within population heterozygosity and the effective population size, one can easily estimate the mutation rate (Deng and Lynch 1996; Messer 2009). However, the effective population size is itself a value that is very hard to estimate, and thus usually the mutation rate is used to estimate the effective population size.

Mutation accumulation lines. With the development of the NGS methods, it became possible to directly estimate the spontaneous mutation rates from whole genome sequences (Haag-Liautard et al. 2007; Lynch et al. 2008; Keightley et al. 2009; Ossowski et al. 2010; Ness et al. 2012; Sung et al. 2012; Schrider et al. 2013; Zhu et al. 2014). The most convenient experimental approach to do that is so called mutation accumulation (MA) lines. This

approach uses extremely inbred lines that evolve under laboratory conditions during a large number of generations (Fig. 5, (Lynch et al. 2016)). The artificial extreme limitation of the population size leads to (almost) complete inefficiency of natural selection (see Chapter 2.3). This leads to fixation of all newly emerged mutations except for lethal ones or those leading to complete infertility. Thus, one can assume that mutations observed at the end of the experiment are a result of mutational process only and are not affected by natural selection. This allows one to directly measure the spontaneous mutation rate (Lynch et al. 2016).
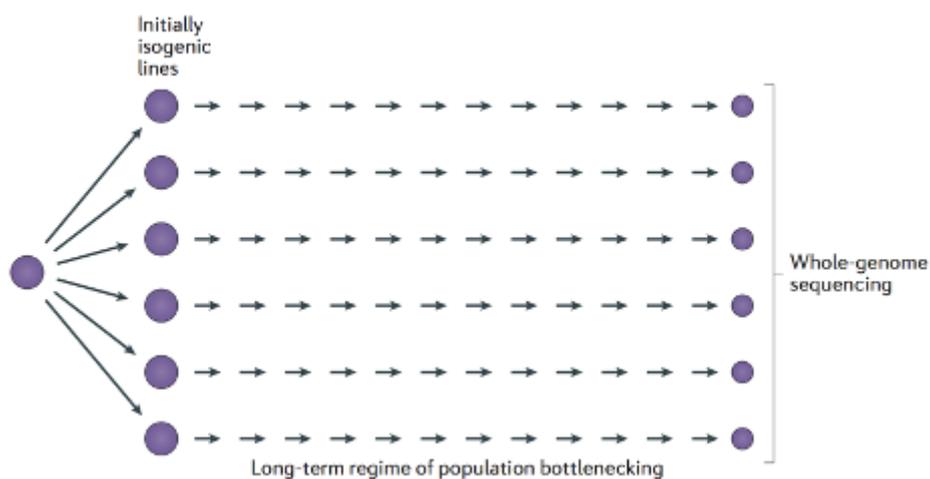
Fig. 5. The scheme of a MA lines experiment (Lynch et al. 2016).

MA lines approach was first used in (Mukai et al. 1972), and was further used for direct estimation of mutation rates as well as studying of mutational spectra and other parameters of the mutational process.

Values of mutation rates across different species. Nowadays, dozens of works in which mutation rates are estimated are available. In a recent review (Lujan and Kunkel 2021), the results of most of the studies were summarised in a beautiful table (Table 2.1).

Generational mutation rates vary from 0.00761 substitutions/Gbp/generation in the *Tetrahymena thermophila* (ciliate) (Long et al. 2016) to 3380 substitutions/Gbp/generation in the *Neurospora crassa* (hyphal fungus) (Wang et al. 2020), with mean being approximately 75 substitutions/Gbp/generation (~9 substitutions/Gbp/generation if minimum and maximum outliers excluded).

Somatic mutation rates vary from 0.0038 substitutions/Gbp/cell division in the *Marasmius oreades* (Hiltunen et al. 2019) to 8.6 substitutions/Gbp/cell division in the *Homo sapiens* (Lujan and Kunkel 2021), with mean being approximately 0.8 substitutions/Gbp/cell division (~0.5 substitutions/Gbp/cell division if minimum and maximum outliers excluded).

Table 2.1. Somatic and generational mutation rates in different species. From (Lujan and Kunkel 2021), with changes.

| | | | | | Mutation rates | | | |
Species	Clade	Cellu-larity	Ploidy	Germ V. Soma	Gbp^{-1} $gen.^{-1}$	Gbp^{-1} $div.^{-1}$	Lines	Mutations
Phaeodactylum tricornutum	Stramenopiles	uni-	2n	g	0.49	0.49	36	156
Paramecium tetraurelia	Ciliophora	uni-	2n	g	0.03	0.03	7	29
Tetrahymena thermophila	Ciliophora	uni-	2n	g	0.0076	0.0076	8	5
Plasmodium falciparum	Apicomplexa	uni-	1n	g	0.25	0.25	279	85
Bathycoccus prasinos	Chlorophyta	uni-	1n	g	0.44	0.44	37	32
Chlamydomonas reinhardtii	Chlorophyta	uni-	1n	g	0.18	0.18	91	6890
Micromonas pusilla	Chlorophyta	uni-	1n	g	0.98	0.98	36	85

Ostreococcus mediterraneus	Chlorophyta	uni-	1n	g	0.59	0.59	37	65
Ostreococcus tauri	Chlorophyta	uni-	1n	g	0.48	0.48	40	104
Arabidopsis thaliana	Embryophyta	multi-	2n	g	6.7	0.26	156	2324
Arabidopsis thaliana	Embryophyta	multi-	2n (het.)	g	27	-	99	299
Eucalyptus melliodora	Embryophyta	multi-	2n	g	62	-	1	90
Lemna minor	Embryophyta	multi-	2n	g	0.087	-	16	29
Oryza sativa	Embryophyta	multi-	2n	g	3.2	-	5	10
Oryza sativa	Embryophyta	multi-	2n (het.)	g	11	-	11	55
Picea sitchensis	Embryophyta	multi-	2n	s	27	-	20	5
Populus trichocarpa	Embryophyta	multi-	2n	g	2	-	2	186
Prunus hybrid	Embryophyta	multi-	2n (het.)	g	14	-	30	171
Prunus persica	Embryophyta	multi-	2n	g	8.6	-	32	114
Quercus robur	Embryophyta	multi-	2n	s	47	-	1	17
Silene latifolia	Embryophyta	multi-	2n	g	7.3	-	10	39
Spirodela polyrhiza	Embryophyta	multi-	2n	g	0.082	-	47	46
Dictyostelium discoideum	Mycetozoa	alternate s	1n	g	0.029	0.029	3	1
Neurospora crassa	Ascomycota	multi-	1n	g	3400	-	268	10493
Neurospora crassa	Ascomycota	multi-	1n	s	-	0.6	10	90
Saccharomyces cerevisiae	Ascomycota	uni-	1n	g	0.39	0.35	68	475
Saccharomyces cerevisiae	Ascomycota	uni-	2n	g	0.23	0.23	392	3194
Schizosaccharomyces pombe	Ascomycota	uni-	1n	g	0.37	0.37	180	1308
Marasmius oreades	Basidomycota	multi-	2n	s	73	0.0038	40	111
Schizophyllum commune	Basidomycota	multi-	2n	g	20	-	17	9
Caenorhabditis elegans	Nematoda	multi-	2n	g	3.1	0.57	57	3553
Caenorhabditis species	Nematoda	multi-	2n	g	1.3	0.12	25	448
Pristionchus pacificus	Nematoda	multi-	2n	g	2	-	22	802
Apis mellifera	Arthropoda	multi-	1n	g	4.5	-	46	35
Bombus terrestris	Arthropoda	multi-	1n	g	3.9	-	32	23
Chironomus riparius	Arthropoda	multi-	2n	g	4.2	-	10	51
Daphnia pulex	Arthropoda	multi-	2n	g	3.1	-	30	1210
Drosophila melanogaster	Arthropoda	multi-	2n	g	5.1	0.13	175	3539
Heliconius melpomene	Arthropoda	multi-	2n	g	2.9	0.073	30	9
Aotus nancymaae	Chordata	multi-	2n	g	8.1	-	8	283
Canis lupus	Chordata	multi-	2n	g	4.5	-	4	27
Chlorocebus aethiops	Chordata	multi-	2n	g	9.4	-	3	8
Clupea harengus	Chordata	multi-	2n	g	2	-	12	19
Ficedula albicollis	Chordata	multi-	2n	g	4.6	-	7	55

Gallus gallus domesticus	Chordata	multi-	2n	s	-	0.91	6	384
Gorilla gorilla	Chordata	multi-	2n	g	11	-	1	83
Homo sapiens	Chordata	multi-	2n	g	12	0.17	3062	156475
Homo sapiens	Chordata	multi-	2n	s	-	8.6	388	86157
Macaca mulatta	Chordata	multi-	2n	g	5.8	-	14	307
Mus musculus	Chordata	multi-	2n	g	5.1	0.11	50	1614
Mus musculus	Chordata	multi-	2n	s	-	4.2	30	3697
Pan troglodytes	Chordata	multi-	2n	g	13	-	7	283
Papio anubis	Chordata	multi-	2n	g	6.2	-	12	475
Pongo abelii	Chordata	multi-	2n	g	17	-	1	51

To conclude, the mutational process is ambiguous as it simultaneously produces the opportunity for adaptation and is the cause of deleterious mutations, thus reducing fitness. Because of that, the main parameter of the mutational process - mutation rate - is subject of great interest when studying the evolution of any species.

2.3 Natural selection and the effective population size

Effective population size is a fundamental term introduced by Write in 1931 (Wright 1931). This concept is used to estimate the rate of evolution that is caused by random changes in allele frequencies in a population of a finite size - the process called the genetic drift (Wright 1970).

Effective population size is one of two factors that define the within-population polymorphism in neutral sites. Moreover, the effective population size N_e affects the probability Q of fixation of a non-neutral mutation that has a selective coefficient s (Charlesworth 2009):

$$Q = \frac{N_e s}{N} \frac{1}{1 - e^{-2N_e s}}$$

Nearly neutral evolutionary theory states that slightly deleterious and slightly beneficial mutation ($|N_e s| \ll 1$) act as neutral and may be randomly fixed in a population due to genetic drift (Ohta 1992). For example, the probability of fixation of a mutation for which $|N_e s| < 0.25$ almost equals that for a neutral mutation, while deleterious mutations for which $|N_e s| > 2$ have almost no chance to be fixed in a population (Charlesworth 2009). Thus, natural selection is more efficient in populations with larger effective population size, as with the increase of the effective population size mutations with less module if the selection coefficient become effectively non-neutral and prone to the natural selection. At the same time, in small populations the natural selection against deleterious mutations may become non-effective, and some deleterious mutations may randomly fix. The smaller the population is, the more deleterious mutations may fix in it.

The relationship between the effective population size and the effectiveness of natural selection was studied in a number of works. In (Popadin et al. 2007) the relationship between natural selection in mitochondria and the body mass as a proxy for the effective population size was studied. In (Prado-Martinez et al. 2013; Romiguier et al. 2014; de Valles-Ibáñez et al. 2016; Chen et al.), whole genome and transcriptome data was used to study the relationship between natural selection and neutral polymorphism as well as different life-history traits of species.

The impact of the effective population size on molecular evolution was directly studied in (Katju et al. 2015). 35 MA lines of *C. elegans* with different population sizes were studied across 410 generations. The spectrum of the selection coefficients of accumulation mutations is shown in Fig. 6.

Fig. 6. The spectrum of the selection coefficients of accumulation mutations (Katju et al. 2015).

A significant decrease of fitness was observed in populations with the lowest effective population size (Fig. 7).
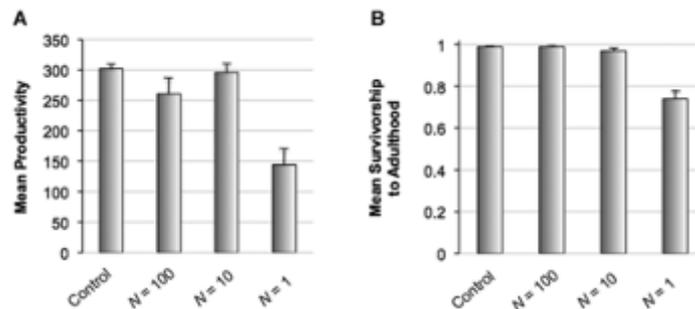
Fig. 7. The decrease of fitness in MA lines with population size of 1 (Katju et al. 2015).

2.4 Homologous recombination

Homologous recombination is an exchange between maternal and paternal DNA from homologous chromosomes. This process is one of the key aspects of sexual reproduction that allows species to overcome the process known as Muller's ratchet (Felsenstein 1974) - the inevitable decrease of haplotype fitness with time due to the accumulation of deleterious

mutations. Recombination provides the opportunity to break linkage between beneficial and deleterious mutations, thus setting beneficial mutations free from the burden of linked deleterious mutations. However, recombinations may also break combinations of beneficial variants, thus reducing the overall fitness (Barton 1995; Charlesworth and Barton 1996; Otto 2009). This makes the homologous recombination itself and its rates a subject of interest not less than that for the mutational process.

It is now clear that the recombination rates vary between taxa, populations, sexes, individuals and genome regions (Stapley et al. 2017).

Variation between taxa. The compilation of different studies and the comparison of the genome-wide recombination rates (GwRR, the sum of distances between loci in cM divided by the genome length in Mb) has revealed that the recombination rates may have at least one order of magnitude difference between distant taxa (Table 2.2 and Fig. 8, (Stapley et al. 2017)). In particular, microorganisms (SAR group) and fungi have the recombination rates in 1.4 - 120 cM/Mb, with means being 39 and 49 cM/Mb respectively. Moreover, fungi have the recombination rate in approximately 20 - 120 cM/Mb range when the outlier with the unusually low recombination rate is not considered (Fig. 8). In contrast, animals and plants have the recombination rates in 0.03 - 28.1 cM/Mb range, with means being 2.52 and 1.85 cM/Mb respectively.

Table 2.2. Recombination characteristics for large taxa (from (Stapley et al. 2017) with changes).

group	*n*	linkage map length (cM)			genome size (Mb)			recombination rate (cM/Mb)		
		mean	min	max	mean	min	max	mean	min	max
SAR	9	1782	653	2884	189	18.87	560	38.67	3.24	108
fungi	15	2068	86	5860	49.26	19.05	170.2	48.68	1.4	119.9
animals	140	1813	90	5961	1538	43.15	30880	2.52	0.12	28.1
plants	189	1567	309	8184	2956	120.4	29280	1.85	0.03	9.22
total or mean	353	1807.5			1183			22.93		
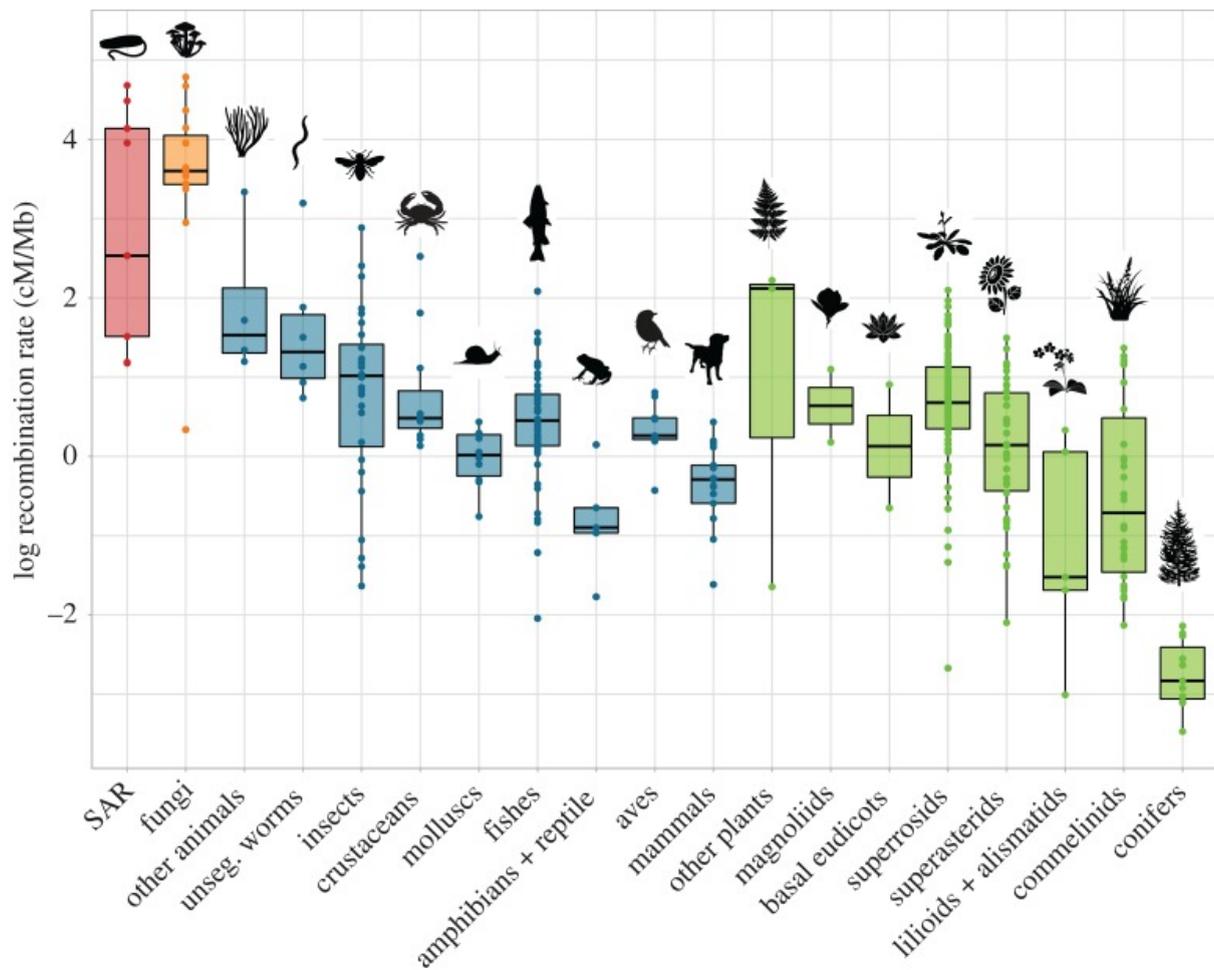
Fig. 8. Log recombination rates across large taxa (Stapley et al. 2017).

Patterns of recombination within a genome. Recombination events are known to happen non-randomly across the genome, both in terms of single or multiple crossing-over (CO) events. First, in many species recombination tends to happen within narrow regions called recombination hotspots, where the probability of the recombination increases drastically (Gerton et al. 2000; Choi and Henderson 2015; Croll et al. 2015; Singhal et al. 2015; Latrille et al. 2017); this includes *S. commune*, where CO events tend to occur within more conserved regions (Seplyarskiy et al. 2014). However, some species lack recombination hotspots (Rockman and Kruglyak 2009; Comeron et al. 2012; Smukowski Heil et al. 2015; Wallberg et al. 2015). Second, a process called CO interference is known to suppress the formation of chiasma and following CO events near already formed chiasma (Hillers 2004). This means that the probability of two closely located CO events is lower than expected by chance given the probability of one CO event.

Chapter 3. Accumulation of somatic mutations in growing mononuclear haploid mycelia of *Schizophyllum commune in vitro*

3.1 Introduction

The per generation mutation rate in a multicellular organism is a product of the mutation rate per cell division and the number of mitoses between two consecutive eiosis. Thus, the per generation mutation rate can be modulated by two, not mutually exclusive, mechanisms. The first one is to reduce the per cell division mutation rate, and the second one is to reduce the number of mitoses between consecutive meiosis. For brevity, we will refer to them as "fidelity" and "economy" mechanisms, respectively. Both can be implemented in a variety of ways.

The fidelity mechanism can involve reduction of the mutation rate in all cells. However, in this case the cost of fidelity (Blomberg 1987) is incurred across-the-board. Thus, in species with a dedicated germline, the per cell division mutation rate may be specifically reduced, by as much as an order of magnitude, only in germline cells, as it is the case in mammals (Milholland et al. 2017).

The economy mechanism can also depend on the existence of a dedicated germline, if it is shielded from repetitive divisions during the lifetime of an organism, as it is the case in females, although not in males, of mammals (Kong et al. 2012; Jónsson et al. 2017). However, there are ways to reduce the number of mitoses between consecutive meiosis even in the absence of a germline. First, shoots or hyphae of an organism may possess apical cells that divide only rarely because most of the growth occurs due to intercalary cell divisions (Lanfear et al. 2013; Anderson and Catona 2014). Another potential scenario is an increase of

38

cell size, which reduces the number of cell divisions required for a given amount of linear growth.

Obviously, a reduced per cell division mutation rate still leads to a linear accumulation of mutations with the number of mitoses between consecutive eiosis. Similarly, if the economy of cell divisions is achieved by the increased size of cells, mutations still should accumulate linearly with somatic growth of an organism. In contrast, a shielded germline or apical cells are likely to lead to decelerated accumulation of mutations with age or in the course of somatic growth. The number of cell divisions before gametogenesis and, thus, the number of mutations accumulated per generation was found to be independent of the life span and the extent of vegetative growth in *Arabidopsis thaliana* (Watson et al. 2016), indicating that the economy mechanism is operating, perhaps through intercalary growth.

Mutation rate-reducing mechanisms can be particularly salient in species where individuals can reach huge sizes, such as some plants and fungi. Indeed, in several such species the number of genetic differences between even remote parts of the same individual are surprisingly low. This is the case for the oak *Quercus robur* (Schmid-Siegert et al. 2017), a giant honey mushroom *Armillaria gallica* (Anderson et al. 2018), and the fairy-ring fungus *Marasmius oreades* (Hiltunen et al. 2019).

While the mutation rate per generation can be easily measured by comparing individuals, measuring the mutation rate per cell division is harder. In multicellular organisms, this could be achieved by either direct sequencing of a cell and its offspring, or of two cells separated by a known number of cell divisions. However, single-cell sequencing is still in its infancy, and it is hard to track cell lineages within an individual, which precludes precise estimates of the number of cell divisions separating two locations within the organism.

Mycelial fungi are characterized by linear mycelial growth, possibly simplifying this task. Still, making use of this advantage is difficult. First, the exact linear distance between locations within a mycelium can only be measured in a lab, and many fungi cannot be easily cultivated. Second, it remains unknown how the number of cell divisions scales with the linear distance. Third, fungi often have multinuclear cells, complicating measurements and interpretation of data.

However, *S. commune* is a model organism that lacks all these disadvantages. The mycelium of *S. commune* grows linearly and apically in cell-thick hyphae (Gooday 1995); the cell length is known, and comprises approximately 100 μm (Essig 1922), and mycelium of the monokaryon stage can be relatively easily cultivated on solid media, where it grows vegetatively without producing fruit bodies, and where the distance between to samples of the mycelium can be measured. The distance and the cell length produces the number of cell divisions between two samples of the mycelium, and knowing the number of cell divisions between two points, it is easy to estimate the mutation rate per cell division. Moreover, it becomes possible to study the dynamics of somatic mutation accumulation with the linear growth of the mycelium.

3.2 Experimental layout

We developed an experimental system that allows us to cultivate haploid mycelia of *S. commune* for a long period of time, maintaining a strictly vegetative mode of growth and an approximately constant number of growing hyphae. Each culture was started from a single haplospore which gave rise to a haploid mycelium with mononuclear cells (Stankis et al. 1990), and was then cultivated in glass tubes of a fixed diameter on solid medium. We regularly measured growth rates of the mycelia along the tube, and took samples for

sequencing. We also sequenced all founding cultures and performed genome assembly for each culture individually. Sequenced samples of derived cultures were then mapped to corresponding assemblies; this was done to achieve good mapping quality, as mapping on an assembly of a different individual is difficult because of the high genetic diversity of *S. commune* (Baranova et al. 2015).

We used tubes of two different diameters. Narrow tubes had an inner diameter of approximately 0.8 mm, with its width partially filled with solid medium. Thick tubes comprised a cylinder of solid medium 4 mm in diameter, placed within a glass tube with a slightly larger inner diameter (Fig. 9A). The tubes were 15-20 cm long, and cultures were transferred to the next tube as soon as the hyphae reached the end of the tube. In the case of narrow tubes, this procedure by itself did not always yield successful replanting because the number of transferred cells was too small. Therefore, before the transfer to the next tube, cultures were cultivated on Petri dishes for some time to obtain enough material. The overall period of growth on the Petri dish was ~20 times shorter than that in the tubes, and was not counted towards the overall growth time of the corresponding mycelium. For transfer, we then attempted to sample cells from the same position of the Petri dish where the culture was planted, minimizing the number of mutations accumulated on the Petri dish. At the time of transfer, mycelial samples were also collected for sequencing. The overall experimental layout is shown in Fig. 9B.
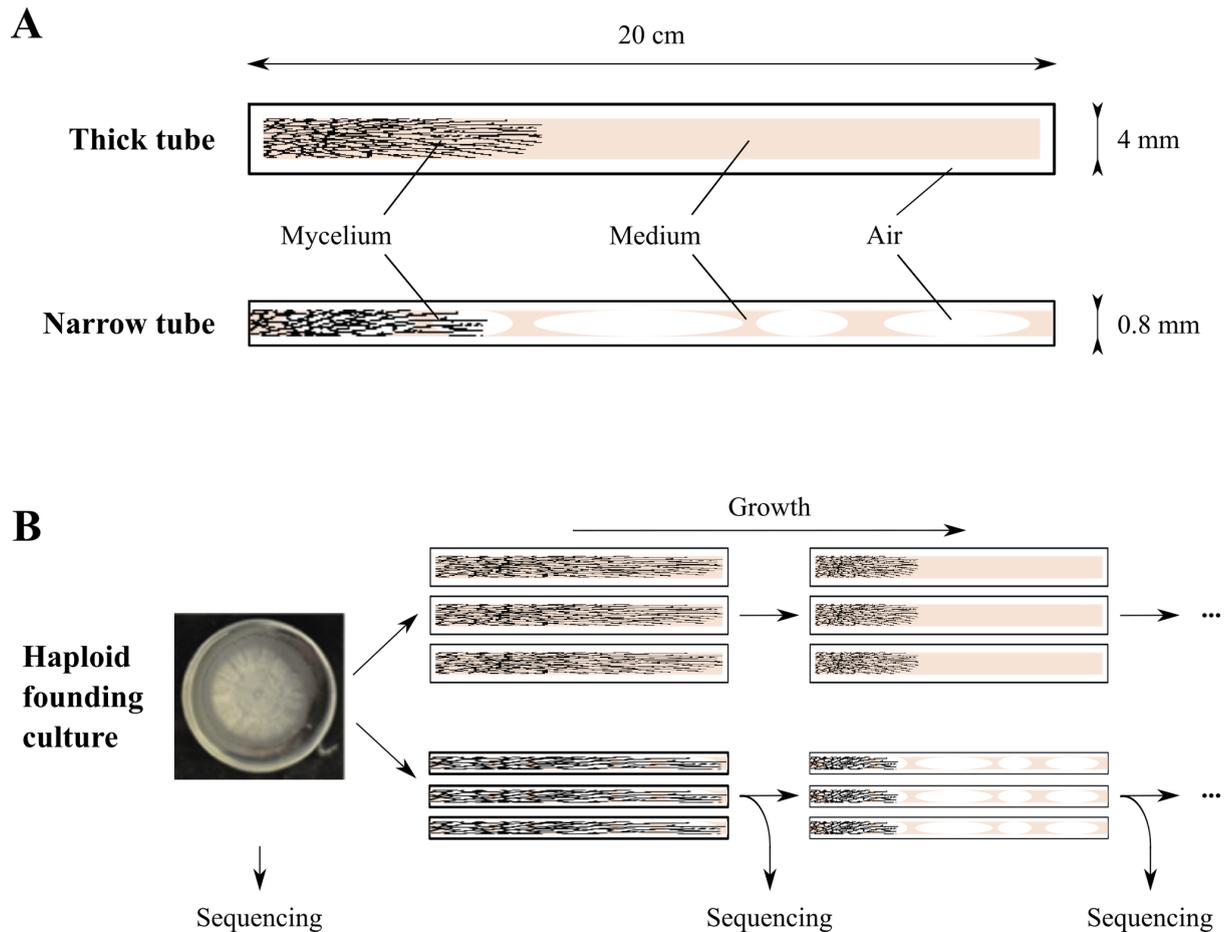
Fig. 9. Experimental system. (A) Schematic representation of the tubes used in the experiment (not to scale). (B) Overall experimental layout.

We used four founding haploid cultures, each originated from a single haplospore. Three of the cultures (sh01, sh02, sh03; specimen vouchers WS-M203, WS-M222, WS-M276) were obtained from fruit bodies collected in Ann Arbor, Michigan, USA, and one culture (sh04; specimen voucher WS-M45), from a fruit body collected in Moscow, Russia. Each founding culture was used to start six experimental lines in tubes of two different diameters (three replicates in each), for a total of 24 experimental lines.

3.3 Materials & Methods

Obtaining original haploid mononuclear cultures. *S. commune* fruit bodies were collected from tree trunks in autumn. Each fruit body was then attached to the Petri dish lid over the solid agar medium (see below), and the Petri dish was placed in a diagonal manner. The Petri dish was exposed to the light at room temperature, and under such conditions some fruit bodies released haploid spores that turned out on the surface of the solid medium. At the periphery of the area where spores were located, it was possible to visually locate individual spores. Such spores were cut out from the solid medium and transferred to new Petri dishes where they originated mononuclear haploid cultures.

Cultivation and preservation. Cultures were cultivated on solid medium (beer wort Maltax10 – 25.6 g, water – 1 l, agar – 40 g) in the light at room temperature. Collected samples and founding cultures were stored at 4ºC and -20ºC.

Cell size measurement. Founding cultures were separately cultivated in thick and narrow (2 replicates) tubes until mycelia reached a length of 5 cm. Apical mycelia were sampled and longitudinal sections were prepared. Length of apical cells was measured using Altami Bio 1 microscope in transmitted light using 40X/0.65 objective, U3CMOS05100KPA camera and ToupView 3.7.5 ToupTek Photonics software with 0.1 μm precision.

Whole-genome sequencing. Before DNA extraction, samples of mycelium were first grown in liquid medium (beer wort Maltax10 – 8 g, water – 1 l) on shaker to reach sufficient mass, and then were lyophilized. DNA was extracted using the CTAB method (Doyle and Doyle 1987). Libraries were prepared using NEBNext® Ultra™ II DNA Library Prep Kit for Illumina with 5 PCR cycles (or Accel-NGS® 2S Plus DNA Library Kit with 6 PCR cycles) and sequenced on Illumina HiSeq2000 platform with 127 bp pair-end reads. Two biological replicates were sequenced independently for samples sh01 - sh39.

***De novo* genome assembling and annotation.** Although a *S. commune* reference genome is available (Ohm et al. 2010), it is difficult to map reads from other *S. commune* individuals onto it due to extreme genetic diversity (Baranova et al. 2015). Thus, we obtained *de novo* genome assemblies for each founding culture. Pair-end reads were trimmed using Trimmomatic (Bolger et al. 2014) with options (ILLUMINACLIP:adapters:2:30:10 LEADING:3 TRAILING:3 SLIDINGWINDOW:4:20 MINLEN:36). *De novo* genome assemblies were obtained using SPAdes (Bankevich et al. 2012) (with -k 21,33,55,77 --careful --only-assembler options). Assemblies were filtered of contamination using Blobology (Keightley et al. 2009). We aligned our assemblies and reference genome using Lastz (Harris 2007), removed overlapped regions using single_cov2 program from Multiz package (Blanchette et al. 2004), and used the existing annotation of the reference genome *S. commune* H4-8 v3.0 (JGI) to annotate coding sequences. Assembly and annotation statistics are presented in Tables A1 and A2.

Variant calling. Pair-end reads trimmed using Trimmomatic were mapped onto corresponding reference assemblies using bowtie2 (Langmead and Salzberg 2012). Only reads with properly mapped pair and with mapping quality 42 were kept. Duplicate reads were removed using Picard Tools (Broad Institute). *De novo* single nucleotide mutations in experimental lines were called as follows. First, all positions with at least one read supporting the non-reference base were listed, and a total of 32280 positions were obtained. At these positions, we called variants that had the following properties: (i) at least in one sample, coverage in the 10-90% range and non-reference variant frequency >30%, or coverage in the 15-85% range and non-reference variant frequency >20% (13962 variants); (ii) not supported by any read in the reference sequence (289 variants). For these variants, we assessed their frequencies in all samples. For samples sh01 - sh39 variant frequency was calculated as mean between two sequenced replicas. Short indels were called using samtools mpileup and

freebayes software, with the same filters as described above applied. Called *de novo* mutations are listed in Tables A3 and A4.

Dn/Ds ratio and expected distributions of the number of nonsynonymous and coding mutations. Dn/Ds ratio was calculated using codeml program from PAML software (Yang 2007) with the following options: runmode = 0, seqtype = 1, CodonFreq = 2, clock = 0, model = 0, NSsites = 0, icode = 0, fix_kappa = 0, kappa = 2, fix_omega = 0, omega = 2, fix_alpha = 1, alpha = .0, Malpha = 0, ncatG = 4.

ANOVA. To see how mutation rate correlates with time, we used a two-way ANOVA with genotype and tube size as a categorical fixed effects and mycelium length (which reflects time) as a continuous predictor. To see how mutation rate correlates with tube sizes and founding cultures, we used a two-way ANOVA with genotype and tube size as categorical fixed effects. To check that average sample coverage does not correlate with the inferred mutation rates, we also included mean coverage as a covariate in both ANOVA tests and saw no correlation between coverage and mutation rate

3.4 Results

Cell size. Measured cell lengths for different cultures and different tube sizes are presented in Table 3.1; we estimated mean cell length at 163 (95% CI: 154.75 - 171.25) and 165 µm (95% CI: 157.48 - 172.52) in narrow and thick tubes respectively.

Table 3.1. Cell sizes.

Size	Culture	Mean cell length, um	SE, um	N
Thick	sh01	174,53	8,71	57
	sh02	161,07	6,09	99
	sh03	172,07	8,4	59
	sh04	149,83	8,04	55
Thick		164,57	3,82	270
Narrow	sh01	178,44	9,65	56
	sh02	148,92	6,48	97
	sh03	156,08	7,94	82
	sh04	170,08	10,91	64
Narrow		162,91	4,19	299

Mycelia cultivation. We cultivated 24 experimental haploid mononuclear lines of *S. commune* in glass tubes for between 220 and 360 days. The mean growth rate in the thick tubes (5.9 mm/day) was almost twice as high as in the narrow tubes (3.5 mm/day). The cultures have grown up to 96 cm in narrow tubes, with the mean 78 cm (corresponding to approximately 4 800 cell divisions), and up to 247 cm in thick tubes, with the mean 198 cm (approximately 12 000 cell divisions). The growth rate remained constant in the thick tubes, while in the narrow tubes, it decreased slightly but significantly (Fig. 10).
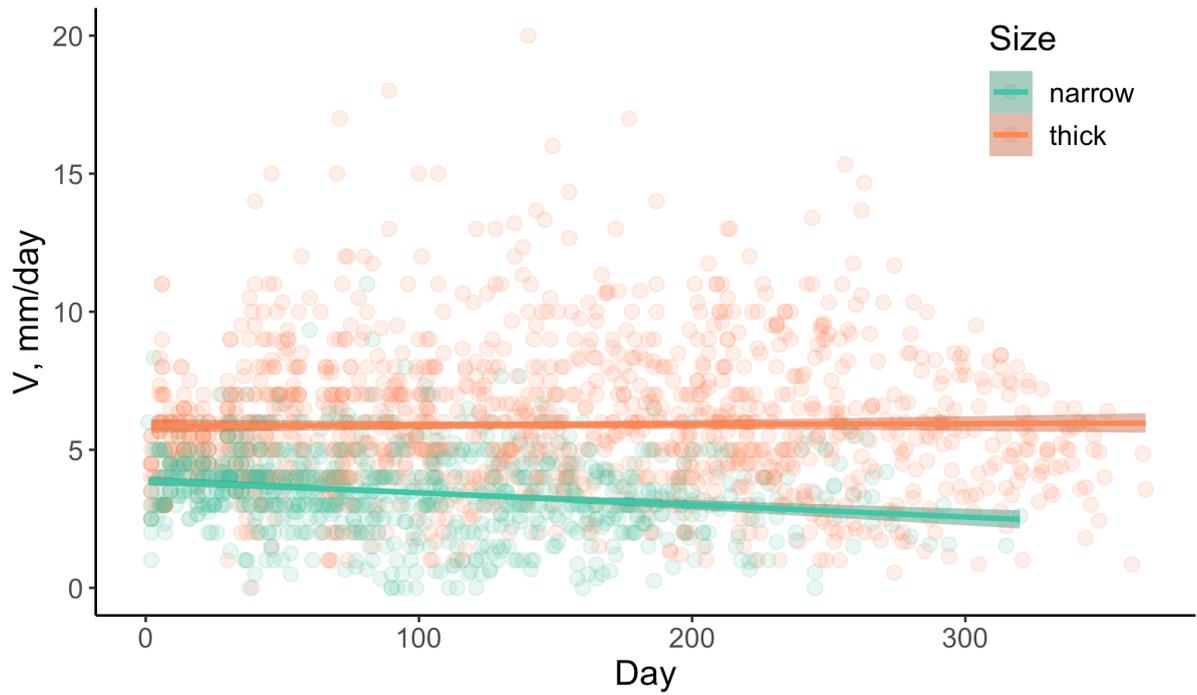
Fig. 10. Growth rates in thick and narrow tubes during the experiment. Data for all lines are pooled together. Linear regression for narrow tubes: R2 = -0.04, P-value = $3.7 \cdot 10^{-9}$. Linear regression for thick tubes: R2 = $1.2 \cdot 10^{-4}$, P-value = 0.68.

Accumulation of *de novo* mutations. We obtained and sequenced a total of 112 samples of growing mycelium. Each of the 24 lineages was successively sampled from 4 to 7 times (Fig. 11).
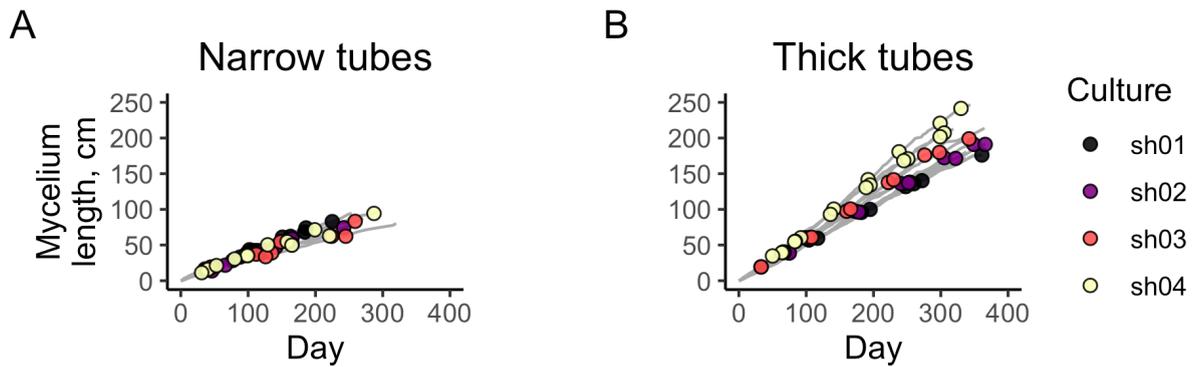
Fig. 11. Growth of the mycelia during the experiment in narrow (A) and thick (B) tubes. Sequenced points are marked with circles.

Each sample was sequenced with the average coverage 135x, and a total of 300 *de novo* mutations was detected (Table A3 and A4); the mutational spectrum is shown on Fig. 12. Among these mutations, 63 were coding, including 45 nonsynonymous mutations, 2 nonsense mutations, 3 frameshifts and 1 stopgain insertion. Most of these mutations were fixed in the mycelium, i.e., have been present in all or nearly all reads in all subsequent time points; still, a number of mutations have reached high frequencies but were later lost, and some mutations have never reached high frequencies (Table 3.2). In each line, the vast majority of mutations that were observed at the last time point (72-100%) were fixed. We observed no parallel mutations between different founding cultures. Among coding mutations, the overall dN/dS ratio was somewhat lower in thick tubes (0.7) than in narrow tubes (1.0), although the difference was not significant.

Table 3.2. Number of different types of *de novo* mutations.

# *de novo* mutations	Narrow tubes				Thick Tubes				Total
	sh0 1	sh0 2	sh0 3	sh0 4	sh0 1	sh0 2	sh0 3	sh0 4	
Total	25	31	20	55	15	27	30	97	300
Single nucleotide variants	25	29	20	54	12	26	30	93	289
Indels	0	2	0	1	3	1	0	4	11
Categorized by fate									
Never reached frequency of 70%	8	3	0	17	3	4	4	22	61
Reached frequency of 70% but then lost	3	2	0	12	1	0	2	5	25
Fixed	14	26	20	26	11	23	24	70	214
Categorized by type									
Nonsense	0	0	0	0	0	1	1	1	3
Nonsynonymous	1	6	5	10	3	4	5	11	45
Synonymous	1	2	2	2	0	0	3	6	16
Frameshift	0	0	0	0	1	1	0	1	3
Intronic	1	2	2	3	0	0	1	5	14
Other non-coding	22	21	11	40	11	21	20	73	219
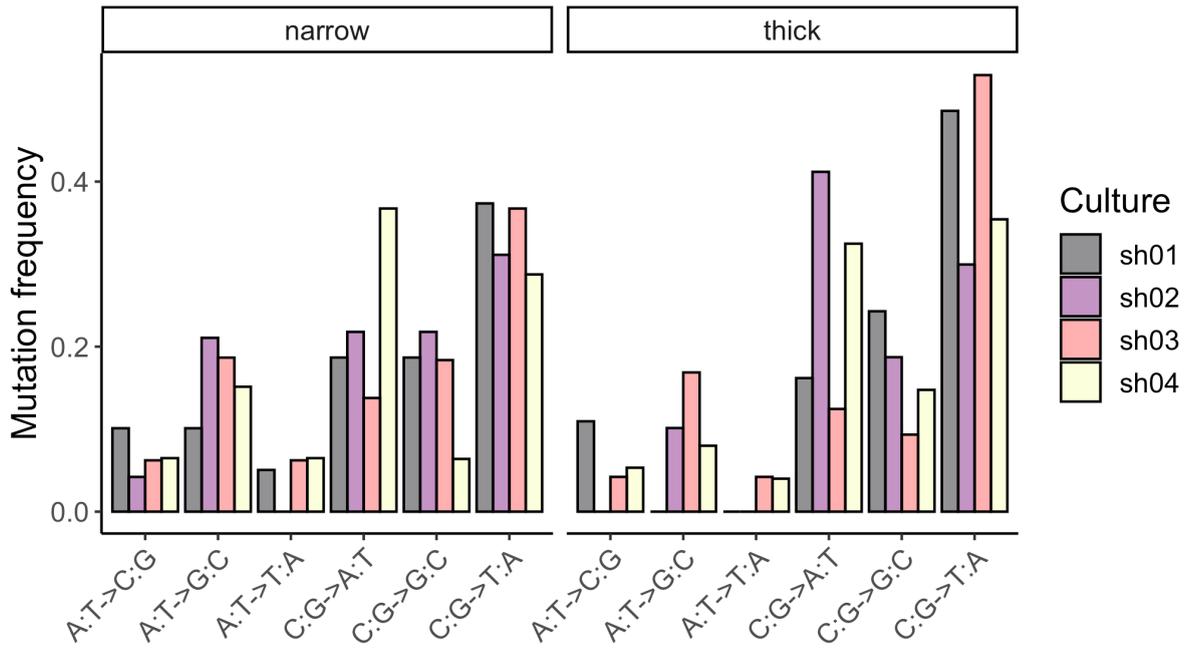
Fig. 12. Mutational spectrum for narrow and thick tubes.

By the end of the experiment, between 2 and 29 mutations have reached frequencies over 70% in each line, with the mean value of 9 mutations. The dynamics of this accumulation is shown in Fig. 13. We saw no significant change in the mutation rate over the course of mycelial growth (ANOVA test, P-value = 0.084; Fig. 14).
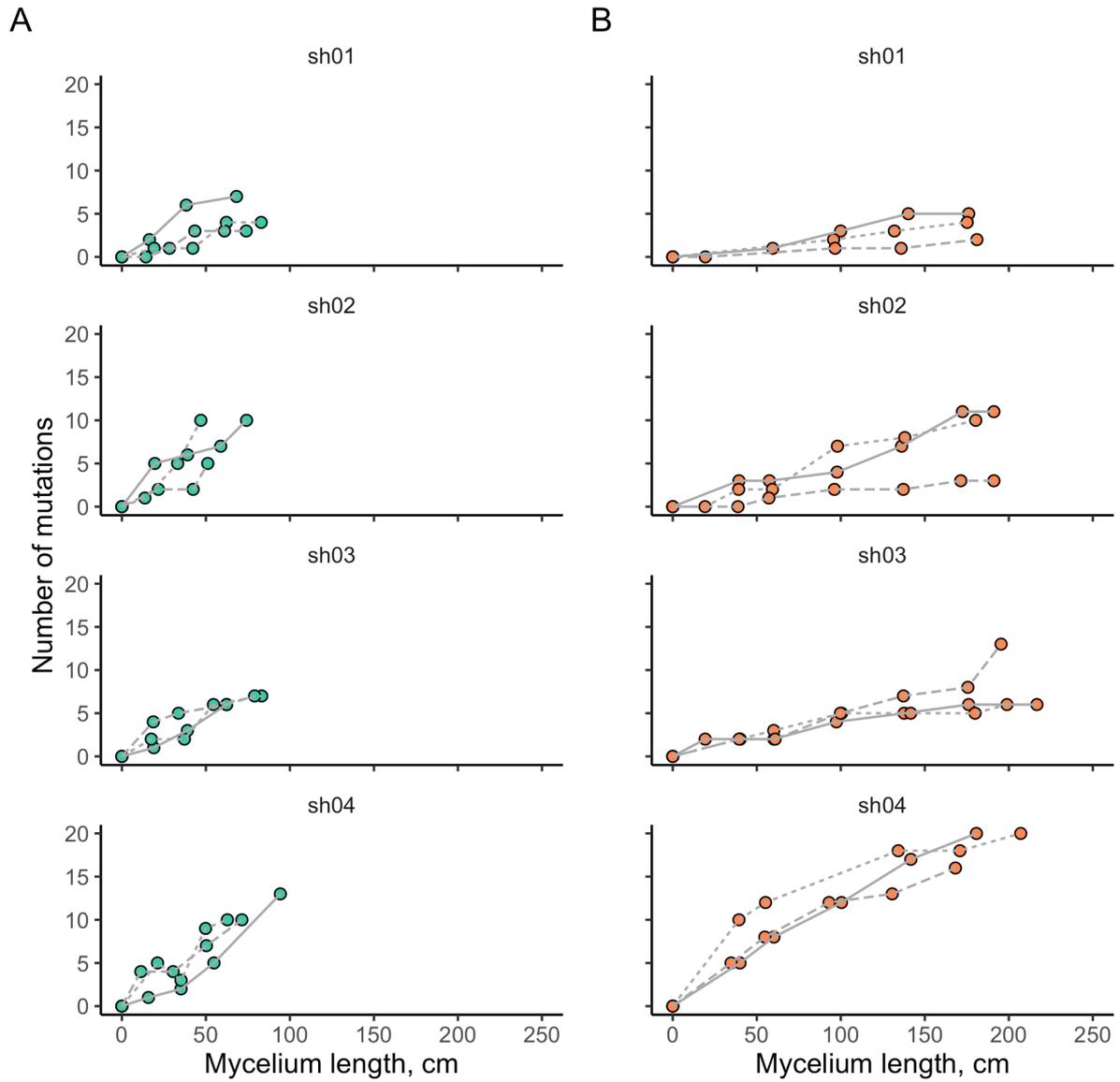
Fig. 13. Accumulation of mutations during the growth of the mycelium. Number of mutations that have reached 70% frequency in sequenced samples are shown. Replicas are displayed with different line types. (A) Narrow tubes. (B) Thick tubes.
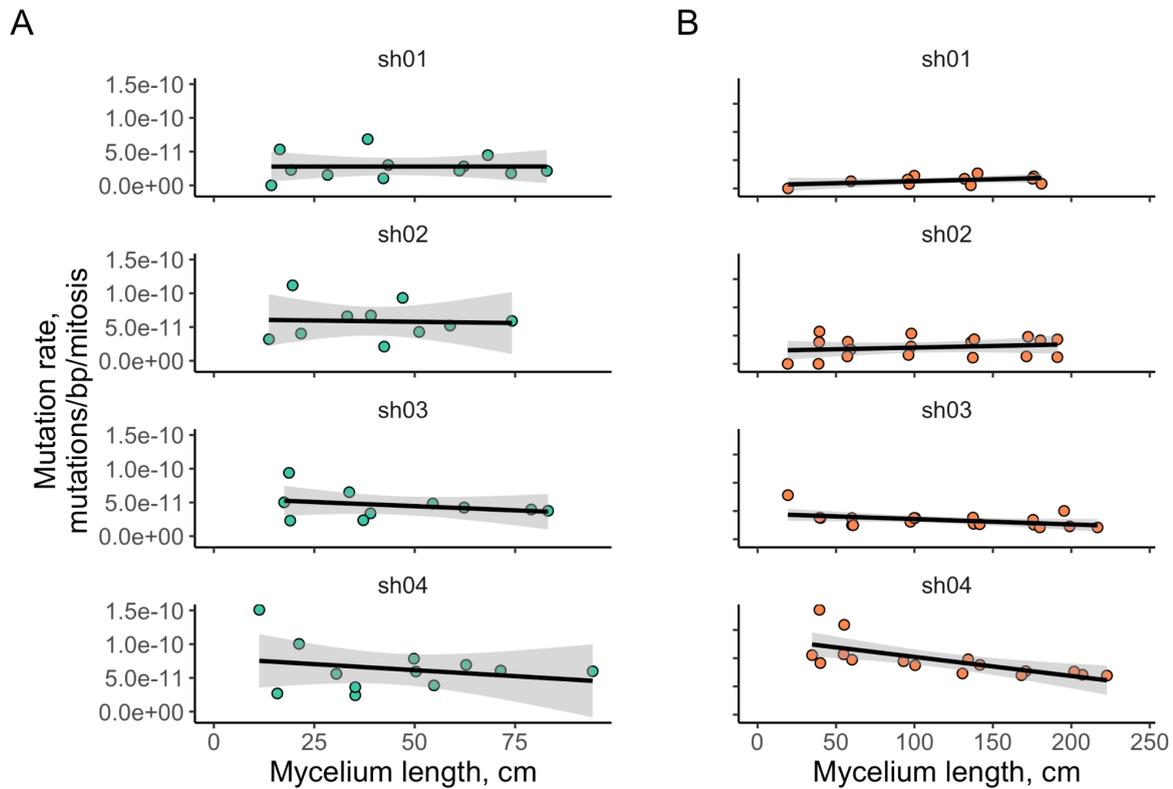
Fig. 14. The relationship between the mutation accumulation rate and mycelium length. (A) Narrow tubes. (B) Thick tubes.

We used the mean cell length estimates of 163 μm in narrow tubes and 165 μm in thick tubes to estimate the rate at which new mutations fix in a growing mycelium per cell division of linear growth. This rate in the narrow tubes ($4.99 \cdot 10^{-11}$ substitutions/nucleotide/cell division, 95% CI: $3.62 \cdot 10^{-11} - 6.36 \cdot 10^{-11}$) was more than twice as high as that in the thick tubes ($2.04 \cdot 10^{-11}$, 95% CI: $1.14 \cdot 10^{-11} - 2.93 \cdot 10^{-11}$; ANOVA test, P-value $= 4.82 \cdot 10^{-5}$) (Fig. 15A). The mutation rate differed significantly between founding cultures (ANOVA test, P-value $= 0.0024$) (Fig. 15B).
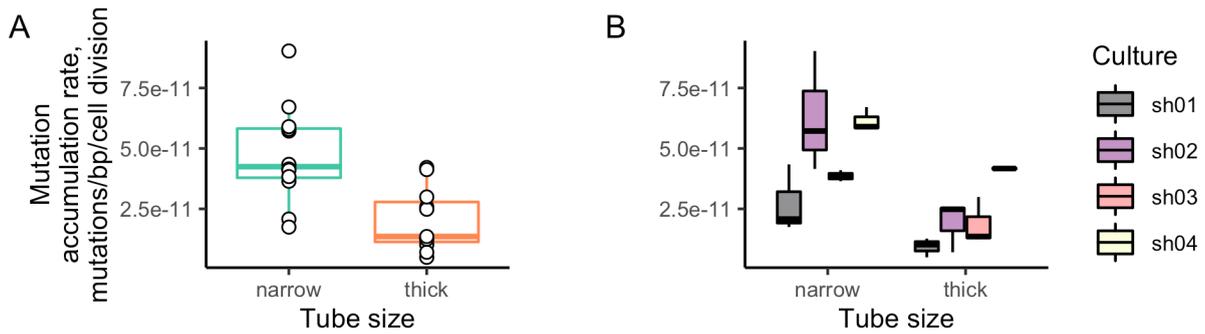
Fig. 15. Mutation accumulation rates in narrow and thick tubes (A) and for individual founding cultures (B).

3.5 Discussion

S. commune is a mycelial fungus and can grow through distances of the order of several meters, occupying whole tree trunks. One can expect *S. commune* to have some mechanism that will minimize the number of mutations accumulated during vegetative somatic growth in order to reduce the per generation mutation rate. Both "fidelity" and "economy" mechanisms of this reduction are well-known for mammals (Kong et al. 2012; Jónsson et al. 2017)(Kong et al. 2012; Jónsson et al. 2017) and have been recently reported for plants (Watson et al. 2016; Milholland et al. 2017) and fungi ((Anderson et al. 2018), see our analysis of their data in Fig. 16, and (Hiltunen et al. 2019)). If *S. commune* were employing an economy mechanism similar to that found in several species with extensive vegetative growth, this would likely lead to a slower-than-linear accumulation of mutations with its growth.
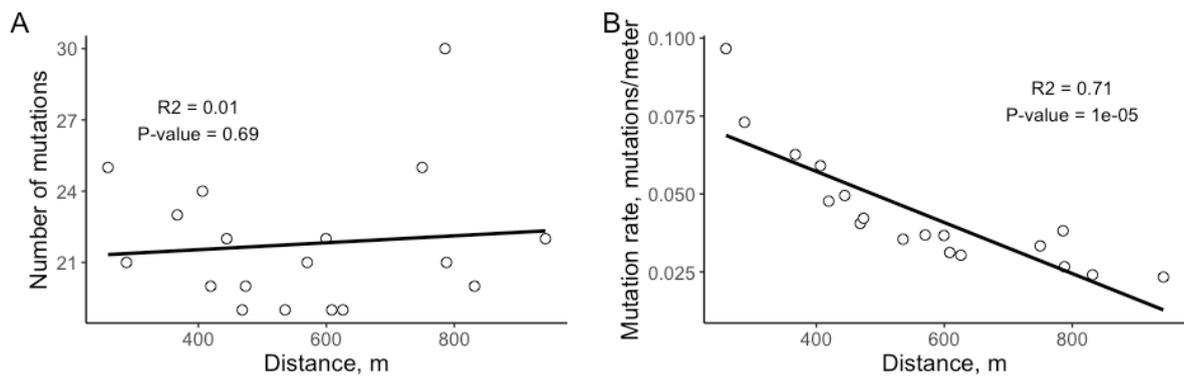
Fig. 16. Relationship between the number of mutations (A) and mutation rate, and the distance between sequenced samples in Armillaria fungus. Obtained based on data from (Anderson et al. 2018).

In our experiment, however, mutations accumulated linearly with the number of cell divisions, so that the number of mutations was proportional to the mycelium length (Fig. 13, 14). This is what allows us to report just a single per cell division mutation rate for a mycelium. The mutation accumulation rate varied both between the lines and the tube sizes. Assuming that the process of mycelial growth in nature, as well as on a Petri dish, is better represented by growth in thick than in narrow tubes, we estimate the mutation accumulation rate at $2.04 \cdot 10^{-11}$ mutations/nucleotide/cell division, or $1.24 \cdot 10^{-7}$ mutations/nucleotide/m.

This estimate is broadly consistent with that obtained by Baranova et al. (2015). In that work, the per generation mutation rate during growth on a Petri dish was estimated as $2 \cdot 10^{-8}$ mutations/nucleotide/generation. Although the exact amount of mycelial growth between generations was not measured in that experiment, it was roughly ~10 cm, giving the mutation accumulation rate of $\sim 2 \cdot 10^{-7}$ mutations/nucleotide/m, which is similar to our result. It is hard to compare per cell division estimates of mutation rates obtained in different studies, as the

number of cell divisions is usually unknown. Still, the mutation rate per unit linear growth in *S. commune* seems high. In oak, a comparison of parts of the same tree yielded the mutation rate estimate of ~$3.3 \cdot 10^{-10}$ mutations/nucleotide/m (Schmid-Siegert et al. 2017), or ~$3.3 \cdot 10^{-9}$ mutations/nucleotide/generation for an oak 10 meters high. The per meter mutation rate in *A. gallica* is lower than $5 \cdot 10^{-10}$ mutations/nucleotide/m (Anderson et al. 2018). The per mitosis mutation rate in *Marasmius oreades* fungus was found to be approximately one order of magnitude lower than that in *S. commune* (Hiltunen et al. 2019).

Although higher than in most of previously studied fungi and plants (except for the striking example of *Neurospora crassa* in which the somatic mutation rate appeared to be extremely high (Wang et al. 2020), the per cell division mutation accumulation rate in our study is lower than the somatic mutation rates in humans and mice, being closer to their germline mutation rates. In (Milholland et al. 2017), the median germline mutation rates were estimated at $3.3 \cdot 10^{-11}$ and $1.2 \cdot 10^{-10}$ mutations per nucleotide per mitosis for humans and mice, respectively, while the somatic mutation rates (in fibroblasts) were estimated at $2.66 \cdot 10^{-9}$ and $8.1 \cdot 10^{-9}$.

Even though the per mitosis mutation rate in *S. commune* appeared to be quite moderate, the linear scaling of the number of accumulated mutations with distance may result in very large per generation mutation rates if the mycelium growth spans large distances. If the mutations continue to accumulate linearly, a distance between fruiting bodies of ~1 m can result in a per generation mutation rate of the order of 10^{-7} substitutions/nucleotide, which is at the top of the known mutation rate range except for *Neurospora crassa* (Wang et al. 2020; Lujan and Kunkel 2021); and if this distance is larger, this rate can be even higher.

Such a high per generation mutation rate might contribute to the extreme genetic diversity of *S. commune*. In addition, if the variability in mycelial length between fruiting bodies in *S.*

commune is high, which is observed in other basidiomycetes (Anderson et al. 2018), linear accumulation of mutations may result in high variability of the per generation mutation rate between parent-offspring pairs. Moreover, this puts previous estimation of *S. commune* N_e (Baranova et al. 2015) on the high end of the spectrum. If *S. commune* is, indeed, characterized by both high N_e and high per generation mutation rate, this would imply that a high mutation rate does not need to be explained through inefficient selection in small populations (Lynch et al. 2016). Still, (Xu et al. 2019) observed a low mutation rate in a duckweed *Spirodela polyrhiza* which has high N_e. Thus, it is not clear if there is any causal connection between evolution of mutation rates and the strength of random drift, and this issue warrants further study.

The mutation rate differs strongly between founding cultures, and these differences are consistent between replicas (Fig. 15), implying that they are at least partly determined by the genotype of the fungus. The rate in the culture from the Russian population (sh04) was larger than those in cultures collected from North American populations (sh01, sh02, sh03). This is unexpected, since the genetic diversity in the Russian population is lower than that in the North American populations (Baranova et al. 2015). The differences in diversity levels between the two populations are therefore not explainable by their different mutation rates per unit length, and may instead arise from differences in other factors such as effective population size or length of mycelia.

The rate at which mutations accumulate can be affected by selection discriminating between the growing hyphae. As selection is expected to be more efficient in larger populations (Kimura 1983), we expect its effect to be more pronounced in thick than in narrow tubes. Our data provide some evidence for such selection. First, the mutation accumulation rate is lower in thick tubes than in narrow tubes, consistent with negative selection ridding the population of some of the hyphae carrying deleterious mutations in thick tubes. Second, the mycelium

56

growth rate decreases over the course of the experiment in the narrow tubes, consistent with accumulation of deleterious mutations in them that decrease the growth rate; but it remains constant in thick tubes, consistent with negative selection purging these mutations. Third, the dN/dS ratio among the accumulated mutations in thick tubes (but not in narrow tubes) appears to be lower than 1, although the difference is not statistically significant. If such selection indeed operates in nature, then the actual per-generation number of mutations distinguishing the parental and offspring individuals of *S. commune* can be shaped not just by the mutation rate and the number of cell divisions, but also by the extent of competition between hyphae within a mycelium. Selection between germ line cell lineages is not unprecedented and has been observed before, for example, as competition between sperm lines in multiple species including humans, other mammals and birds (Ramm et al. 2014) and purifying selection reducing mitochondrial heteroplasmy in mammalian female germ lines. This selection is an interesting field for further research.

Chapter 4. Accumulation of somatic mutations in growing dikaryon mycelia of

Schizophyllum commune in vivo

4.1 Introduction

S. commune is the record holder for the level of genetic diversity among studied species (Baranova et al. 2015). Genetic diversity, or virtual heterozygosity, is proportional to the generational mutation rate and the effective population size. Thus, one can expect either or both of these values for *S. commune* to be high. It was previously shown that the gerational mutation rate of *S. commune in vitro* is although high, but not extremely high, being of the same order magnitude (10^{-8} substitutions/nucleotide/generation) as generational mutation rate for humans . However, as we show in Chapter 3 of this work, the generational mutation rate of *S. commune* has the potential to be much higher in nature than *in vitro*. *S. commune* can occupy territory both via mono- and dikaryon mycelia (Palmer and Horton 2006), and both of these stages can accumulate somatic mutations that can further translate into generational mutations. Here, we aim to study how somatic mutations are accumulated during vegetative growth of *S. commune* mycelia in nature.

4.2 Experimental layout

We collected terminal fruit bodies of *S. commune* from trunks occupied by visually continuous series of fruit bodies. Thus, we hoped that these fruit bodies were either produced by a single dikaryon, or were produced by a single monokaryon that crossed with two different monokaryons. We obtained whole genome sequences of these fruit bodies and *de novo* assembled one of the samples from each trunk. Given the high genetic distance between homologous chromosomes, they did not assemble into one contig but rather into two separate

contigs. Thus, we obtained *de novo* assemblies with length almost twice as large as the length of a haploid *S. commune* genome. Then we mapped reads from another fruit body from each trunk onto the obtained assemblies, separately for each trunk. If reads were (almost) fully mapped onto the assembly, we concluded that the fruit bodies were produced by a single dikaryon and shared the diploid genome. If approximately half of reads were mapped and covered approximately half of the assembly, we concluded that the fruit bodies shared one of the parens (and thus shared a haploid genome). In both cases, we called substitutions in covered parts of the assemblies, and estimated the rate at which substitutions were accumulated between fruit bodies, and how this rate correlated with distances between fruit bodies.

4.3 Materials & Methods

Collection of fruit bodies. *S. commune* fruit bodies were collected from tree trunks in autumn, in Kostromskaya oblast', Russia. Distances between fruit bodies from the same trunk were measured (see Table A5). Each fruit body was stored in a paper envelope.

Whole-genome sequencing. Fruit bodies were destroyed in liquid nitrogen. DNA was extracted using the CTAB method (Doyle and Doyle 1987). Libraries were prepared using NEBNext® Ultra™ II DNA Library Prep Kit for Illumina with 5 PCR cycles and sequenced on Illumina NextSeq platform with 150 bp pair-end reads.

***De novo* genome assembling and annotation.** Pair-end reads were trimmed using Trimmomatic (Bolger et al. 2014) with options (ILLUMINACLIP:adapters:2:30:10 LEADING:3 TRAILING:3 SLIDINGWINDOW:4:20 MINLEN:36). *De novo* genome assemblies were obtained using SPAdes (Bankevich et al. 2012) (with -k 21,33,55,77

--careful --only-assembler options). Assemblies were filtered from contamination as follows. Assembled contigs were checked for similarity with *S. commune* reference assembly (Ohm et al. 2010) using BLASTN software. Contigs for which hits with evalue < 1e-05 were found were kept. Reads were mapped back at the assemblies using BWA software, and reads that were mapped at the kept contigs were kept. These reads were used during the second *de novo* assembly using SPAdes (with -k 21,33,55,77 --careful --only-assembler options). The resulting assemblies were used in the following analysis. Assembly statistics are shown in Table A6.

Variant calling. Pair-end reads from one fruit body from a trunk, trimmed using Trimmomatic, were mapped onto corresponding assembly from another fruit body from the same trunk (further called reference) using BWA software. Only reads with properly mapped pair and with mapping quality >=42 were kept. Single nucleotide substitutions between two fruit bodies from the same trunk were called as follows. First, all positions with at least one read supporting the non-reference base were listed. At these positions, we called high frequency variants that had the following properties: (i) coverage in the 10-90% range and >=10X; (ii) non-reference variant frequency >70%; (iii) non-reference variant supported by at least one forward and one reverse read; (iii) coverage in reference sample >= 5x; (iv) non-reference variant frequency in reference sample <= 20%; (v) distance from the end of contig not less than 100bp. For a trunk variants were called in both directions, using both fruit bodies as reference samples. Then variants were compared to check if they are actually the same substitutions called twice using the comparison of the 100bp context of the substitutions.

4.4 Results

Sample collection and sequencing. We collected 26 fruit bodies from 13 trunks; distances between fruit bodies from the same trunk were in the 32 - 190 cm range and are shown in Fig. 17 and Table A5. These fruit bodies were sequenced with 23x - 56x mean depth range. For each fruit body *de novo* genome assemblies were obtained; assembly statistics are shown in Table A6.
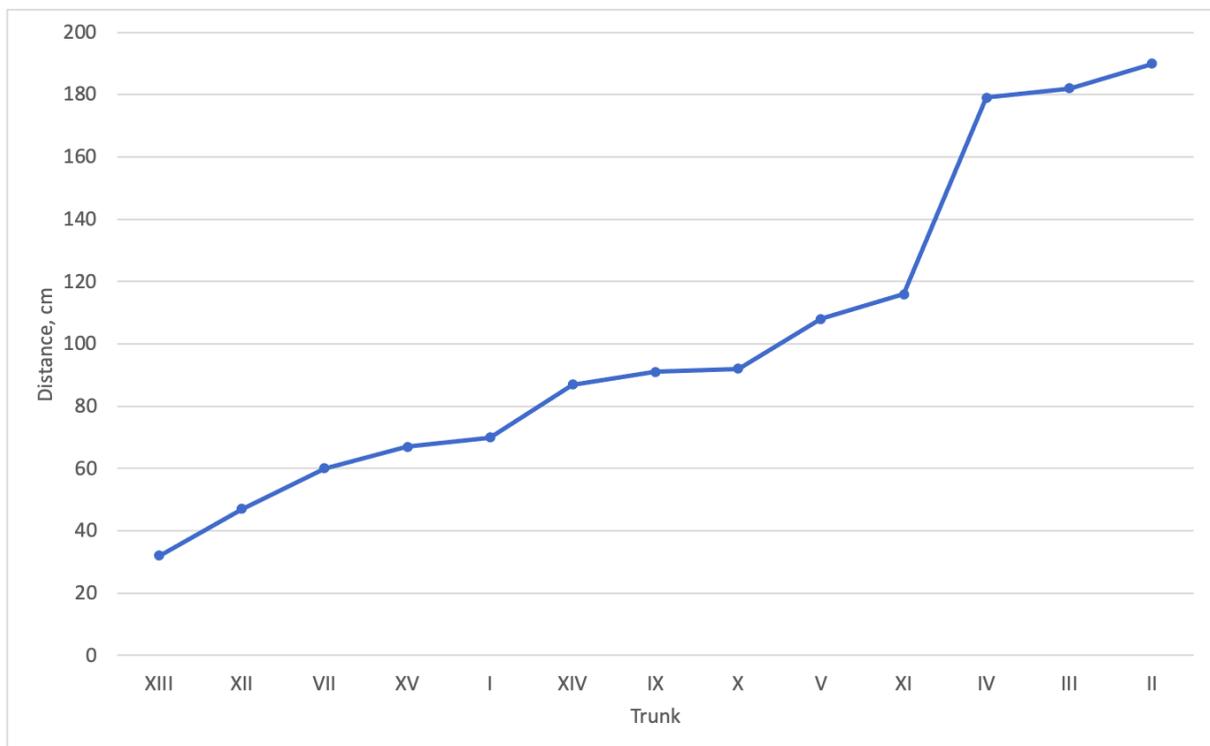
Fig. 17. Distances between pairs of fruit bodies from the same tree trunk.

Comparison of fruit bodies from the same trunk. For each tree trunk, reads from one fruit body were mapped onto the genome assembly obtained for another fruit body (in both directions; thus, two read alignments were obtained for each trunk), and substitutions

between fruit bodies were called. Out of 13 pairs of fruit bodies, two pairs (from trunks I and XIII) appeared to be produced by a single dikaryon. One other pair (trunk XII) shared one haplotype but had different second haplotypes.

We called substitutions between fruit bodies from trunks I (samples Shiz1 and Shiz2) and XIII (samples Shiz21 and Shiz22) that reached high frequency in the mycelium (>=70%) and detected a total of 13 substitutions for trunk I (Table 4.2) and 5 substitutions for trunk XIII (Table 4.3). This gave us the estimations of the accumulation rate of high frequency substitutions at 2.6×10^{-7} substitutions/nucleotide/m for trunk I and 2.3×10^{-7} substitutions/nucleotide/m for trunk XIII, which did not differ significantly from the results obtained in Chapter 3 (Mann-Whitney test, P-value=0.1). The mean substitution accumulation rate was estimated at 2.5×10^{-7} substitutions/nucleotide/m.

Table 4.1. List of substitutions that reach high (>=70% frequency) for trunk I.

Reference Sample	Contig	Pos	Base Shiz1	Base Shiz2	DP Shiz1	DP Shiz2	Non-reference base frequency Shiz1	Non-reference base frequency Shiz2
	NODE_3322	190	A	G	8	17	0.125	0.706
	NODE_30800	174	T	C	18	16	0.056	0.75
Shiz1	NODE_27573	121	T	C	27	14	0.111	0.714
	NODE_96	7367	A	G	17	12	0.176	0.917
	NODE_2088	3053	C	T	10	12	0.200	0.833

Reference Sample	Contig	Pos	Base Shiz2	Shiz1	DP Shiz2	Shiz1	Non-reference base frequency Shiz2	Shiz1
	NODE_1410	1413	T	G	7	30	0.143	0.833
	NODE_359	361	C	G	22	18	0.045	0.722
	NODE_16621	756	C	T	6	15	0.000	0.733
	NODE_32488	262	C	T	8	15	0.125	0.733
Shiz2	NODE_455	1357	C	T	12	14	0.167	0.714
	NODE_4657	380	T	A	9	11	0.111	0.818
	NODE_2002	5573	C	T	15	10	0.133	0.7
	NODE_40750	171	G	T	6	10	0.000	0.8

Table 4.2. List of substitutions that reach high (>=70% frequency) for trunk XIII.

Reference Sample	Contig	Pos	Base Shiz21	Shiz22	DP Shiz21	Shiz22	Non-reference base frequency Shiz21	Shiz22
	NODE_3034	3580	T	G	28	13	0.179	0.769
Shiz21	NODE_1324	3335	C	A	13	10	0.000	0.7
	NODE_2685	388	A	G	7	10	0.143	0.7

Reference Sample	Contig	Pos	Base		DP		Non-reference base frequency	
			Shiz22	Shiz21	Shiz22	Shiz21	Shiz22	Shiz21
Shiz21	NODE_1626	1891	G	T	6	25	0.167	0.76
	NODE_39501	113	C	T	17	22	0.176	0.727

4.5 Discussion

In Chapter 3 we show the potential of the generational mutation rate for *S. commune* to be high given the following conditions: the fungus occupies large territories in nature (of the order up to 1 m) via vegetative state, either monokaryon or dikaryon; and if in natural monokarions and dikaryons somatic substitutions are accumulated in the same manner as in monokaryons *in vitro*.

Here, we show that *S. commune* may indeed occupy territories via vegetative state, as fruit bodies at the ends of two trunks share diploid genomes, and for one trunk the fruit bodies share haploid genome.

We estimate the high frequency substitution accumulation rate between fruit bodies at 2.5×10^{-7} substitutions/nucleotide/m, which translates to the *in vivo* generational mutation rate of 2.6×10^{-7} substitutions/nucleotide/generation for trunk I and 2.3×10^{-7} substitutions/nucleotide/generation for trunk XIII. This is indeed at the top of the known mutation rate, being the second highest rate above the 0.73×10^{-7} substitutions/nucleotide/generation for *Marasmius oreades* but below the 34×10^{-7} substitutions/nucleotide/generation for *Neurospora crassa* (Wang et al. 2020; Lujan and

Kunkel 2021). Moreover, these observations were made for fruit bodies at a 70 and 32 cm distance. However, if the occupied distance is larger, the generational mutation rate may reach even higher values

Chapter 5. Accumulation of generational *de novo* mutations in *Schizophyllum commune in vitro*

5.1 Introduction

The generational mutation rate for *S. commune in vitro* has been previously studied and estimated at $2 \cdot 10^{-8}$ mutations/nucleotide/generation (Baranova et al. 2015). However, the generational mutation process can be further studied. In particular there is a question how the mutation rate depends on the level of genetic heterozygosity. This question arises given two facts. First, it was previously shown that homologous recombination is associated with the elevated mutation rate (Halldorsson et al. 2019). Second, the crossing-over events in *S. commune* tend to occur within more conserved regions (Seplyarskiy et al. 2014). Thus, one can hypothesize that the mutation rate in *S. commune* may be elevated given less heterozygous genome segments. To test this hypothesis, we performed back crossings of offsprings with their parents, and tried to compare the mutation rate in completely homozygous and highly heterozygous genome regions.

5.2 Experimental layout

We crossed two non-relative haploid mononuclear individuals and obtained haploid mononuclear F1 offsprings. Some of these offsprings were back crossed with parents, and one crossing that produced sufficient amount of offspring was selected for further analysis (this crossing will be further referred to as a BC crossing). Moreover, the F1 offspring involved in this crossing also crossed with one of it's siblings (the crossing that will be further referred to as F2 crossing). We obtained whole genome sequences of all individuals involved in F1, BC and F2 crossings and their offsprings, and assessed parental genotypes along the offspring genomes. Then, we called *de novo* single nucleotide mutations and

determined whether they happened in homo- or hetero parts of the genome during the crossing. Thus, we were able to try to estimate if the level of heterozygosity has a significant impact on the *de novo* mutation rate in *S. commune*.

5.2 Materials & Methods

Obtaining original haploid mononuclear cultures. *S. commune* fruit bodies were collected from tree trunks in autumn. Each fruit body was then attached to the Petri dish lid over the solid agar medium (see below), and the Petri dish was placed in a diagonal manner. The Petri dish was exposed to the light at room temperature, and under such conditions some fruit bodies released haploid spores that turned out on the surface of the solid medium. At the periphery of the area where spores were located, it was possible to visually locate individual spores. Such spores were cut out from the solid medium and transferred to new Petri dishes where they originated mononuclear haploid cultures.

Cultivation and preservation. Cultures were cultivated on solid medium (beer wort Maltax10 – 25.6 g, water – 1 l, agar – 40 g) in the light at room temperature, and stored at 4ºC.

Crossing of two individuals. Two haploid mononuclear cultures were put on the same Petri dish on solid agar medium (see above), and exposed to the light at room temperature. If mating types of the cultures were compatible, fruit bodies were produced when two mycelia met each other at the center of the Petri dish. These fruit bodies were collected and exposed to the same procedure as described in the "Obtaining haploid mononuclear cultures" segment.

Whole-genome sequencing. Before DNA extraction, samples of mycelium were first grown in liquid medium (beer wort Maltax10 – 8 g, water – 1 l) on shaker to reach sufficient mass,

and then were destroyed in liquid nitrogen. DNA was extracted using the CTAB method (Doyle and Doyle 1987). Libraries were prepared using NEBNext® Ultra™ II DNA Library Prep Kit for Illumina with 5 PCR cycles and sequenced on Illumina HiSeq2000 platform with 76 bp pair-end reads.

***De novo* genome assembling.** Pair-end reads were trimmed using Trimmomatic (Bolger et al. 2014) with options (ILLUMINACLIP:adapters:2:30:10 LEADING:3 TRAILING:3 SLIDINGWINDOW:4:20 MINLEN:36). *De novo* contig genome assemblies were obtained using SPAdes (Bankevich et al. 2012) (with -k 21,33,55,77 --careful --only-assembler options). These contig assemblies were aligned to the reference genome (Ohm et al. 2010) using Lastz (Harris 2007), overlapped regions were removed using single_cov2 program from Multiz package (Blanchette et al. 2004), and the alignment was used to construct the scaffold assembly for the sample (gaps were replaced with N-s). Assembly statistics are presented in Table A7.

Contig genotyping. Parental reads were mapped to the offspring contig assemblies using Bowtie2 software. Only reads with paired read properly mapped, mapping quality >=42 and not more than one miss-match were kept. Given the large genetic distance between different genotypes, parental reads were only mapped to the segments of the offspring genome that had the same genotype. Then the coverages when mapping parental reads was calculated in 1000bp windows, and the genotype was assigned using the following rule: if both coverages were less than 10x, the 'unknown' genotype was assigned; if both coverages were greater than 10x, the 'both' genotype was assigned, meaning that this locus the parental genotypes have small genetic distance from each other and could not be distinguish; if parent 1 coverage was greater than 10x and parent 2 coverage was less than 10x, the 'parent 1' genotype was assigned; if parent 2 coverage was greater than 10x and parent 1 coverage was less than 10x, the 'parent 2' genotype was assigned.

SNP calling. For each offspring, parental reads were mapped to the offspring contig assembly as well as offspring reads themselves. SNPs were called for each of the mappings using samtools mpileup software (the reference and non-reference nucleotides were swapped). The following filters were further applied to the called variants: i) read depth in 10x-500x range; ii) contig length >=1000bp; iii) quality >=30. SNP in a parental mapping was called as *de novo* if i) read depth in the offspring mapping itself was >=10x; ii) not a single read in both parental mappings supported non-reference nucleotide in case of 'both' genotype, or not a single read supported non-reference nucleotide in parent 1(2) in case of 'parent 1(2)' genotype.

5.3 Results

Crossings, sequencing and *de novo* assembling. We crossed two individuals from USA (sh01) and Russia (sh04) and obtained 24 F1 offsprings. One of these offsprings - f1_26 - was back crossed with sh01 (BC crossing) and with its sibling f1_7 (F2 crossing). We obtained and sequenced 24 F1 offsprings, 8 BC offsprings and 16 F2 offsprings with 65X (34X - 91X), 30X (16X - 43X) and 25X (10X - 44X) mean coverage correspondingly. The statistics of the *de novo* assemblies are shown in Table A7.

***De novo* mutations.** We called *de novo* single nucleotide mutations and attributed them as happened in homo- or heterozygous context. We detected a total of 89 mutations (Table A8). However, it appeared that some of the mutations were presented in clusters of offsprings, most likely resulting from single premiotic mutational events that were spread to multiple offsprings (Thompson et al. 1998; Yang et al. 2001). Thus, we obtained a list of 39 unique mutational events.

F1 offsprings were obtained from four different fruit bodies (all collected from a single crossing). Two of these fruit bodies (fb1 and fb2) were collected soon after the initial crossing, however, another two fruit bodies (fb3 and fb4) were collected after a series of passages of the dykarion. Interestingly, the cluster mutations were only observed in fb3 and fb4, while in fb1 and fb 2 we only observed singleton mutations. Unfortunately, the data about fruit bodies for BC and F2 was not tracked during the experiment.

The number of detected mutations is shown in Table 5.1.

Table 5.1. Number of detected *de novo* single nucleotide mutations in F1, F2 and BC crossings.

Crossing		# Mutations		# Mutational events		Callable length, bp
	Fruit bodies fb1-2	9		9		292 515 018
			54		26	
F1	**Fruit bodies fb3-4**	45		17		291 575 363
	Homozygous parts of the genome	8		4		112 706 475
			8		4	
BC	**Heterozygous parts of the genome**	0		0		99 705 462
	Homozygous parts of the genome	18		6		201 419 630
			27		9	
F2	**Heterozygous parts of the genome**	9		3		222 074 971

Unfortunately, due to the clustered mutations, it became impossible to estimate the *de novo* spontaneous mutation rates in homo- and heterozygous contexts, as it was impossible to estimate the callable length of the genomes. We were only able to estimate the spontaneous mutation rate in F1, fruit bodies fb1 and fb2, where there were no clustered mutations. The

mutation rate was estimated at 2.86×10^{-8} substitutions/nucleotide/generation (95% CI: 1.18×10^{-8} - 4.54×10^{-8}), which is in line with the estimates obtained in (Baranova et al. 2015), where the same experiment was also conducted *in vitro*.

However, we were able to estimate the generational mutation accumulation rates. This rate was estimated at 8.29×10^{-8} substitutions/nucleotide/generation (95% CI: 5.72×10^{-8} - 1.08×10^{-7}) in homozygous regions, and at 2.70×10^{-8} substitutions/nucleotide/generation (95% CI: 7.63×10^{-9} - 4.64×10^{-8}) in heterozygous regions, with the difference being significant (Mann-Whitney one-side paired test, P-value = 0.001). The overall mutation accumulation rate for F1 offsprings was estimated at 7.16×10^{-8} substitutions/nucleotide/generation (95% CI: 4.08×10^{-8} - 1.02×10^{-7}), with difference between homozygous BC regions and F1 being insignificant (Mann-Whitney one-side test, P-value = 0.18), and difference between heterozygous BC regions and F1 being significant (Mann-Whitney one-side test, P-value = 0.003).

5.4 Discussion

Unfortunately, we were not able to estimate and compare the spontaneous mutation rates within completely homozygous and highly heterozygous regions. This was due to mutations shared between individuals - cluster mutations most likely resulting from a single premiotic mutational event and spread to multiple offsprings (Thompson et al. 1998; Yang et al. 2001). These clusters were found among offsprings obtained from fruit bodies that were obtained not immediately after the mating of the monokaryons, but after some time of cultivation of the dikaryons. It is likely that somatic mutations appeared during the cultivation of the mycelium inside growing hyphae that gave birth to fruit bodies. Thus, such mutations

occupied fruit bodies (as a fruit body originates from a single hyphae) and were transferred to all offsprings of such fruit bodies.

The mutation rate estimated from offsprings that were collected from fruit bodies that were not cultivated for some prolonged time was inline with the previous *in vitro* estimation (Baranova et al. 2015). However, the estimations of the mutation accumulation rates for fruit bodies that originated from mycelia that were cultivated for some time were higher, and at first sight ambiguous. In particular, the rate for F1 was close to that in homozygous regions in BC, and differed from tha rate in heterozygous regions. However, this is probably due to the longer time of cultivation of some dykarions in F1 crossing. In the meantime, the difference between mutation accumulation rates in homozygous and heterozygous regions in BC was significant, and is probably reliable as the significance test is paired and both homo- and heterozygous regions undergone the same time of cultivation.

Chapter 6. The dependence of homologous recombination rate on the level on heterozygosity in *Schizophyllum commune in vitro*

6.1 Introduction

Homologous recombination is one of the key processes during sexual reproduction. As long as it involves homologous chromosomes, it usually happens when the genetic distance between to exchanging parts of the genome is very small, usually less than 2% (Leffler et al. 2012). However, the question is how homologous recombination might operate in hypervariable species where the heterozygosity may easily exceed 10-15%.

It was previously shown that the level of heterozygosity may indeed affect the recombination rate. In particular, in (Waldman and Liskay 1988) it was shown that uninterrupted homologous tracts resulted in an elevated recombination rate. In (Datta et al. 1997) it was also shown that with the increase of the genetic divergence the recombination becomes less efficient. Approximately log-linear relationship between the recombination rate and the level of heterozygosity was shown.

However, these studies have not operated on the level of sequence divergence available in *S. commune*. In (Seplyarskiy et al. 2014), it was shown that in *S. commune* the CO events tend to occur within more conserved regions, in particular exons. Here we aim to further explore the homologous recombination in *S. commune*. In particular, we aim to directly compare the recombination rate when the recombining region is completely homozygous and highly heterozygous.

6.2 Experimental layout

We developed an experimental system in which we can directly measure the recombination rate in a completely homozygous region of a chromosome, and compare it to the normal recombination rate in regions with high heterozygosity in *S. commune*.

First, we obtained F1 offsprings from two non-relative mononuclear haploid *S. commune* individuals. Then, we obtained whole genome sequences of parents and all the F1 offsprings, and determined points of crossing over events in F1 individuals with high precision, taking advantage of the high genetic distance between parents. Then, we aimed to determine an F1 individual that carried a chromosome of interest in which two CO events occured, thus giving the following structure of the chromosome: central segment having genotype of one parent, shoulders having genotype of another parent. Then, this F1 individual of interest was back crossed with both parents. Obviously, this individual wouldn't cross with one of the parents (say, parent 2) with which it shares the mating type. To overcome this obstruction, we replaced this parent 2 with another F1 offspring with which the F1 individual of interest would cross and which has the same genotype of the chromosome of interest as parent 2. To simplify the narrative, we will sometimes refer to the back crosses with both parents, meaning that one of the parents was replaced by the appropriate F1 offspring.

Back crosses with both parents gave us the following genetic state of the chromosome of interest during the crossings: the central segment was in completely homozygous state when crossing with one of the parents, and in normal highly heterozygous state when crossing with another parent, whereas the shoulders had the opposite to the central cegment state.

Thus, we obtained backcross offsprings from both parents, and determined the number of CO events in the central segment of the chromosome of interest using Sanger sequencing and
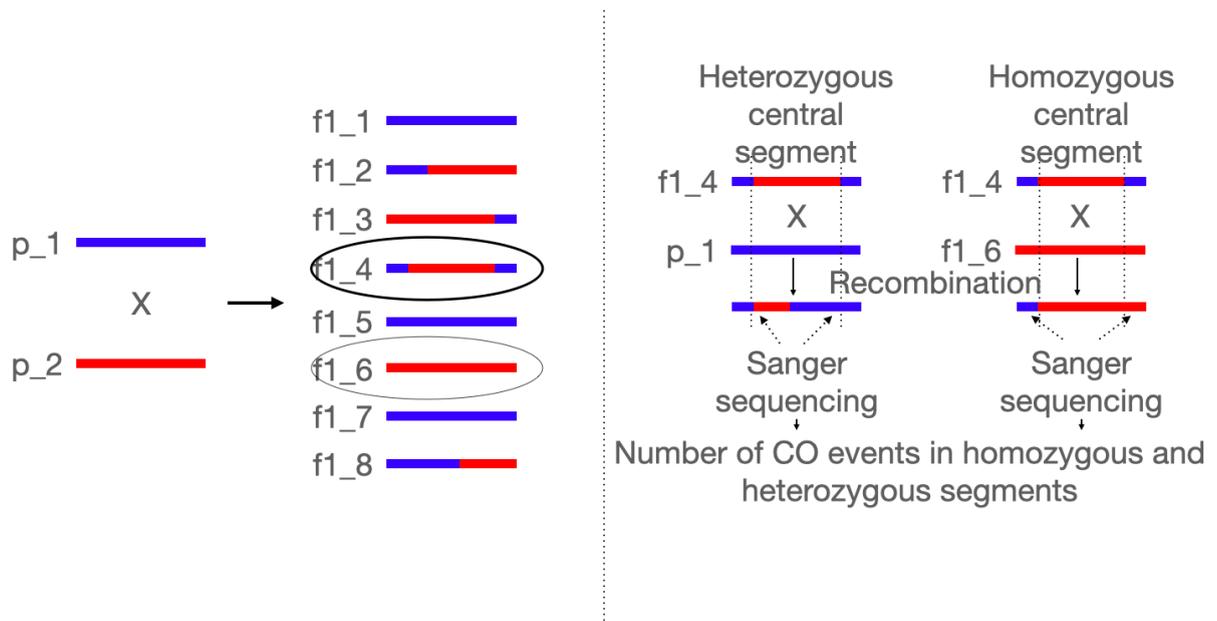
Fig. 19. Experimental layout. Colors represent two parental genotypes of the chromosome of interest.

genotyping of the small loci near the ends of the central segment. Then, we compared the numbers in case of homo- and heterozygous state of the segment.

The overall experimental layout is shown in Fig. 19

6.3 Materials & Methods

Obtaining original haploid mononuclear cultures. *S. commune* fruit bodies were collected from tree trunks in autumn. Each fruit body was then attached to the Petri dish lid over the solid agar medium (see below), and the Petri dish was placed in a diagonal manner. The Petri dish was exposed to the light at room temperature, and under such conditions some fruit bodies released haploid spores that turned out on the surface of the solid medium. At the

periphery of the area where spores were located, it was possible to visually locate individual spores. Such spores were cut out from the solid medium and transferred to new Petri dishes where they originated mononuclear haploid cultures.

Cultivation and preservation. Cultures were cultivated on solid medium (beer wort Maltax10 – 25.6 g, water – 1 l, agar – 40 g) in the light at room temperature, and stored at 4ºC.

Crossing of two individuals. Two haploid mononuclear cultures were put on the same Petri dish on solid agar medium (see above), and exposed to the light at room temperature. If mating types of the cultures were compatible, fruit bodies were produced when two mycelia met each other at the center of the Petri dish. These fruit bodies were collected and exposed to the same procedure as described in the "Obtaining haploid mononuclear cultures" segment.

Whole-genome sequencing. Before DNA extraction, samples of mycelium were first grown in liquid medium (beer wort Maltax10 – 8 g, water – 1 l) on shaker to reach sufficient mass, and then were destroyed in liquid nitrogen. DNA was extracted using the CTAB method (Doyle and Doyle 1987). Libraries were prepared using NEBNext® Ultra™ II DNA Library Prep Kit for Illumina with 5 PCR cycles and sequenced on Illumina HiSeq2000 platform with 76 bp pair-end reads.

***De novo* genome assembling.** Pair-end reads were trimmed using Trimmomatic (Bolger et al. 2014) with options (ILLUMINACLIP:adapters:2:30:10 LEADING:3 TRAILING:3 SLIDINGWINDOW:4:20 MINLEN:36). *De novo* genome assemblies were obtained using SPAdes (Bankevich et al. 2012) (with -k 21,33,55,77 --careful --only-assembler options). These contig assemblies were aligned to the reference genome(Ohm et al. 2010) using Lastz software (Harris 2007), overlapped regions were removed using single_cov2 program from Multiz package (Blanchette et al. 2004), and the resulting alignment was used to construct the

scaffold assembly for the sample (gaps were replaced with N-s). Assembly statistics for parents and F1 offsprings are presented in Table A7.

Scaffold genotyping and determination of the CO events in F1. Parental reads were mapped to the F1 assemblies using Bowtie2 software. Only reads with paired read properly mapped, mapping quality >=42 and not more than one miss-match were kept. Given the large genetic distance between parental genotypes, parental reads were only mapped to the segments of the F1 genome that had the same genotype. Then the coverages when mapping parental reads was calculated in 1000bp windows, and the genotype was assigned using the following rule: if both coverages were less than 10x, the 'unknown' genotype was assigned; if both coverages were greater than 10x, the 'both' genotype was assigned, meaning that it this locus the parental genotypes have small genetic distance from each other and could not be distinguish; if parent 1 coverage was greater than 10x and parent 2 coverage was less than 10x, the 'parent 1' genotype was assigned; if parent 2 coverage was greater than 10x and parent 1 coverage was less than 10x, the 'parent 2' genotype was assigned. Scaffolds were depicted using color code for parental genotypes, and visually examined to determine potential CO events. The difference of parental coverages in 1bp windows was calculated in the 60kb areas that included each potential CO event. These differences were plotted against scaffold coordinates, and visually examined. The scaffold coordinates of the intersections with 0 were considered as the coordinates of CO events.

Sanger sequencing and CO events determination in the chromosome of interest. We picked random ~1kb loci in the ~+-5kb region around CO spots - one locus on one side of the CO spot, one primer on the other side of the CO spot. PCR primers for these loci were generated using primer3 software. For each locus, two pairs of primers were generated - based on two parental genotypes at those loci. Thus, eight primers were generated for one chromosome of interest (Fig. 20).
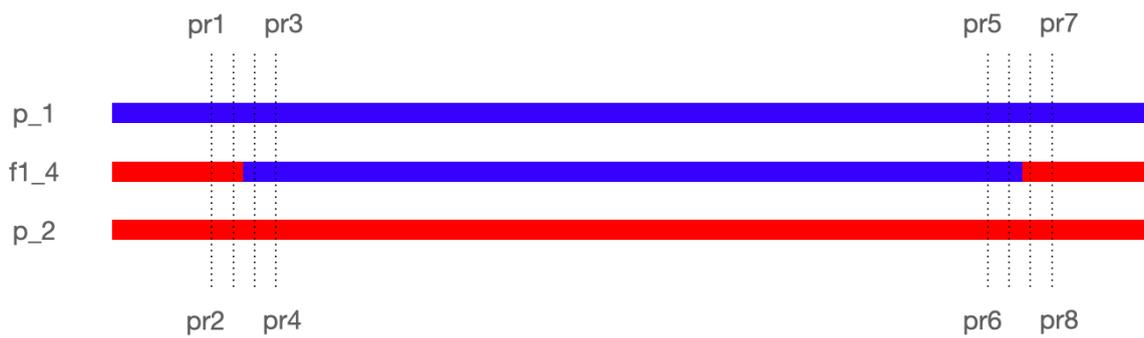
77

Fig. **20.** Scheme of the primers used for determination of CO events in the central segment of the chromosome. Coordinates in the chromosome are not to scale

Then, we used primers pr1, pr2, pr7 and pr8 for determination of the CO events during f1_4 x p_1 back cross, when the central segment is homozygous; and we used primers pr3, pr4, pr5 and pr6 for determination of the CO events during f1_4 x p_2 back cross, when the central segment is heterozygous.

DNA from BC offsprings was extracted using DNeasy Plant pro Kit ("Qiagen") according to the DNEasy Plant pro handbook instruction. To determine whether a CO event happened in a given offspring, we amplified appropriate loci (see above, and obtained their nucleotide sequences using Sanger sequencing using the following protocol. Polymerase chain reaction (PCR) amplifications were carried out in a 20-μL reaction volume, which included 4 μL of 5x Screen Mix (Evrogen Joint Stock Co., Moscow, Russia), 0.5 μL of each primer (10 μL stock), 1 μL of genomic DNA and 14 μL of sterile water. The condition of PCR: 940 – 5 min ,then cicles 940 – 15 s, annealing temperature (57-590) – 20 s and elongation 720 – 40s. The number of cycles varied from 35 to 40.DNA The Promega PCR Purification Kit protocol (Promega) was used to purify the amplification products. Amplification of products proceeded in both directions. Each sequencing reaction mixture contained 1 μL of BigDye

(Applied Biosystems, PerkinElmer Corporation, Foster City, CA), 1 μL of 1 μM primer, and 1 μL of DNA template; sequencing reactions were run for 40 cycles of 96 °C (15 s), 50 °C (30 s), and 60 °C (4 min). Sequences were subjected to ethanol precipitation to remove unincorporated primers and dyes. The products were resuspended in 12 μL of formamide and subjected to electrophoresis in an ABI Prism 3500 Genetic. Cromatogramma and sequences were visually inspected using CodonCode Aligner software.

We then obtained the nucleotide simmilarity of these sequences to both of the F1 parental genomes (p_1 and p_2) using BLASTN software. We compared these similarities, and assigned the genotype that showed higher similarity to the given locus (if the difference was less than 5%, 'unknown' genotype was assigned). If both primers that were generated for the locus showed the same genotype, this genotype was decisively attributed to the locus. If two primers showed different genotypes, 'unknown' genotype was attributed. If in an offspring the genotypes of two loci on the both ends of the central segment of the chromosome of interest were known and were different, we concluded that a CO event happened in the given offspring.

6.4 Results

Obtaining and sequencing of F1 offsprings. We crossed two individuals from the USA (sh01) and Russia (sh04) and obtained 24 F1 offsprings. We obtained whole genome sequences of these individuals with 34x - 91x mean coverage range, and assembled them into scaffold assemblies. The assembly statistics are presented in Table A8. Chromosomes (scaffolds) of F1 individuals were genotyped, 190 CO events were determined, and individual f1_13 (with scaffold 3 of interest) was selected for further analysis. Z14 offspring from

crossing of f1_26 with sh01 (see chapter 5) was selected as a representative of the sh01 parent (Fig. 21). Primer names and their attribution to loci (ends of the central segment) are shown in Fig. 22; primer coordinates are shown in Table 6.1.
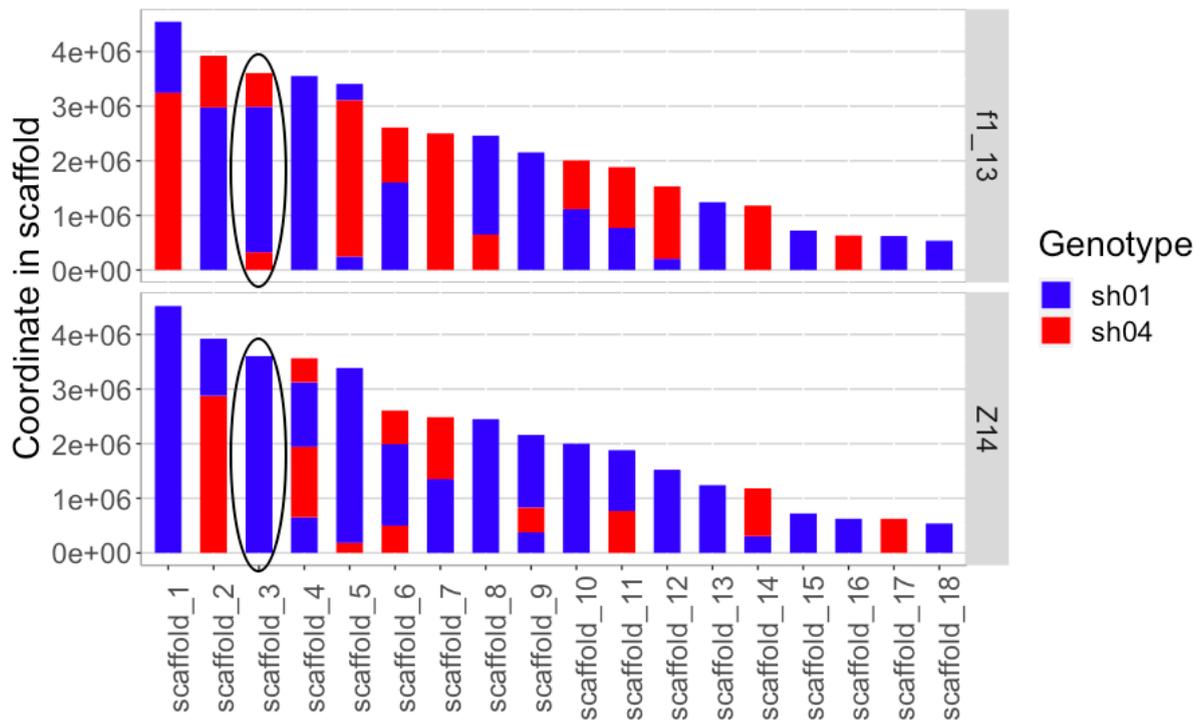
Fig. 21. Scaffold genotypes of f1_13 and Z14 individuals. Scaffold of interest is circled.
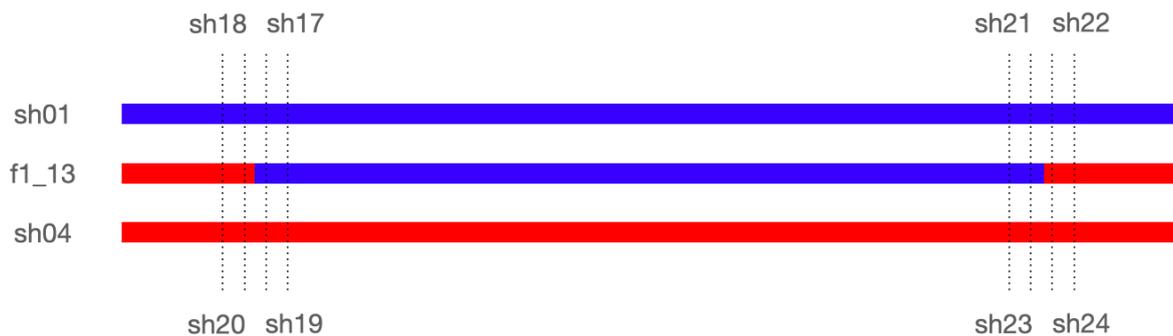
Fig. 22. Primer names for the loci of interest in scaffold 3. Coordinates in the scaffold are not to scale.

Table. 6.1 Primer coordinates for the loci of interest in scaffold 3, fl_13 assembly.

				Primer start coordinate			
ID	**Scaffold**	**CO coordinates range, bp**		**sh18**	**sh20**	**sh22**	**sh24**
fl_13	scaffold_3	453622	454154	447359	448539	458463	472658
				sh17	**sh19**	**sh21**	**sh23**
fl_13	scaffold_3	3431892	3431936	3559244	3437480	3427523	3427375

Table 6.2. Description of back crosses.

Crossing	ID	Scaffold	What does it actually mean	Scaffold genotype	Center segment is	Number of offsprings
fl_13 x Z14	WZ5	scaffold_3	fl_13 x sh01	sh04-sh01-sh04	Homozygous	88
fl_13 x sh04	WK21	scaffold_3	fl_13 x sh04	sh04-sh01-sh04	Heterozygous	80

Back crosses. We obtained 88 offsprings from fl_13 x sh01 back cross (ID WZ5), and 80 offsprings from fl_13 x sh04 back cross (ID WK21) (see Table).

Recombination rates in homo- and heterozygous genome segments. We obtained Sanger sequences of the loci near the ends of the central segment in scaffold 3 for back cross offsprings. For some offsprings, we were not able to obtain sequences with sufficient quality. For the minority of offsprings (17%) two pairs of primers for the same end of the segment produced contradicting results. In one third of offsprings (37%), we were able to determine the genotype of both ends of the central segment (Table 6.3).

Table 6.3. Number of analyzed samples and loci.

Primer	Total samples	# sequences with sufficient quality	At least one pair of the primers produced sequences with sufficient quality	Two primers produced different results	Both ends of a central segment were genotyped
Sh18		32			
Sh20	88	21	45	2	
Sh22		32			26
Sh24		25	52	5	
Sh17		40			
Sh19	80	44	64	9	
Sh21		55			38
Sh23		46	82	13	

We determined 17 CO in 26 back cross offsprings events when the central segment was in completely homozygous state, and 14 CO events in 38 back cross offsprings events when the central segment was in heterozygous state. These numbers alongside with the target genome length and resulting recombination rates are shown in Table 6.4 (data for F1 as well as for F1 from (Seplyarskiy et al. 2014) is also shown).

Table 6.4. Recombination rates.

Crossing	Genome state	# Samples	Target length in one sample, bp	# Recombi-nations	Recombi-nation rate, cM/Mb
WZ5	Homozygous	26	2 978 314	17	2.24
WK21	Heterozygous	38	2 978 314	14	1.24

The probability of recombination in the central segment of scaffold 3 in case of homozygosity of the segment was almost 2 times higher than that in case of heterozygosity, with the difference being significant (Fisher's exact test P-value = 0.04).

6.5 Discussion

Back crossing allowed us to compare the recombination rates in case of complete homozygosity and high heterozygosity of the genome region. Usually, it is impossible to determine the recombination events inside homozygous regions. However, we designed an experiment that allowed us to do so. As expected from previous works (Waldman and Liskay 1988; Datta et al. 1997; Seplyarskiy et al. 2014), the recombination rate in the homozygous region was higher than in the heterozygous region, and the difference was quite large (almost two times) and significant.

Fungi are characterized by high recombination rate compared to other eukaryotes except for the SAR group (Stapley et al. 2017). Interestingly, both homo- and heterozygous recombination rates in *S. commune* appeared to be low for fungi, being one order of magnitude less than the mean recombination rate for this group.

Chapter 7. Conclusions

In this work we analyzed the evolutionary factors that affect and are affected by the extreme level of genetic polymorphism in *S. commune.* This includes somatic and generational mutational processes *in vitro* and *in vivo*, natural selection within mycelium and homologous recombination. The conclusions of this work are:

1. Somatic mutation rate in monokaryotic mycelia *in vitro* is quite moderate; however, it does not change with the length of the mycelium, and thus has the potential to be translated to a high generational mutation rate. The somatic accumulation rate is affected by the effectiveness of natural selection in the growing mycelium.

2. *S. commune* in nature occupies territories via vegetative state, both monokaryotic and dikaryotic, and indeed accumulates somatic mutations under the same mode as *in vitro*. The detected and potential generational mutation rates are indeed at the top of the known mutation range for living organisms. This might contribute to the extreme genetic diversity of this species.

3. Homologous recombination rate in *S. commune* is suppressed by the high level of the heterozygosity and can increase given less heterozygous genetic regions.

Our results shed light on the potential causes of the extreme genetic diversity of *S. commune*. We hypothesize that extensive vegetative growth and lack of mechanisms of preserving the genetic material during this growth (which is not always the case for fungi), resulting in a high generational mutation rate, contribute to the high polymorphism. However, the estimated mutation rate is still not the highest known and is not extremely higher than mutation rates estimated for some other species including fungi. Thus, we further hypothesize that *S. commune* might have a high effective population size.

Bibliography

Anderson JB, Bruhn JN, Kasimer D, Wang H, Rodrigue N, Smith ML. 2018. Clonal evolution and genome stability in a 2500-year-old fungal individual. *Proc. R. Soc. B Biol. Sci.* 285:20182233.

Anderson JB, Catona S. 2014. Genomewide mutation dynamic within a long-lived individual of Armillaria gallica. *Mycologia* 106:642–648.

Anon. MycoBank. Available from: http://www.mycobank.org/MycoTaxo.aspx?Link=T&Rec=208403

Bankevich A, Nurk S, Antipov D, Gurevich AA, Dvorkin M, Kulikov AS, Lesin VM, Nikolenko SI, Pham S, Prjibelski AD, et al. 2012. SPAdes: a new genome assembly algorithm and its applications to single-cell sequencing. *J. Comput. Biol. J. Comput. Mol. Cell Biol.* 19:455–477.

Baranova MA, Logacheva MD, Penin AA, Seplyarskiy VB, Safonova YY, Naumenko SA, Klepikova AV, Gerasimov ES, Bazykin GA, James TY, et al. 2015. Extraordinary Genetic Diversity in a Wood Decay Mushroom. *Mol. Biol. Evol.* 32:2775–2783.

Barton NH. 1995. A general model for the evolution of recombination. *Genet. Res.* 65:123–145.

Blanchette M, Kent WJ, Riemer C, Elnitski L, Smit AFA, Roskin KM, Baertsch R, Rosenbloom K, Clawson H, Green ED, et al. 2004. Aligning Multiple Genomic Sequences With the Threaded Blockset Aligner. *Genome Res.* 14:708–715.

Blomberg C. 1987. Free energy and time economy for the mutual selection of monomers in biosynthesis, primarily protein synthesis. *J. Theor. Biol.* 128:87–107.

Bolger AM, Lohse M, Usadel B. 2014. Trimmomatic: a flexible trimmer for Illumina sequence data. *Bioinformatics* 30:2114–2120.

Broad Institute. Picard Tools - By Broad Institute. Available from: http://broadinstitute.github.io/picard/

Charlesworth B. 2009. Effective population size and patterns of molecular evolution and variation. *Nat. Rev. Genet.* 10:195–205.

Charlesworth B, Barton NH. 1996. Recombination load associated with selection for increased recombination. *Genet. Res.* 67:27–41.

Chen J, Glémin S, Lascoux M. Genetic Diversity and the Efficacy of Purifying Selection across Plant and Animal Species. *Mol. Biol. Evol.* [Internet]. Available from: https://academic.oup.com/mbe/article-abstract/doi/10.1093/molbev/msx088/3049540/Genetic-Diversity-and-the-Efficacy-of-Purifying

Choi K, Henderson IR. 2015. Meiotic recombination hotspots - a comparative view. *Plant J. Cell Mol. Biol.* 83:52–61.

Chowdhary A, Kathuria S, Agarwal K, Meis JF. 2014. Recognizing filamentous basidiomycetes as agents of human disease: A review. *Med. Mycol.* 52:782–797.

Chowdhary A, Randhawa HS, Gaur SN, Agarwal K, Kathuria S, Roy P, Klaassen CH, Meis JF. 2013. Schizophyllum commune as an emerging fungal pathogen: a review and report of two cases. *Mycoses* 56:1–10.

Clark TA, Anderson JB. 2004. Dikaryons of the basidiomycete fungus Schizophyllum commune: evolution in long-term culture. *Genetics* 167:1663–1675.

Comeron JM, Ratnappan R, Bailin S. 2012. The many landscapes of recombination in Drosophila melanogaster. *PLoS Genet.* 8:e1002905.

Cooke WmB. 1961. The Genus Schizophyllum. *Mycologia* 53:575–599.

Croll D, Lendenmann MH, Stewart E, McDonald BA. 2015. The Impact of Recombination Hotspots on Genome Evolution of a Fungal Plant Pathogen. *Genetics* 201:1213–1228.

Crow JF. 1993. How much do we know about spontaneous human mutation rates? *Environ.*

Mol. Mutagen. 21:122–129.

Datta A, Hendrix M, Lipsitch M, Jinks-Robertson S. 1997. Dual roles for DNA sequence identity and the mismatch repair system in the regulation of mitotic crossing-over in yeast. *Proc. Natl. Acad. Sci.* 94:9757–9762.

Deng HW, Lynch M. 1996. Estimation of Deleterious-Mutation Parameters in Natural Populations. *Genetics* 144:349–360.

Dey A, Chan CKW, Thomas CG, Cutter AD. 2013. Molecular hyperdiversity defines populations of the nematode Caenorhabditis brenneri. *Proc. Natl. Acad. Sci. U. S. A.* 110:11056–11060.

Doyle JJ, Doyle JL. 1987. A rapid DNA isolation procedure for small quantities of fresh leaf tissue. *Phytochem. Bull.* 19:11–15.

Drake JW, Charlesworth B, Charlesworth D, Crow JF. 1998. Rates of spontaneous mutation. *Genetics* 148:1667–1686.

Essig FM. 1922. The morphology, development, and economic aspects of Schizophyllum commune Fries. Berkeley, Calif. : University of California Press Available from: http://archive.org/details/morphologydevelo00essirich

Felsenstein J. 1974. The Evolutionary Advantage of Recombination. *Genetics* 78:737–756.

Gerton JL, DeRisi J, Shroff R, Lichten M, Brown PO, Petes TD. 2000. Global mapping of meiotic recombination hotspots and coldspots in the yeast Saccharomyces cerevisiae. *Proc. Natl. Acad. Sci. U. S. A*. 97:11383–11390.

Gooday GW. 1995. The dynamics of hyphal growth. *Mycol. Res.* 99:385–394.

Haag-Liautard C, Dorris M, Maside X, Macaskill S, Halligan DL, Houle D, Charlesworth B, Keightley PD. 2007. Direct estimation of per nucleotide and genomic deleterious mutation rates in Drosophila. *Nature* 445:82–85.

Halldorsson BV, Palsson G, Stefansson OA, Jonsson H, Hardarson MT, Eggertsson HP, Gunnarsson B, Oddsson A, Halldorsson GH, Zink F, et al. 2019. Characterizing mutagenic effects of recombination through a sequence-level genetic map. *Science* 363.

Hanafusa Y, Hirano Y, Watabe H, Hosaka K, Ikezawa M, Shibahara T. 2016. First isolation of Schizophyllum commune in a harbor seal (Phoca vitulina). *Med. Mycol.* 54:492–499.

Harris RS. 2007. Improved pairwise alignment of genomic DNA.

Hillers KJ. 2004. Crossover interference. *Curr. Biol. CB* 14:R1036-1037.

Hiltunen M, Grudzinska-Sterno M, Wallerman O, Ryberg M, Johannesson H. 2019. Maintenance of High Genome Integrity over Vegetative Growth in the Fairy-Ring Mushroom Marasmius oreades. *Curr. Biol.* 29:2758-2765.e6.

JGI. Joint Genome Institute. Available from: http://genome.jgi.doe.gov/Schco3/Schco3.home.html

Jónsson H, Sulem P, Kehr B, Kristmundsdottir S, Zink F, Hjartarson E, Hardarson MT, Hjorleifsson KE, Eggertsson HP, Gudjonsson SA, et al. 2017. Parental influence on human germline de novo mutations in 1,548 trios from Iceland. *Nature* 549:519–522.

Katju V, Packard LB, Bu L, Keightley PD, Bergthorsson U. 2015. Fitness decline in spontaneous mutation accumulation lines of Caenorhabditis elegans with varying effective population sizes. *Evol. Int. J. Org. Evol.* 69:104–116.

Keightley PD, Trivedi U, Thomson M, Oliver F, Kumar S, Blaxter ML. 2009. Analysis of the genome sequences of three Drosophila melanogaster spontaneous mutation accumulation lines. *Genome Res.* 19:1195–1201.

Kimura M. 1983. The Neutral Theory of Molecular Evolution. Cambridge University Press

Kondrashov AS. 2003. Direct estimates of human per nucleotide mutation rates at 20 loci causing Mendelian diseases. *Hum. Mutat.* 21:12–27.

Kondrashov FA, Kondrashov AS. 2010. Measurements of spontaneous rates of mutations in

the recent past and the near future. *Philos. Trans. R. Soc. B Biol. Sci.* 365:1169–1176.

Kong A, Frigge ML, Masson G, Besenbacher S, Sulem P, Magnusson G, Gudjonsson SA, Sigurdsson A, Jonasdottir Aslaug, Jonasdottir Adalbjorg, et al. 2012. Rate of de novo mutations and the importance of father's age to disease risk. *Nature* 488:471–475.

Kothe E. 1999. Mating Types and Pheromone Recognition in the Homobasidiomycete Schizophyllum commune. *Fungal Genet. Biol.* 27:146–152.

Lanfear R, Ho SYW, Jonathan Davies T, Moles AT, Aarssen L, Swenson NG, Warman L, Zanne AE, Allen AP. 2013. Taller plants have lower rates of molecular evolution. *Nat. Commun.* 4:1–7.

Langmead B, Salzberg SL. 2012. Fast gapped-read alignment with Bowtie 2. *Nat. Methods* 9:357–359.

Latrille T, Duret L, Lartillot N. 2017. The Red Queen model of recombination hot-spot evolution: a theoretical investigation. *Philos. Trans. R. Soc. Lond. B. Biol. Sci.* 372:20160463.

Leffler EM, Bullaughey K, Matute DR, Meyer WK, Ségurel L, Venkat A, Andolfatto P, Przeworski M. 2012. Revisiting an Old Riddle: What Determines Genetic Diversity Levels within Species? *PLoS Biol.* [Internet] 10. Available from: http://www.ncbi.nlm.nih.gov/pmc/articles/PMC3439417/

Long H, Winter DJ, Chang AY-C, Sung W, Wu SH, Balboa M, Azevedo RBR, Cartwright RA, Lynch M, Zufall RA. 2016. Low Base-Substitution Mutation Rate in the Germline Genome of the Ciliate Tetrahymena thermophila. *Genome Biol. Evol.* 8:3629–3639.

Lujan SA, Kunkel TA. 2021. Stability across the Whole Nuclear Genome in the Presence and Absence of DNA Mismatch Repair. *Cells* 10:1224.

Lynch M, Ackerman MS, Gout J-F, Long H, Sung W, Thomas WK, Foster PL. 2016. Genetic drift, selection and the evolution of the mutation rate. *Nat. Rev. Genet.* 17:704–714.

Lynch M, Sung W, Morris K, Coffey N, Landry CR, Dopman EB, Dickinson WJ, Okamoto K, Kulkarni S, Hartl DL, et al. 2008. A genome-wide view of the spectrum of spontaneous mutations in yeast. *Proc. Natl. Acad. Sci. U. S. A.* 105:9272–9277.

Messer PW. 2009. Measuring the Rates of Spontaneous Mutation From Deep and Large-Scale Polymorphism Data. *Genetics* 182:1219–1232.

Milholland B, Dong X, Zhang L, Hao X, Suh Y, Vijg J. 2017. Differences between germline and somatic mutation rates in humans and mice. *Nat. Commun.* 8:15183.

Mukai T, Chigusa SI, Mettler LE, Crow JF. 1972. Mutation Rate and Dominance of Genes Affecting Viability in DROSOPHILA MELANOGASTER. *Genetics* 72:335–355.

Nachman MW, Crowell SL. 2000. Estimate of the mutation rate per nucleotide in humans. *Genetics* 156:297–304.

Ness RW, Morgan AD, Colegrave N, Keightley PD. 2012. Estimate of the spontaneous mutation rate in Chlamydomonas reinhardtii. *Genetics* 192:1447–1454.

Niederpruem DJ, Wessels JG. 1969. Cytodifferentiation and morphogenesis in Schizophyllum commune. *Bacteriol. Rev.* 33:505–535.

Ohm RA, de Jong JF, Lugones LG, Aerts A, Kothe E, Stajich JE, de Vries RP, Record E, Levasseur A, Baker SE, et al. 2010. Genome sequence of the model mushroom Schizophyllum commune. *Nat. Biotechnol.* 28:957–963.

Ohta T. 1992. The Nearly Neutral Theory of Molecular Evolution. *Annu. Rev. Ecol. Syst.* 23:263–286.

Ossowski S, Schneeberger K, Lucas-Lledó JI, Warthmann N, Clark RM, Shaw RG, Weigel D, Lynch M. 2010. The rate and molecular spectrum of spontaneous mutations in Arabidopsis thaliana. *Science* 327:92–94.

Otto SP. 2009. The evolutionary enigma of sex. *Am. Nat.* 174 Suppl 1:S1–S14.

Palmer GE, Horton JS. 2006. Mushrooms by magic: making connections between signal transduction and fruiting body development in the basidiomycete fungus Schizophyllum commune. *FEMS Microbiol. Lett.* 262:1–8.

Popadin K, Polishchuk LV, Mamirova L, Knorre D, Gunbin K. 2007. Accumulation of slightly deleterious mutations in mitochondrial protein-coding genes of large versus small mammals. *Proc. Natl. Acad. Sci. U. S. A.* 104:13390–13395.

Prado-Martinez J, Sudmant PH, Kidd JM, Li H, Kelley JL, Lorente-Galdos B, Veeramah KR, Woerner AE, O'Connor TD, Santpere G, et al. 2013. Great ape genetic diversity and population history. *Nature* 499:471–475.

Puhalla JE. 1970. Genetic studies of the b incompatability locus of Ustilago maydis. *Genet. Res.* 16:229–232.

Ramm SA, Schärer L, Ehmcke J, Wistuba J. 2014. Sperm competition and the evolution of spermatogenesis. *Mol. Hum. Reprod.* 20:1169–1179.

Raper J. 1996. Genetics of Sexuality in Higher Fungi. Available from: https://www.abebooks.com/9780826072955/Genetics-Sexuality-Higher-Fungi-Raper-082607295X/plp

Rockman MV, Kruglyak L. 2009. Recombinational landscape and population genomics of Caenorhabditis elegans. *PLoS Genet.* 5:e1000419.

Romiguier J, Gayral P, Ballenghien M, Bernard A, Cahais V, Chenuil A, Chiari Y, Dernat R, Duret L, Faivre N, et al. 2014. Comparative population genomics in animals uncovers the determinants of genetic diversity. *Nature* 515:261–263.

Scally A, Dutheil JY, Hillier LW, Jordan GE, Goodhead I, Herrero J, Hobolth A, Lappalainen T, Mailund T, Marques-Bonet T, et al. 2012. Insights into hominid evolution from the gorilla genome sequence. *Nature* 483:169–175.

Schmid-Siegert E, Sarkar N, Iseli C, Calderon S, Gouhier-Darimont C, Chrast J, Cattaneo P, Schütz F, Farinelli L, Pagni M, et al. 2017. Low number of fixed somatic mutations in a long-lived oak tree. *Nat. Plants* 3:926–929.

Schrider DR, Houle D, Lynch M, Hahn MW. 2013. Rates and genomic consequences of spontaneous mutational events in Drosophila melanogaster. *Genetics* 194:937–954.

Seplyarskiy VB, Logacheva MD, Penin AA, Baranova MA, Leushkin EV, Demidenko NV, Klepikova AV, Kondrashov FA, Kondrashov AS, James TY. 2014. Crossing-over in a hypervariable species preferentially occurs in regions of high local similarity. *Mol. Biol. Evol.* 31:3016–3025.

Singer R. 1949. The «Agaricales» (Mushrooms) in modern taxonomy. XXII.

Singhal S, Leffler EM, Sannareddy K, Turner I, Venn O, Hooper DM, Strand AI, Li Q, Raney B, Balakrishnan CN, et al. 2015. Stable recombination hotspots in birds. *Science* 350:928–932.

Smukowski Heil CS, Ellison C, Dubin M, Noor MAF. 2015. Recombining without Hotspots: A Comprehensive Evolutionary Portrait of Recombination in Two Closely Related Species of Drosophila. *Genome Biol. Evol.* 7:2829–2842.

Sommer SS. 1995. Recent human germ-line mutation: inferences from patients with hemophilia B. *Trends Genet.* 11:141–147.

Stankis M, Specht C, Giasson L. 1990. Sexual incompatibility in Schizophyllum commune: from classical genetics to a molecular view. In: Seminars in Developmental Biology. Vol. 1. Philadelphia, PA: Saunders Scientific Publishers. p. 195–206.

Stapley J, Feulner PGD, Johnston SE, Santure AW, Smadja CM. 2017. Variation in recombination frequency and distribution across eukaryotes: patterns and processes. *Philos. Trans. R. Soc. B Biol. Sci.* 372:20160455.

Sung W, Ackerman MS, Miller SF, Doak TG, Lynch M. 2012. Drift-barrier hypothesis and mutation-rate evolution. *Proc. Natl. Acad. Sci.* 109:18488–18492.

Takemoto S, Nakamura H, Erwin, Imamura Y, Shimane T. 2010. Schizophyllum commune as a Ubiquitous Plant Parasite. *Jpn. Agric. Res. Q. JARQ* 44:357–364.

The 1000 Genomes Project Consortium. 2015. A global reference for human genetic variation. *Nature* 526:68–74.

Thompson JN, Woodruff RC, Huai H. 1998. Mutation rate: a simple concept has become complex. *Environ. Mol. Mutagen.* 32:292–300.

de Valles-Ibáñez G, Hernandez-Rodriguez J, Prado-Martinez J, Luisi P, Marquès-Bonet T, Casals F. 2016. Genetic Load of Loss-of-Function Polymorphic Variants in Great Apes. *Genome Biol. Evol.* 8:871–877.

Waldman AS, Liskay RM. 1988. Dependence of intrachromosomal recombination in mammalian cells on uninterrupted homology. *Mol. Cell. Biol.* 8:5350–5357.

Wallberg A, Glémin S, Webster MT. 2015. Extreme recombination frequencies shape genome variation and evolution in the honeybee, Apis mellifera. *PLoS Genet.* 11:e1005189.

Wang L, Sun Y, Sun X, Yu L, Xue L, He Z, Huang J, Tian D, Hurst LD, Yang S. 2020. Repeat-induced point mutation in Neurospora crassa causes the highest known mutation rate and mutational burden of any cellular life. *Genome Biol.* 21:142.

Watling R, Sweeney J. 1974. Observations on Schizophyllum commune fries. *Sabouraudia J. Med. Vet. Mycol.* 12:214–226.

Watson JM, Platzer A, Kazda A, Akimcheva S, Valuchova S, Nizhynska V, Nordborg M, Riha K. 2016. Germline replications and somatic mutation accumulation are independent of vegetative life span inArabidopsis. *Proc. Natl. Acad. Sci.* 113:12226–12231.

Wright S. 1931. Evolution in Mendelian Populations. *Genetics* 16:97–159.

Wright S. 1970. Random Drift and the Shifting Balance Theory of Evolution. In: Kojima K, editor. Mathematical Topics in Population Genetics. Biomathematics. Berlin, Heidelberg: Springer. p. 1–31. Available from: https://doi.org/10.1007/978-3-642-46244-3_1

Xu S, Stapley J, Gablenz S, Boyer J, Appenroth KJ, Sree KS, Gershenzon J, Widmer A, Huber M. 2019. Low genetic variation is associated with low mutation rate in the giant duckweed. *Nat. Commun.* 10:1–6.

Yang HP, Tanikawa AY, Kondrashov AS. 2001. Molecular nature of 11 spontaneous de novo mutations in Drosophila melanogaster. *Genetics* 157:1285–1292.

Yang Z. 2007. Phylogenetic Analysis by Maximum Likelihood (PAML). Available from: http://abacus.gene.ucl.ac.uk/software/paml.html

Zhu YO, Siegal ML, Hall DW, Petrov DA. 2014. Precise estimates of mutation rate and spectrum in yeast. *Proc. Natl. Acad. Sci. U. S. A.* 111:E2310–E2318.

Table A1. Assembly statistics for the founding cultures in Chapter 3.

Founding culture	# Contigs	Total length, bp	N50, bp
sh01	5720	37542547	283586
sh02	5735	37427046	251820
sh03	5311	37288394	261670
sh04	7824	38456804	95672

Table A2. Annotation statistics for the founding cultures in Chapter 3.

Founding culture	# CDS	CDS total length, bp	Length of the reference genome covered by lastz alignment	Number of contigs mapped by lastz alignment
sh01	10738	14200866	33077082	1362
sh02	10727	14153358	32909134	1263
sh03	10721	14189031	32911843	1331
sh04	9400	12556173	31930663	1812

Table A3. Frequencies of *de novo* variants in sequenced samples of mycelia (Chapter 3).

Color code of frequency of reads that support alternative variant:		
1=		
0.9 - 1		
0-0.9		
0=		

sh01, narrow tubes, line 1

Contig	Position	Mycelium length ->	Fraction of reads supporting alternative NT		
			164	383	682
NODE_56_length_194026_cov_108.451046	85635		1	1	1
NODE_60_length_186599_cov_115.368809	122372		1	1	1
NODE_230_length_13520_cov_103.278212	7894		0	1	1
NODE_60_length_186599_cov_115.368809	7898		0	1	1
NODE_197_length_21860_cov_113.060690	16042		0	0.99333333	1
NODE_19_length_372591_cov_109.177537	175753		0	0.99315068	1
NODE_703_length_889_cov_329.447044	318		0	0.29859485	0.19832041
NODE_62_length_185817_cov_110.307494	82248		0	0	1
NODE_240_length_12473_cov_153.008712	12081		0	0	0.30748286
NODE_240_length_12473_cov_153.008712	12085		0	0	0.3002997
NODE_240_length_12473_cov_153.008712	12082		0	0	0.29924242
NODE_240_length_12473_cov_153.008712	12083		0	0	0.29761905
NODE_240_length_12473_cov_153.008712	12084		0	0	0.2960373

sh01, narrow tubes, line 2

Contig	Position	Mycelium length ->	Fraction of reads supporting alternative NT			
			192	422	622	829
NODE_124_length_61910_cov_104.730969	21703		0.98913043	1	1	1
NODE_215_length_16154_cov_113.175903	4827		0	0	1	0.99186992
NODE_215_length_16154_cov_113.175903	4828		0	0	1	0.99166667
NODE_215_length_16154_cov_113.175903	4829		0	0	1	0.99166667
NODE_16_length_426697_cov_110.114132	212189		0	0	0	0.5

sh01, narrow tubes, line 3

Contig	Position	Mycelium length ->	Fraction of reads supporting alternative NT				
			143	283	434	611	740
NODE_152_length_38432_cov_104.007431	22923		0	1	1	0	0
NODE_22_length_352799_cov_110.258870	71564		0	0.4197901	0	0	0
NODE_269_length_9972_cov_108.466397	5136		0	0	1	0	0
NODE_43_length_244796_cov_112.842677	109334		0	0	1	0	0
NODE_208_length_18061_cov_155.033029	1285		0	0	0	1	1
NODE_208_length_18061_cov_155.033029	1286		0	0	0	1	1
NODE_208_length_18061_cov_155.033029	1287		0	0	0	1	1

sh01, thick tubes, line 1

Contig	Position	Mycelium length ->	Fraction of reads supporting alternative NT			
			595	999	1402	1761
NODE_16_length_426697_cov_110.114132	86811		1	1	1	1
NODE_267_length_10254_cov_103.595952	8029		0.02238806	1	1	1
NODE_207_length_18330_cov_117.204240	1094		0.01515152	0	1	1
NODE_224_length_14456_cov_86.193338	3670		0	1	0.99212598	1
NODE_2_length_1781848_cov_110.977338	446486		0	0.496	1	1

sh01, thick tubes, line 2

Contig	Position	Mycelium length ->	Fraction of reads supporting alternative NT		
			957	1319	1752
NODE_6_length_671173_cov_108.044111	248290		1	1	1
NODE_67_length_164601_cov_110.809712	77115		0.99047619	0.98305085	1
NODE_6_length_671173_cov_108.044111	115275		0.2295082	1	1
NODE_182_length_26832_cov_111.036031	18562		0	0	0.99363057
NODE_203_length_18792_cov_110.920866	2765		0	0	0.69166667

sh01, thick tubes, line 3

Contig	Position	Mycelium length ->	Fraction of reads supporting alternative NT			
			194	966	1359	1810
NODE_286_length_8606_cov_107.452222	8108		0	1	1	1
NODE_301_length_7453_cov_101.916079	1708		0	0	0	1

sh02, narrow tubes, line 1

Contig	Position	Mycelium length ->	Fraction of reads supporting alternative NT			
			196	391	588	742
NODE_112_length_89523_cov_59.429287	65425		1	1	1	1
NODE_24_length_375303_cov_59.165575	237616		1	1	1	1
NODE_51_length_222525_cov_59.142645	214675		1	1	1	1
NODE_6_length_691203_cov_59.449205	555265		0.98863636	0	0	0
NODE_24_length_375303_cov_59.165575	277181		0.96345029	0	0	0
NODE_19_length_409335_cov_59.158890	294929		0	1	1	1
NODE_51_length_222525_cov_59.142645	118716		0	1	1	1
NODE_61_length_199244_cov_56.432833	83226		0	1	1	1
NODE_216_length_17922_cov_61.459905	14655		0	0	1	1
NODE_10_length_531351_cov_58.599517	349996		0	0	0	1
NODE_15_length_443409_cov_59.075542	163632		0	0	0	1
NODE_18_length_421947_cov_58.771842	153371		0	0	0	0.98461538
NODE_79_length_152252_cov_59.812512	151402		0	0	0	0.21111111

sh02, narrow tubes, line 2

Contig	Position	Mycelium length ->	Fraction of reads supporting alternative NT		
			137	332	470
NODE_50_length_229808_cov_59.045205	27530		1	1	1
NODE_299_length_7948_cov_87.295769	1889		1	1	1
NODE_38_length_268583_cov_60.155196	219152		0.32191781	0	0
NODE_4_length_837680_cov_59.567738	129431		0	1	1
NODE_45_length_242809_cov_59.591059	31011		0	1	1
NODE_84_length_137933_cov_58.905945	67974		0	1	1
NODE_62_length_195584_cov_58.978538	193063		0	1	0.97959184
NODE_58_length_202251_cov_58.040628	26961		0	0	1
NODE_87_length_137488_cov_58.539222	123111		0	0	1
NODE_299_length_7948_cov_87.295769	2447		0	0	0.78396573
NODE_6_length_691203_cov_59.449205	548096		0	0	0.77633627
NODE_6_length_691203_cov_59.449205	555147		0	0	0.74519231

sh02, narrow tubes, line 3

Contig	Position	Mycelium length ->	Fraction of reads supporting alternative NT		
			217	424	511
NODE_50_length_229808_cov_59.045205	27530		1	1	1
NODE_22_length_379479_cov_59.185769	5502		1	1	1
NODE_49_length_230831_cov_61.293065	25329		1	1	1
NODE_4_length_837680_cov_59.567738	630854		0	0	1
NODE_51_length_222525_cov_59.142645	156609		0	0	1
NODE_7_length_684464_cov_59.019493	315546		0	0	0.99342105

sh02, thick tubes, line 1

Contig	Position	Mycelium length ->	Fraction of reads supporting alternative NT					
			394	575	978	1362	1724	1911
NODE_43_length_246384_cov_59.724129	171522		1	1	1	1	1	1
NODE_251_length_12194_cov_63.722539	6852		0.99441341	1	1	1	1	1
NODE_193_length_24245_cov_61.698361	8491		0.99350649	1	1	1	1	1
NODE_198_length_22018_cov_51.587485	19677		0	0	1	1	1	1
NODE_131_length_69637_cov_58.077746	59742		0	0	0	1	1	1
NODE_1_length_1111563_cov_59.791513	597526		0	0	0	1	1	1
NODE_134_length_60610_cov_59.577371	24361		0	0	0	0.98888889	1	1
NODE_290_length_8537_cov_57.916667	3698		0	0	0	0	1	1
NODE_72_length_166475_cov_58.422968	15206		0	0	0	0	1	1
NODE_109_length_90695_cov_58.695833	6871		0	0	0	0	1	0.99354839

sh02, thick tubes, line 2

Contig	Position	Mycelium length ->	Fraction of reads supporting alternative NT					
			192	392	593	980	1381	1803
NODE_104_length_102579_cov_57.628876	69577		0.63366337	0	0	0	0	0
NODE_9_length_554609_cov_58.646058	49442		0.34042553	0.99324324	1	1	1	1
NODE_138_length_57529_cov_58.079179	21580		0.29113924	0.99447514	1	1	1	1
NODE_2_length_939082_cov_58.669906	46278		0	0	0	1	0.99305556	0.99305556
NODE_34_length_286189_cov_59.197196	221928		0	0	0	1	0.97979798	1
NODE_194_length_24053_cov_55.964631	23041		0	0	0	0.98507463	1	1
NODE_13_length_461234_cov_58.934838	127638		0	0	0	0.91208791	0.98130841	0.99264706
NODE_4_length_837680_cov_59.567738	53382		0	0	0	0.87155963	1	1
NODE_99_length_109423_cov_58.283348	61290		0	0	0	0	0.94	1
NODE_24_length_375303_cov_59.165575	276322		0	0	0	0	0	1
NODE_37_length_269810_cov_58.303192	3989		0	0	0	0	0	1

sh02, thick tubes, line 3

Contig	Position	Mycelium length ->	Fraction of reads supporting alternative NT					
			388	572	962	1372	1714	1911
NODE_2_length_939082_cov_58.669906	687093		0.30357143	0	0	0	0	0
NODE_84_length_137933_cov_58.905945	39789		0	0.99453552	1	1	1	1
NODE_11_length_485163_cov_58.782096	149754		0	0.01111111	1	1	1	1
NODE_68_length_172554_cov_58.695797	157879		0	0	0.4076087	0	0	0
NODE_51_length_222525_cov_59.142645	167403		0	0	0	0	1	1

sh03, narrow tubes, line 1

Contig	Position	Mycelium length ->	Fraction of reads supporting alternative NT		
			190	390	623
NODE_107_length_100614_cov_92.404737	29437		1	1	1
NODE_206_length_18696_cov_95.198453	6640		0.07738095	1	1
NODE_34_length_299130_cov_90.075632	250625		0.05527971	1	1
NODE_3_length_784502_cov_87.514203	487495		0	0	1
NODE_51_length_229122_cov_93.229715	145110		0	0	1
NODE_92_length_134134_cov_91.109304	105742		0	0	1

sh03, narrow tubes, line 2

Contig	Position	Mycelium length ->	Fraction of reads supporting alternative NT			
			175	372	545	832
NODE_172_length_34994_cov_83.067331	29973		1	1	1	1
NODE_44_length_254151_cov_89.986350	226182		0.77330508	1	1	1
NODE_11_length_480772_cov_89.019389	462036		0	0	1	1
NODE_154_length_45895_cov_107.543782	9878		0	0	1	1
NODE_2_length_871388_cov_89.347957	605976		0	0	1	1
NODE_81_length_167162_cov_89.380686	130522		0	0	1	1
NODE_159_length_40808_cov_91.974319	15713		0	0	0	1

sh03, narrow tubes, line 3

Contig	Position	Mycelium length ->	Fraction of reads supporting alternative NT			
			187	337	623	790
NODE_16_length_419152_cov_90.779051	145004		1	1	1	1
NODE_3_length_784502_cov_87.514203	395044		1	1	1	1
NODE_64_length_191584_cov_86.525146	181779		1	1	1	1
NODE_8_length_512774_cov_87.426445	357261		1	1	1	1
NODE_132_length_63401_cov_90.203477	42376		0	0.99342105	1	1
NODE_30_length_345374_cov_87.789590	73745		0	0.0375	1	0.99230769
NODE_72_length_178528_cov_93.712476	100370		0	0	0	1

sh03, thick tubes, line 1

Contig	Position	Mycelium length ->	Fraction of reads supporting alternative NT						
			193	401	602	974	1377	1761	2167
NODE_37_length_288349_cov_87.903001	44435		1	0	0	0	0	0	0
NODE_132_length_63401_cov_90.203477	40138		1	0	0	0	0	0	0
NODE_48_length_242291_cov_88.381654	90244		0	1	1	1	1	1	1
NODE_57_length_214349_cov_89.362082	168862		0	1	1	1	1	1	1
NODE_99_length_115198_cov_86.340781	91854		0	0	0.42758621	0	0	0	0
NODE_113_length_86671_cov_86.983244	63928		0	0	0	1	1	1	1
NODE_80_length_169892_cov_92.563537	140747		0	0	0	1	1	1	1
NODE_151_length_50361_cov_83.292141	12565		0	0	0	0	0.98518519	1	1
NODE_1_length_920302_cov_89.049444	80192		0	0	0	0	0	1	1

sh03, thick tubes, line 2

Contig	Position	Mycelium length ->	Fraction of reads supporting alternative NT					
			399	601	1005	1416	1799	1989
NODE_111_length_91966_cov_92.826116	10244		1	1	1	1	1	0.99315068
NODE_50_length_229890_cov_89.936997	120489		1	1	1	1	0.99137931	1
NODE_1226_length_332_cov_197.309804	154		0.39393939	0.36842105	0.32653061	0.39823009	0.38853503	0.31612903
NODE_95_length_126860_cov_90.462459	24458		0.28191489	0	0	0	0	0
NODE_4_length_673840_cov_88.325125	560204		0	1	1	1	1	1
NODE_20_length_389471_cov_90.237641	202344		0	0.01515152	1	1	1	1
NODE_17_length_414073_cov_89.543935	60868		0	0	1	1	0.99342105	1
NODE_9_length_501571_cov_89.433796	407908		0	0	0	0	0	1

sh03, thick tubes, line 3

Contig	Position	Mycelium length ->	Fraction of reads supporting alternative NT					
			395	609	998	1373	1756	1954
NODE_37_length_288349_cov_87.903001	44435		1	0.99401198	1	1	1	1
NODE_363_length_4585_cov_143.594942	3755		1	0.98347107	1	0.99166667	1	1
NODE_3_length_784502_cov_87.514203	289400		0.45255474	0	0	0	0	0
NODE_167_length_36664_cov_90.107169	13148		0	0.01492537	0	1	1	1
NODE_89_length_139449_cov_88.387187	36624		0	0.01129944	1	0.99264706	1	1
NODE_30_length_345374_cov_87.789590	207538		0	0	1	1	0.98765432	1
NODE_27_length_365797_cov_87.174532	152511		0	0	0.97802198	1	1	0.992
NODE_63_length_193509_cov_86.717218	174566		0	0	0	1	1	1
NODE_85_length_157728_cov_89.759247	141170		0	0	0	0	1	1
NODE_214_length_17348_cov_94.284755	15370		0	0	0	0	0	1
NODE_31_length_332802_cov_85.563748	323286		0	0	0	0	0	1
NODE_69_length_181663_cov_89.094958	94263		0	0	0	0	0	1
NODE_73_length_176697_cov_88.246173	42772		0	0	0	0	0	1
NODE_74_length_174647_cov_89.512940	32290		0	0	0	0	0	1

sh04, narrow tubes, line 1

Contig	Position	Mycelium length ->	Fraction of reads supporting alternative NT		
			158	352	548	943
NODE_85_length_113433_cov_91.374616	58526		1	1	1	1
NODE_294_length_40043_cov_88.676050	36664		0.33403805	0	0	0
NODE_111_length_99812_cov_89.937284	52057		0.29487179	0	0	0
NODE_413_length_18999_cov_98.051316	8631		0.24108734	0	0	0
NODE_283_length_42416_cov_83.990411	38333		0	1	1	1
NODE_129_length_89866_cov_90.624330	82605		0	0	1	1
NODE_49_length_149651_cov_91.609083	135720		0	0	0.98299028	1
NODE_149_length_81007_cov_86.108761	21268		0	0	0.88727273	0
NODE_112_length_97955_cov_90.974979	69640		0	0	0	1
NODE_153_length_77540_cov_89.495888	11884		0	0	0	1
NODE_164_length_74363_cov_94.833912	43403		0	0	0	1
NODE_16_length_232479_cov_94.081100	159260		0	0	0	1
NODE_20_length_214542_cov_92.399911	25111		0	0	0	1
NODE_285_length_42048_cov_92.110815	21599		0	0	0	1
NODE_3_length_334555_cov_91.847359	8726		0	0	0	1
NODE_62_length_135693_cov_92.181498	106051		0	0	0	1
NODE_728_length_4022_cov_100.706464	3449		0	0	0	1
NODE_1471_length_705_cov_130.109873	178		0	0	0	0.43434343
NODE_7_length_292489_cov_91.473975	15599		0	0	0	0.34545455
NODE_7_length_292489_cov_91.473975	236954		0	0	0	0.28125

sh04, narrow tubes, line 2

Contig	Position	Mycelium length ->	Fraction of reads supporting alternative NT		
			212	352	498	628
NODE_182_length_67240_cov_91.872177	10727		1	1	0	0
NODE_374_length_23648_cov_99.775869	1782		1	0	0	0
NODE_26_length_201227_cov_91.579901	55662		0.98148148	0	0	0
NODE_26_length_201227_cov_91.579901	55663		0.98076923	0	0	0
NODE_143_length_83357_cov_85.103482	48361		0.9537037	0	0	0
NODE_263_length_47415_cov_106.081182	22463		0.34158986	0	0	
NODE_431_length_17068_cov_94.695604	7860		0.21746032	0	0	0
NODE_9_length_287798_cov_92.323772	33736		0	1	0	0
NODE_664_length_5195_cov_102.194998	2574		0	1	0	0
NODE_375_length_23550_cov_94.093682	457		0	0.20362158	0	0
NODE_672_length_4970_cov_110.962395	1222		0	0	1	1
NODE_307_length_36352_cov_91.697312	9185		0	0	1	1
NODE_2_length_357413_cov_90.721066	52533		0	0	1	1
NODE_1_length_560659_cov_93.433251	38428		0	0	1	1
NODE_10_length_271403_cov_92.920026	117327		0	0	1	1
NODE_14_length_232828_cov_92.949268	165283		0	0	1	1
NODE_69_length_127640_cov_96.265093	19826		0	0	1	1
NODE_115_length_97191_cov_91.380058	10340		0	0	1	0.99056604
NODE_19_length_216373_cov_92.703836	161108		0	0	1	0.9893617
NODE_784_length_3308_cov_123.537295	2043		0	0	0.46728972	0.45631068
NODE_271_length_45070_cov_90.939613	29503		0	0	0	1

sh04, narrow tubes, line 3

Contig	Position	Mycelium length ->	Fraction of reads supporting alternative NT		
			113	305	503	714
NODE_19_length_216373_cov_92.703836	161108		1	0	0	0
NODE_672_length_4970_cov_110.962395	1222		1	0	0	0
NODE_307_length_36352_cov_91.697312	9185		0.98684211	0	0	0
NODE_2_length_357413_cov_90.721066	52533		0.95652174	0	0	0
NODE_1_length_560659_cov_93.433251	38428		0.20187166	0	0	0
NODE_183_length_66665_cov_95.152310	4941		0.25	0	0	0
NODE_113_length_97782_cov_89.884049	14244		0.23076923	0	0	0
NODE_115_length_97191_cov_91.380058	69914		0.22222222	0	0	0
NODE_382_length_22559_cov_93.187039	11395		0.22222222	0	0	0
NODE_757_length_3605_cov_85.023243	3399		0.22222222	0	0	0
NODE_83_length_114648_cov_94.820068	101824		0.20833333	0	0	0
NODE_1505_length_672_cov_131.914286	189		0.2	0	0	0
NODE_280_length_43058_cov_89.714223	36190		0.2	0	0	0
NODE_347_length_26977_cov_92.980112	21413		0.2	0	0	0
NODE_481_length_12624_cov_88.474137	6981		0.2	0	0	0
NODE_182_length_67240_cov_91.872177	10727		0	1	1	1
NODE_9_length_287798_cov_92.323772	33736		0	1	1	1
NODE_664_length_5195_cov_102.194998	2574		0	0.99019608	1	1
NODE_208_length_60143_cov_91.349099	33045		0	1	1	1
NODE_23_length_206721_cov_91.395337	15090		0	0	1	1
NODE_342_length_27784_cov_94.655610	2525		0	0	1	1
NODE_803_length_3174_cov_85.736842	171		0	0	1	1
NODE_417_length_18468_cov_86.702083	8598		0	0	0	1
NODE_431_length_17068_cov_94.695604	6225		0	0	0	1
NODE_35_length_162610_cov_93.149607	41240		0	0	0	0.99212598

sh04, thick tubes, line 1

Contig	Position	Mycelium length ->	Fraction of reads supporting alternative NT						
			400	602	1004	1417	1803	2210	2424
NODE_363_length_25055_cov_92.084074	1860		1	1	1	1	1	1	1
NODE_23_length_206721_cov_91.395337	30558		1	1	1	1	1	1	1
NODE_12_length_245569_cov_92.250705	65198		1	1	1	1	1	0.99342105	0.99230769
NODE_459_length_14321_cov_91.459351	13474		1	1	0.99310345	1	1	1	1
NODE_546_length_9202_cov_107.517589	399		0.76811594	0	0	0	0	0	0
NODE_162_length_75624_cov_94.281004	48119		0	1	1	1	1	0.99193548	1
NODE_158_length_76771_cov_91.643140	62625		0	1	0.99382716	1	1	1	1
NODE_15_length_232821_cov_89.193891	215377		0	1	0.99315068	1	1	1	1
NODE_297_length_38529_cov_92.123609	24014		0	0.91156463	0	0	0	0	0
NODE_137_length_85349_cov_89.585151	34488		0	0.32773109	0	0	0	0	0
NODE_149_length_81007_cov_86.108761	40233		0	0	1	1	1	1	1
NODE_391_length_21794_cov_98.707602	18165		0	0	1	1	1	1	1
NODE_498_length_11947_cov_104.530413	9146		0	0	1	1	1	1	1
NODE_184_length_66406_cov_91.959912	23351		0	0	1	1	1	1	0.99074074
NODE_96_length_103962_cov_91.695442	77446		0	0	1	1	1	0.97826087	1
NODE_4_length_312741_cov_91.696271	192814		0	0	0.63636364	0	0	0	0
NODE_12_length_245569_cov_92.250705	236231		0	0	0.5840708	0	0	0	0
NODE_11_length_257355_cov_92.096767	164745		0	0	0.32089552	0	0	0	0
NODE_14_length_232828_cov_92.949268	60926		0	0	0.3115942	0	0	0	0
NODE_92_length_107494_cov_92.027947	15401		0	0	0.27586207	0	0	0	0
NODE_180_length_67739_cov_101.735051	56434		0	0	0.24031008				
NODE_250_length_49795_cov_92.872159	25207		0	0	0.23966942	0	0	0	0
NODE_10_length_271403_cov_92.920026	90726		0	0	0.21487603	0	0	0	0
NODE_222_length_57256_cov_92.440302	40136		0	0	0	1	1	1	1
NODE_221_length_57387_cov_95.807416	52905		0	0	0	1	1	1	1
NODE_222_length_57256_cov_92.440302	40137		0	0	0	1	1	1	1
NODE_408_length_19119_cov_92.701975	481		0	0	0	1	1	1	1
NODE_458_length_14380_cov_98.966511	8277		0	0	0	1	1	0.99166667	1
NODE_211_length_60031_cov_90.510942	24720		0	0	0	0	1	1	1
NODE_274_length_44173_cov_88.579599	34404		0	0	0	0	1	1	1
NODE_99_length_102718_cov_91.987023	14575		0	0	0	0	1	1	1
NODE_193_length_64454_cov_97.101605	3230		0	0	0	0	0.04040404	1	1
NODE_242_length_51997_cov_93.282473	9853		0	0	0	0	0.02608696	1	1
NODE_258_length_48375_cov_91.087416	37551		0	0	0	0	0	1	1
NODE_275_length_43677_cov_91.269748	36304		0	0	0	0	0	1	1
NODE_292_length_40995_cov_94.135564	36489		0	0	0	0	0	1	1
NODE_292_length_40995_cov_94.135564	36490		0	0	0	0	0	1	1
NODE_9_length_287798_cov_92.323772	148563		0	0	0	0	0	1	1
NODE_33_length_174611_cov_88.248628	53782		0	0	0	0	0	1	0.98058252
NODE_6_length_306897_cov_91.785007	67934		0	0	0	0	0	0.48275862	0
NODE_142_length_83555_cov_92.536980	43955		0	0	0	0	0	0.47096774	0
NODE_100_length_102613_cov_91.907174	21825		0	0	0	0	0	0	1
NODE_127_length_91007_cov_95.027219	40054		0	0	0	0	0	0	0.48623853
NODE_106_length_101452_cov_92.352641	39687		0	0	0	0	0	0	0.39805825
NODE_10_length_271403_cov_92.920026	251198		0	0	0	0	0	0	0.34920635

Table A4. Annotation of *de novo* variants in sequenced samples of mycelia (Chapter 3).

sh01, narrow tubes, line 1

Contig	Position	Reference NT	Alternative NT	Type	Class	AA substitution	ProteinID*	Description*
NODE_56_length_194026_cov_108.451046	85635	C	T	intronic				
NODE_60_length_186599_cov_115.368809	122372	G	C	intergenic				
NODE_230_length_13520_cov_103.278212	7894	G	A	exonic	nonsynonymous	R1328W	2579920	Uncharacterized conserved low complexity protein
NODE_60_length_186599_cov_115.368809	7898	G	A	intergenic				
NODE_197_length_21860_cov_113.060690	16042	C	A	intergenic				
NODE_19_length_372591_cov_109.177537	175753	G	T	upstream				
NODE_703_length_889_cov_329.447044	318	C	T	intergenic				
NODE_62_length_185817_cov_110.307494	82248	C	T	exonic	synonymous	A624A	2637167	
NODE_240_length_12473_cov_153.008712	12081	C	A	intergenic				
NODE_240_length_12473_cov_153.008712	12085	A	C	intergenic				
NODE_240_length_12473_cov_153.008712	12082	G	T	intergenic				
NODE_240_length_12473_cov_153.008712	12083	T	C	intergenic				
NODE_240_length_12473_cov_153.008712	12084	G	A	intergenic				

sh01, narrow tubes, line 2

Contig	Position	Reference NT	Alternative NT	Type	Class	AA substitution	ProteinID*	Description*
NODE_124_length_61910_cov_104.730969	21703	C	T	upstream				
NODE_215_length_16154_cov_113.175903	4827	A	C	intergenic				
NODE_215_length_16154_cov_113.175903	4828	G	C	intergenic				
NODE_215_length_16154_cov_113.175903	4829	G	T	intergenic				
NODE_16_length_426697_cov_110.114132	212189	C	G	intergenic				

sh01, narrow tubes, line 3

Contig	Position	Reference NT	Alternative NT	Type	Class	AA substitution	ProteinID*	Description*
NODE_152_length_38432_cov_104.007431	22923	C	T	upstream				
NODE_22_length_352799_cov_110.258870	71564	C	G	intergenic				
NODE_269_length_9972_cov_108.466397	5136	C	T	intergenic				
NODE_43_length_244796_cov_112.842677	109334	T	A	exonic	nonsynonymous	T139S	1124667	
NODE_208_length_18061_cov_155.033029	1285	G	A	intergenic				
NODE_208_length_18061_cov_155.033029	1286	G	C	intergenic				
NODE_208_length_18061_cov_155.033029	1287	T	C	intergenic				

sh01, thick tubes, line 1

Contig	Position	Reference NT	Alternative NT	Type	Class	AA substitution	ProteinID*	Description*
NODE_16_length_426697_cov_110.114132	86811	G	T	exonic	nonsynonymous	A140E	2605117	Mitotic checkpoint protein PRCC
NODE_267_length_10254_cov_103.595952	8029	C	T	intergenic				
NODE_207_length_18330_cov_117.204240	1094	G	A	intergenic				
NODE_224_length_14456_cov_86.193338	3670	G	C	intergenic				
NODE_2_length_1781848_cov_110.977338	446486	G	C	downstream				

sh01, thick tubes, line 2

Contig	Position	Reference NT	Alternative NT	Type	Class	AA substitution	ProteinID*	Description*
NODE_6_length_671173_cov_108.044111	248290	T	G	exonic	nonsynonymous	T233P	2497064	
NODE_67_length_164601_cov_110.809712	77115	G	C	exonic	nonsynonymous	A672P	2497614	Sirtuin 4 and related class II sirtuins (SIR2 family)
NODE_6_length_671173_cov_108.044111	115275	C	T	upstream; downstream				
NODE_182_length_26832_cov_111.036031	18562	C	T	intergenic				
NODE_203_length_18792_cov_110.920866	2765	C	T	intergenic				

sh01, thick tubes, line 3

Contig	Position	Reference NT	Alternative NT	Type	Class	AA substitution	ProteinID*	Description*
NODE_286_length_8606_cov_107.452222	8108	G	A	intergenic				
NODE_301_length_7453_cov_101.916079	1708	C	A	downstream				

sh02, narrow tubes, line 1

Contig	Position	Reference NT	Alternative NT	Type	Class	AA substitution	ProteinID*	Description*
NODE_112_length_89523_cov_59.429287	65425	T	C	upstream				
NODE_24_length_375303_cov_59.165575	237616	T	C	downstream				
NODE_51_length_222525_cov_59.142645	214675	G	A	exonic	synonymous	A364A	2753567	
NODE_6_length_691203_cov_59.449205	555265	C	G	exonic	nonsynonymous	N127K	2613111	G-protein beta subunit
NODE_24_length_375303_cov_59.165575	277181	C	T	upstream				
NODE_19_length_409335_cov_59.158890	294929	G	T	exonic	nonsynonymous	L294F	1159597	Cytochrome P450 CYP4/CYP19/CYP26 subfamilies
NODE_51_length_222525_cov_59.142645	118716	T	C	downstream				
NODE_61_length_199244_cov_56.432833	83226	G	A	intergenic				
NODE_216_length_17922_cov_61.459905	14655	G	A	intergenic				
NODE_10_length_531351_cov_58.599517	349996	G	A	exonic	nonsynonymous	P524S	2635346	Choline transporter-like protein
NODE_15_length_443409_cov_59.075542	163632	C	T	intronic				
NODE_18_length_421947_cov_58.771842	153371	G	A	upstream				
NODE_79_length_152252_cov_59.812512	151402	G	C	upstream				

sh02, narrow tubes, line 2

Contig	Position	Reference NT	Alternative NT	Type	Class	AA substitution	ProteinID*	Description*
NODE_50_length_229808_cov_59.045205	27530	C	A	downstream				
NODE_299_length_7948_cov_87.295769	1889	C	G	intergenic				
NODE_38_length_268583_cov_60.155196	219152	C	G	upstream				
NODE_4_length_837680_cov_59.567738	129431	C	T	exonic	synonymous	F185F	2747457	Uncharacterized conserved protein
NODE_45_length_242809_cov_59.591059	31011	G	A	intronic				
NODE_84_length_137933_cov_58.905945	67974	C	A	intergenic				
NODE_62_length_195584_cov_58.978538	193063	C	G	intergenic				
NODE_58_length_202251_cov_58.040628	26961	G	T	exonic	nonsynonymous	T1360K	2593296	Dystonin, GAS (Growth-arrest-specific protein), and related proteins
NODE_87_length_137488_cov_58.539222	123111	C	A	intergenic				
NODE_299_length_7948_cov_87.295769	2447	G	A	intergenic				
NODE_6_length_691203_cov_59.449205	548096	A	G	exonic	synonymous	R378R	80616	Spindle pole body protein - Sad1p
NODE_6_length_691203_cov_59.449205	555147	T	G	exonic	nonsynonymous	I88S	2613111	G-protein beta subunit

sh02, narrow tubes, line 3

Contig	Position	Reference NT	Alternative NT	Type	Class	AA substitution	ProteinID*	Description*
NODE_50_length_229808_cov_59.045205	27530	C	A	downstream				
NODE_22_length_379479_cov_59.185769	5502	C	G	downstream				
NODE_49_length_230831_cov_61.293065	25329	C	G	intergenic				
NODE_4_length_837680_cov_59.567738	630854	C	A	upstream				
NODE_51_length_222525_cov_59.142645	156609	C	A	exonic	nonsynonymous	S191Y	2644330	Reductases with broad range of substrate specificiti
NODE_7_length_684464_cov_59.019493	315546	A	G	exonic	nonsynonymous	N44S	2562768	Nitrogen permease regulator NLRG/NPR2

100

sh02, thick tubes, line 1

Contig	Position	Reference NT	Alternative NT	Type	Class	AA substitution	ProteinID*	Description*
NODE_43_length_246384_cov_59.724129	171522	C	A	exonic	nonsynonymous	A1408S	2488968	SWI-SNF chromatin-remodeling complex protein
NODE_251_length_12194_cov_63.722539	6852	C	G	intergenic				
NODE_193_length_24245_cov_61.698361	8491	G	A	intergenic				
NODE_198_length_22018_cov_51.587485	19677	G	T	intergenic				
NODE_131_length_69637_cov_58.077746	59742	C	A	downstream				
NODE_1_length_1111563_cov_59.791513	597526	C	A	upstream; downstream				
NODE_134_length_60610_cov_59.577371	24361	G	T	downstream				
NODE_290_length_8537_cov_57.916667	3698	G	C	intergenic				
NODE_72_length_166475_cov_58.422968	15206	C	G	exonic	nonsynonymous	G183R	2509890	G-protein alpha subunit (small G protein superfamil
NODE_109_length_90695_cov_58.695833	6871	C	T	upstream				

sh02, thick tubes, line 2

Contig	Position	Reference NT	Alternative NT	Type	Class	AA substitution	ProteinID*	Description*
NODE_104_length_102579_cov_57.628876	69577	C	T	exonic	nonsynonymous	R444C	2704852	Splicing coactivator SRm160/300, subunit SRm300
NODE_9_length_554609_cov_58.646058	49442	G	T	exonic	stopgain	Y259X	2608678	Monocarboxylate transporter
NODE_138_length_57529_cov_58.079179	21580	G	C	intergenic				
NODE_2_length_939082_cov_58.669906	46278	C	A	downstream				
NODE_34_length_286189_cov_59.197196	221928	G	A	intergenic				
NODE_194_length_24053_cov_55.964631	23041	G	A	intergenic				
NODE_13_length_461234_cov_58.934838	127638	G	T	upstream				
NODE_4_length_837680_cov_59.567738	53382	G	A	exonic	nonsynonymous	V467I	1165810	Protein involved in vacuole import and degradation
NODE_99_length_109423_cov_58.283348	61290	T	C	downstream				
NODE_24_length_375303_cov_59.165575	276322	C	T	intergenic				
NODE_37_length_269810_cov_58.303192	3989	A	G	downstream				

sh02, thick tubes, line 3

Contig	Position	Reference NT	Alternative NT	Type	Class	AA substitution	ProteinID*	Description*
NODE_2_length_939082_cov_58.669906	687093	C	A	downstream				
NODE_84_length_137933_cov_58.905945	39789	C	G	intergenic				
NODE_11_length_485163_cov_58.782096	149754	C	T	intergenic				
NODE_68_length_172554_cov_58.695797	157879	C	A	exonic	nonsynonymous	P213T	2502581	
NODE_51_length_222525_cov_59.142645	167403	G	T	exonic	nonsynonymous	L117I	2753584	

sh03, narrow tubes, line 1

Contig	Position	Reference NT	Alternative NT	Type	Class	AA substitution	ProteinID*	Description*
NODE_107_length_100614_cov_92.404737	29437	G	C	exonic	nonsynonymous	P249A	2635873	
NODE_206_length_18696_cov_95.198453	6640	T	G	intergenic				
NODE_34_length_299130_cov_90.075632	250625	C	T	intronic				
NODE_3_length_784502_cov_87.514203	487495	T	C	exonic	nonsynonymous	V702A	2618282	Nucleolar GTPase/ATPase p130
NODE_51_length_229122_cov_93.229715	145110	A	G	exonic	synonymous	P77P	2614389	von Willebrand factor and related coagulation proteins
NODE_92_length_134134_cov_91.109304	105742	C	T	exonic	nonsynonymous	A133T	2645469	

sh03, narrow tubes, line 2

Contig	Position	Reference NT	Alternative NT	Type	Class	AA substitution	ProteinID*	Description*
NODE_172_length_34994_cov_83.067331	29973	A	T	upstream				
NODE_44_length_254151_cov_89.986350	226182	C	T	intergenic				
NODE_11_length_480772_cov_89.019389	462036	G	A	upstream				
NODE_154_length_45895_cov_107.543782	9878	C	T	intergenic				
NODE_2_length_871388_cov_89.347957	605976	C	G	exonic	nonsynonymous	G266A	2601543	Cyclin-dependent kinase WEE1
NODE_81_length_167162_cov_89.380686	130522	C	T	intronic				
NODE_159_length_40808_cov_91.974319	15713	C	T	intergenic				

sh03, narrow tubes, line 3

Contig	Position	Reference NT	Alternative NT	Type	Class	AA substitution	ProteinID*	Description*
NODE_16_length_419152_cov_90.779051	145004	A	G	upstream; downstream				
NODE_3_length_784502_cov_87.514203	395044	C	A	upstream				
NODE_64_length_191584_cov_86.525146	181779	C	T	intergenic				
NODE_8_length_512774_cov_87.426445	357261	G	T	exonic	nonsynonymous	A18D	2023027	Uncharacterized conserved protein
NODE_132_length_63401_cov_90.203477	42376	C	G	upstream				
NODE_30_length_345374_cov_87.789590	73745	C	G	exonic	synonymous	S427S	1216168	Serine/threonine protein kinase
NODE_72_length_178528_cov_93.712476	100370	C	A	upstream				

sh04, narrow tubes, line 1

Contig	Position	Reference NT	Alternative NT	Type	Class	AA substitution	ProteinID*	Description*
NODE_85_length_113433_cov_91.374616	58526	G	T	intergenic				
NODE_294_length_40043_cov_88.676050	36664	T	A	intergenic				
NODE_111_length_99812_cov_89.937284	52057	G	A	upstream; downstream				
NODE_413_length_18999_cov_98.051316	8631	C	T	intergenic				
NODE_283_length_42416_cov_83.990411	38333	G	C	upstream				
NODE_129_length_89866_cov_90.624330	82605	C	T	intronic				
NODE_49_length_149651_cov_91.609083	135720	A	T	intergenic				
NODE_149_length_81007_cov_86.108761	21268	C	T	exonic	synonymous	G743G	2704936	SWI/SNF-related matrix-associated actin-dependent regulator of chromatin
NODE_112_length_97955_cov_90.974979	69640	G	A	exonic	nonsynonymous	R865C	2523785	Transcription elongation factor SPT6
NODE_153_length_77540_cov_89.495888	11884	C	T	exonic	nonsynonymous	P85S	2623742	Collagens (type IV and type XIII), and related proteins
NODE_164_length_74363_cov_94.833912	43403	G	A	upstream				
NODE_16_length_232479_cov_94.081100	159260	A	G	exonic	nonsynonymous	E678G	2739833	Dual-specificity tyrosine-phosphorylation regulated kinase
NODE_20_length_214542_cov_92.399911	25111	G	A	exonic	synonymous	V42V	2678167	NADH:flavin oxidoreductase/12-oxophytodienoate reductase
NODE_285_length_42048_cov_92.110815	21599	T	C	exonic	nonsynonymous	N88S	2583795	Dehydrogenases with different specificities (related to short-chain alcohol dehydrogenases)
NODE_3_length_334555_cov_91.847359	8726	G	T	exonic	nonsynonymous	A96E	2568898	
NODE_62_length_135693_cov_92.181498	106051	T	C	upstream; downstream				
NODE_728_length_4022_cov_100.706464	3449	G	A	intergenic				
NODE_1471_length_705_cov_130.109873	178	G	A	intergenic				
NODE_7_length_292489_cov_91.473975	15599	G	A	intergenic				
NODE_7_length_292489_cov_91.473975	236954	G	C	exonic	synonymous	S190S	2544657	

sh04, narrow tubes, line 2

Contig	Position	Reference NT	Alternative NT	Type	Class	AA substitution	ProteinID*	Description*
NODE_182_length_67240_cov_91.872177	10727	C	T	downstream				
NODE_374_length_23648_cov_99.775869	1782	C	A	exonic	nonsynonymous	V179F	2509890	G-protein alpha subunit (small G protein superfamily)
NODE_26_length_201227_cov_91.579901	55662	T	G	intronic				
NODE_26_length_201227_cov_91.579901	55663	T	G	intronic				
NODE_143_length_83357_cov_85.103482	48361	G	C	intergenic				
NODE_263_length_47415_cov_106.081182	22463	C	T	intergenic				
NODE_431_length_17068_cov_94.695604	7860	T	C	intergenic				
NODE_9_length_287798_cov_92.323772	33736	C	A	upstream				
NODE_664_length_5195_cov_102.194998	2574	G	A	intergenic				
NODE_375_length_23550_cov_94.093682	457	T	C	upstream				
NODE_672_length_4970_cov_110.962395	1222	C	A	intergenic				
NODE_307_length_36352_cov_91.697312	9185	C	A	intergenic				
NODE_2_length_357413_cov_90.721066	52533	G	T	exonic	nonsynonymous	K97N	2613111	G-protein beta subunit
NODE_1_length_560659_cov_93.433251	38428	C	A	downstream				
NODE_10_length_271403_cov_92.920026	117327	A	G	upstream				
NODE_14_length_232828_cov_92.949268	165283	G	A	intergenic				
NODE_69_length_127640_cov_96.265093	19826	T	G	exonic	nonsynonymous	L71W	1166919	Multidrug resistance-associated protein/mitoxantrone resistance protein, ABC superfamily
NODE_115_length_97191_cov_91.380058	10340	G	C	downstream				
NODE_19_length_216373_cov_92.703836	161108	G	T	intronic				
NODE_784_length_3308_cov_123.537295	2043	C	A	intergenic				
NODE_271_length_45070_cov_90.939613	29503	G	T	upstream				

sh04, narrow tubes, line 3

Contig	Position	Reference NT	Alternative NT	Type	Class	AA substitution	ProteinID*	Description*
NODE_19_length_216373_cov_92.703836	161108	G	T	intronic				
NODE_672_length_4970_cov_110.962395	1222	C	A	intergenic				
NODE_307_length_36352_cov_91.697312	9185	C	A	intergenic				
NODE_2_length_357413_cov_90.721066	52533	G	T	exonic	nonsynonymous	K97N	2613111	G-protein beta subunit
NODE_1_length_560659_cov_93.433251	38428	C	A	downstream				
NODE_183_length_66665_cov_95.152310	4941	G	T	intergenic				
NODE_113_length_97782_cov_89.884049	14244	C	A	intergenic				
NODE_115_length_97191_cov_91.380058	69914	G	T	intergenic				
NODE_382_length_22559_cov_93.187039	11395	C	A	exonic	nonsynonymous	A60S	2606133	Serine carboxypeptidases
NODE_757_length_3605_cov_85.023243	3399	G	T	intergenic				
NODE_83_length_114648_cov_94.820068	101824	C	A	downstream				
NODE_1505_length_672_cov_131.914286	189	G	T	intergenic				
NODE_280_length_43058_cov_89.714223	36190	C	A	exonic	nonsynonymous	A263E	2617051	Ca2+ transporting ATPase
NODE_347_length_26977_cov_92.980112	21413	G	T	intergenic				
NODE_481_length_12624_cov_88.474137	6981	C	A	intergenic				
NODE_182_length_67240_cov_91.872177	10727	C	T	downstream				
NODE_9_length_287798_cov_92.323772	33736	C	A	upstream				
NODE_664_length_5195_cov_102.194998	2574	G	A	intergenic				
NODE_208_length_60143_cov_91.349099	33045	T	A	exonic	nonsynonymous	V472D	2627190	Multidrug resistance-associated protein/mitoxantrone resistance protein, ABC superfamily
NODE_23_length_206721_cov_91.395337	15090	A	G	exonic	nonsynonymous	Y932C	2662880	Multidrug resistance-associated protein/mitoxantrone resistance protein, ABC superfamily
NODE_342_length_27784_cov_94.655610	2525	C	T	intergenic				
NODE_803_length_3174_cov_85.736842	171	C	T	intergenic				
NODE_417_length_18468_cov_86.702083	8598	C	A	exonic	nonsynonymous	T65N	2500894	Predicted membrane protein
NODE_431_length_17068_cov_94.695604	6225	G	A	intergenic				
NODE_35_length_162610_cov_93.149607	41240	G	T	intronic				

sh04, thick tubes, line 1

Contig	Position	Reference NT	Alternative NT	Type	Class	AA substitution	ProteinID*	Description*
NODE_363_length_25055_cov_92.084074	1860	G	A	intergenic				
NODE_23_length_206721_cov_91.395337	30558	C	T	upstream; downstream				
NODE_12_length_245569_cov_92.250705	65198	C	A	upstream; downstream				
NODE_459_length_14321_cov_91.459351	13474	C	T	upstream				
NODE_546_length_9202_cov_107.517589	399	C	G	intergenic				
NODE_162_length_75624_cov_94.281004	48119	C	A	downstream				
NODE_158_length_76771_cov_91.643140	62625	C	A	intronic				
NODE_15_length_232821_cov_89.193891	215377	G	A	downstream				
NODE_297_length_38529_cov_92.123609	24014	C	A	exonic	nonsynonymous	Q354H	2743676	
NODE_137_length_85349_cov_89.585151	34488	C	A	intergenic				
NODE_149_length_81007_cov_86.108761	40233	C	T	intergenic				
NODE_391_length_21794_cov_98.707602	18165	C	G	intergenic				
NODE_498_length_11947_cov_104.530413	9146	G	T	intergenic				
NODE_184_length_66406_cov_91.959912	23351	G	C	intronic				
NODE_96_length_103962_cov_91.695442	77446	C	T	intergenic				
NODE_4_length_312741_cov_91.696271	192814	A	C	intronic				
NODE_12_length_245569_cov_92.250705	236231	G	A	downstream				
NODE_11_length_257355_cov_92.096767	164745	T	G	upstream				
NODE_14_length_232828_cov_92.949268	60926	C	G	upstream; downstream				
NODE_92_length_107494_cov_92.027947	15401	G	C	intergenic				
NODE_180_length_67739_cov_101.735051	56434	T	A	intergenic				
NODE_250_length_49795_cov_92.872159	25207	G	T	exonic	nonsynonymous	M213I	2628798	Nuclear receptors of the nerve growth factor-induced protein B type

NODE_10_length_271403_cov_92.920026	90726	C	A	upstream; downstream				
NODE_222_length_57256_cov_92.440302	40136	G	T	downstream				
NODE_221_length_57387_cov_95.807416	52905	C	A	intergenic				
NODE_222_length_57256_cov_92.440302	40137	A	T	downstream				
NODE_408_length_19119_cov_92.701975	481	A	C	intergenic				
NODE_458_length_14380_cov_98.966511	8277	C	G	exonic	synonymous	P361P	2586820	
NODE_211_length_60031_cov_90.510942	24720	T	C	splicing				
NODE_274_length_44173_cov_88.579599	34404	G	A	exonic	synonymous	S755S	2197600	Septin family protein (P-loop GTPase)
NODE_99_length_102718_cov_91.987023	14575	G	T	intronic				
NODE_193_length_64454_cov_97.101605	3230	C	A	intergenic				
NODE_242_length_51997_cov_93.282473	9853	C	A	intergenic				
NODE_258_length_48375_cov_91.087416	37551	G	T	intergenic				
NODE_275_length_43677_cov_91.269748	36304	C	T	exonic	synonymous	R166R	2587980	
NODE_292_length_40995_cov_94.135564	36489	C	A	downstream				
NODE_292_length_40995_cov_94.135564	36490	A	G	downstream				
NODE_9_length_287798_cov_92.323772	148563	C	A	exonic	nonsynonymous	S526R	2637598	Predicted transporter (major facilitator superfamily)
NODE_33_length_174611_cov_88.248628	53782	G	A	intergenic				
NODE_6_length_306897_cov_91.785007	67934	C	A	exonic	nonsynonymous	D684E	2607375	
NODE_142_length_83555_cov_92.536980	43955	A	C	upstream				
NODE_100_length_102613_cov_91.907174	21825	C	G	upstream; downstream				
NODE_127_length_91007_cov_95.027219	40054	T	A	intergenic				
NODE_106_length_101452_cov_92.352641	39687	T	C	downstream				
NODE_10_length_271403_cov_92.920026	251198	G	A	intergenic				

Contig	Position	Reference NT	Alternative NT	Type	Class	AA substitution	ProteinID*	Description*
sh04, thick tubes, line 2								
NODE_100_length_102613_cov_91.907174	56293	C	A	exonic	nonsynonymous	V22L	2630488	
NODE_105_length_101768_cov_93.063083	76556	G	C	exonic	nonsynonymous	P46A	2556348	
NODE_225_length_56749_cov_95.118524	52792	G	T	intergenic				
NODE_22_length_207203_cov_91.980480	13222	T	C	exonic	nonsynonymous	N411D	2486189	mRNA splicing factor
NODE_4_length_312741_cov_91.696271	142952	G	C	upstream				
NODE_98_length_102797_cov_95.022702	3751	G	C	intergenic				
NODE_215_length_59221_cov_94.871382	7963	G	A	intergenic				
NODE_76_length_120449_cov_90.873841	113974	C	A	intergenic				
NODE_193_length_64454_cov_97.101605	8884	C	G	exonic	synonymous	V86V	2631390	Multifunctional pyrimidine synthesis protein CAD (includes carbamoyl-phophate synthetase, aspartate transcarbamylase, and glutamine amidotransferase)
NODE_525_length_9996_cov_87.230769	926	C	T	intergenic				
NODE_197_length_61978_cov_94.241725	56287	C	A	upstream				
NODE_279_length_43127_cov_88.153330	5275	C	T	intergenic				
NODE_200_length_61483_cov_94.304091	58916	G	C	exonic	nonsynonymous	C565W	1207856	
NODE_112_length_97955_cov_90.974979	20798	C	A	upstream; downstream				
NODE_240_length_52442_cov_94.319183	27350	G	A	exonic	nonsynonymous	V412I	2608117	Predicted E3 ubiquitin ligase
NODE_66_length_132311_cov_88.949173	90160	C	A	downstream				
NODE_32_length_175704_cov_92.502918	102787	G	C	intergenic				
NODE_352_length_26595_cov_85.267592	11951	C	A	upstream				
NODE_934_length_2244_cov_93.312796	1178	G	A	intergenic				
NODE_982_length_1917_cov_98.096739	635	G	A	intergenic				
NODE_301_length_37602_cov_90.296762	28145	A	G	upstream				
NODE_521_length_10316_cov_100.048247	5452	G	A	intergenic				
NODE_278_length_43166_cov_94.678387	25568	C	T	exonic	synonymous	S202S	2526554	Growth factor receptor-bound proteins (GRB7, GRB10, GRB14)
NODE_238_length_54561_cov_90.199508	23552	G	T	downstream				
NODE_465_length_13644_cov_88.872264	5555	G	A	intergenic				
NODE_71_length_126715_cov_91.477455	11363	G	A	exonic	nonsynonymous	R237C	2611100	Mitochondrial translation elongation factor Tu
NODE_391_length_21794_cov_98.707602	13892	A	G	intergenic				
NODE_12_length_245569_cov_92.250705	65312	G	A	upstream; downstream				

sh04, thick tubes, line 3								
Contig	Position	Reference NT	Alternative NT	Type	Class	AA substitution	ProteinID*	Description*
NODE_345_length_27033_cov_91.022036	21606	C	T	downstream				
NODE_630_length_6028_cov_153.771299	3370	G	A	intergenic				
NODE_69_length_127640_cov_96.265093	3053	G	C	intronic				
NODE_218_length_57825_cov_92.013628	47591	C	A	exonic	nonsynonymous	A167E	2624635	
NODE_20_length_214542_cov_92.399911	4363	C	T	upstream				
NODE_249_length_50122_cov_91.808432	7729	C	A	upstream				
NODE_74_length_122341_cov_93.978784	24878	G	T	upstream; downstream				
NODE_149_length_81007_cov_86.108761	42099	C	T	upstream				
NODE_46_length_152718_cov_93.775611	12635	G	T	intronic				
NODE_27_length_192673_cov_91.219247	176989	G	T	upstream				
NODE_1_length_560659_cov_93.433251	327237	C	T	exonic	synonymous	I1095I	2744705	Nuclear pore complex, Nup155 component (D Nup154, sc Nup157/Nup170)
NODE_114_length_97512_cov_93.418915	96181	C	T	exonic	nonsynonymous	E98K	2606814	mRNA splicing protein CDC5 (Myb superfamily)
NODE_43_length_156944_cov_91.858976	151358	C	T	upstream				
NODE_513_length_10751_cov_59.953064	5251	C	G	intergenic				
NODE_187_length_66028_cov_91.257479	13853	C	T	exonic	nonsynonymous	A60V	2615415	
NODE_39_length_159138_cov_90.840748	23759	C	A	upstream				
NODE_70_length_127247_cov_92.627208	1338	C	T	exonic	nonsynonymous	A932T	2626105	Positive cofactor 2 (PC2), subunit of a multiprotein coactivator of RNA polymerase II
NODE_25_length_204537_cov_91.888139	119516	C	T	exonic	nonsynonymous	P317L	2634380	Splicing coactivator SRm160/300, subunit SRm300
NODE_61_length_137219_cov_91.734655	124699	G	A	exonic	nonsynonymous	G107D	2639414	Transcriptional coactivator p100
NODE_279_length_43127_cov_88.153330	5275	C	T	intergenic				
NODE_183_length_66665_cov_95.152310	33189	G	A	intergenic				
NODE_164_length_74363_cov_94.833912	65437	C	T	exonic	synonymous	E208E	2068975	
NODE_310_length_35454_cov_90.250869	7422	G	A	intergenic				
NODE_89_length_109783_cov_92.474678	8572	G	T	upstream				
NODE_393_length_21752_cov_94.480323	11858	C	A	exonic	nonsynonymous	Q74K	2521109	von Willebrand factor and related coagulation proteins

*From http://genome.jgi.doe.gov/Schco3/Schco3.home.html

Table A5. Distances between collected fruit bodies, Chapter 4.

Trunk	Fruit bodies		Distance, cm
I	Shiz1	Shiz2	70
II	Shiz3	Shiz4	190
III	Shiz5	Shiz6	182
IV	Shiz7	Shiz8	179
V	Shiz9	Shiz10	108
VII	Shiz11	Shiz12	60
IX	Shiz13	Shiz14	91
X	Shiz15	Shiz16	92
XI	Shiz17	Shiz18	116
XII	Shiz19	Shiz20	47
XIII	Shiz21	Shiz22	32
XIV	Shiz23	Shiz24	87
XV	Shiz25	Shiz26	67

Table A6. Assembly statistics, Chapter 4.

Sample	# contigs	Total length, bp	GC, %	N50. bp	# N's per 100 kbp
Shiz1	53391	63149569	57.48	3149	884.72
Shiz11	52777	64450997	57.35	3443	936.81
Shiz14	70004	78516207	57.02	4110	740.8
Shiz15	56691	64793962	57.53	3220	788.76
Shiz16	64279	70078415	57.58	3501	707.83
Shiz17	56238	64783527	57.4	3239	783.61
Shiz18	77882	79643181	57.01	3323	959.85
Shiz19	64100	72202554	57.83	3555	874.52
Shiz2	61413	67528185	57.42	3443	742.22
Shiz21	58127	64898260	57.57	2913	874.89
Shiz22	50401	62164389	57.46	3187	832.74
Shiz23	74005	73858210	57.87	3172	900.62

Shiz24	59693	66609803	57.62	3148	809.25
Shiz25	64033	67164782	57.59	3109	829.25
Shiz26	66156	71136504	57.72	3475	736.16
Shiz3	58575	66346535	57.61	3045	904.78
Shiz5	55271	70454560	57.58	3788	825.05
Shiz7	80093	74684720	57.83	3145	873.74
Shiz9	50307	64330433	57.45	3428	931.95

Table A7. Statistics for scaffold assemblies for parents, F1, BC and F2 offsprings (Chapters 5 and 6).

Crossing	Sample	# contigs	Total length, bp	Largest contig, bp	GC (%)	N50, bp	# N's per 100 kbp
-	sh01	25	39298859	4520496	57.59	2600426	10043.9
	sh04	25	39626086	4552179	57.73	2628537	13390.89
	fl_1	25	39425227	4543380	57.63	2606682	11073.07
	fl_10	25	39508978	4521665	57.66	2629077	11640.85
	fl_11	25	39380835	4523702	57.62	2602195	10442.42
	fl_12	25	39457553	4528023	57.66	2619879	11742.87
	fl_13	25	39441210	4546165	57.64	2609368	11203.59
	fl_14	25	39450636	4544676	57.63	2629014	11275.33
	fl_15	25	39482240	4529189	57.66	2618873	11468.64
F1 (sh01 x sh04)	fl_16	25	39453967	4520024	57.64	2616939	11184.79
	fl_17	25	39466501	4520266	57.63	2629536	11430.53
	fl_18	25	39440593	4546741	57.65	2606154	11475.02
	fl_19	25	39505433	4534322	57.62	2628173	11282.2
	fl_20	25	39385275	4525878	57.62	2601344	10767.87
	fl_21	25	39454912	4521107	57.65	2624894	11827.83
	fl_22	25	39489202	4540973	57.68	2626046	11888.65

	f1_24	25	39341342	4519871	57.6	2608040	10283.91
	f1_25	25	39406450	4529128	57.63	2606426	10840.88
	f1_26	25	39445890	4534093	57.66	2607699	11730.43
	f1_27	25	39470579	4537330	57.66	2627394	12114.03
	f1_3	25	39444000	4545413	57.64	2613623	11319.51
	f1_5	25	39526688	4536934	57.65	2627447	11846
	f1_6	25	39505518	4537087	57.67	2609879	11991.58
	f1_7	25	39477148	4533845	57.66	2615895	12137.95
	f1_8	25	39479332	4538969	57.62	2608038	11486.13
	f1_9	25	39452519	4548385	57.65	2607220	11405.37
	Z1	25	39405141	4527703	57.65	2607304	11228
	Z12	25	39350207	4523922	57.61	2606893	10736.38
	Z14	25	39365782	4520796	57.61	2608346	10830.69
BC (sh01 x f1_26)	Z17	25	39379381	4526411	57.61	2607790	9844.69
	Z19	25	39376910	4530089	57.63	2608140	10683.09
	Z22	25	39419142	4522789	57.64	2607998	11308.86
	Z29	25	39340383	4519573	57.61	2604495	10585.97
	Z35	25	39376408	4520741	57.63	2604173	11230.32
	C11	25	39479246	4533300	57.66	2618450	12319.32
	C12	25	39449575	4534021	57.66	2615397	11577.67
	C14	25	39431459	4533229	57.62	2609540	11428.16
	C15	25	39476821	4533694	57.64	2617843	12155.89
	C17	25	39434152	4532121	57.64	2609495	11565.78
	C18	25	39455893	4526984	57.63	2606657	11137.26
	C20	25	39443710	4522364	57.65	2609632	11893.81
F2 (f1_26 x f1_7)	C22	25	39489310	4528811	57.68	2617177	12266.3
	C26	25	39476927	4533918	57.66	2608286	11743.47
	C27	25	39372942	4535222	57.62	2602907	10802
	C28	25	39457143	4531474	57.67	2613713	11487.05
	C31	25	39453700	4526152	57.65	2615880	12105.22
	C33	25	39468213	4529033	57.66	2607903	12733.83
	C34	25	39499648	4531343	57.67	2615556	12827.88

						Ref	Alt	DP in	DP in	Geno-
C5	25		39450965	4523439	57.66	2609592	11805.14			
C7	25		39445835	4530852	57.65	2617355	11932.53			

e

Genome state	Crossing	Fruit body	Sample	Scaffold	Position	Ref base	Alt base	DP in sample	DP in ref	Geno-type
		4	fl_9	scaffold_1	3699827	C	T	258	104	sh01
		3	fl_17	scaffold_1	4189757	T	A	233	100	sh01
		1	fl_22	scaffold_10	826024	G	A	258	95	sh01
		3	fl_7	scaffold_10	1667820	G	A	238	127	sh04
		3	fl_17	scaffold_10	1679604	G	A	238	69	sh04
		3	fl_11	scaffold_10	1690861	G	A	246	83	sh04
		4	fl_19	scaffold_10	1707190	G	A	238	77	sh04
		4	fl_9	scaffold_10	1717804	G	A	202	75	sh04
		2	fl_14	scaffold_11	56708	C	T	84	63	sh01
		4	fl_8	scaffold_11	1208840	T	C	215	121	sh04
		1	fl_1	scaffold_12	1117582	A	G	289	117	sh04
Hetero-zygous	F1	4	fl_8	scaffold_14	935390	T	G	225	127	sh01
		4	fl_21	scaffold_15	154599	G	A	222	49	sh04
		3	fl_15	scaffold_15	174261	G	A	222	93	sh04
		3	fl_17	scaffold_15	174826	G	A	222	80	sh04
		4	fl_10	scaffold_15	177938	G	A	229	106	sh04
		4	fl_18	scaffold_15	178146	G	A	222	66	sh04
		4	fl_9	scaffold_15	179368	G	A	222	93	sh04
		4	fl_8	scaffold_15	179768	G	A	229	96	sh04
		2	fl_14	scaffold_17	500350	T	G	263	118	sh01
		3	fl_15	scaffold_2	491100	G	C	292	104	sh01
		1	fl_20	scaffold_2	2494297	C	G	213	108	sh04
		4	fl_9	scaffold_3	2552112	T	A	206	88	sh04

4	f1_8	scaffold_3	2567406	T	A	210	109	sh04
4	f1_18	scaffold_3	2594192	T	A	206	93	sh04
4	f1_8	scaffold_3	2819079	T	C	191	105	both
4	f1_18	scaffold_3	2844862	T	C	188	62	sh04
4	f1_19	scaffold_5	1103054	G	A	285	77	sh01
4	f1_21	scaffold_5	2319087	C	T	276	60	sh01
4	f1_9	scaffold_5	2328713	C	T	276	70	sh01
3	f1_24	scaffold_5	2334814	C	T	276	94	sh01
3	f1_11	scaffold_5	2336170	C	T	278	91	sh01
3	f1_17	scaffold_5	2337783	C	T	276	87	sh01
4	f1_8	scaffold_5	2340739	C	T	278	99	sh01
4	f1_18	scaffold_5	2340793	C	T	276	91	sh01
4	f1_19	scaffold_5	2367374	C	T	276	80	sh01
1	f1_16	scaffold_5	3022349	T	C	205	67	sh04
1	f1_3	scaffold_5	3024603	T	G	254	102	sh01
1	f1_26	scaffold_5	3032079	G	C	16	40	sh01
3	f1_17	scaffold_5	3052354	A	C	52	76	both
4	f1_21	scaffold_5	3127152	A	G	290	38	sh01
4	f1_9	scaffold_5	3132914	A	G	290	82	sh01
3	f1_24	scaffold_5	3143332	A	G	290	95	sh01
4	f1_18	scaffold_5	3146823	A	G	290	95	sh01
3	f1_17	scaffold_6	743325	G	A	232	82	sh04
3	f1_15	scaffold_6	783897	G	A	232	108	sh04
4	f1_10	scaffold_6	793073	G	A	234	89	sh04
4	f1_21	scaffold_6	820843	G	A	232	48	sh04
4	f1_9	scaffold_6	1580141	T	C	71	24	sh01
4	f1_18	scaffold_7	133318	A	C	329	85	sh01
4	f1_9	scaffold_7	134151	A	C	329	95	sh01
4	f1_10	scaffold_7	144618	A	C	335	126	sh01
3	f1_11	scaffold_7	145227	A	C	335	69	sh01
1	f1_3	scaffold_9	723876	C	T	223	118	sh04
Homo- F2	C14	scaffold_1	401077	C	T		19	both

		ID	scaffold	position			%	depth	type
zygous		C26	scaffold_10	136363	G	A		16	both
		C31	scaffold_10	136405	G	A		12	both
		C5	scaffold_10	476546	C	G		14	both
		C7	scaffold_10	477023	C	G		22	both
		C12	scaffold_10	477417	C	G		49	both
		C28	scaffold_10	477464	C	G		20	both
		C20	scaffold_10	477475	C	G		10	both
		C22	scaffold_10	477575	C	G		24	both
		C17	scaffold_10	477697	C	G		21	both
		C34	scaffold_10	477735	C	G		34	both
		C11	scaffold_5	343804	G	C		58	both
		C20	scaffold_5	1476007	A	C		16	both
		C12	scaffold_6	836323	T	G		29	both
		C31	scaffold_6	836428	T	G		16	both
		C28	scaffold_6	836737	T	G		14	both
		C26	scaffold_6	837699	T	G		27	both
	BC	C34	scaffold_6	837810	T	G		23	both
		Z22	scaffold_10	860	C	G		60	both
		Z35	scaffold_10	860	C	G		62	both
		Z35	scaffold_16	505272	T	G		88	both
		Z12	scaffold_20	8751	C	T		22	both
		Z1	scaffold_20	8945	C	T		27	both
		Z14	scaffold_20	8947	C	T		43	both
		Z19	scaffold_20	8947	C	T		37	both
		Z14	scaffold_3	3035766	C	T		55	both
Hetero-zygous	F2	C14	scaffold_1	3481640	A	T	32	26	fl_26
		C11	scaffold_1	3481931	A	T	32	27	fl_26
		C11	scaffold_3	2273340	T	G	63	44	fl_26
		C5	scaffold_3	2459104	A	G	100	13	fl_7
		C26	scaffold_3	2462087	A	G	100	61	fl_7
		C14	scaffold_3	2462236	A	G	100	31	fl_7
		C33	scaffold_3	2462519	A	G	100	19	fl_7
		C15	scaffold_3	2462811	A	G	100	18	fl_7

C28   scaffold_3   2463615 A   G   100   20   fl_7