Fungi contain multitudes: Tempo and modes of evolution in the purplepore bracket fungus *Trichaptum*

Dabao Sun Lü

Dissertation presented for the degree of
Philosophiae Doctor (PhD)
2023

Department of Biosciences
Faculty of Mathematics and Natural Sciences
University of Oslo

oslo mycology group

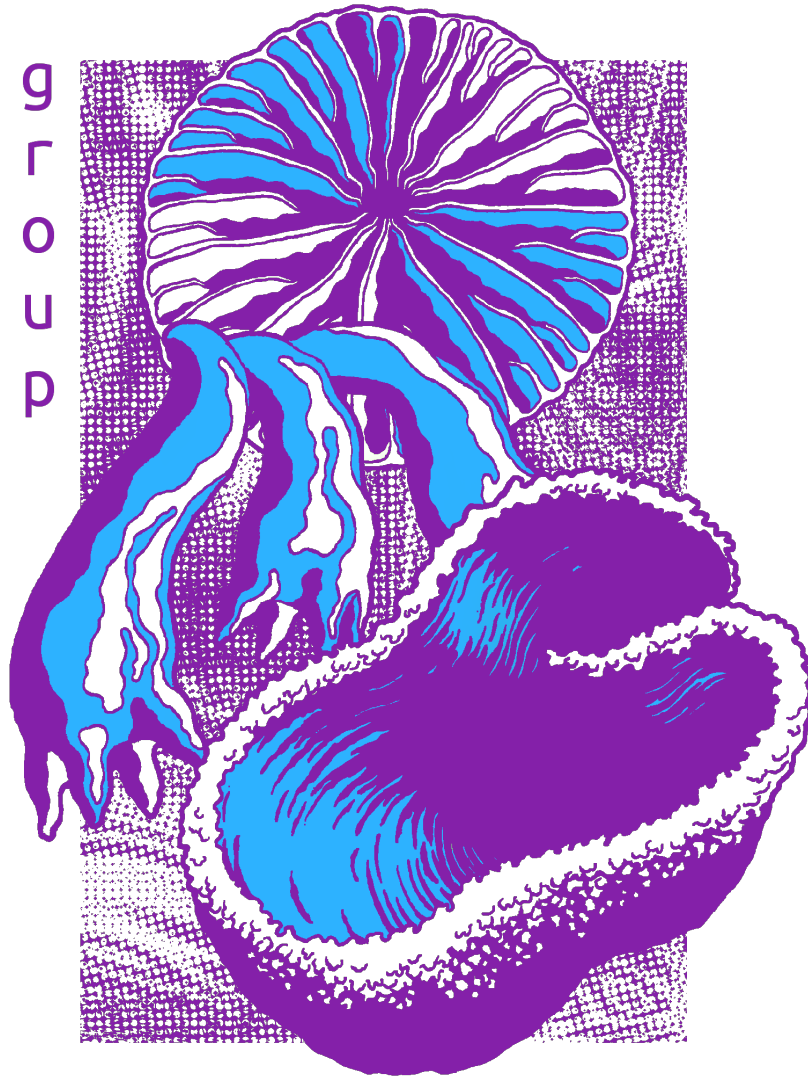*Dedicated to my friends Brittny and Thomas,*
who shared my appreciation for fungi,
and to my Godfather Bend,
who showed me how to appreciate life at large.

Acknowledgements

If it´s so that it takes a village to raise a child, then I think, it must take both a village and a faculty to raise a PhD. I will focus my gratitude here towards the faculty part. But to everyone else, all my friends and family: I hope you realize that your love is what has kept me together, and that you therefore are an integral part of this work. Foremost then, I would like to thank my main advisor Inger for giving me the opportunity to embark on this undertaking, and for seeing it safely ashore. Like the best of captains, you have set the main course and steered me away from the most dangerous waters, but at the same time given me room to navigate and explore along the way. I thank you for the guidance and foresight your keen mind has provided, all the trust and support disseminating from your kind nature, and perhaps most crucially, for understanding and knowing how to keep at bay my more obsessive and uncritical impulses. Next in line I wish to thank each of my co-supervisors: Håvard; once upon a time more than a decade ago, you were the one who introduced this mushroom enthusiast to mycology, and your passion and vision for the subject have been vital for my development as a mycologist and the fruition of this project. Thank you for your continued encouragement and invaluable help with figure design. Sandy; you were the one who took me on to the mycology group in Oslo, and I have learned a lot from working for you. I admire your deep knowledge of mycology and ability to think outside the box. Thank you for being so genuine and bringing more warmth to our latitudes. Peris; your zeal has been an inspiration and a fresh breath, and we all miss you! Thank you for always being up to discuss and troubleshoot analyses, and for introducing to me important principles on open science and reproducible research. Jørn Henrik; thank you for remaining well-grounded and being an immutable force of calm, and for providing me with on point help with population genetic analyses at the most critical moments. To Mark; I feel privileged to have been able to regularly consult someone with your level of expertise on evolutionary genomics. Thank you for generously hosting me in Nottingham during a difficult time of the pandemic, and for giving me the opportunity to interact with some of the greatest minds in evolutionary biology. It became an important period for ripening many of my ideas, and I look back at my time there with fondness. Glenn-Peter Sætre; learning and eventually teaching evolutionary genetics for you while trying to get a grip on your more challenging study questions has been decisive for my own understanding of the subject. I respect your integrity and courage to speak up for scientific truth, and thank you for everything you and your book have taught me.

Table of Contents

List of Papers

Paper I.

Large-scale fungal strain sequencing unravels the molecular diversity in mating loci maintained by long-term balancing selection. David Peris, <u>Dabao Sun Lu</u>, Vilde Bruhn Kinneberg, Ine-Susanne Methlie, Malin Stapnes Dahl, Timothy Y. James, Håvard Kauserud, Inger Skrede. (2022) *PLOS Genetics.*

Paper II.

Introgression between highly divergent fungal sister species. Vilde Bruhn Kinneberg, <u>Dabao Sun Lü</u>, David Peris, Mark Ravinet, Inger Skrede. *Manuscript accepted in Journal of Evolutionary Biology.*

Paper III.

Reticulate evolution and rapid development of reproductive barriers upon secondary contact pose challenges for species delineation in a forest fungus. <u>Dabao Sun Lu</u>, David Peris, Jørn Henrik Sønstebø, Timothy Y. James, Loren H. Rieseberg, Sundy Maurice, Håvard Kauserud, Mark Ravinet, Inger Skrede. *Manuscript intended for Nature Ecology and Evolution.*

Paper IV.

Postglacial history of a widespread forest fungus in Europe: Migration out of multiple refugia followed by admixture. <u>Dabao Sun Lu</u>*, Ine-Susanne Hopland Methlie*, Jørn-Henrik Sønstebø, David Peris Navarro, Sundy Maurice, Håvard Kauserud, Inger Skrede. *Manuscript intended for Molecular Ecology.*

Summary

Evolution has brought to life an immense diversity of forms, lifestyles and species in the fungi. The presence of fungi pervades every ecosystem, yet many aspects of their biology remain obscure to us as they spend most of their lives hidden within substrates and elude direct observation. Speciation, the process by which life diversifies, can unfold in a multitude of ways, but how this happens in fungi is largely unknown. Thus, the broader aim of this thesis is to improve our understanding of speciation in fungi. To this end, whole genome DNA sequencing and laboratory experiments have been conducted on the wood decay fungus *Trichaptum abietinum* and allies. Together the four papers in this thesis cover the speciation continuum from undifferentiated populations to completely reproductively isolated lineages. Paper I and II investigate evolutionary dynamics across longer time spans that have happened within and between *T. abietinum* and the sister clade *T. fuscoviolaceum*, whereas paper III considers the processes shaping the diversity within the *T. abietinum* morphospecies complex. Finally, paper IV narrows down to the more recent changes in *T. abietinum* occurring within Europe, and examines local changes in the genome.

The development of reproductive barriers is pivotal to the speciation process, and a recurrent test throughout this thesis is the ability of different isolates to mate with each other. Paper I examines the genetics of the mating loci in *Trichaptum*. The peculiar genetic system which controls mating in mushrooms and their allies is characterized by an enormous allelic diversity at two unlinked mating loci, where typically only different alleles at both loci result in compatible matings. About half of the sequenced *Trichaptum* genomes were annotated for the mating alleles, and corresponding mating tests between the isolates were carried out to verify the *in silico* predictions on the ability to mate. By calculating different population genetic measures on diversity at the mating loci and comparing it to a genome wide baseline, this study is the first to leverage population genomic data to demonstrate the action of balancing selection at the mating genes. In paper II mating experiments are carried out between two different populations each of *T. abietinum* and *T. fuscoviolaceum* to investigate hybridization between these two morphospecies. There has been speculation on whether they hybridize as they have overlapping distributions and can also intergrade morphologically, but the mating experiments found the two morphospecies to be completely intersterile. Nonetheless, D-statistic analyses suggest that introgression events likely occurred in the past, indicating that these taxa did not split cleanly, but that some degree of geneflow has been part of the lineage diversification here.

The findings of introgression are supported by demographic modelling, which indicated geneflow through an unsampled or extinct population, and these analyses showcase the sensitivity of whole genome analyses and their applicability even in cases of incomplete sampling.

Incipient levels of divergence can be more informative for studying speciation processes, and in paper III the focus is confined to the *T. abietinum* morphospecies where reproductive barriers are more recent, and in many cases incomplete. Including all the sequenced genomes of *T. abietinum*, the population structure is mapped with phylogeographic analyses and intersterility groups are delineated with systematic mating tests. The fitness of the hybrids from the mated cultures were assessed by measuring their ability to decay wood, and revealed a strong negative correlation between genetic distance between isolates and the fitness of their hybrids. In comparison, a weaker correlation was observed between increase of genetic distance and ability to mate. These trends and the occurrences of stronger mating barriers in sympatry are interpreted as evidence of reinforcement selection. Moreover, demographic modelling, coalescence and phylogenetic network analyses detected evidence of past geneflow between reproductively isolated lineages in North America and Asia. Viewed together, these results suggest that genetic incompatibilities from independent divergence initiated the development of reproductive barriers in *T. abietinum*, but that reinforcement selection can complete the divergence process by selecting for strong pre-mating barriers. The postglacial recolonization of animals and plants, including that of conifer trees, has been extensively studied in Europe. In comparison, we have limited information about postglacial migration of fungi. In paper IV, the finer scale sampling of *T. abietinum* in Europe is leveraged to investigate how its current genetic structure can be explained in light of postglacial history. Two clearly differentiated lineages are found in the Mediterranean basin and Eastern Europe and Fennoscandia, and these are connected with glacial survival in southern and eastern refugia, respectively. These lineages form an admixture zone in central Europe, and closer inspection of the genomes of the admixed individuals reveal that they are not a homogenous mix of their parental lineages. Rather, the second half of their largest chromosome exclusively come from the Boreal lineage, indicating strong selection or genomic incompatibilities. In addition, a fourth Atlantic lineage was detected in western Europe. Analysis of linkage decay patterns indicate that this lineage is of older origin, and its origin and glacial survival is connected to a western refugia.

In sum, these studies demonstrate how multiple evolutionary forces across different temporal and genetic scales interact to shape the current population structure and genomic landscape in *Trichaptum* species. Moreover, the prevalence of introgression observed in Paper II-IV suggests that this process is common among these fungi.

Sammendrag

Et enormt mangfold av former, levesett og arter av sopp har oppstått gjennom evolusjonens løp. Soppenes tilstedeværelse gjennomsyrer ethvert økosystem, men det er mye ved deres biologi vi fremdeles ikke kjenner til da de tilbringer mesteparten av livet sitt nedgravd i substrater og dermed ikke lar seg observere. Artsdannelse, som er prosessen der livet diversifiserer, kan utarte seg på atskillige vis. Hvordan artsdannelse utspiller seg i sopp er i stor grad ukjent, og det overordnede målet med denne avhandlingen er å øke forståelsen vår på dette feltet. For overnevnte formål har helgenom DNA-sekvensering og eksperimenter blitt utført på den vedboende soppen *Trichaptum abietinum* og beslektede arter. Sammen dekker de fire artiklene i denne avhandlingen hele spekteret av artsdannelsesprosessen fra udifferensierte populasjoner til fullstendig reproduktivt isolerte linjer. Artikkel I og II undersøker evolusjonære prosesser som har funnet sted over lengre tid innad og mellom *T. abietinum* og søstergruppen *T. fuscoviolaceum*, mens artikkel III tar fatt på prosessene som har formet diversiteten i den morfologiske arten *T. abietinum*. Artikkel IV smalner inn til nylige endringer som har skjedd i Europa i *T. abietinum*, og undersøker disse endringene lokalt i genomet.

Dannelsen av reproduktive barrierer står sentralt i artsdannelsesprosessen, og evnen til ulike isolater til å pare seg med hverandre benyttes som en gjennomgående test i denne avhandlingen. Artikkel I undersøker sexgenene i *Trichaptum*. Det spesielle genetiske systemet som kontrollerer sex i flertallet av storsoppene karakteriseres av en enorm allelisk diversitet på to ukoblede gener. Kompatible paringer krever som regel ulike alleler på begge sexgenene. Rundt halvparten av de sekvenserte genomene av *Trichaptum* ble annotert for sexgenene, og krysningstester mellom utvalgte isolater ble utført for å verifisere *in silico* prediksjonene på evnen til å pare seg. Ved å regne ut ulike populasjonsgenetiske mål på diversitet i sexgenene og sammenlikne de mot et genomisk gjennomsnitt, er dette studiet det første til å ta i bruk populasjonsgenomiske data til å demonstrere hvordan balanserende seleksjon opprettholder diversiteten i disse genene.

I artikkel II blir krysningseksperimenter utført mellom *T. abietinum* og *T. fuscoviolaceum* for å undersøke hybridisering mellom dem. Det er blitt spekulert i hvorvidt disse to morfologiske artene kan hybridisere da de har overlappende utbredelser og til tider også overlappende morfologi. Krysningseksperimentene mellom to ulike populasjoner av *T. abietinum* og *T. fuscoviolaceum* viser imidlertid at de er fullstendig intersterile seg imellom. D-statistiske

analyser viser likevel at introgresjonshendelser trolig har skjedd i fortiden, hvilket tyder på at disse artene ikke splittet tvers av, men at en viss grad av genflyt har vært en del av diversifiseringsprossesen her. Introgresjon støttes også av demografiske modelleringer, som angir at genflyt skjedde gjennom en ukjent eller utdødd populasjon. Disse analysene viser sensitiviteten til helgenomanalyser, og anvendbarheten deres selv i tilfeller der ikke alle populasjonene er inkludert i studiet.

Begynnende nivåer av divergens kan være mer informative for å trekke slutninger om artsdannelsesprosessen, og artikkel III fokuserer på artskomplekset *T. abietinum* der reproduktive barrierer mellom grupper er mer nylige, og i mange tilfeller ufullstendige. Ved å inkludere alle de sekvenserte genomene av *T. abietinum* kartlegges populasjonsstrukturen med fylogeografiske analyser, mens systematiske krysningstester brukes for å avgrense intersterilitetsgrupper. Hybridenes fitness i de kryssede kulturene ble vurdert ved å måle evnen deres til å bryte ned trevirke, og dette avdekket en sterk negativ korrelasjon mellom genetisk distanse mellom isolater og fitnessen til hybridene deres.

Den postglasiale rekolonialiseringshistorien for dyr og planter, inkludert bartrær, er blitt grundig studert i Europa. Kunnskapen om sopp på dette feltet er imidlertid svært begrenset. Artikkel IV drar fordel av fin-skala innsamlingen av *T. abietinum* i Europa for å undersøke hvordan den nåværende populasjonsstrukturen her kan forklares i lys av postglasial historie. To klart differensierte linjer ble funnet i Middelhavsområdet og Øst-Europa pluss Fennoskandia, og disse kobles til istidsoverlevelse i henholdsvis sørlige og østlige refugia. Disse linjene danner en hybridsone i Sentral-Europa, og nærmere undersøkelse av genomene til hybridene viser at de ikke er en homogen blanding av foreldrelinjene. For eksempel så stammer siste halvdel av det største kromosomet deres eksklusivt fra den boreale linjen, noe som enten kan indikere sterk seleksjon eller genomiske inkompatibiliteter. En fjerde atlantisk linje ble funnet i Vest-Europa, og analyse av koblingsulikevektmønstre indikerer at denne linjen er eldre enn de sentraleuropeiske hybridene, og dens overlevelse knyttes til et vestlig refugium.

I alt viser disse studiene hvordan flere evolusjonære krefter på tvers av ulike tids- og genetiske skalaer har virket sammen for å danne det genomiske landskapet og den nåværende populasjonsstrukturen blant *Trichaptum* arter. Hyppigheten av introgresjon mellom linjer som ble observert i artikkel II-IV indikerer videre at denne prosessen er vanlig blant disse soppene.

"All visible objects, man, are but as pasteboard masks. But in each event—in the living act, the undoubted deed—there, some unknown but still reasoning thing puts forth the mouldings of its features from behind the unreasoning mask. If man will strike, strike through the mask!" (Ahab, Herman Melville's *Moby-Dick*)

Introduction

The units of life

No matter where we look in nature, whether it is the huge or the very small, we find discrete discernable units. Some of these units are similar to each other, and we recognize that they are of the same kind. The planets on the night sky differ from the stars, atoms can be classified according to their composition of elementary particles. So it appears to be for life on earth, and our propensity to categorize and name the diversity of life has found expression in every known culture and tribe (Mayr, 1982). The formal system of classification in biology that we use to this day was introduced by Carl Von Linné, and in this hierarchical system, the species represents the basic unit of diversity. "God created, Linnaeus set in order" he famously declared (Blunt, 2004), and following his example much of the focus in biology became directed towards describing species and classifying them. But the living differs from the inanimate matter it consists of: As Darwin and Wallace came to realize, the diversity of life is not a given constant, but rather it changes through time and adapts to the environment by the means of natural selection (Wallace, 1858, Darwin, 1859). In light of this discovery, the history of life can be pictured as a giant tree, springing forth from a common origin at the base and then branching out successively through time, each branch tip representing a species. But what makes new species bud on the tree of life, and how does this happen? Evolution by natural selection is a gradual process, and thus it is not evident how it can create the distinct and discontinuous units we recognize as species. Speciation research seeks to answer this apparent paradox. This thesis takes a closer look at one very specific twig within the fungal limb in the tree of life, with the aim of gaining more insight about the processes underlying diversification here. To do so, I have examined evolution in genus *Trichaptum* at multiple scales, from the relatively recent changes occurring within Europe to changes that have happened between continents and distinctly differentiated species across larger time spans. To appreciate how the small and the large in my four separate studies fit and work together, I will begin by introducing basic concepts on evolutionary change and speciation, before I move on to consider these concepts in fungi and introduce my study system. A separate section is then dedicated to the methods I have applied before I present my results and discuss them.

Boundaries help us discern and recognize units. A gene is bounded by a start and stop codon, membranes delimit single cells and keep their contents together, an individual tree is encased in bark that sets it apart from its surroundings. What are the boundaries between species? To

understand the heart of this matter we must consider another special property of life; reproduction. It is no coincidence that the first paper of my thesis is centered around this subject. Species consists of individuals, though sooner or later individuals perish. But before they do so, their DNA may be passed on to new generations through reproduction. It is the DNA that lives on in the unbroken chain that extends from each present individual back to the dawn of life. And it is in the DNA where each individuals´ characteristics are encoded. When individual members of a species reproduce with each other, a mixing of DNA happens, resulting in a combination of the parents in their offspring. In order for a species to be distinct and discernable from other species then, it follows that its DNA must remain distinct. Consequently, individuals within a species should reproduce only with each other and not with members of other species. As such, a species can be understood as reproductively isolated units separated by reproductive barriers. As we shall see, reproductive barriers can manifest in many ways and during different stages of an organism's life cycle. To complicate matters further, it turns out that these barriers need not be absolute, but they can be rather porous. Consequently, species are, upon closer inspection, not encircled with solid lines, and finding a universal species definition that fits all of life's endless forms has so far proven to be impossible (Coyne and Orr, 2004). For that matter, it turns out that genes can overlap, cell membranes are porous, and the inside of a tree trunk teems with microbial life. Even the tree of life metaphor I laid out has holes; branches can grow back into each other, and thus the grand tree may appear more like thicket. Life is messy, and after all there is no reason for why it would conform to our disposition for order. This unruliness seems to be particularly pronounced in fungi, as some of the findings in my thesis will demonstrate.

Before I move on to speciation there are a few more general concepts on evolution that need to be introduced. So far I have focused on the similarity between individuals in a given species, but it is the variation between individuals within a species that drives evolution by natural selection (Darwin, 1859). No two individuals are the same, and this fact made such an impression on the natural historian Buffon that he wrote "Nature knows no species, genera, and other categories; it knows only individuals" (Mayr, 1982). What constitutes this variation and how does it arise? The answer sits at the DNA, which is the very fabric of evolutionary change. Each time DNA copies itself, it is prone to make a few errors, it mutates. A few of these mutations may happen to confer an advantage to the individual carrying it that increase the chances of survival or reproductive output. In that case we say they have greater fitness, and consequently we expect these mutations to increase in frequency over generations. In the same

vein, the original version of the gene that did not receive the beneficial mutation is expected to decrease in frequency. The different variants of the genes are called alleles, and they are subject to competition for survival just as individuals (Dawkins, 2006). We say that an allele has reached fixation when it is the only one around, having obtained genetic monopoly. But life is not fair, and selection is not the only force that shapes patterns of DNA sequence variation. Because populations are finite in size, there is an element of luck, and alleles can reach fixation or be lost through sheer chance (Hartl and Clark, 2007). Moreover, it turns out that most mutations have little to no effect on fitness (Kimura, 1991), and can be considered as neutral. Thus evolution itself at the DNA level is a largely stochastic process that proceeds even in the absence of natural selection (Kimura, 1983). As I will return to in the methods section, this property of DNA sequence evolution makes it possible to make historical inferences on a species´ past.

The tempos and modes of speciation

Reproductive barriers maintain, and in many cases help us define species. But how did these barriers come to be in the first place? Understanding the origin and assembly of reproductive barriers goes a long way towards understanding speciation as the two are intimately linked. Reproductive barriers can be classified according to whether they occur prior to or after fertilization and the subsequent formation of a zygote. Another way to classify reproductive barriers is to determine whether they are extrinsic, that is external to the physical organism but embodied in behavior or the ecology, or intrinsic, that is, inherent to the physiological traits of the organism. To trace the origin of reproductive barriers, a natural place to begin is the population, the level of biological organization that sits between individuals and species. A population can broadly be defined as a group of interbreeding individuals in a species (Hartl and Clark, 2007). Resources and habitats in nature are clustered, and species tend to be subdivided accordingly into individual populations. Nonetheless, most lifeforms have the ability to disperse at some stage during their life cycle, and a population is rarely completely reproductively isolated. Some number of individuals migrating and reproducing across populations can be expected. Consequently, there is movement, or flow, of genetic material, and this geneflow has a homogenizing effect on the genetic differences between populations.

Geological processes ensure that the environment and habitats are constantly changing, and distributions of populations also change accordingly as evidenced by pollen and fossil records (Webb, 1995). For instance, landmasses move, and albeit slowly, populations can literally drift

apart on separating continents. Another example are the giant ice sheets that emerge with the ice ages, dividing populations and acting as a wall to geneflow between them. In the absence of reproduction with each other, the DNA of these populations can diverge both neutrally and as a consequence of adapting to different conditions. At any rate, they evolve independently, and given sufficient time the two populations might have become distinct enough for a taxonomist to recognize them as different species. This geographic mode of speciation is called allopatric speciation. The reproductive isolation here is caused by extrinsic geographic barriers, and reproductive barriers per se are therefore not necessary to maintain the distinctness of species. But that is only true as long as they don´t meet.

Geographic barriers are not permanent, for instance, the ice sheets that kept populations apart may melt away during warmer periods, or the lowering of sea levels during glaciation can connect landmasses that were previously separated. Even mountains erode away. Moreover, the globalization of our own species has also in effect broken down geographic barriers for many other lifeforms as we bring them around, either intentionally or by accident. What happens when populations that have diverged in allopatry are brought back together? The outcome of this secondary contact will depend on how far the populations have drifted apart. In cases of low divergence, they may simply fuse back together e.g. (Seierstad et al., 2013, Rudman and Schluter, 2016). In cases where they don´t, reproductive barriers can be inferred. The Bateson-Dobzhansky-Muller (BDM) model of reproductive isolation by genetic incompatibility is a simple but widely accepted model for how evolution within populations can lead to reproductive barriers between populations in the absence of geneflow (Coyne and Orr, 2004). This model pictures two or more genes that evolve separately in allopatric populations that each become fixed for new alleles in the two populations. No gene works in a vacuum, but rather it must interact in a carefully orchestrated symphony with other genes to generate and maintain an organism. The new alleles have never been tested against each other, and as a consequence they may be incompatible. In that case, a reproductive barrier will exert itself between these populations when they meet again because hybrids will be inviable or have low fitness (Orr, 1996). BDM incompatibilities are therefore postzygotic barriers as they take effect after fertilization when the genes from the two parental population have come together. The model can be extended to more than two loci, which is probably more realistic (Palopoli and Wu, 1994). Comparative studies on animals e.g. (Coyne and Orr, 1989, Presgraves, 2002, Sasa et al., 1998) and plants (Moyle et al., 2004) have shown that postzygotic reproductive barriers increase with divergence time, consistent with the accumulation of BDM

incompatibilities. Reproductive barriers caused by BDM incompatibilities can be considered a by-product of other evolutionary changes, and not as adaptive in their own right. Speciation can simply happen as a consequence of populations evolving separately for long enough.

Although it can be shown that BDM incompatibilities are expected to accumulate exponentially with time (Orr and Turelli, 2001), the strength of these incompatibilities must be weak given that reproductive isolation in the form of complete hybrid inviability typically takes on the order of millions of years to evolve in the animals where it has been examined (Price and Bouvier, 2002, Fitzpatrick, 2004) But natural selection can, in theory, also work directly against the formation of hybrids between diverged populations. By favoring traits that reinforce prezygotic reproductive barriers, unfit or inviable hybrids never get to become a problem (Liou and Price, 1994, Kirkpatrick and Ravigné, 2002). This process is known as reinforcement, and could explain the pattern observed in many taxa of stronger pre-mating barriers between closely related species with overlapping ranges e.g. (Coyne and Orr, 1997, Howard, 1993), and also why divergence times between sister species in sympatry tend to be shorter (McCune and Lovejoy, 1998). It is important to note, as part of the critique against reinforcement theory has brought to attention, that such patterns can arise from a number of other processes which I will return to in the discussion. Consequently, the theory has been somewhat controversial historically, but a number of convincing examples are known today, reviewed in Servedio and Noor (2003), although the frequency and importance of reinforcement is still being debated. BDM incompatibilities and reinforcement are not mutually exclusive, on the contrary, it has been suggested that reinforcement could be an important pathway for completing allopatric speciation where BDM incompatibilities already exist but some level of maladaptive hybridization still occurs (Servedio, 2004).

Could geographic isolation alone have generated the enormous diversity of some taxa we observe, such as the 400,000 described species of beetles? It turns out that reproductive barriers can also arise without spatial segregation, and in some cases instantaneously. In plants there are many known cases of polyploidization events where the accidental doubling of chromosome number can lead to instant reproductive isolation with the parental population(s) and the origin of a new species (Wood et al., 2009). In many cases such polyploid speciation events arise from hybridization between species, but hybridization without chromosome doubling can also result in hybrids that display reproductive barriers with both parental species, referred to as homoploid hybrid or recombinational speciation (Abbott et al., 2010). In ecological speciation, adaptation

to different niches result in extrinsic reproductive barriers and strong reproductive isolation. For instance, consider parasites that come to live on different hosts, the populations will in effect be separated and rarely meet and exchange genes even if they appear to have overlapping distributions and no intrinsic barriers that prevent mating. And even if they were to mate, the hybrids would likely be maladapted. As such, ecological barriers can be both pre- and postzygotic (Schluter, 2009). A more special case occurs if populations of these parasites adapted to different hosts while they were freely interbreeding. Or to reframe the question with regards to population genetics: can reproductive barriers arise between populations in the face of strong geneflow? Theoretical work suggests that there are special cases, often involving differentiation in resource use or sexual selection, where selection can be sufficiently strong to outweigh the homogenizing effects of geneflow and cause differentiation (Bolnick and Fitzpatrick, 2007). But this mode of speciation in sympatry is probably not common and it remains hard to prove unequivocally (Coyne and Orr, 2004). The consensus holds that allopatric speciation is the most common mode of speciation (Coyne and Orr, 2004). Nevertheless, as we have seen, geneflow can be reduced by both intrinsic and extrinsic barriers, that can occur both prior to or after mating, and consequently there is a plethora of ways in which new species can arise.

The speciation process is most commonly classified according to the geographic scenario in which it plays out. But these are conceptual categories, and in actuality, it is likely that speciation more often progresses as an intermediate or combination of these extremes (Butlin et al., 2008). For instance, parapatric speciation describes a scenario where new species evolve in contiguous, yet spatially segregated habitats (Coyne and Orr, 2004). Furthermore, different reproductive barriers can work together, and the coupling of reproductive barriers may result in stronger reproductive isolation (Butlin and Smadja, 2017). Nonetheless, geography is undoubtedly a useful framework to understand speciation, and phylogeography, the study of genetic structure across time and space has proven informative (Hewitt, 2001). The Pleistocene glacial cycles that begun 2.4 million years ago, with periodic advancement and retreat of ice sheets on the northern hemisphere (Batchelor et al., 2019), isolated populations into allopatric refugia and would have provided opportunity for secondary contact when the diverged populations were brought together again during the interglacials. A number of phylogeographic studies have shown how the Pleistocene glacial cycles have had a great impact on the genetic substructure of plant and animal species (Taberlet et al., 1998, Hewitt, 2000), in many cases resulting in hybrid zones (Hewitt, 1999, Swenson and Howard, 2005, Li et al., 2016), and

presumably also speciation (Hewitt, 2000, Kadereit and Abbott, 2021). However, some researchers have argued that the repeated colonization of deglaciated regions from the same refugial source population in the interglacial periods would instead work against genetic differentiation and prevent speciation (Baker, 2008). A compilation of data by Avise et al. (1998) of mitochondrial DNA of 140 different vertebrate species pairs found that more than half of these speciation events can be dated to the Pleistocene using molecular clocks. However, as noted by Hewitt (2000), the genetic divergence between these sister groups has not progressed significantly more in the Pleistocene than in the Pliocene, its preceding epoch. In spite of the difficulties in making broad generalizations on speciation patterns in the Pleistocene, the glacial cycles have undoubtedly left an imprint on the genetic structure of extant populations on the Northern Hemisphere and played a part in speciation for many individual taxa.

Speciation is a slow process when seen from the human perspective. A study by Hedges et al. (2015) that calibrated the tree of life to time, found that new tips on the tree, that is species, on average take ~2 million years to form. In order to study the speciation process, a comparative approach has been employed on groups where different levels of divergence and reproductive barriers exist. But reproductive barriers are not always easy or even possible to quantify, and there are uncertainties tied to estimating divergence time from DNA sequence data alone. Moreover, identifying the genetic underpinnings of reproductive barriers have been out of reach up until recently, or extremely tedious at best, involving extensive crossing schemes and linkage mapping (Coyne and Orr, 2004). Consequently, speciation research has largely revolved around a few model systems. However, recent improvements in DNA sequencing technology have changed the research and launched what has become a burgeoning field of "speciation genomics" (Campbell et al., 2018). Genome scale sequence data of population samples, should in principle, make it possible to characterize the genetic underpinnings of the speciation process (Jiggins, 2019). Connecting phenotype to genotype can be considered an overarching goal in contemporary biology, and much of the initial focus in speciation genomics has been on identifying speciation genes, that is specific genes that cause reproductive isolation. Localizing genetic differentiation between species pairs were therefore a major focus in these studies, as such "genomic islands" of differentiation were thought to contain the speciation genes (Nosil and Schluter, 2011). However, it has been shown that these regions of high genomic differentiation may in fact result from other processes such as variation in recombination rate (Noor and Bennett, 2009), and the validity of a purely genic view of speciation has also been

called into question as intrinsic features in genome architecture such as inversions, gene duplications, recombination patterns and higher order genomic architecture have been implicated in speciation (Campbell et al., 2018).

On the whole, the development of speciation research can be seen as progressing from patterns to processes, and now to mechanisms, with new questions arising along the way. Notwithstanding this progress, our understanding of patterns and processes in speciation is far from complete. Fortunately, the availability of population level genome data can also greatly aid and improve mapping the patterns of lineage diversification and inferences of the processes that shaped them. DNA sequencing of archeological samples represent a direct time capsule that provides insight on the genetic variation and structure of past populations (Librado et al., 2021, Slatkin and Racimo, 2016), but in the cases where sequencing of past samples are not possible, a great deal of information can still be inferred from population genomic data. I will outline some of these approaches in more detail in the methods section.

All the speciation studies discussed so far have been on animals or plants, that is no coincidence, as these dominate the literature. Kingdom specific speciation patterns have emerged from this body of work, such as polyploid speciation in plants, or reproductive barriers in animals related to the heterogametic sex (Haldane´s rule) (Coyne and Orr, 2004). Could there also be such patterns in fungi? To evaluate this question I will now present fungi as group and provide a brief review of existing speciation studies on fungi. I then zoom in on the phylum Basidiomycota where mating and phylogeographic studies are discussed at some length before I consider these in *Trichaptum*, the study organism in this thesis.

The secret lives of fungi

Fungi can be viewed as both microbes and macro-organisms. Many fungi are unicellular, and the multicellular taxa typically form a mycelium, a diffuse filamentous growth form consisting of strands that are only a single cell layer thick and hardly visible for the naked eye. Yet, mycelium from a single individual can collectively cover large areas, and in most species in the phylum Basidiomycota they aggregate to form macroscopic fruit bodies, where the largest ones can reach several hundred kilograms (Dai and Cui, 2011). The honey mushroom is the biggest recorded organism on earth, the mycelium of a genetic individual have been estimated to cover almost 1000 hectares of forest (Ferguson et al., 2003). Nonetheless, the presence of this

organism may not be apparent even if one were to walk around the forest with a magnifying lens. The majority of the life of these fungi is spent buried within their substrates, and only when conditions are right, do they form fruit bodies that spread billions of spores. As fungi are largely hidden and elude observation, they have been misunderstood historically, and many aspects of their biology remain obscure to this day. Linnaeus classified fungi as plants, and it was not until 1959 that they were officially recognized as a separate kingdom (Whittaker, 1959), though it was already proposed in 1897 by the Norwegian mycologist Johan Olav Sopp (Høiland, 2003). Modern phylogenetic analyses have revealed that fungi are more closely related to us than to plants, and current estimates using fossil and genome data suggest that the common ancestor of fungi and animals lived 1.2 billion years ago (Berbee et al., 2020). Like animals, fungi are heterotrophic, but rather than ingesting food to process internally through a digestive tract, fungi excrete enzymes that break down and mobilize food externally. Accordingly, they have evolved an impressive ability to break down and feed on a wide array of substrates, including even tough materials such as solid wood (Fukasawa, 2021), and overall play an indispensable part of nutrient cycling in terrestrial ecosystems (Dighton, 1995). Fungi have also evolved many intimate associations with living lifeforms, to their detriment as pathogens, or benefitting them as mutualistic symbionts, and the whole spectrum in between (Peay et al., 2016).

Kingdom fungi are among the most diverse organismal groups on earth, and together with animals and plants they represent one of the three most successful multicellular groups (Chen and Wiens, 2021). Nonetheless, the attention they have received is modest, and does not stand in proportion to their enormous diversity. A consensus by Antonelli et al. (2020) counts 148000 described species, but the true diversity is certainly much higher due to their microbial and elusive nature, and the prevalence of cryptic species hiding under a common morphology. Current estimates of species number in fungi range from 1.5 to 7 million species (Hawksworth and Lucking, 2017, Blackwell, 2011). In spite of their diversity, most groups of fungi remain completely unchartered by speciation studies. The few fungi that have been more extensively studied include model organisms for eukaryote genetics, such as the yeast *Saccharomyces cerevisiae* (Louis, 2011) and the mold *Neurospora crassa* (Gladieux et al., 2020), or organisms that in some way threaten human societies, such as the plant pathogens *Zymoseptoria* (Feurtey et al., 2023) and *Magnaporthe* (Zhong et al., 2016) that jeopardize food security, the tree pathogen *Heterobasidion* (Garbelotto and Gonthier, 2013) that cause considerable losses for forestry, and the human pathogen *Cryptococcus* which can be fatal (Casadevall et al., 2017).

Disregarding their immediate impact on us, these systems are still interesting in their own right because they may have undergone rapid adaptation and speciation as a response to the recent changes brought about by human civilization such as the cultivation of crops or trees in monocultures. A considerable amount of genetic resources are therefore available for these systems, and detailed studies on the genetics underlying reproductive isolation between these and closely related species have been carried out. Altogether these studies on model fungi and pathogens have demonstrated that genetic incompatibilities (Dettman et al., 2008, Anderson et al., 2010, Giordano et al., 2018), chromosomal rearrangements (Charron et al., 2014, Yadav et al., 2020), hybridization (Leducq et al., 2016, Stukenbrock et al., 2012) and ecological specialization (Stukenbrock et al., 2011, Zaffarano et al., 2008) can be involved in fungal speciation. In addition to these systems, the plant parasite *Microbotryum* have also been established as a model to study fungal speciation (Gladieux et al., 2010). Excellent reviews have been written about fungal speciation by Kohn (2005) and Giraud et al. (2008), and the interested reader are referred to these for a more complete overview.

Most of the species in the abovementioned studies belong to the phylum Ascomycota, and it is not clear if the speciation mechanisms identified here apply to phylum Basidiomycota beyond those on *Heterobasidion, Microbotryum* and *Crypotococcus*. Moreover, the somewhat atypical life cycles of *Microbotryum* and Cryptococcus could also set them apart from the majority of forest fungi in Basidiomycota where populations may be structured differently. In the rest of my introduction on speciation in fungi I will focus on the class Agaricomycetes within Basidiomycota, as this is where the study organism of this thesis belongs. This class comprise the mushroom forming fungi, and make up roughly a quarter of the described fungal diversity (Sánchez-García et al., 2020). Despite the paucity of modern speciation research on Agaricomycetes, a great number of studies assessing reproductive barriers are available here since many of the saprotrophic species can readily be cultured and mated in the lab. It is possible to make inferences on the speciation processes by interpreting these mating studies in light of phylogeography. The majority of these findings are presented and discussed in Petersen and Hughes (1999), and summarized with regards to patterns of reproductive barriers in Gac and Giraud (2008) and modelled in Giraud and Gourbière (2012). In short, species descriptions in Agaricomycetes based on morphological similarity initially implied wide geographic distributions spanning multiple continents. Early work using simple molecular markers seemed to comply with this to some extent e.g. (James et al., 1999, Vilgalys and Sun, 1994), in line with Baas Beckings famous tenet on microbes that "everything is everywhere, but the

environment selects" (De Wit and Bouvier, 2006). This is intuitively alluring when considering how most agaricomycetes produce billions of air-borne spores that, in theory, can be carried for long distances with air currents, although the extent to which they do so remains uncertain (Golan and Pringle, 2017), and several studies indicate that they are dispersal limited (Peay et al., 2012, Norros et al., 2015). Cryptic and geographically structured diversity within agaricomycete morphospecies has since been detected by both DNA based methods and mating studies that have revealed pre-mating barriers within several morphospecies e.g. (Haight et al., 2019, Anderson et al., 1980, Vilgalys and Sun, 1994). Perhaps this should come as no surprise when considering how the morphospecies are based on fruit body morphology, which are relatively simple structures that offer few taxonomic characters to work with, and in addition display a high degree of phenotypic plasticity. Moreover, as emphasized, the fungal fruit body represent only a small part of the fungal life cycle, but other morphological traits were hard to incorporate in the traditional taxonomy due to their hidden nature. Nevertheless, there are several examples where taxonomists have been able to detect consistent morphological or ecological differences between cryptic lineages in a morphospecies after they have been delineated by either phylogeny or reproductive barriers e.g. (Luoma et al., 2011, Petersen and Hughes, 1998, Gordon and Petersen, 1997).

Prior to formal description as species, reproductively isolated lineages within morphospecies have been referred to as intersterility groups (ISGs). Vilgalys (1991) and Vilgalys and Sun (1994) found an overall increase of pre-mating barriers when relating ISGs to genetic and geographic distance within *Pleurotus* and *Collybia*, suggesting that speciation in mushrooms predominantly happens in allopatry. Moreover, a systematic review by (Gac and Giraud, 2008) found that reproductive isolation in Agaricomycetes, as measured by pre-mating barriers, increase faster in sympatry, indicating that there is also a role for reinforcement here. Indeed, so-called ABC patterns of intersterility where there is intersterility between sympatric lineages, but at least partial interfertility with a third allopatric lineage, often from a different continent, have been reported for numerous Agaricomycetes e.g. (Haight et al., 2019, Stenlid and Karlsson, 1991, Anderson et al., 1980, Moncalvo and Buchanan, 2008, Menolli et al., 2022). Cases of cross-continent interfertility between genetically diverged lineages in North America and Europe have intrigued researchers as they could imply that allopatric splits dating back to when these continents where connected in the mid tertiary, can still mate (Dalman et al., 2010). However, it is generally not known whether reproductive barriers such as hybrid inviability or infertility occur after these matings, as they have typically not been assessed. Many

Agaricomycetes display a high degree of host specificity with plants, whether it is as symbionts or saprotrophs, and there are indications that these can co-speciate with their plant hosts (Kennedy et al., 2011, Mujic et al., 2019). Additionally, recent studies leveraging whole genome data have found consistent genetic differences between populations residing in different environments, indicating ongoing local adaptation that may promote ecological speciation (Branco et al., 2017, Branco et al., 2015).

I will now outline the agaricomycete life cycle to establish what is meant by mating in these organisms before I discuss the genetic system underlying it, as it differs from the animal and plant systems people tend to be more familiar with. In most eukaryotes, including animals and plants, mating results in the fusion of cytoplasm and nuclei of gametes in rapid succession to produce a zygote with a diploid nucleus. But in Agaricomycetes these two events are separated in time, and the predominant stage of the organism is a state of two separate nuclei per cell (Figure 1). This stage is known as a dikaryon, and here karyogamy, the fusion of nuclei, is delayed up until the final stages of fruit body development when meiotic spores are to be produced in the hymenium. These haploid spores germinate into monokaryotic mycelia, and two monokaryotic mycelia with compatible mating types may mate and undergo plasmogamy, the fusion of cytoplasm, to form a dikaryon (Anderson and Kohn, 2007). Pre- and postzygotic reproductive barriers are therefore more appropriately characterized as pre- and postmating barriers in these fungi since a zygote is not formed immediately after mating. A few more notable distinctions exist between the sex life of these fungi and those of plants and animals. Instead of having two dimorphic sexes, the hyphae which mate are not morphologically different. Therefore, it is more appropriate to speak of mating types than sexes in fungi. Although mating in nature is thought to frequently involve a spore which germinates onto an established monokaryotic mycelium (Anderson and Kohn, 2007), these roles are interchangeable and do therefore not represent sexual dimorphism.
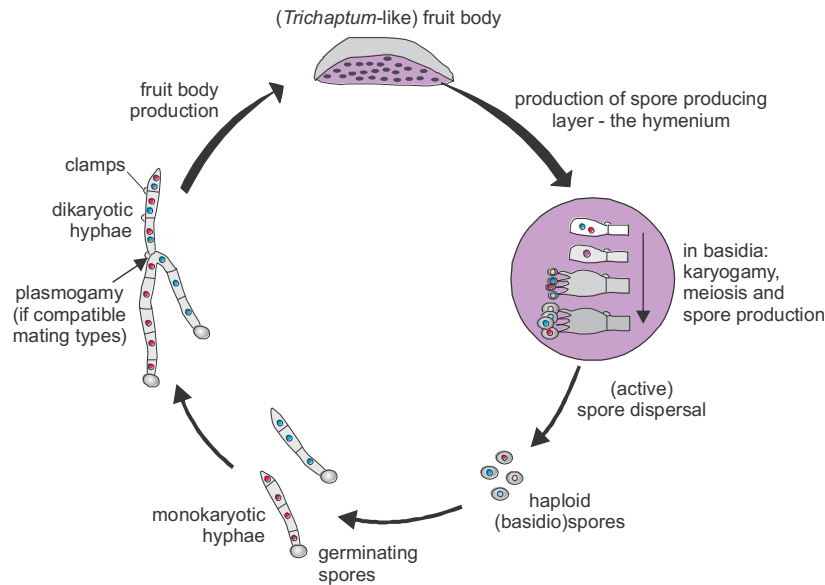
Figure 1. Overview of the typical agaricomycete life cycle with *Trichaptum* as example. Figure made by Håvard Kauserud.

What these fungi lack in sexual dimorphism, they more than make up with their diversity of mating types. Mate compatibility in Agaricomycetes is determined by homogenic incompatibility factors, meaning that the same mating types cannot mate, analogous to how the same sex cannot mate. The difference is that there can be a lot more than two mating types in these fungi. In fact there has been estimated that they can exceed 20000 (Heitman et al., 2013), making chances of a compatible mating with unrelated individuals extremely high. It is easy to see how this can be advantageous for a gamete that is haphazardly carried away by air currents as a spore and has limited mobility once it lands and germinates. This trend can be self-reinforcing as pointed out by Nieuwenhuis et al. (2013a), since novel mating types initially has an advantage when their frequency is low, and the diversity gets to be maintained by frequency dependent selection (May et al., 1999). What constitutes this breathtaking diversity at the mating loci? Crossing studies have demonstrated that there are typically two unliked loci that govern mate compatibility and that there can be a great abundance of alleles at each locus, in effect generating the multitude of unique mating types (Whitehouse, 1949, Raper, 1966). This is known as a tetrapolar mating system and considered ancestral for Agaricomycetes. It is estimated that in about one third of the species, one of the mat loci have stopped to function in determining mate compatibility (James, 2007), and these mating systems are known as bipolar. Tetrapolar mating promotes outcrossing to a greater degree in that only 25% of sibling spores are compatible, while 50% are compatible in a bipolar system (Heitman et al., 2013). How does

an allele function in self-incompatibility? Molecular biology has revealed that the two mating loci, named the mat A and mat B, consist of homeodomain proteins and pheromone receptors coupled with pheromones, respectively. The mat A contains two homeodomain proteins, HD1 and HD2, and it is the failure of HD1 and HD2 from the same allele to form a heterodimer which leads to incompatibility (Casselton, 2002). Analogously, pheromone receptors at the mat B can only bind a pheromone of a different allele (Casselton and Betts, 1998). In cases of compatibility, the heterodimer from the mat A activates transcription that triggers clamp cell formation, and pheromone binding at the mat B activates pathways that result in nuclear migration and clamp cell fusion. Thus the two mating loci work together in ensuring the development and maintenance of a functional dikaryon (Brown and Casselton, 2001).

Trichaptum the purplepore brackets

Trichaptum (Murrill, 1904*)* is a genus of wood decay fungi in the order Hymenochaetales (Ko et al., 1997, Larsson et al., 2006). A survey on Index fungorum (http://www.indexfungorum.org/names/names.asp) shows 41 accepted species as of May 2023. All species are white rot fungi (Ryvarden and Melo, 2017), meaning that they have the capacity to break down both the cellulose and lignin in wood. The majority of species grow on hardwood and are found at lower latitudes (Vlasák and Vlasák, 2017), but three morphospecies are common across the northern hemisphere where they grow on conifers, predominantly of the Pinaceae (Macrae, 1967). Among these species, *Trichaptum laricinum* (P. Karst) Ryvarden is less common in Europe and appears to be confined to a continental climate (Ryvarden and Melo, 2017), whereas *Trichaptum abietinum* (Dicks.:Fr.) Ryvarden, and *Trichaptum fuscoviolaceum* (Ehrenb.:Fr.) Ryvarden are more common and widespread (Ryvarden and Melo, 2017). In information compiled by Macrae (1967), and reproduced in table 1, some adaptation to various conifer hosts seem to occur among the three morphospecies, although this varies between continents, and none of them appear to have high host specificity. Previous crossing experiments performed on monokaryotic cultures of *T. abietinum* have revealed the ABC pattern of intersterility: there are two sympatric intersterility groups (ISG) in Eastern North America that are unable to mate with each other, but both of these two ISGs were found to be able to mate with isolates from Europe (Magasi, 1976, Macrae, 1967). Moreover, molecular studies by Kauserud and Schumacher (2003) detected highly divergent haplotypes in the nuclear ribosomal DNA (rDNA) of *T. abietinum*, within northern Europe. Similar patterns of divergent rDNA in *T. abietinum* was also observed in a study by Ko and Jung (2002)

where Korean samples were included. In a recent study, Seierstad et al. (2021) included some of the original isolates Macrae used in her crossings, and demonstrated that the rDNA of the two ISGs in North America consists of a group containing only North American samples and another group where the North American samples grouped together with samples from Europe and Asia, which was interpreted as a widespread circumboreal lineage. In addition, there was a third group consisting entirely of samples from Europe. In contrast to *T. abietinum*, less geographic structure was observed within the *T. fuscoviolaceum* clade.
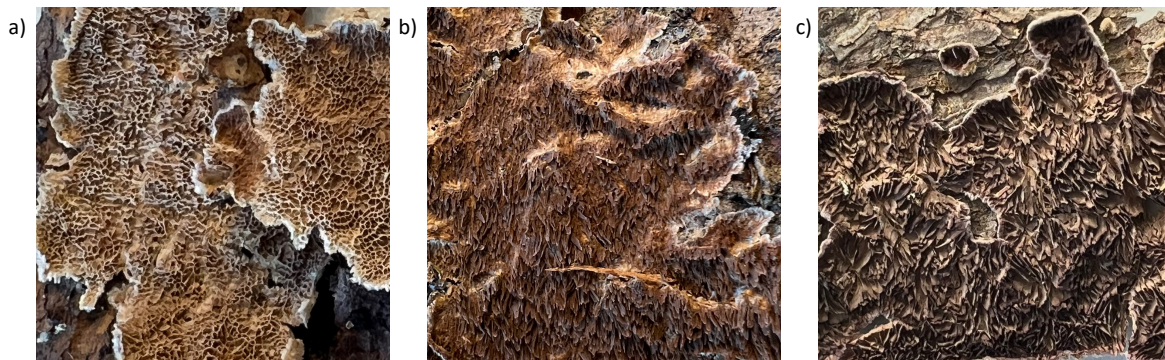
Figure 2) Different hymeniums in *Trichaptum* morphospecies on conifers. **a)** Pores ("poroid") in *T. abietinum* **b)** Toothlike (irpicoid) in *T. fuscoviolaceum*. **c)** Gills ("lamelloid") in *T. laricinum*. The difference between poroid and irpicoid, and irpicoid and lamelloid can be difficult to tell in cases where the fruit bodies tend to a resupinate growth form, which can occur in all three forms. When fruit bodies form small conks, *T. laricinum* appears more robust, and those of *T. abietinum* are the thinnest. The purple hue in the hymenium tends to be darker in *T. laricinum* and *T. fuscoviolaceum* than in *T. abietinum*, though considerable variation exists. Color of the hymenium fade to tan or brown with age, as seen in the dried specimens here. Photo: Daxiang Lu.

Table 1. Reproduced from Macrae 1967, counting the occurrence of the different morphospecies on host trees. por. = *T. abietinum*, irp. = *T. fuscoviolaceum*, lam. = *T. laricinum.*

Host	North America			Asia			Europe		
	por.	irp.	lam.	por.	irp.	lam.	por.	irp.	lam.
Abies	10	53	13	17	94				
Chamaecyparis	1								
Larix	2				2	3			
Picea	44	1	4		15	3	8	2	1
Pinus	22	2	6	55	9		11	7	
Pseudotsuga	12	1	4	43					
Thuja	8		1						
Tsuga	10	4		2	2				
Alnus									
Arbutus	1								
Betula	1								
Prunus	1								
Total	112	61	28	117	122	6	20	9	1

Methods

In this section an overview is provided of the workflow and main methods that have been used to obtain the results in the four papers. Particular emphasis is placed on the population genomic data and analyses as these have been in rapid development. The descriptions are kept at the conceptual level, whereas a more detailed account can be found within the corresponding materials and methods sections in each paper.

Field collections

Fruit bodies of *Trichaptum* are generally common and can be found throughout the year, though they form during fall or the wet season. They can be collected by slicing off a piece of the wood substrate containing a cluster of fruit bodies. As a rule, samples were collected on separate logs or branches, but in cases where few samples were found at a locality, they were collected on the same log, but then at least 2 m apart, as studies have shown that multiple individuals can reside on the same log (Kauserud and Schumacher, 2003). Metadata as location, coordinates, and substrate were recorded during collection. After collection, fruit bodies were dried in room temperature or at maximum 30 °C in a fungal drier before they were stored at room temperature.

Work with fungal cultures

To isolate monokaryotic cultures in the lab, the dried fruit bodies were re-wetted in a wet paper towel and placed at 4 °C overnight. This activates spore shooting, and spores were collected by attaching fruit bodies inside the lid of a petri dish with malt extract agar including antibiotics and the fungicide benomyl. From this petri dish, germinated monokaryons from single spores can be picked after a few days. Since two spores may land next to each other and form a dikaryon if they have compatible mating types (as 25% does in the case of *T. abietinum*), slides from each isolated culture were thoroughly checked for clamps under the microscope. It was decided to work with monokaryotic cultures as these can be used for mating experiments and to generate hybrids. In addition, there are certain advantages of working with sequences data of monokaryons as they are haploid. Isolated cultures were Sanger sequenced for the ITS marker in order to verify that they were indeed *Trichaptum*, as other opportunistic fungi in the environment or in the fruit bodies (Maurice et al., 2021) can sometimes grow out instead. The ITS sequence was also used to tentatively group the collections since *T. abietinum* and *T. fuscoviolaceum* cannot always be distinguished reliably on morphology.

Mating tests between monokaryons in paper I, II and III, were carried out following a standard procedure in mycology (Raper, 1966, Macrae, 1967, Pirozynski et al., 1988): Two monokaryons are first plated on each side of a petri dish, and after they have grown out to meet, after about two weeks, a new subculture is taken from the meeting zone. This culture is allowed to grow for one week before the mycelium is examined under the microscope to identify clamps. The proportion of septa with clamps was used to determine the degree of reproductive pre-mating barriers, where 10 septa were inspected in each of three replicates for each mating. The decay ability of established dikaryons were assessed in paper III by growing them on wood blocks and measuring the mass loss after three months. Both dikaryons produced by mating monokaryons from within and between populations of *T. abietinum* were included in the experiment, the former referred to as "regular" and the latter as "hybrids". The mass loss of the wood blocks was used as a measure of fitness, with greater mass loss interpreted as higher fitness. The ability of dikaryons to fruit were also assessed, and these preliminary results are brought up in the discussion but not included in any of the papers. To induce fruiting, hybrid and regular dikaryons were inoculated onto sterilized twigs of *Picea abies* with bark and allowed to grow for 4 months at 19 °C in the dark before they were placed at 4 °C for one week to reflect a cold shock, and moved over to 19 °C with 12 hours light cycles.

Genome sequencing and its data

Sequencing technologies can broadly be classified into long and short read sequencing based on the length of the sequenced reads, that is the fragments of DNA. Each method has their own advantages and drawbacks, and both approaches were used in this thesis. Long read sequencing was carried out on the PacBio platform for one selected strain each of *T. abietinum* and *T. fuscoviolaceum* from North America. This type of long read sequencing has proven valuable in assembling sequences back into chromosomes because the long reads provide information on how different parts of the genome are oriented relative to each other (Mascher et al., 2021, Gordon and Petersen, 1997). In addition, all the samples included in this thesis were short read sequenced on the Illumina platform, which has the advantage of being more parallelized and as such can produce more data at a lower cost. Short reads alone can struggle with assembling a genome *de novo*, typically resulting in thousands of scaffolds, but can be informative in aiding the assembly of sequenced PacBio genomes. For papers II, III and IV, the Illumina sequences were mapped to the Pacbio genome assemblies. From these mappings, variable sites in the DNA in the form of single nucleotide polymorphisms (SNPs) were extracted and analyzed further. As it is the change of allele frequencies that underlies population genetic models of evolutionary

change, this SNP data is sufficient for most population analyses. However, different analyses rest on different assumptions, and the SNP data must therefore be filtered accordingly. For example, many analyses only use bi-allelic sites, and some require the data to be linkage pruned. Notwithstanding the power of SNP data, the whole DNA sequence is necessary for prediction of the gene product. In paper I, the illumina sequences were therefore also assembled into genomes by referring to the PacBio assemblies to investigate the mating genes and their flanking regions.

From molecules to models: Reconstructing evolutionary relationships

Given how all life is related through a common ancestor, it is possible to reconstruct evolutionary relationships between organisms, or even specific genes within them. Tree like phylogenies are most common, and a number of different approaches exist to construct these from DNA sequence data. Most of these require a model on how DNA sequence data evolves. For example, mutational changes from a given nucleotide to another may not be equally probable, and vary both between sites and taxa. Statistical criteria are typically used to compare and select the most probable phylogeny. The phylogenies in paper II and III are built using maximum likelihood on SNP data. Put very briefly, likelihood is the probability of observing the data given a model, i.e. what are the chances of a specific phylogenetic tree given our sequence data and a specified model of its evolution. However, clustering algorithms without the need to specify a model of DNA sequence evolution also exist. The neighbor joining tree in paper I is an example of this type of phylogeny. To evaluate the robustness of a given phylogenetic tree, a bootstrap can be added where the same phylogenetic analysis is repeatedly run on subsamples of the data to assess how frequently the same branching pattern comes up, and in this way it is possible to quantify the support for each branch in the tree. Apart from the simplifying assumptions in the models of sequence evolution, it is also commonly assumed in phylogenetic models that splits between lineages are bifurcating and that they do not fuse once they have split. But as introduced in the section on speciation, these assumptions often do not hold within shorter evolutionary timespans where reproductive isolation is not complete. It is therefore common to employ additional methods to investigate evolutionary relationships when working with populations. For instance, phylogenetic networks using Treemix were explored in paper II, III, and IV. Here geneflow and hybridization between lineages can be modelled by adding migration events between branches that are poor fits to a bifurcating phylogenetic tree.

Principal Component Analysis (PCA) is a multivariate method that has become popular in assessing population structure as it is inherently model free, not resting on any evolutionary model or being constrained to bifurcating evolutionary patterns. This method is used in paper II, III, and IV. Given a SNP dataset, a PCA summarizes the patterns of genetic variation between samples across all SNPs in samples by extracting principal components. The principal components (PC) are uncorrelated with each other and capture successively less of the genetic variation in the data (Patterson et al., 2006). Typically, the value for each sample on the first two PC axes are used for plotting, where samples that cluster more closely together on a PC axis share more of the genetic variation for the respective PC, or put in plain language, they are more similar. It is possible to do PCA in local windows along genomes to capture their heterogeneous nature, and this type of local PCA analysis is used in paper III and IV.

Structure type analysis is another popular approach to assess population structure from SNP data. Like PCA, the method makes no assumptions on DNA sequence evolution, but requires a prior specification of number of clusters to divide samples in to. From this basis the analysis seeks to assess the sample´s ancestry, including the possibility that they can contain ancestry from multiple clusters. The original version used Bayesian statistics with MCMC sampling (Pritchard et al., 2000), but likelihood based approaches as implemented in Admixture (Alexander et al., 2009) has since become more popular because it can accommodate genome scale data. Admixture analyses are used in paper II, III and IV. Analyses were run for a range of cluster numbers, and by also running these analyses with subsamples it is possible to get a cross validation error for each run, and thereby obtain an indication on which cluster number that fits the data better. Analogous to a local PCA, it is also possible to run a local admixture analysis to assess how the ancestry can vary along the genome. This type of local ancestry inference is used in paper IV through PCAdmix and Ancestry HMM. PCAdmix is PCA based (Brisbin et al., 2012) whereas Ancestry_hmm (Corbett-Detig and Nielsen, 2017) utilizes a hidden markov model, and in addition estimates time since admixture. Both programs require setting a priori which samples are admixed and the specification of their source populations.

After ascribing populations to samples using the more exploratory methods described above, it is possible to calculate specific statistics on population differentiation and diversity. Common measures of population genetic differentiation include Fst and Dxy, which quantify relative and absolute divergence respectively. Tajima´s D and pi are common measures of nucleotide diversity, where pi measures the diversity in a population, and Tajima´s D captures the relative

composition of the diversity in terms of whether it is mostly represented by common or rare variants. Using genome data it is possible to calculate these measures locally so that genomic regions that are highly differentiated between populations can be identified, putatively representing "genomic islands" of adaptation or genomic regions involved in reproductive barriers. However, these measures were designed for alignments of DNA sequence data, and Dxy, pi and Tajima´s D ideally also require information on the invariable sites in the DNA. Including these for large genomes inflate computational time and storage, and methods more adapted to genome scale SNP data include site frequency spectra (SFS) which efficiently summarizes the patterns of variation in a population genomic dataset.

With the basis of SFS it is possible to test evolutionary models of divergence and geneflow between populations and other demographic changes. This can be done by simulating SFS according to a specified model and then obtaining a likelihood measure based on the fit to the actual SFS of the data. This approach is known as demographic modelling, and the demographic model space quickly grows out of hand if all possible migration and divergence orders between taxa are to be tested. The method is therefore usually applied to choose between specific hypotheses rather than exploratory analyses. The method performs best with a specified mutation rate, and in our demographic modelling with fastsimcoal2 in paper II, III and IV this was set to 1e-7 per generation, based on specific studies on mutation rate in other Agaricomycetes.

Coalescence analysis, which is what demographic modelling in fastsimcoal2 simulates, is another retrospective way to investigate population histories. Rather than considering the branching out of lineages as in a phylogeny, it is the coming together of them, the coalescence, of two branches back in time that is being modelled. Time to coalescence is frequently used to indicate divergence, time but the two are not interchangeable as coalescence will predate divergence in the absence of geneflow, but in its presence, introgressed sites will coalesce after the population split. The approach is powerful because it inherently captures change of allele frequency through time and as such encompass both demographic changes, geneflow and selection without the need to assume this a priori such as in demographic modelling. In addition, it is possible to estimate changes in effective population size through time as this variable has a direct connection to coalescence rate (Wang et al., 2016). Because of recombination, different parts of the genome can have different coalescence patterns, and inferring the ancestral

recombination graph remains challenging and computationally costly (Brandt et al., 2022). Coalescence analyses on *T. abietinum* is explored in paper III with the software Relate.

Another more specific and sensitive test that can detect ancient introgression is the family of D-statistic analyses, which assess patterns of allele sharing between taxa given a phylogeny. The ABBA-BABA test, also known as Patterson´s D-statistic, is probably the most known among these as it was used to detect and quantify introgression between Neanderthals and modern humans (Green et al., 2010). The D-statistic in the ABBA-BABA test consists of comparing the frequency of ancestral ("A") and derived ("B") alleles across genomes of four taxa, which include two sister taxa that are being probed for introgression, a closely related outgroup, and a more distantly related outgroup. In the absence of introgression, "ABBA" and "BABA" allelic patterns are expected to occur equally frequently in the two sister taxa, whereas an excess of either ABBA or BABA in the two will result in a non-zero D-statistic, which indicate introgression (Soraggi et al., 2018, Mughal and DeGiorgio, 2022). The ABBA-BABA test can be considered a special case of the more general estimate of admixture fraction quantified by the f_4 ratio tests (Patterson et al., 2012). Admixture can also be assessed through three population f_3 statistics which in addition provides a relative estimate of when the splits happened (Reich et al., 2009).

Another informative method to infer demographic processes such as hybridization from genome data, is to investigate patterns of linkage decay (LD), that is the linkage of sites, in our case SNPs, along the chromosome. Due to the action of recombination, linkage falls with distance since two sites close to each other are less likely to be broken up by recombination as there are fewer possible ways for it to happen. However, how fast linkage decays is affected by both geneflow and changes in population size, and more locally in the genome by selection. Recent hybridization events are expected to cause high levels of LD over long distances. This analysis was used to address the admixture in paper IV.

Results

Paper I

In this paper the population genomic data of *Trichaptum* is harnessed to study the mode of evolution in the highly polymorphic mating loci (mat A and mat B). Chromosome level PacBio assemblies of *T. abietinum* and *T. fuscoviolaceum* were generated for the annotation of the mat loci. Both assemblies consisted of 12 chromosome level scaffolds, as evidenced by telomere sequences, and the mat A and B loci were located on scaffold2 and scaffold9, respectively. These assemblies revealed a genome size of 49 Mb in *T. abietinum* and 59 Mb in *T. fuscoviolaceum*, and also served as references for subsequent studies of *Trichaptum*. Illumina sequenced and assembled genomes of 138 *T. abietinum*, 41 *T. fuscoviolaceum*, and 1 *T. biforme* were annotated for the mat A and mat B loci by BLASTing to the mat loci and the conserved flanking genes in the Pacbio reference genomes. These mating gene annotations revealed a duplication at the mat A locus resulting in an additional homeodomain complex that were present in some, but not all isolates, and sometimes inverted. Four pheromone receptors and two pheromones were detected at the mat B locus. Comparisons against a genome wide baseline revealed that the diversity was significantly higher in the mat A locus, and for two of the pheromone receptors at the mat B locus. These genes showed low Fst, elevated pi/Dxy ratio, high Tajima´s D, and high dS values, indicative of balancing selection, which a multilocus HKA-test provided additional support for. The low diversity measures at the other two pheromone receptors suggest that they do not function in mate specificity, and this was verified in vitro by the successful crosses of isolates that should have been incompatible due to similar alleles at these two pheromone receptors. Additionally, crossing experiments indicated that more than 86% amino acid identity is needed to function as the same mat allele and display self-incompatibility. Using this threshold it could be estimated that 270 different mat A alleles and 65 different mat B alleles are segregating across *Trichaptum* species, which can combine to form 17,550 mating types, consistent with previous predictions. The different mating type alleles were segregating across species, and showed no geographic structure. Moreover, phylogenetic analyses including more distantly related genomes of Agaricomycetes, revealed that these trans-specific polymorphisms are ancient, and attest to the consistent and long-term action of balancing selection. From the results in this study, it could also be inferred that reproductive barriers between species in *Trichaptum* are not governed by the same loci as those that control mating within species, as mating that should be compatible based on alleles at the mating loci fails between divergent populations of *T. abietinum*.

Paper II

In this paper the aim was to investigate hybridization or introgression between *Trichaptum abietinum* and *T. fuscoviolaceum* using both mating experiments and population genomic analyses. The samples included Italian and North American *T. fuscoviolaceum*, and two North American lineages of *T. abietinum*. Mating experiments did not detect any successful mating between *T. abietinum* and *T. fuscoviolaceum*, even when they were predicted to mate based on the mating type annotations in paper I. Unlike the two lineages of *T. abietinum* in North America, which are known to be intersterile, the Italian and North American *T. fuscoviolaceum* were found to be able to mate. Maximum likelihood phylogeny separated the *T. abietinum* and *T. fuscoviolaceum* clades and each lineage within them with 100% bootstrap support. Reads from *T. fuscoviolaceum* samples were mapped to both the *T. abietinum* and *T. fuscoviolaceum* reference genome, and a comparison did not reveal any samples mapping equally well to both reference genomes, which would be expected from a hybrid. Consistent with this result, Admixture and PCA analyses did not indicate hybridization either, but instead revealed high divergence between the two morphospecies. *Trichaptum abietinum* and *T. fuscoviolaceum* were clearly separated on each side of the PC1 which explained 58% of the variation, and sliding window analyses of Fst, Dxy and pi revealed that this divergence is genome wide. However, ABBA-BABA test statistics detected significant D values between the Italian *T. fuscoviolaceum* and the North American *T. abietinum* lineages. Further analyses with f_4 statistics supported introgression between these lineages, and the f_3 statistic indicated that the *T. abietinum* lineages split later than the *T. fuscoviolaceum* lineages. A sliding window analysis based on the f_3 statistic localized several genomic regions with a significant signal of introgression. Scaffold 1 and 5 contained more of these introgressed regions, but these are also among the largest scaffolds. Enrichment analyses for Gene Ontologies did not find any significant terms associated with the introgressed regions. Phylogenetic network analyses with Treemix supported introgression from Italian *T. fuscoviolaceum* into North America B of *T. abietinum*. Demographic modelling further supported this finding, and found that a scenario of two introgression events, between ancestral *T. abietinum* and *T. fuscoviolaceum*, and more recently, between an extinct or unsampled population and North America B in *T. abietinum,* to be most likely. These results highlight the importance of including scenarios of extinct lineages and ghost populations when testing different models.

Paper III

Incipient levels of divergence can be more informative for inferring speciation processes, and here the attention is turned to the pattern and processes of evolutionary divergence in the *T. abietinum* morphospecies. Phylogeographic analyses like PCA, Admixture and ML phylogeny, all supported the division of the morphospecies into six geographically restricted major lineages. Three of the lineages occur in North America, and are referred to as North America A, B, and C. North America A and B were recovered throughout the continent, whereas North America C is represented by only two samples from high elevation sites in Western North America. Two of the major lineages were found in Europe, with one lineage confined to the Mediterranean Basin, and the other lineage distributed across Northern and Eastern Europe extending all the way into Siberia. Samples in western and central Europe appeared to be admixtures of the two major lineages in Europe. A distinct lineage was found in East Asia, containing several populations with signs of some admixture from the European and North American lineages. Mating tests indicate that North America A is reproductively isolated from the other lineages in North America, and that the East Asian sublineage from Southwestern China has reduced ability to mate with the admixed lineages in East Asia, as well as most of the other major lineages. The fitness of the hybrids from the mated cultures were assessed by measuring their ability to decay wood, which revealed a strong correlation between genetic distance between isolates and reduction of fitness in their corresponding hybrids. In comparison, a weaker correlation was observed between increase of genetic distance and ability to mate. The stronger pre- than post-mating barriers, and occurrences of stronger pre-mating barriers in sympatry are interpreted as evidence of reinforcement selection. Demographic modelling, coalescence, and phylogenetic network analyses were carried out to investigate geneflow between reproductively isolated lineages, and detected evidence for this in both North America and East Asia. Viewed together, these results suggest that genetic incompatibilities from independent divergence in allopatry could have set the stage for reinforcement to complete the speciation process by selecting for strong pre-mating barriers once highly diverged lineages came together in secondary contact. As a greater number of reproductive barriers occurs in North America and Asia, it is speculated that geographical characteristics affecting the glacial cycles and migration routes played a role in shaping these barriers and consequently the genetic structure in the morphospecies. A first stab at identifying the genes causing the pre-mating barriers using genome scans and genome wide association studies hint at a role for metacaspases and controlled cell death, similar to what have been found in certain ascomycetes.

Paper IV

The phylogeographic patterns of *T. abietinum* in Europe is examined closer in this paper, with a particular focus on the admixed populations that were detected. A re-analysis of the European data of *T. abietinum* in paper III found that this morphospecies can be divided into three lineages in Europe. In addition to the Eurasian and European lineage identified in paper III (which in this study are referred to as Boreal and Mediterranean, respectively), a third Atlantic lineage was identified in western Europe. Samples of the Atlantic lineage were indicated as admixtures of the Boreal and Mediterranean lineages in paper III, but the PCA here including only samples from Europe indicates that these samples contain a unique genetic component akin to, but different from the Mediterranean lineage. Otherwise, the Atlantic lineage share large levels of genetic variation with the Boreal lineage, likely as a result of geneflow. Phylogenetic network analyses in Treemix support that the central European and Atlantic populations resulted from two independent events of admixture. Linkage decay analyses and inference of these using Ancestry_hmm further show that admixture happened more recently in central Europe, as evidenced by longer segments of haplotypes and higher levels of LD, consistent with recent admixture where recombination have not had enough time to break down the LD. A local PCA revealed that these patterns are not evenly dispersed throughout the genome, instead the second half of Scaffold 5 stands out with the central European samples falling closer to the Boreal lineage, whereas the Atlantic lineage falls very close to the Portuguese population of the Mediterranean lineage for this region of the genome. A PCAdmix analysis further revealed that the second half of Scaffold 5 in all central European samples consist exclusively of the Boreal haplotype, suggesting either strong selection or some reproductive barrier in the form of genomic incompatibility between the Boreal and Mediterranean lineage on this part of scaffold 5. In light of known postglacial history of host trees, the Boreal lineage can be connected to the recolonization of Norway spruce from a refugium in eastern Europe, whereas the Atlantic lineage could have arrived with Scots Pine from the west and south before it admixed with the Boreal lineage resulting in the present-day populations. Additionally, the now submerged landmass of Doggerland could have supported both Scots pine and Norway spruce, thus a possible refugial area for the Atlantic lineage. The southern European refugia of Pine and other conifer species would likely have supported the Mediterrenean lineage, and the low levels of differentiation within them could indicate some ongoing geneflow here. In sum, it can be concluded that the postglacial history of *T. abietinum* in Europe resemble what has been found in numerous plants and animals, with survival and divergence in multiple separate glacial refugia and the formation of suture zones in central Europe.

Discussion

The tempo of evolution in Trichaptum

What can be inferred about the diversification and evolution of reproductive isolation in *Trichaptum* based on the speciation continuum that paper II, III and IV have examined? Before making any inferences of the processes involved, a summary of the patterns they have generated is warranted, and I will begin by considering the timescale these diversifications have occurred on. The phylogenies in paper I, II and III are not time calibrated due to the lack of fossils in *Trichaptum,* moreover only one fossil is known from the order Hymenochaetales (Smith et al., 2004). In a genome wide phylogeny of the Agaricomycetes by Varga et al. (2019) the origin of Hymenochaetales was dated to the mid Jurassic 167 MYA, which opens the possibility for genus *Trichaptum* being old. Notwithstanding the difficulties of absolute dating of phylogenies, it is possible to assess relative divergence times. From the phylogenies in paper I and II it is evident that the divergence between *T. abietinum* and *T. fuscoviolaceum* have happened over a different order of timespan than the divergence of lineages within each morphospecies. Time estimates from the demographic modelling corroborate these patterns in the phylogenies, with paper II suggesting that the split between *T. abietinum* and *T. fuscoviolaceum* started some 500,000 generations ago, whereas the split between the reproductively isolated North America A and B lineages is placed to around 100,000 generations ago in paper III, and finally, in paper IV, the split between lineages that are fusing in central Europe is placed to around 30,000 generations ago. Moreover, the phylogenies in paper I and III indicate that divergence between the Asian lineages in *T. abietinum* occurred within roughly the same timescale as in Europe, indicating that reproductive barriers can also evolve rapidly.

How does generation time in *Trichaptum* relate to geological time? The fruitification experiments that were set up for *T. abietinum* provides some clues. In these experiments, a few dikaryons fruited after 5 months, while the majority required 6-8 months. Based on the ability of *T. abietinum* to fruit in the lab in less than a year, and the fact that *Trichaptum* species are primary colonizer of recently dead wood substrates, it seems reasonable to assume a generation time on the order of one to a few years. However, there are additional challenges in converting the demographic modelling estimates to geological time as these estimates greatly hinge on the mutation rate, which was set to 1e-7 per generation based on studies of *Schizophyllym commune* (Bezmenova et al., 2020), a common wood decay fungus in order Agaricales with a similar ecology. It is also possible to use known geological events for time calibration of phylogenies

and demographic modelling. For instance, in the case of the recent radiation of Cichlid fish species in the great African lakes, the lower limit of divergence could be set to 17-12,000 years ago when the water levels in the lakes were low or completely dried up (Sturmbauer et al., 2001). Could the ice sheets that covered large parts of the northern hemisphere during the last glacial maximum (LGM) represent a comparable scenario that could also be used for dating? The more recent range expansions of *T. abietinum* and *T. fuscoviolaceum* must have happened concurrent with or after the spread of conifers after the end of the LGM 20,000 years ago. The vast area that is covered by conifer trees in the taiga likely represents a great increase in population size for the lineages there, and the coalescence analysis in paper III did indeed observe a rapid increase in effective population size around 20,000 generations ago for all analyzed lineages of *T. abietinum*. It is tempting then, to link this to the end of the last glacial maximum and infer a generation time of 1 year. However, effective population size is a function of genetic diversity and demographic processes, and would not be expected to follow the exponential population increase of an expanding population, as this would initially lead to many genetically similar individuals and low levels of genetic diversity.

Another question pertaining to the tempo of evolution is whether diversification happens evenly through time, or if it is concentrated to certain epochs. The idea of punctuated equilibria, which propose that evolutionary change is generally restricted to speciation events, was put forward to explain why there are periods with apparent stasis in the fossil record interrupted by rapid change (Gould and Eldredge, 1993). Although the theory is contentious in the context of macroevolution as inferred from changes in morphology (Voje et al., 2019), a comparative study by Pagel et al. (2006) analysing DNA sequence data, found that overall 22% of nucleotide substitutions could be localized within speciation events as inferred from the phylogenies. Interestingly, this trend appeared to be more common in fungi and plants than in animals for the data that was included. Could it be that the Pleistocene glacial cycles prompted the microevolution of *Trichaptum* in a fashion of punctuated equilibrium, with an influx of genetic diversity following introgression events in the interglacials when diverged lineages met through secondary contact? The demographic modelling in paper III could indicate so, but studies looking further back in time would need to be carried out in order to probe whether there is support for the same pattern for interglacials beyond the most recent one.

Patterns in space

What spatial, or biogeographical patterns can be summarized from the four papers? As a more comprehensive geographic sampling was not conducted on *T. fuscoviolaceum* it is difficult to make any inferences on this morphotaxon. But in *T. abietinum* a few patterns seem to stand out, including a higher diversity of lineages in North America and admixture in East Asia from different lineages. Moreover, some lineages in *T. abietinum*, such as the Eurasian and North America B lineage have wide distributions, whereas others, such as North America C appears to be more local. Although many factors could have contributed to these patterns, such as local adaptation or historical contingency, habitat connectivity appears to be important as evidenced by the wide distribution of the Eurasian lineage across much of the taiga.

An intriguing question to ask in speciation research is why some lineages diversify more than others. Based on the sampling in this thesis and Seierstad et al. (2021), there seems to be more lineages in *T. abietinum* than *T. fuscoviolaceum*. But before speculating on what ecological or genetic properties that allowed *T. abietinum* to diversify more, it should be kept in mind that diversification rates are not necessarily reflected in the current diversity. Extinct lineages and extinction rates must be taken into consideration. There is an inherent limitation in assessing extinction patterns in fungi as they do not tend to fossilize, and even if they did, their relatively simple morphological structures would likely not offer much to work with in the cases of closely related species. However, as demonstrated in paper II, the inclusion of so-called ghost lineages in demographic modelling may indicate their existence if they have introgressed into extant populations. With the potential for extinction in mind, it could be contemplated, and even tentatively assessed through demographic modelling of ghost lineages, if the observed patterns could be due to more extinctions in *T. fuscoviolaceum*, rather than higher diversifications within *T. abieitinum*. Granted that there is evidence of higher extinction rates in *T. fuscoviolaceum*, one could speculate on whether there is something about the ecology or geographic origin that has contributed to higher extinction rates. For example, could a boreal ecology have contributed to the higher extinction rates of lineages as these areas have been more impacted by the ice ages?

Mode of speciation

Now turning from patterns to processes, we can ask if these studies on *Trichaptum* suggest that they speciate in the same allopatric manner that have been proposed for other agaricomycetes

(Vilgalys, 1991, Vilgalys and Sun, 1994, Petersen and Hughes, 1999). If we consider pre-mating barriers within *T. abietinum* they seem to comply with the expected patterns from speciation in allopatry in that i) ability to mate generally decreases with increased genetic divergence and ii) genetic divergence tends to increase with geographic distance. But the parapatric lineages in East Asia that are reproductively isolated speaks against i), whereas the diverged sympatric lineages of *T. abietinum* in North America that also are reproductively isolated contradicts ii), suggesting that other mechanisms are involved in generating strong pre-mating barriers in sympatry, in line with has been observed in many other fungi e.g. (Dettman et al., 2003, Menolli et al., 2022, Moncalvo and Buchanan, 2008, Aanen et al., 2000, Murphy and Zeyl, 2015). For *T. fuscoviolaceum,* neither pre-mating barriers nor sympatric lineages were observed here or in previous studies (Macrae, 1967, Seierstad et al., 2021). No signal of introgression was identified between the European or North American *T. fuscoviolaceum* lineages in the Treemix analyses in paper II, and the lack of premating barriers is not so surprising when considering how secondary contact seems to frequently be involved in generating strong pre-mating barriers in these fungi, and taking into account that mating can still be possible in cases of considerable genetic divergence. It is tempting then, to use this apparent lack of geneflow and reproductive barriers as support for dispersal limitation in *Trichaptum* in that there is no evidence for transatlantic geneflow within *T. fuscoviolaceum* and *T. abietinum*. Many agaricomycetes appear to retain the ability to mate across large evolutionary distances. For instance, allopatric populations of *Heterobasidion* species that have been proposed to have split in the mid Tertiary, more than 20 MYA, can still be mated in the lab (Dalman et al., 2010). However, this does not appear to be the case for *Trichaptum*. Corroborating the findings of Macrae, the crosses in paper II between *T. fuscoviolaceum* from Europe and *T. abietinum* lineages from North America A and B did not detect any compatibility between allopatric lineages of the morphospecies.

The strong correlation between genetic distance and wood decay ability in paper III that have been interpreted as reduced hybrid fitness in *T. abietinum,* suggests the accumulation of BDM incompatibilities in line with speciation in allopatry. As the Italian and North American population of *T. fuscoviolaceum* are more diverged than any two *T. abietinum* lineage, it would be interesting to assess the fitness of these hybrids. Fruiting experiments are still in the early phases of trial, but preliminary results in *T. abietinum* have shown that the North America A x Eurasia and North America B x Eurasia lineages can both produce fruit bodies, although the amount and viability of spores remains to be systematically assessed. At any rate, it would not

seem out of place to infer, based on the papers here, that pre-mating barriers in *Trichaptum* tend to be stronger than post-mating barriers, as expected in reinforcement. However, the theory of reinforcement has received much critique, and in the next section I will therefore subject the interpretation of reinforcement in paper III to a critical re-examination.

Reinforcement re-examined

Reinforcement of pre-mating barriers has been invoked to explain the observed patterns of stronger pre-mating barriers in sympatry for agaricomycetes (Giraud and Gourbière, 2012, Gac and Giraud, 2008, Petersen and Hughes, 1999). But there are several other explanations for these patterns. Among them is differential fusion, which reasons that reproductive barriers can be observed more frequently in sympatry because lineages would already have fused in cases where there are no reproductive barriers between them (Templeton, 1981). However, differential fusion does predict stronger pre- than postmating barriers. In considering the alleles causing pre-mating barriers that are being selected for in reinforcement, theory predicts that they must either be neutral or slightly deleterious outside the zone of contact. If they were positive in allopatry, they would have been expected to be selected for in the absence of secondary contact, and consequently there would be no room for reinforcement selection to increase their frequency. Conversely, if these alleles are slightly deleterious in the absence of secondary contact, a challenge arises in explaining how they are maintained and not "swamped" out by alternative alleles (Coyne and Orr, 2004).

Reinforcement alleles are expected to be dominant (Coyne and Orr, 2004), and interestingly, in a model on ISGs in *Heterobasidion* put forward by Chase and Ullrich (1990), homozygosity at any one of five loci is sufficient for interfertility between isolates. This model would be consistent with a dominant effect of alleles causing failure to mate. The patchwork of intersterility between isolates of different lineages of *T. abietinum* in paper III seem to imply that the observed intersterility patterns have a multigenic basis. How could a genetic architecture as the one proposed in *Heterobasidion* come to be in the first place? Recent findings in the ascomycete *Podospora anserina* might provide some clues: Ament-Velásquez et al. (2022) discovered that genes normally functioning in allorecognition could also cause intersterility between populations. Allorecognition, or vegetative compatibility, is the more basal ability to tell self from non self, and is common to all domains in the tree of life (Buss, 1982, Gibbs et al., 2008). This ability must be particularly important for filamentous fungi, as a hypha encountering another hypha must decide on whether it is itself, prompting fusion, or

non-self, warranting defense and rejection. In order to increase the discriminatory ability, the involvement of several loci through balancing selection can be expected, conceptually similar to the mat loci in paper I, but different in that it is expected to be a heterogenic incompatibility system, rejecting non-self alleles. It is not clear what selection pressures that brought about the intersterility in *Podospora* (Ament-Velásquez et al., 2022), but considering how co-option of biological systems appear to be common, it can be speculated that the allorecognition system in basidiomycetes could also have been recruited by reinforcement selection to cause intersterility. A function in allorecognition would explain how diversity at these alleles could be maintained in the absence of reinforcement. Additionally, it would predict that demographic processes such as loss of genetic variation through bottlenecks as observed in *Serpula* by Kauserud et al. (2006) could change intersterility patterns. Unlike the Ascomycota, the genes causing vegetative incompatibility have not been characterized for any species in the Basidiomycota. Moreover, their dikaryotic state makes it challenging to formulate a theory for their mechanism, although Auxier et al. (2021) proposed a model involving cytoplasmic mixing and cell death. A characterization of these genes will be necessary before the possibility of their co-option as "speciation genes" could be tested.

Mating through the looking glass

The patterns in paper III which suggest that there have been several cases of reinforcement in *T. abietinum,* could indicate that mating is costly. Are there characteristics of the *Trichaptum*, or more generally speaking, Agaricomycete biology that implies a high cost of mating? If a monokaryon could only mate once, mating with the wrong hypha would indeed be the end. However, Buller´s phenomenon, where the monokaryons from a formed dikaryon can mate again with another monokaryon, seems to suggest some degree of reversibility (Anderson and Kohn, 2007), or at least ability to re-mate. Nieuwenhuis et al. (2013b) found in natural *Schizophyllum* populations that monokaryons appeared to have engaged in several dikaryons. Buller´s phenomenon is known to occur in *Trichaptum,* Kauserud and Schumacher (2003) showed that in 82% of cases, crosses between dikaryons and monokaryons resulted in dikaryotization of the monokaryotic isolates. Notwithstanding the potential for Buller´s phenomenon, the chances for a mal-mated dikaryon to be rescued by encountering another monokaryon may not be very high as mobility in these organisms is limited, though there seems to be growing support for insect mediated dispersal of spores for wood decay fungi (Lunde et al., 2023).

The tetrapolar mating system in *Trichaptum* selects for diversity at the gene level and promotes outcrossing, whereas local adaptation acting on the individual may favor mating within the population. Indeed, sexual reproduction itself can be a double edged sword. It may scramble well-adapted genotypes through the influx of new genetic variation and recombination of the existing material. Avoiding sex altogether, as many fungi that reproduce mostly clonally do, circumvents this problem. However, asexual lineages seem to suffer from higher extinction rates (Crow, 1994). The novel variation generated by sexual reproduction can become useful in a dynamic environment, thus sex is believed to be advantageous in the long run. However, mating with the too genetically divergent is associated with disadvantages like BDM incompatibilities, and there is a fine tuned balance between inbreeding and outcrossing. In *Trichaptum* and many other fungi, reproductive barriers can be porous. Does this openness to mate with diverged lineages also come with any advantages? Considering the role of sex, it is tempting to draw the analogy of introgression as a higher order process for generating novel genotypes that offer adaptive advantages. It would be interesting to assess if lineages where introgression is common have lower extinction rates, though this will be difficult to test as we are largely unable observe the extinct lineages that did not partake in introgression. Moreover, speaking of these processes as adaptative at the species level is problematic. Rather than an adaptation, their prevalence should be understood as the consequence of a sorting mechanism, where lineages with the given trait have higher survival rates and thereby pass it on. However, this cannot explain how the trait came to be in the first place.

Ecological speciation

Allopatric speciation is believed to commonly produce some degree of ecological divergence as a consequence of adaptation to different local conditions in the divided populations (Schluter, 2009). Moreover, classical ecological theory predicts that species need to have different niches in order to coexist (Hardin, 1960), Ernst Mayr has gone as far as to suggest that "the process of speciation is not completed by the acquisition of isolating mechanisms but requires also the acquisition of adaptations that permit co-existence with potential competitors" (Mayr, 1982). Irrespective of Mayr's statement and niche theory, ecological character displacement is common when closely related lineages occur in sympatry (Schluter, 2000). This appears to be

the case for *T. fuscoviolacum* and *T. abietinum*, where *T. fuscoviolaceum* prefers *Abie*s in North America, and *Pinus* in Europe, whereas *T. abietinum* is more common on *Picea* and *Larix (Macrae, 1967)*. *Trichaptum abietinum* can also can be common on *Pinus*, although it is unknown to which degree local adaptation to host trees in specific lineages are involved here. What about sympatric lineages within *T. abietinum*? The finer scale sampling of both North America A and B lineages in Eastern North America allows for a statistical assessment on the relation between lineage and host substrate (table 2). A Fisher´s exact test on this data did not detect any significant differences in host preference. Wood decay fungi are to a greater or lesser extent adapted to specific substrates (Rustøen et al., 2023), for example the different sister species in the *Hetereobasidion* complex have specialized on different conifer species (Garbelotto and Gonthier, 2013), however these fungi can attack living trees, and it is possible that adaptation to the host is more important when it is alive.

Table 2. Substrate North America A and B lineages were collected on in Eastern North America (New Brunswick, Canada, and New York, US) where both lineages occur in sympatry at the regional level.

	North America A	North America B
Host		
Abies	4	20
Picea	21	48
Pinus	1	3
Larix	3	6
Tsuga		3

More samples were recovered of North America B in Northeastern North America, and a geographic examination reveals only 3 of 18 samples from New York to be North America A. In New Brunswick, the two lineages were also found at different frequencies throughout the province, with North America A being recovered from only certain sites, generally towards the east or the south, which could indicate differentiation to habitat in some other way than host tree (figure 3). Another explanation is that North America B is displacing North America A in New Brunswick, although there could be many other explanations to these patterns, as trivial as that these lineages, like many other fungi, fluctuate in frequency from year to year, or fruit

at different times. Despite the apparent lack of ecological character displacement between these two lineages, local adaptation is likely to have occurred to some degree in the *Trichaptum* as they can be found in a wide range of different environments with vastly different humidity and temperature conditions. In an examination of *T. abietinum* populations across climatic gradients in Norway a significant signal of local adaptation was indeed observed (Methlie, 2021). Nevertheless, the results were difficult to interpret as the population in western Norway, which represents the extreme in several climatic gradients, also have a different genetic composition as shown in paper IV, meaning that population structure is confounded with the climatic gradients.

Figure 3. Distribution of North America A and B in New Brunswick, Canada.
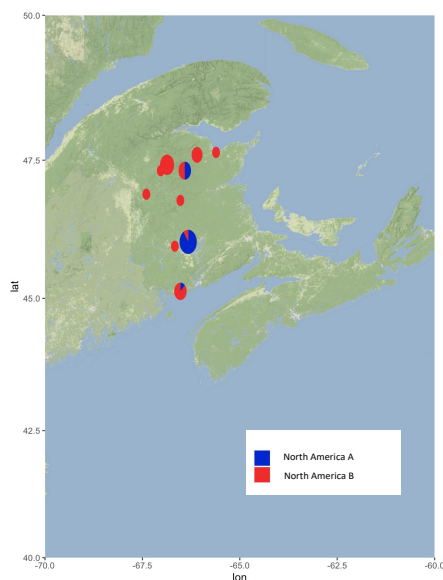
Applying species concepts in Trichaptum

Up to this point I have focused on the speciation process and the evolution of reproductive barriers, but deliberately avoided the question of how to define their end result, the species. The so-called species problem is inherently difficult, recognized by Darwin as "undefinable" (Darwin, 1856), and has been historically contentious (Hey, 2001). In a survey by Stankowski and Ravinet (2021), the most popular species concept was found to be some version of the biological species concept (BSC), which defines species according to their potential to interbreed. There are differing views on whether this interbreeding should be understood purely as ability to interbreed, or interbreeding taking place in nature (Coyne and Orr, 2004). Likewise, the frequency of interbreeding with other groups that should be allowed, is also open to

interpretation in the BSC. Should occasional hybridization events be allowed, or does the occurrence of hybrids among two groups mean that they should be viewed as one species? The idea here, as in most species concepts, is to encapsulate an independently evolving lineage, but paper II, III, IV have demonstrated that few lineages in *T. abietinum* seem to be truly separated and independently evolving.

The split between the European and North American lineages of *T. fuscoviolaceum* is considerably older than reproductively isolated lineages of *T. abietinum*. Should these *T. fuscoviolaceum* lineages be recognized as different species? Assuming that pre-mating barriers in these fungi generally require secondary contact for completion, the BSC may not be sufficient to separate these lineages. The phylogenetic species concept (PSC) has therefore been advocated by many mycologists (Taylor et al., 2006), as a way to tackle how allopatric lineages seem to retain the ability to mate across long evolutionary timescales in kingdom fungi (Peris et al., 2023). But while the BSC can struggle in the cases of long divergence in allopatry, the PSC is expected to run into difficulties when dealing with incipient species with incomplete reproductive isolation, as in paper III. Moreover a species tree, as inferred from phylogenies built on genome wide data can contain a multitude of conflicting genealogies (Avise et al., 1998). Or put differently, individual gene trees do not need to accord with population trees, as demonstrated by the trans-species polymorphisms in paper I, and the coalescence analyses in paper III. Viewed in this way, a species could be understood as the most probable container of a gene history (Maddison, 1997). Some have even argued, echoing Buffon, that species are not real but purely human constructs (Coyne and Orr, 2004).

Does defining species matter then, beyond a philosophical exercise? For one thing it matters because they are the common currency of biodiversity that is being used to direct efforts in conservation biology, although policies, such as the local or regional responsibility to protect a species, implicitly recognizes some aspect of diversity within a defined species. Furthermore, as we find ourselves in the middle of an unprecedented mass extinction event (Ceballos et al., Cowie et al., 2022), conservation biology and understanding biodiversity have become the most urgent of matters if we are to mitigate the massive shedding of branches in the tree of life. Of course, protecting species goes well beyond defining them according to some measure of genetic distance or ability to breed with others individuals, it requires knowledge of their ecology. But to even begin with an understanding of ecology, it is imperative to know which units to deal with. In this regard the PSC is insightful because it recognizes the evolutionary

history of organisms, and this information could be used as basis to look for ecological differences between cryptic species. For instance, in *T. abietinum,* North America C appears to live in a very different environment than North America B. Even though they appear to be able to mate and hide behind a common morphology, and as such could be defined as the same species, they may have adapted to very different conditions. Recognizing this diversity is important, and the conservation of this genetic diversity could be central for these fungi to adapt to the changing climate that is predicted, and thereby also for the continued function of the ecosystems they are a part of. On the other hand, the BSC can be useful in the planning of conservation because it provides an idea of what we can expect when distributions of organisms change with the climate and evolved lineages come back into contact.

Looking ahead

We have seen that the common conifer associated morphospecies of *T. abietinum* and *T. fuscoviolaceum* contain a multitude of lineages shaped by different evolutionary forces across different spatiotemporal scales. This thesis has only touched the surface of several of these evolutionary mechanisms, including post-mating barriers, whereas other aspects, such as local adaptation and mitonuclear incompatibility, remain completely untouched. For the areas that have been explored more in depth, such as the mating loci, introgression, and pre-mating barriers and phylogeography within *T. abietinum*, the results have stimulated new questions: How is the multitude of mating alleles generated? What keeps recombination from generating self-compatible mating alleles? Accordingly, there are also a multitude of areas and ways in which future research on *Trichaptum* could be carried out. But regardless of which avenue the research goes down, there are presently common roadblocks in the form of a limited understanding of basic biology and life-history traits of *Trichaptum* such as generation time, durability of the monokaryon phase, mutation rate, and recombination. If the promising system of *Trichaptum* presented here should flourish, determining these aspects of its biology should be a priority. The sequencing of both parent monokaryons along with their spore family should make it possible to pin down both mutation rate and a recombination map. Ascertaining the generation time in the wild would require more work, but inoculating logs (with a local dikaryon) and measuring the time to fruit, and continuing with sampling the subsequent rounds of fruiting on the logs, to assess if the same individuals continue to produce viable spores, should give a fairly good indication. Ideally, observational studies of substrates in the forest should also be carried out in parallel, as artificially inoculating a log with a dikaryon could change the temporal dynamics.

The many instances of reproductive barriers in *T. abietinum* would be ideal to probe if the same genes are involved in reproductive isolation in independently evolved reproductive barriers. In the event that different genes are involved, there could still be a common genetic architecture underlying reproductive isolation in various lineages. Moreover, as introgression appear to have been common in this system, it also begs the question on whether the origin of the genes or alleles causing reproductive isolation, in some cases, such as in East Asia, could be linked to geneflow. The preliminary fruit body experiments have verified the possibility of generating a hybrid fruit body where the ability to mate with a given lineage are expected to be segregating in the F1 hybrid monokaryons. Sequencing a large spore family of these F1 hybrid monokaryons and subjecting them to crosses with a third lineage that one of the parents are not able to mate with, should make it possible to pin down the genes causing intersterility in that case when paired with genome wide association studies. Moreover, by comparing gene expression in intersterile with interfertile matings involving different lineages, it should be possible to verify these findings, and also investigate the role of allorecognition that have been indicated in paper III. These studies could be expanded to also investigate dikaryon-monkaryon matings, as well as studying the gene expression at different time points after mating, to verify how the intersterility works together with the regular mat-loci machinery to regulate clamp cell formation and nuclear migration.

The phylogeography of Europe has been considered simpler, and less exciting compared to that of North America or Asia. But the closer examination of paper IV reveals that this is far from the case in *T. abietinum*. The admixture in central Europe represents a natural experiment to understand if, and how, selection plays out after hybridization, and an opportunity to probe for genomic incompatibities. A closer sampling surrounding this zone is warranted, and similarly for the Atlantic lineage in western Norway. Here too, generating lab hybrids would be interesting, as they would allow for the comparison with potential natural hybrids that have been admixing for some time. Would the lab F1 hybrids show the same pattern on scaffold 5, inheriting it exclusively from the Boreal lineage? And how about mitochondrial inheritance? Is there a pattern where mitochondria are inherited in an asymmetric manner like what have been found in several other fungi (Nieuwenhuis et al., 2013b, Anderson and Kohn, 2007)? Common garden experiments varying substrate, temperature, humidity, would be informative here as well, to assess fitness of hybrids, but also for investigating local adaptation. In a common garden experiment (Methlie, 2021) found Atlantic lineage in western Norway to grow faster

than the populations of the Boreal lineage in Norway. Temperatures rarely fall below freezing here, and thus it would be natural to test their tolerance to freezing.

Finally, a closer phylogeographic study of *T. fuscoviolaceum* along with an assessment of post-mating barriers in addition to pre-mating barriers, would allow for comparative studies within these *Trichaptum* species that could more conclusively say something about diversification patterns here, or possibly the lack thereof. So far, the patterns of genetic diversity within *T. abietinum* indicate that isolated mountains ranges can harbor unique diversity, and this should be kept in mind for planning future sampling. Additionally, including other taxa of wood decay fungi with known ABC-patterns of pre-mating barriers could ascertain if these patterns originated as a consequence of the same biogeographic processes, such as a Beringian lineage entering North America and resulting in secondary contact and reproductive barriers with a residential North American lineage, similar to what has been proposed for *T. abietinum*. And ultimately, such comparative studies would allow the assessment of the extent to which geography could be channelling evolution in these organisms in a collective fashion within the Quaternary.

On the origin of speciation

I will close this thesis with a few remarks on the initial question on how come the tree of life have bud into the units we recognize as species. First of all, as Darwin pointed out, it is not evident why life would be organized into discrete species rather than existing as a continuum (Mayr, 1982). In light of what we now know about the evolution of reproductive barriers, a continuum of organic life seems to be out of place when taking into account the variety of habitats that are either discontinuous or separated by steep environmental gradients. We could instead ponder if speciation would have been a phenomenon at all given a perfectly homogenous earth. Moving on to the speciation process, we can try to understand it, like any biological phenomenon, both in terms of how it works and why it came to be. Our explanations can be categorized accordingly, into proximate and ultimate causes that are concerned with the how and why respectively (Mayr, 1961). Next I will consider the speciation question for each of them separately.

Proximate explanations deal with mechanisms, and their characterization require controlled experiments. Here, speciation studies are facing an exciting time with the advent of gene editing technologies like CRIPSR-Cas9 that have radically eased the manipulation of genetic material.

The creators of the first synthetic biological cell encoded into its DNA sequence: "What I cannot build I cannot understand" (Saenz, 2010). Indeed, tinkering with the DNA in such a way that we can build or even tear down reproductive barriers in the lab would prove our mechanical understanding of them. Yet such experiments alone would not be able to speak to why these barriers came to be in nature. Moreover, they would fall short of addressing evolutionary change over longer timespans. A different form of experiments, carried out in the computer as simulations, have gained more traction as they can simulate the evolution of populations, individuals and their DNA sequences under different scenarios and identify the conditions that facilitate or promote evolution of reproductive isolation. Even so, these computer simulations will be caricatures of the natural world at best, and neither simulation nor experiment can replace natural history in the foreseeable future. Instead, they could be recruited as tools to form an integrative approach for the study of evolution. Systems like *Trichaptum* are promising in this regard as they have large natural populations, are amenable to experimental manipulation, and their small genomes simplify computer simulations.

The question pertaining to ultimate cause in speciation may at first glance seem straight forward to answer, considering, as Dobzhansky summarized, how "nothing in biology makes sense except is in the light of evolution" (Dobzhansky, 1973). But digging into the ultimate cause of evolutionary change itself confronts us with an apparent paradox: we query a reason for a process that is essentially aimless. In his thought experiment "replaying life´s tape" Stephen Jay Gould asks whether we can expect to have the same organisms today if we went back hundreds of millions of years and let it play out once again (Gould, 1989), making the point that it may very well turn out differently due to stochastic processes. Evolution is woven together by both adaptation and historical contingency, making every speciation event unique. In the sense that one can speak of cause in speciation more specifically than evolution at large, there is not one cause, but multitudes entangled together. Teasing them apart may be nearly impossible, but our legacy of categorical thinking prompts us to make the attempt nonetheless. We classify speciation as either being sympatric or allopatric, we contrast selection against genetic drift, and characterize reproductive isolation as pre- or postzygotic. These frameworks direct, but also constrain, the questions we ask and what observations we choose to make. Nevertheless, they are undeniably useful, and may even help us appreciate the unity of life on earth by revealing how vastly different lifeforms separated by billions of years of evolution, all originated from, and are subject to the same basic evolutionary processes by which "endless forms most beautiful and most wonderful have been, and are being evolved" (Darwin, 1859).

References

AANEN, D. K., KUYPER, T. W., MES, T. H. & HOEKSTRA, R. F. 2000. The evolution of reproductive isolation in the ectomycorrhizal Hebeloma crustuliniforme aggregate (Basidiomycetes) in northwestern Europe: a phylogenetic approach. *Evolution,* 54**,** 1192-206.

ABBOTT, R. J., HEGARTY, M. J., HISCOCK, S. J. & BRENNAN, A. C. 2010. Homoploid hybrid speciation in action. *Taxon,* 59**,** 1375-1386.

ALEXANDER, D. H., NOVEMBRE, J. & LANGE, K. 2009. Fast model-based estimation of ancestry in unrelated individuals. *Genome Res,* 19**,** 1655-64.

AMENT-VELÁSQUEZ, S. L., VOGAN, A. A., GRANGER-FARBOS, A., BASTIAANS, E., MARTINOSSI-ALLIBERT, I., SAUPE, S. J., DE GROOT, S., LASCOUX, M., DEBETS, A. J. M., CLAVÉ, C. & JOHANNESSON, H. 2022. Allorecognition genes drive reproductive isolation in Podospora anserina. *Nature Ecology & Evolution,* 6**,** 910-923.

ANDERSON, J. B., FUNT, J., THOMPSON, D. A., PRABHU, S., SOCHA, A., SIRJUSINGH, C., DETTMAN, J. R., PARREIRAS, L., GUTTMAN, D. S., REGEV, A. & KOHN, L. M. 2010. Determinants of divergent adaptation and Dobzhansky-Muller interaction in experimental yeast populations. *Curr Biol,* 20**,** 1383-8.

ANDERSON, J. B. & KOHN, L. M. 2007. Dikaryons, Diploids, and Evolution. *Sex in Fungi.*

ANDERSON, J. B., KORHONEN, K. & ULLRICH, R. C. 1980. Relationships between European and North American biological species of Armillaria mellea. *Experimental Mycology,* 4**,** 78-86.

ANTONELLI, A., FRY, C., SMITH, R. J., SIMMONDS, M. S. J., KERSEY, P. J., PRITCHARD, H. W., ABBO, M. S., ACEDO, C., ADAMS, J., AINSWORTH, A. M., ALLKIN, B., ANNECKE, W., BACHMAN, S. P., BACON, K., BÁRRIOS, S., BARSTOW, C., BATTISON, A., BELL, E., BENSUSAN, K., BIDARTONDO, M. I., BLACKHALL-MILES, R. J., BORRELL, J. S., BREARLEY, F. Q., BREMAN, E., BREWER, R. F. A., BRODIE, J., CÁMARA-LERET, R., CAMPOSTRINI FORZZA, R., CANNON, P., CARINE, M., CARRETERO, J., CAVAGNARO, T. R., CAZAR, M.-E., CHAPMAN, T., CHEEK, M., CLUBBE, C., COCKEL, C., COLLEMARE, J., COOPER, A., COPELAND, A. I., CORCORAN, M., COUCH, C., COWELL, C., CROUS, P., DA SILVA, M., DALLE, G., DAS, D., DAVID, J. C., DAVIES, L., DAVIES, N., DE CANHA, M. N., DE LIRIO, E. J., DEMISSEW, S., DIAZGRANADOS, M., DICKIE, J., DINES, T., DOUGLAS, B., DRÖGE, G., DULLOO, M. E., FANG, R., FARLOW, A., FARRAR, K., FAY, M. F., FELIX, J., FOREST, F., FORREST, L. L., FULCHER, T., GAFFOROV, Y., GARDINER, L. M., GÂTEBLÉ, G., GAYA, E., GESLIN, B., GONÇALVES, S. C., GORE, C. J. N., GOVAERTS, R., GOWDA, B., GRACE, O. M., GRALL, A., HAELEWATERS, D., HALLEY, J. M., HAMILTON, M. A., HAZRA, A., HELLER, T., HOLLINGSWORTH, P. M., HOLSTEIN, N., HOWES, M.-J. R., HUGHES, M., HUNTER, D., HUTCHINSON, N., HYDE, K., IGANCI, J., JONES, M., KELLY, L. J., KIRK, P., KOCH, H., GRISAI-GREILHUBER, I., LALL, N., LANGAT, M. K., LEAMAN, D. J., LEÃO, T. C., et al. 2020. State of the World's Plants and Fungi 2020. Royal Botanic Gardens, Kew.

AUXIER, B., SCHOLTMEIJER, K., VAN PEER, A. F., BAARS, J. J. P., DEBETS, A. J. M. & AANEN, D. K. 2021. Cytoplasmic Mixing, Not Nuclear Coexistence, Can Explain Somatic Incompatibility in Basidiomycetes. *Microorganisms,* 9.

AVISE, J. C., WALKER, D. & JOHNS, G. C. 1998. Speciation durations and Pleistocene effects on vertebrate phylogeography. *Proceedings. Biological sciences,* 265**,** 1707-1712.

BAKER, A. J. 2008. Islands in the sky: the impact of Pleistocene climate cycles on biodiversity. *Journal of biology,* 7**,** 32-32.

BATCHELOR, C. L., MARGOLD, M., KRAPP, M., MURTON, D. K., DALTON, A. S., GIBBARD, P. L., STOKES, C. R., MURTON, J. B. & MANICA, A. 2019. The configuration of Northern Hemisphere ice sheets through the Quaternary. *Nature Communications,* 10**,** 3713.

BERBEE, M. L., STRULLU-DERRIEN, C., DELAUX, P.-M., STROTHER, P. K., KENRICK, P., SELOSSE, M.-A. & TAYLOR, J. W. 2020. Genomic and fossil windows into the secret lives of the most ancient fungi. *Nature Reviews Microbiology,* 18**,** 717-730.

BEZMENOVA, A. V., ZVYAGINA, E. A., FEDOTOVA, A. V., KASIANOV, A. S., NERETINA, T. V., PENIN, A. A., BAZYKIN, G. A. & KONDRASHOV, A. S. 2020. Rapid Accumulation of Mutations in Growing Mycelia of a Hypervariable Fungus Schizophyllum commune. *Molecular Biology and Evolution,* 37**,** 2279-2286.

BLACKWELL, M. 2011. The Fungi: 1, 2, 3 ... 5.1 million species? *American Journal of Botany,* 98**,** 426-38.

BLUNT, W. 2004. *Linnaeus: The Compleat Naturalist*, Frances Lincoln.

BOLNICK, D. I. & FITZPATRICK, B. M. 2007. Sympatric Speciation: Models and Empirical Evidence. *Annual Review of Ecology, Evolution, and Systematics,* 38**,** 459-487.

BRANCO, S., BI, K., LIAO, H. L., GLADIEUX, P., BADOUIN, H., ELLISON, C. E., NGUYEN, N. H., VILGALYS, R., PEAY, K. G., TAYLOR, J. W. & BRUNS, T. D. 2017. Continental-level population differentiation and environmental adaptation in the mushroom Suillus brevipes. *Mol Ecol,* 26**,** 2063-76.

BRANCO, S., GLADIEUX, P., ELLISON, C. E., KUO, A., LABUTTI, K., LIPZEN, A., GRIGORIEV, I. V., LIAO, H.-L., VILGALYS, R., PEAY, K. G., TAYLOR, J. W. & BRUNS, T. D. 2015. Genetic isolation between two recently diverged populations of a symbiotic fungus. *Molecular Ecology,* 24**,** 2747-2758.

BRANDT, Y. C., WEI, X., DENG, Y., VAUGHN, A. H. & NIELSEN, R. 2022. Evaluation of methods for estimating coalescence times using ancestral recombination graphs. *Genetics,* 221.

BRISBIN, A., BRYC, K., BYRNES, J., ZAKHARIA, F., OMBERG, L., DEGENHARDT, J., REYNOLDS, A., OSTRER, H., MEZEY, J. G. & BUSTAMANTE, C. D. 2012. PCAdmix: principal components-based assignment of ancestry along each chromosome in individuals with admixed ancestry from two or more populations. *Hum Biol,* 84**,** 343-64.

BROWN, A. J. & CASSELTON, L. A. 2001. Mating in mushrooms: increasing the chances but prolonging the affair. *Trends Genet,* 17**,** 393-400.

BUSS, L. W. 1982. Somatic cell parasitism and the evolution of somatic tissue compatibility. *Proc Natl Acad Sci U S A,* 79**,** 5337-41.

BUTLIN, R., GALINDO, J. & GRAHAME, J. 2008. Review. Sympatric, parapatric or allopatric: The most important way to classify speciation? *Philosophical transactions of the Royal Society of London. Series B, Biological sciences,* 363**,** 2997-3007.

BUTLIN, R. K. & SMADJA, C. M. 2017. Coupling, Reinforcement, and Speciation. *The American Naturalist,* 191**,** 155-172.

CAMPBELL, C. R., POELSTRA, J. W. & YODER, A. D. 2018. What is Speciation Genomics? The roles of ecology, gene flow, and genomic architecture in the formation of species. *Biological Journal of the Linnean Society,* 124**,** 561-583.

CASADEVALL, A., FREIJ, J. B., HANN-SODEN, C. & TAYLOR, J. 2017. Continental Drift and Speciation of the Cryptococcus neoformans and Cryptococcus gattii Species Complexes. *mSphere,* 2.

CASSELTON, L. & BETTS, N. 1998. Casselton, LA, Olesnicky, NL. Molecular genetics of mating recognition in basidiomycete fungi. Microbiol Mol Biol Rev, 62: 55-70. *Microbiology and molecular biology reviews : MMBR,* 62**,** 55-70.

CASSELTON, L. A. 2002. Mate recognition in fungi. *Heredity,* 88**,** 142-147.

CEBALLOS, G., EHRLICH, P. R., BARNOSKY, A. D., GARCÍA, A., PRINGLE, R. M. & PALMER, T. M. Accelerated modern human–induced species losses: Entering the sixth mass extinction. *Science Advances,* 1**,** e1400253.

CHARRON, G., LEDUCQ, J.-B. & LANDRY, C. R. 2014. Chromosomal variation segregates within incipient species and correlates with reproductive isolation. *Molecular Ecology,* 23**,** 4362-4372.

CHASE, T. E. & ULLRICH, R. C. 1990. Five Genes Determining Intersterility in Heterobasidion annosum. *Mycologia,* 82**,** 73-81.

CHEN, L. & WIENS, J. J. 2021. Multicellularity and sex helped shape the Tree of Life. *Proc Biol Sci,* 288**,** 20211265.

CORBETT-DETIG, R. & NIELSEN, R. 2017. A Hidden Markov Model Approach for Simultaneously Estimating Local Ancestry and Admixture Time Using Next Generation Sequence Data in Samples of Arbitrary Ploidy. *PLOS Genetics,* 13**,** e1006529.

COWIE, R. H., BOUCHET, P. & FONTAINE, B. 2022. The Sixth Mass Extinction: fact, fiction or speculation? *Biological Reviews,* 97**,** 640-663.

COYNE, J. A. & ORR, H. A. 1989. PATTERNS OF SPECIATION IN DROSOPHILA. *Evolution,* 43**,** 362-381.

COYNE, J. A. & ORR, H. A. 1997. "PATTERNS OF SPECIATION IN DROSOPHILA" REVISITED. *Evolution,* 51**,** 295-303.

COYNE, J. A. & ORR, H. A. 2004. *Speciation*, Sinauer Associates Sunderland, MA.

CROW, J. F. 1994. Advantages of sexual reproduction. *Developmental Genetics,* 15**,** 205-213.

DAI, Y.-C. & CUI, B.-K. 2011. Fomitiporia ellipsoidea has the largest fruiting body among the fungi. *Fungal Biology,* 115**,** 813-814.

DALMAN, K., OLSON, A. & STENLID, J. 2010. Evolutionary history of the conifer root rot fungus Heterobasidion annosum sensu lato. *Mol Ecol,* 19**,** 4979-93.

DARWIN, C. 1859. *On the Origin of Species by Means of Natural Selection, or Preservation of Favoured Races in the Struggle for Life,* London.

DARWIN, C. R. 1856. On the variety of species definitions prevalent among naturalists. *Letter from C. R. Darwin to J. D. Hooker 24 December [1856]*

DAWKINS, R. 2006. *The Selfish Gene,* London, England, Oxford University Press.

DE WIT, R. & BOUVIER, T. 2006. 'Everything is everywhere, but, the environment selects'; what did Baas Becking and Beijerinck really say? *Environmental Microbiology,* 8**,** 755-758.

DETTMAN, J. R., ANDERSON, J. B. & KOHN, L. M. 2008. Divergent adaptation promotes reproductive isolation among experimental populations of the filamentous fungus Neurospora. *BMC Evolutionary Biology,* 8**,** 35.

DETTMAN, J. R., JACOBSON, D. J., TURNER, E., PRINGLE, A. & TAYLOR, J. W. 2003. Reproductive Isolation and Phylogenetic Divergence in Neurospora: Comparing Methods of Species Recognition in a Model Eukaryote. *Evolution,* 57**,** 2721-2741.

DIGHTON, J. 1995. Nutrient cycling in different terrestrial ecosystems in relation to fungi. *Canadian Journal of Botany,* 73**,** 1349-1360.

DOBZHANSKY, T. 1973. Nothing in Biology Makes Sense except in the Light of Evolution. *The American Biology Teacher,* 35**,** 125-129.

FERGUSON, B. A., DREISBACH, T. A., PARKS, C. G., FILIP, G. M. & SCHMITT, C. L. 2003. Coarse-scale population structure of pathogenic*Armillaria*species in a mixed-conifer forest in the Blue Mountains of northeast Oregon. *Canadian Journal of Forest Research,* 33**,** 612-623.

FEURTEY, A., LORRAIN, C., MCDONALD, M. C., MILGATE, A., SOLOMON, P. S., WARREN, R., PUCCETTI, G., SCALLIET, G., TORRIANI, S. F. F., GOUT, L., MARCEL, T. C., SUFFERT, F., ALASSIMONE, J., LIPZEN, A., YOSHINAGA, Y., DAUM, C., BARRY, K., GRIGORIEV, I. V., GOODWIN, S. B., GENISSEL, A., SEIDL, M. F., STUKENBROCK, E. H., LEBRUN, M.-H., KEMA, G. H. J., MCDONALD, B. A. & CROLL, D. 2023. A thousand-genome panel retraces the global spread and adaptation of a major fungal crop pathogen. *Nature Communications,* 14**,** 1059.

FITZPATRICK, B. M. 2004. Rates of Evolution of Hybrid Inviability in Birds and Mammals. *Evolution,* 58**,** 1865-1870.

FUKASAWA, Y. 2021. Ecological impacts of fungal wood decay types: A review of current knowledge and future research directions. *Ecological Research,* 36**,** 910-931.

GAC, M. L. & GIRAUD, T. 2008. Existence of a pattern of reproductive character displacement in Homobasidiomycota but not in Ascomycota. *Journal of Evolutionary Biology,* 21**,** 761-772.

GARBELOTTO, M. & GONTHIER, P. 2013. Biology, Epidemiology, and Control of Heterobasidion Species Worldwide. *Annual Review of Phytopathology,* 51**,** 39-59.

GIBBS, K. A., URBANOWSKI, M. L. & GREENBERG, E. P. 2008. Genetic determinants of self identity and social recognition in bacteria. *Science,* 321**,** 256-9.

GIORDANO, L., SILLO, F., GARBELOTTO, M. & GONTHIER, P. 2018. Mitonuclear interactions may contribute to fitness of fungal hybrids. *Scientific Reports,* 8**,** 1706.

GIRAUD, T. & GOURBIÈRE, S. 2012. The tempo and modes of evolution of reproductive isolation in fungi. *Heredity,* 109**,** 204-214.

GIRAUD, T., REFREGIER, G., LE GAC, M., DE VIENNE, D. M. & HOOD, M. E. 2008. Speciation in fungi. *Fungal genetics and biology,* 45**,** 791-802.

GLADIEUX, P., DE BELLIS, F., HANN-SODEN, C., SVEDBERG, J., JOHANNESSON, H. & TAYLOR, J. W. 2020. Neurospora from Natural Populations: Population Genomics Insights into the Life History of a Model Microbial Eukaryote. *Methods Mol Biol,* 2090**,** 313-336.

GLADIEUX, P., VERCKEN, E., FONTAINE, M. C., HOOD, M. E., JONOT, O., COULOUX, A. & GIRAUD, T. 2010. Maintenance of Fungal Pathogen Species That Are Specialized to Different Hosts: Allopatric Divergence and Introgression through Secondary Contact. *Molecular Biology and Evolution,* 28**,** 459-471.

GOLAN, J. J. & PRINGLE, A. 2017. Long-Distance Dispersal of Fungi. *Microbiol Spectr,* 5.

GORDON, S. A. & PETERSEN, R. H. 1997. Infraspecific variation among geographically separated collections of Marasmius androsaceus. *Mycological Research,* 101**,** 365-371.

GOULD, S. J. 1989. *Wonderful Life: The Burgess Shale and the Nature of History,* New York, Norton.

GOULD, S. J. & ELDREDGE, N. 1993. Punctuated equilibrium comes of age. *Nature,* 366**,** 223-227.

GREEN, R. E., KRAUSE, J., BRIGGS, A. W., MARICIC, T., STENZEL, U., KIRCHER, M., PATTERSON, N., LI, H., ZHAI, W., FRITZ, M. H., HANSEN, N. F., DURAND, E. Y., MALASPINAS, A. S., JENSEN, J. D., MARQUES-BONET, T., ALKAN, C., PRÜFER, K., MEYER, M., BURBANO, H. A., GOOD, J. M., SCHULTZ, R., AXIMU-PETRI, A., BUTTHOF, A., HÖBER, B., HÖFFNER, B., SIEGEMUND, M., WEIHMANN, A., NUSBAUM, C., LANDER, E. S., RUSS, C., NOVOD, N., AFFOURTIT, J., EGHOLM, M., VERNA, C., RUDAN, P., BRAJKOVIC, D., KUCAN, Ž., GUŠIC, I., DORONICHEV, V. B., GOLOVANOVA, L. V., LALUEZA-FOX, C., DE LA RASILLA, M., FORTEA, J., ROSAS, A., SCHMITZ, R. W., JOHNSON, P. L. F., EICHLER, E. E., FALUSH, D., BIRNEY, E., MULLIKIN, J. C., SLATKIN, M., NIELSEN, R., KELSO, J., LACHMANN, M., REICH,

D. & PÄÄBO, S. 2010. A draft sequence of the Neandertal genome. *Science,* 328**,** 710-722.

HAIGHT, J. E., NAKASONE, K. K., LAURSEN, G. A., REDHEAD, S. A., TAYLOR, D. L. & GLAESER, J. A. 2019. Fomitopsis mounceae and F. schrenkii—two new species from North America in the F. pinicola complex. *Mycologia,* 111**,** 339-357.

HARDIN, G. 1960. The Competitive Exclusion Principle. *Science,* 131**,** 1292-1297.

HARTL, D. L. & CLARK, A. G. 2007. *Principles of Population Genetics,* Sunderland, MA, USA, Sinauer.

HAWKSWORTH, D. L. & LUCKING, R. 2017. Fungal Diversity Revisited: 2.2 to 3.8 Million Species. *Microbiol Spectrum,* 5.

HEDGES, S. B., MARIN, J., SULESKI, M., PAYMER, M. & KUMAR, S. 2015. Tree of Life Reveals Clock-Like Speciation and Diversification. *Molecular Biology and Evolution,* 32**,** 835-845.

HEITMAN, J., SUN, S. & JAMES, T. Y. 2013. Evolution of fungal sexual reproduction. *Mycologia,* 105**,** 1-27.

HEWITT, G. 2000. The genetic legacy of the Quaternary ice ages. *Nature,* 405**,** 907.

HEWITT, G. M. 1999. Post-glacial re-colonization of European biota. *Biological Journal of the Linnean Society,* 68**,** 87-112.

HEWITT, G. M. 2001. Speciation, hybrid zones and phylogeography — or seeing genes in space and time. *Molecular Ecology,* 10**,** 537-549.

HEY, J. 2001. The mind of the species problem. *Trends in Ecology & Evolution,* 16**,** 326-329.

HØILAND, K. 2003. Olav Sopp, den glemte grunnlegger av soppriket. *Blekksoppen***,** 10-16.

HOWARD, D. J. 1993. Reinforcement: Origin, dynamics, and fate of an evolutionary hypothesis. *In:* HARRISON, R. G. (ed.) *Hybrid Zones and the Evolutionary Process.* Oxford University Press.

JAMES, T. Y. 2007. Analysis of mating-type locus organization and synteny in mushroom fungi-beyond model species. *In:* J. HEITMAN, J. K., J. W. TAYLOR, AND L. A. CASSELTON (ed.) *Sex in fungi: molecular determination and evolutionary implications.* Washington DC: ASM Pres.

JAMES, T. Y., PORTER, D., HAMRICK, J. L. & VILGALYS, R. 1999. EVIDENCE FOR LIMITED INTERCONTINENTAL GENE FLOW IN THE COSMOPOLITAN MUSHROOM, SCHIZOPHYLLUM COMMUNE. *Evolution,* 53**,** 1665-1677.

JIGGINS, C. D. 2019. Can genomics shed light on the origin of species? *PLOS Biology,* 17**,** e3000394.

KADEREIT, J. W. & ABBOTT, R. J. 2021. Plant speciation in the Quaternary. *Plant Ecology & Diversity,* 14**,** 105-142.

KAUSERUD, H., SAETRE, G. P., SCHMIDT, O., DECOCK, C. & SCHUMACHER, T. 2006. Genetics of self/nonself recognition in Serpula lacrymans. *Fungal Genet Biol,* 43**,** 503-10.

KAUSERUD, H. & SCHUMACHER, T. 2003. Regional and local population structure of the pioneer wood-decay fungus Trichaptum abietinum. *Mycologia,* 95**,** 416-425.

KENNEDY, P. G., GARIBAY-ORIJEL, R., HIGGINS, L. M. & ANGELES-ARGUIZ, R. 2011. Ectomycorrhizal fungi in Mexican Alnus forests support the host co-migration hypothesis and continental-scale patterns in phylogeography. *Mycorrhiza,* 21**,** 559-568.

KIMURA, M. 1983. *The Neutral Theory of Molecular Evolution,* Cambridge, Cambridge University Press.

KIMURA, M. 1991. The neutral theory of molecular evolution: a review of recent evidence. *Jpn J Genet,* 66**,** 367-86.

KIRKPATRICK, M. & RAVIGNÉ, V. 2002. Speciation by natural and sexual selection: models and experiments. *Am Nat,* 159 Suppl 3**,** S22-35.

KO, K. S., HONG, S. G. & JUNG, H. S. 1997. Phylogenetic analysis of Trichaptum based on nuclear 18S, 5.8S and ITS ribosomal DNA sequences. *Mycologia,* 89**,** 727-734.

KO, K. S. & JUNG, H. S. 2002. Three nonorthologous ITS1 types are present in a polypore fungus Trichaptum abietinum. *Molecular Phylogenetics and Evolution,* 23**,** 112-22.

KOHN, L. M. 2005. Mechanisms of fungal speciation. *Annu Rev Phytopathol,* 43**,** 279-308.

LARSSON, K. H., PARMASTO, E., FISCHER, M., LANGER, E., NAKASONE, K. K. & REDHEAD, S. A. 2006. Hymenochaetales: a molecular phylogeny for the hymenochaetoid clade. *Mycologia,* 98**,** 926-36.

LEDUCQ, J. B., NIELLY-THIBAULT, L., CHARRON, G., EBERLEIN, C., VERTA, J.-P., SAMANI, P., SYLVESTER, K., HITTINGER, C. T., BELL, G. & LANDRY, C. R. 2016. Speciation driven by hybridization and chromosomal plasticity in a wild yeast. *Nature Microbiology,* 1**,** 15003.

LI, Y., TADA, F., YAMASHIRO, T. & MAKI, M. 2016. Long-term persisting hybrid swarm and geographic difference in hybridization pattern: genetic consequences of secondary contact between two Vincetoxicum species (Apocynaceae–Asclepiadoideae). *BMC Evolutionary Biology,* 16**,** 20.

LIBRADO, P., KHAN, N., FAGES, A., KUSLIY, M. A., SUCHAN, T., TONASSO-CALVIÈRE, L., SCHIAVINATO, S., ALIOGLU, D., FROMENTIER, A., PERDEREAU, A., AURY, J.-M., GAUNITZ, C., CHAUVEY, L., SEGUIN-ORLANDO, A., DER SARKISSIAN, C., SOUTHON, J., SHAPIRO, B., TISHKIN, A. A., KOVALEV, A. A., ALQURAISHI, S., ALFARHAN, A. H., AL-RASHEID, K. A. S., SEREGÉLY, T., KLASSEN, L., IVERSEN, R., BIGNON-LAU, O., BODU, P., OLIVE, M., CASTEL, J.-C., BOUDADI-MALIGNE, M., ALVAREZ, N., GERMONPRÉ, M., MOSKAL-DEL HOYO, M., WILCZYŃSKI, J., POSPUŁA, S., LASOTA-KUŚ, A., TUNIA, K., NOWAK, M., RANNAMÄE, E., SAARMA, U., BOESKOROV, G., LÕUGAS, L., KYSELÝ, R., PEŠKE, L., BĂLĂŞESCU, A., DUMITRAŞCU, V., DOBRESCU, R., GERBER, D., KISS, V., SZÉCSÉNYI-NAGY, A., MENDE, B. G., GALLINA, Z., SOMOGYI, K., KULCSÁR, G., GÁL, E., BENDREY, R., ALLENTOFT, M. E., SIRBU, G., DERGACHEV, V., SHEPHARD, H., TOMADINI, N., GROUARD, S., KASPAROV, A., BASILYAN, A. E., ANISIMOV, M. A., NIKOLSKIY, P. A., PAVLOVA, E. Y., PITULKO, V., BREM, G., WALLNER, B., SCHWALL, C., KELLER, M., KITAGAWA, K., BESSUDNOV, A. N., BESSUDNOV, A., TAYLOR, W., MAGAIL, J., GANTULGA, J.-O., BAYARSAIKHAN, J., ERDENEBAATAR, D., TABALDIEV, K., MIJIDDORJ, E., BOLDGIV, B., TSAGAAN, T., PRUVOST, M., OLSEN, S., MAKAREWICZ, C. A., VALENZUELA LAMAS, S., ALBIZURI CANADELL, S., NIETO ESPINET, A., IBORRA, M. P., LIRA GARRIDO, J., RODRÍGUEZ GONZÁLEZ, E., CELESTINO, S., OLÀRIA, C., ARSUAGA, J. L., KOTOVA, N., PRYOR, A., CRABTREE, P., ZHUMATAYEV, R., et al. 2021. The origins and spread of domestic horses from the Western Eurasian steppes. *Nature,* 598**,** 634-640.

LIOU, L. W. & PRICE, T. D. 1994. Speciation by Reinforcement of Premating Isolation. *Evolution,* 48**,** 1451-1459.

LOUIS, E. J. 2011. Population genomics and speciation in yeasts. *Fungal Biology Reviews,* 25**,** 136-142.

LUNDE, L. F., BODDY, L., SVERDRUP-THYGESON, A., JACOBSEN, R. M., KAUSERUD, H. & BIRKEMOE, T. 2023. Beetles provide directed dispersal of viable spores of a keystone wood decay fungus. *Fungal Ecology,* 63**,** 101232.

LUOMA, D. L., DURALL, D. M., EBERHART, J. L. & SIDLAR, K. 2011. Rediscovery of the vesicles that characterized Rhizopogon vesiculosus. *Mycologia,* 103**,** 1074-9.

MACRAE, R. 1967. Pairing incompatibility and other distinctions among Hirschiporus (Polyporus) abietinus, H. fusco-violaceus, and H. laricinus. *Canadian Journal of Botany,* 45**,** 1371-1398.

MADDISON, W. P. 1997. Gene Trees in Species Trees. *Systematic Biology,* 46**,** 523-536.

MAGASI, L. P. 1976. Incompatibility factors in Polyporus abietinus, their numbers and distribution. *Mem. of the New York Botanical Garden,* 28**,** 163-173.

MASCHER, M., WICKER, T., JENKINS, J., PLOTT, C., LUX, T., KOH, C. S., ENS, J., GUNDLACH, H., BOSTON, L. B., TULPOVÁ, Z., HOLDEN, S., HERNÁNDEZ-PINZÓN, I., SCHOLZ, U., MAYER, K. F. X., SPANNAGL, M., POZNIAK, C. J., SHARPE, A. G., ŠIMKOVÁ, H., MOSCOU, M. J., GRIMWOOD, J., SCHMUTZ, J. & STEIN, N. 2021. Long-read sequence assembly: a technical evaluation in barley. *The Plant Cell,* 33**,** 1888-1906.

MAURICE, S., ARNAULT, G., NORDÉN, J., BOTNEN, S. S., MIETTINEN, O. & KAUSERUD, H. 2021. Fungal sporocarps house diverse and host-specific communities of fungicolous fungi. *Isme j,* 15**,** 1445-1457.

MAY, G., SHAW, F., BADRANE, H. & VEKEMANS, X. 1999. The signature of balancing selection: Fungal mating compatibility gene evolution. *Proceedings of the National Academy of Sciences,* 96**,** 9172-9177.

MAYR, E. 1961. Cause and Effect in Biology. *Science,* 134**,** 1501-1506.

MAYR, E. 1982. *The growth of biological thought : diversity evolution and inheritance,* Cambridge Massachutes, Belknap Press.

MCCUNE, A. R. & LOVEJOY, N. R. 1998. The relative rate of sympatric and allopatric speciation in fishes. *In:* HOWARD, D. J. & BERLOCHER, S. H. (eds.) *Endless Forms: Species and Speciation.* New York: Oxford University Press.

MENOLLI, N., JR., SÁNCHEZ-RAMÍREZ, S., SÁNCHEZ-GARCÍA, M., WANG, C., PATEV, S., ISHIKAWA, N. K., MATA, J. L., LENZ, A. R., VARGAS-ISLA, R., LIDERMAN, L., LAMB, M., NUHN, M., HUGHES, K. W., XIAO, Y. & HIBBETT, D. S. 2022. Global phylogeny of the Shiitake mushroom and related Lentinula species uncovers novel diversity and suggests an origin in the Neotropics. *Molecular Phylogenetic and Evolution,* 173**,** 107494.

METHLIE, I.-S. H. 2021. *Investigating local adaptation in the wood-decay fungus Trichaptum abietinum through common garden experiments and population genommics.* Master of Science, University of Oslo.

MONCALVO, J.-M. & BUCHANAN, P. K. 2008. Molecular evidence for long distance dispersal across the Southern Hemisphere in the Ganoderma applanatum-australe species complex (Basidiomycota). *Mycological Research,* 112**,** 425-436.

MOYLE, L. C., OLSON, M. S. & TIFFIN, P. 2004. PATTERNS OF REPRODUCTIVE ISOLATION IN THREE ANGIOSPERM GENERA. *Evolution,* 58**,** 1195-1208.

MUGHAL, M. R. & DEGIORGIO, M. 2022. Properties and unbiased estimation of F- and D-statistics in samples containing related and inbred individuals. *Genetics,* 220**,** iyab090.

MUJIC, A. B., HUANG, B., CHEN, M.-J., WANG, P.-H., GERNANDT, D. S., HOSAKA, K. & SPATAFORA, J. W. 2019. Out of western North America: Evolution of the Rhizopogon-

Pseudotsuga symbiosis inferred by genome-scale sequence typing. *Fungal Ecology,* 39**,** 12-25.

MURPHY, H. A. & ZEYL, C. W. 2015. A Potential Case of Reinforcement in a Facultatively Sexual Unicellular Eukaryote. *The American Naturalist,* 186**,** 312-319.

MURRILL, W. A. 1904. The Polyporaceae of North America: IX. Inonotus, Sesia and monotypic genera. *Bulletin of the Torrey Botanical Club,* 31**,** 593-610.

NIEUWENHUIS, B. P. S., BILLIARD, S., VUILLEUMIER, S., PETIT, E., HOOD, M. E. & GIRAUD, T. 2013a. Evolution of uni- and bifactorial sexual compatibility systems in fungi. *Heredity,* 111**,** 445-455.

NIEUWENHUIS, B. P. S., NIEUWHOF, S. & AANEN, D. K. 2013b. On the asymmetry of mating in natural populations of the mushroom fungus Schizophyllum commune. *Fungal Genetics and Biology,* 56**,** 25-32.

NOOR, M. A. F. & BENNETT, S. M. 2009. Islands of speciation or mirages in the desert? Examining the role of restricted recombination in maintaining species. *Heredity,* 103**,** 439-444.

NORROS, V., KARHU, E., NORDÉN, J., VÄHÄTALO, A. V. & OVASKAINEN, O. 2015. Spore sensitivity to sunlight and freezing can restrict dispersal in wood-decay fungi. *Ecol Evol,* 5**,** 3312-26.

NOSIL, P. & SCHLUTER, D. 2011. The genes underlying the process of speciation. *Trends Ecol Evol,* 26**,** 160-7.

ORR, H. A. 1996. Dobzhansky, Bateson, and the genetics of speciation. *Genetics,* 144**,** 1331-1335.

ORR, H. A. & TURELLI, M. 2001. The evolution of postzygotic isolation: accumulating Dobzhansky-Muller incompatibilities. *Evolution,* 55**,** 1085-94.

PAGEL, M., VENDITTI, C. & MEADE, A. 2006. Large Punctuational Contribution of Speciation to Evolutionary Divergence at the Molecular Level. *Science,* 314**,** 119-121.

PALOPOLI, M. F. & WU, C. I. 1994. Genetics of hybrid male sterility between drosophila sibling species: a complex web of epistasis is revealed in interspecific studies. *Genetics,* 138**,** 329-41.

PATTERSON, N., MOORJANI, P., LUO, Y., MALLICK, S., ROHLAND, N., ZHAN, Y., GENSCHORECK, T., WEBSTER, T. & REICH, D. 2012. Ancient Admixture in Human History. *Genetics,* 192**,** 1065-1093.

PATTERSON, N., PRICE, A. L. & REICH, D. 2006. Population Structure and Eigenanalysis. *PLOS Genetics,* 2**,** e190.

PEAY, K. G., KENNEDY, P. G. & TALBOT, J. M. 2016. Dimensions of biodiversity in the Earth mycobiome. *Nature Reviews Microbiology,* 14**,** 434-447.

PEAY, K. G., SCHUBERT, M. G., NGUYEN, N. H. & BRUNS, T. D. 2012. Measuring ectomycorrhizal fungal dispersal: macroecological patterns driven by microscopic propagules. *Molecular Ecology,* 21**,** 4122-4136.

PERIS, D., UBBELOHDE, E. J., KUANG, M. C., KOMINEK, J., LANGDON, Q. K., ADAMS, M., KOSHALEK, J. A., HULFACHOR, A. B., OPULENTE, D. A., HALL, D. J., HYMA, K., FAY, J. C., LEDUCQ, J.-B., CHARRON, G., LANDRY, C. R., LIBKIND, D., GONÇALVES, C., GONÇALVES, P., SAMPAIO, J. P., WANG, Q.-M., BAI, F.-Y., WROBEL, R. L. & HITTINGER, C. T. 2023. Macroevolutionary diversity of traits and genomes in the model yeast genus Saccharomyces. *Nature Communications,* 14**,** 690.

PETERSEN, R. H. & HUGHES, K. W. 1998. Mating systems in Omphalotus (Paxillaceae, Agaricales). *Plant Systematics and Evolution,* 211**,** 217-229.

PETERSEN, R. H. & HUGHES, K. W. 1999. Species and Speciation in Mushrooms: Development of a species concept poses difficulties. *BioScience,* 49**,** 440-452.

PIROZYNSKI, K., RAYNER, A., BRASIER, C. & MOORE, D. 1988. Evolutionary Biology of the Fungi. *Bioscience,* 38.

PRESGRAVES, D. C. 2002. PATTERNS OF POSTZYGOTIC ISOLATION IN LEPIDOPTERA. *Evolution,* 56**,** 1168-1183.

PRICE, T. D. & BOUVIER, M. M. 2002. THE EVOLUTION OF F1 POSTZYGOTIC INCOMPATIBILITIES IN BIRDS. *Evolution,* 56**,** 2083-2089.

PRITCHARD, J. K., STEPHENS, M. & DONNELLY, P. 2000. Inference of population structure using multilocus genotype data. *Genetics,* 155**,** 945-59.

RAPER, J. R. 1966. *Genetics of sexuality in higher fungi [by] John R. Raper,* New York, Ronald Press Co.

REICH, D., THANGARAJ, K., PATTERSON, N., PRICE, A. L. & SINGH, L. 2009. Reconstructing Indian population history. *Nature,* 461**,** 489-94.

RUDMAN, S. M. & SCHLUTER, D. 2016. Ecological Impacts of Reverse Speciation in Threespine Stickleback. *Current Biology,* 26**,** 490-495.

RUSTØEN, F., HØILAND, K., HEEGAARD, E., BODDY, L., GANGE, A. C., KAUSERUD, H. & ANDREW, C. 2023. Substrate affinities of wood decay fungi are foremost structured by wood properties not climate. *Fungal Ecology,* 63**,** 101231.

RYVARDEN, L. & MELO, I. 2017. *Poroid fungi of Europe*, Fungiflora.

SAENZ, A. 2010. *Secret Messages Coded Into DNA Of Venter Synthetic Bacteria* [Online]. Singularity hub. Available: https://singularityhub.com/2010/05/24/venters-newest-synthetic-bacteria-has-secret-messages-coded-in-its-dna/ [Accessed 29.05.2023 2023].

SÁNCHEZ-GARCÍA, M., RYBERG, M., KHAN, F. K., VARGA, T., NAGY, L. G. & HIBBETT, D. S. 2020. Fruiting body form, not nutritional mode, is the major driver of diversification in mushroom-forming fungi. *Proc Natl Acad Sci U S A,* 117**,** 32528-32534.

SASA, M. M., CHIPPINDALE, P. T. & JOHNSON, N. A. 1998. PATTERNS OF POSTZYGOTIC ISOLATION IN FROGS. *Evolution,* 52**,** 1811-1820.

SCHLUTER, D. 2000. Ecological Character Displacement in Adaptive Radiation. *The American Naturalist,* 156**,** S4-S16.

SCHLUTER, D. 2009. Evidence for Ecological Speciation and Its Alternative. *Science,* 323**,** 737-741.

SEIERSTAD, K. S., CARLSEN, T., SÆTRE, G.-P., MIETTINEN, O., HELLIK HOFTON, T. & KAUSERUD, H. 2013. A phylogeographic survey of a circumboreal polypore indicates introgression among ecologically differentiated cryptic lineages. *Fungal Ecology,* 6**,** 119-128.

SEIERSTAD, K. S., FOSSDAL, R., MIETTINEN, O., CARLSEN, T., SKREDE, I. & KAUSERUD, H. 2021. Contrasting genetic structuring in the closely related basidiomycetes Trichaptum abietinum and Trichaptum fuscoviolaceum (Hymenochaetales). *Fungal Biology,* 125**,** 269-275.

SERVEDIO, M. R. 2004. The what and why of research on reinforcement. *PLoS Biol,* 2**,** e420.

SERVEDIO, M. R. & NOOR, M. A. F. 2003. The Role of Reinforcement in Speciation: Theory and Data. *Annual Review of Ecology, Evolution, and Systematics,* 34**,** 339-364.

SLATKIN, M. & RACIMO, F. 2016. Ancient DNA and human history. *Proceedings of the National Academy of Sciences,* 113**,** 6380-6387.

SMITH, S. Y., CURRAH, R. S. & STOCKEY, R. A. 2004. Cretaceous and Eocene Poroid Hymenophores from Vancouver Island, British Columbia. *Mycologia,* 96**,** 180-186.

SORAGGI, S., WIUF, C. & ALBRECHTSEN, A. 2018. Powerful Inference with the D-Statistic on Low-Coverage Whole-Genome Data. *G3 Genes|Genomes|Genetics,* 8**,** 551-566.

STANKOWSKI, S. & RAVINET, M. 2021. Quantifying the use of species concepts. *Current Biology,* 31**,** R428-R429.

STENLID, J. & KARLSSON, J.-O. 1991. Partial intersterility in Heterobasidion annosum. *Mycological Research,* 95**,** 1153-1159.

STUKENBROCK, E. H., BATAILLON, T., DUTHEIL, J. Y., HANSEN, T. T., LI, R., ZALA, M., MCDONALD, B. A., WANG, J. & SCHIERUP, M. H. 2011. The making of a new pathogen: insights from comparative population genomics of the domesticated wheat pathogen Mycosphaerella graminicola and its wild sister species. *Genome Research,* 21**,** 2157-66.

STUKENBROCK, E. H., CHRISTIANSEN, F. B., HANSEN, T. T., DUTHEIL, J. Y. & SCHIERUP, M. H. 2012. Fusion of two divergent fungal individuals led to the recent emergence of a unique widespread pathogen species. *Proceedings of the National Academy of Sciences,* 109**,** 10954-10959.

STURMBAUER, C., BARIC, S., SALZBURGER, W., RÜBER, L. & VERHEYEN, E. 2001. Lake Level Fluctuations Synchronize Genetic Divergences of Cichlid Fishes in African Lakes. *Molecular Biology and Evolution,* 18**,** 144-154.

SWENSON, N. & HOWARD, D. 2005. Clustering of Contact Zones, Hybrid Zones, and Phylogeographic Breaks in North America. *The American Naturalist,* 166**,** 581.

TABERLET, P., FUMAGALLI, L., WUST-SAUCY, A. G. & COSSON, J. F. 1998. Comparative phylogeography and postglacial colonization routes in Europe. *Mol Ecol,* 7**,** 453-64.

TAYLOR, J. W., TURNER, E., TOWNSEND, J. P., DETTMAN, J. R. & JACOBSON, D. 2006. Eukaryotic microbes, species recognition and the geographic limits of species: examples from the kingdom Fungi. *Philosophical Transactions of the Royal Society B: Biological Sciences,* 361**,** 1947-1963.

TEMPLETON, A. R. 1981. Mechanisms of Speciation - A Population Genetic Approach. *Annual Review of Ecology and Systematics,* 12**,** 23-48.

VARGA, T., KRIZSÁN, K., FÖLDI, C., DIMA, B., SÁNCHEZ-GARCÍA, M., SÁNCHEZ-RAMÍREZ, S., SZÖLLŐSI, G. J., SZARKÁNDI, J. G., PAPP, V., ALBERT, L., ANDREOPOULOS, W., ANGELINI, C., ANTONÍN, V., BARRY, K. W., BOUGHER, N. L., BUCHANAN, P., BUYCK, B., BENSE, V., CATCHESIDE, P., CHOVATIA, M., COOPER, J., DÄMON, W., DESJARDIN, D., FINY, P., GEML, J., HARIDAS, S., HUGHES, K., JUSTO, A., KARASIŃSKI, D., KAUTMANOVA, I., KISS, B., KOCSUBÉ, S., KOTIRANTA, H., LABUTTI, K. M., LECHNER, B. E., LIIMATAINEN, K., LIPZEN, A., LUKÁCS, Z., MIHALTCHEVA, S., MORGADO, L. N., NISKANEN, T., NOORDELOOS, M. E., OHM, R. A., ORTIZ-SANTANA, B., OVREBO, C., RÁCZ, N., RILEY, R., SAVCHENKO, A., SHIRYAEV, A., SOOP, K., SPIRIN, V., SZEBENYI, C., TOMŠOVSKÝ, M., TULLOSS, R. E., UEHLING, J., GRIGORIEV, I. V., VÁGVÖLGYI, C., PAPP, T., MARTIN, F. M., MIETTINEN, O., HIBBETT, D. S. & NAGY, L. G. 2019. Megaphylogeny resolves global patterns of mushroom evolution. *Nature Ecology & Evolution,* 3**,** 668-678.

VILGALYS, R. 1991. Speciation and Species Concepts in the Collybia dryophila Complex. *Mycologia,* 83**,** 758-773.

VILGALYS, R. & SUN, B. L. 1994. Ancient and recent patterns of geographic speciation in the oyster mushroom Pleurotus revealed by phylogenetic analysis of ribosomal DNA sequences. *Proceedings of the National Academy of Sciences,* 91**,** 4599-4603.

VLASÁK, J. & VLASÁK, J. J. 2017. Trichaptum (Basidiomycota) in tropical America: a sequence study. *Mycosphere,* 8**,** 1217–1227.

VOJE, K. L., DI MARTINO, E. & PORTO, A. 2019. Revisiting a Landmark Study System: No Evidence for a Punctuated Mode of Evolution in Metrarabdotos. *The American Naturalist,* 195**,** 899-917.

WALLACE, A. R. 1858. On the Tendency of Varieties to Depart Indefinitely From the Original Type.

WANG, J., SANTIAGO, E. & CABALLERO, A. 2016. Prediction and estimation of effective population size. *Heredity,* 117**,** 193-206.

WEBB, T. 1995. Pollen Records of Late Quaternary Vegetation Change: Plant Community Rearrangements and Evolutionary Implications. *National Research Council (US) Panel on Effects of Past Global Change on Life. Effects of Past Global Change on Life.* Washington DC: National Academies Press

WHITEHOUSE, H. L. K. 1949. MULTIPLE-ALLELOMORPH HETEROTHALLISM IN THE FUNGI. *New Phytologist,* 48**,** 212-244.

WHITTAKER, R. H. 1959. On the Broad Classification of Organisms. *The Quarterly Review of Biology,* 34**,** 210-226.

WOOD, T. E., TAKEBAYASHI, N., BARKER, M. S., MAYROSE, I., GREENSPOON, P. B. & RIESEBERG, L. H. 2009. The frequency of polyploid speciation in vascular plants. *Proceedings of the National Academy of Sciences of the United States of America,* 106**,** 13875-13879.

YADAV, V., SUN, S., COELHO, M. A. & HEITMAN, J. 2020. Centromere scission drives chromosome shuffling and reproductive isolation. *Proceedings of the National Academy of Sciences,* 117**,** 7917-7928.

ZAFFARANO, P. L., MCDONALD, B. A. & LINDE, C. C. 2008. Rapid speciation following recent host shifts in the plant pathogenic fungus Rhynchosporium. *Evolution,* 62**,** 1418-36.

ZHONG, Z., NORVIENYEKU, J., CHEN, M., BAO, J., LIN, L., CHEN, L., LIN, Y., WU, X., CAI, Z., ZHANG, Q., LIN, X., HONG, Y., HUANG, J., XU, L., ZHANG, H., CHEN, L., TANG, W., ZHENG, H., CHEN, X., WANG, Y., LIAN, B., ZHANG, L., TANG, H., LU, G., EBBOLE, D. J., WANG, B. & WANG, Z. 2016. Directional Selection from Host Plants Is a Major Force Driving Host Specificity in Magnaporthe Species. *Scientific Reports,* 6**,** 25591.

Paper I

PLOS GENETICS
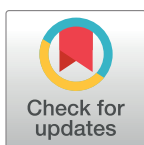
RESEARCH ARTICLE

Large-scale fungal strain sequencing unravels the molecular diversity in mating loci maintained by long-term balancing selection

David Peris[1,2]*, Dabao Sun Lu[1], Vilde Bruhn Kinneberg[1], Ine-Susanne Methlie[1], Malin Stapnes Dahl[1], Timothy Y. James[3], Håvard Kauserud[1], Inger Skrede[1]*

1 Section for Genetics and Evolutionary Biology, Department of Biosciences, University of Oslo, Oslo, Norway, 2 Department of Health, Valencian International University (VIU), Valencia, Spain, 3 Department of Ecology and Evolutionary Biology, University of Michigan, Ann Arbor, Michigan, United States of America

* david.perisnavarro@gmail.com (DP); inger.skrede@ibv.uio.no (IS)

Abstract

Balancing selection, an evolutionary force that retains genetic diversity, has been detected in multiple genes and organisms, such as the sexual mating loci in fungi. However, to quantify the strength of balancing selection and define the mating-related genes require a large number of strains. In tetrapolar basidiomycete fungi, sexual type is determined by two unlinked loci, *MATA* and *MATB*. Genes in both loci define mating type identity, control successful mating and completion of the life cycle. These loci are usually highly diverse. Previous studies have speculated, based on culture crosses, that species of the non-model genus *Trichaptum* (Hymenochaetales, Basidiomycota) possess a tetrapolar mating system, with multiple alleles. Here, we sequenced a hundred and eighty strains of three *Trichaptum* species. We characterized the chromosomal location of *MATA* and *MATB*, the molecular structure of *MAT* regions and their allelic richness. The sequencing effort was sufficient to molecularly characterize multiple *MAT* alleles segregating before the speciation event of *Trichaptum* species. Analyses suggested that long-term balancing selection has generated trans-species polymorphisms. Mating sequences were classified in different allelic classes based on an amino acid identity (AAI) threshold supported by phylogenetics. 17,550 mating types were predicted based on the allelic classes. *In vitro* crosses allowed us to support the degree of allelic divergence needed for successful mating. Even with the high amount of divergence, key amino acids in functional domains are conserved. We conclude that the genetic diversity of mating loci in *Trichaptum* is due to long-term balancing selection, with limited recombination and duplication activity. The large number of sequenced strains highlighted the importance of sequencing multiple individuals from different species to detect the mating-related genes, the mechanisms generating diversity and the evolutionary forces maintaining them.

Author summary

Fungi have complex mating systems, and basidiomycete fungi can encode thousands of mating types. Individuals with the same mating type cannot mate. This sexual system has evolved to facilitate sexual mating with offspring from different parents, increasing the chances to recombine into advantageous allelic combination and prune deleterious alleles. We explored the genomes of hundred and eighty strains, combined with experimental mating studies of selected strains, from a non-model organism (*Trichaptum*). We characterized the genomic regions controlling sex. The mating ability of the strains confirmed the role of the mating alleles observed in the genomic data. The detailed analyses of many strains allowed us to observe gene duplication and rearrangements within the mating loci, increasing the diversity within these loci. We supported previous suggestions of balancing selection in this region, an evolutionary force that maintains genomic diversity. These results supports that fungal strains are prone to outcross, which might facilitate the adaptation to new conditions.

Introduction

Balancing selection is an evolutionary force that maintains genetic diversity [1] receiving long-term attention in evolutionary biology [2]. Heterozygote advantage [1], pleiotropy [3], negative frequency-dependent selection [4], rapid temporal fluctuations in climate [5], and segregation distortion balanced by negative selection [6,7] are modes of balancing selection. These different modes of balancing selection leave similar genomic signatures, such as an increased number of polymorphic sites around the region under balancing selection, and sometimes an enrichment of intermediate-frequency alleles around the selected genomic region [1]. When balancing selection has persisted for a long period, coalescent time of alleles may predate speciation events, and polymorphisms can become shared among distinct species, leading to trans-species polymorphisms [8]. Phylogenetic trees for balanced regions are characterized by the presence of long internal branches [9], and clades with a mixture of species caused by trans-species polymorphisms [10]. The development of methods to detect the genomic footprints of balancing selection [11–13] has unraveled, also with a low number of individuals due to sequencing costs, multiple loci under this type of selection. Well-known examples include: the major histocompatibility locus (MHC) in vertebrates [8]; the ABO histo-blood [14]; non-MHC genes, such as *TRIM5* and *ZC3HAV1* in humans [15,16]; self-incompatibility (SI) loci in plants [17,18] and self/nonself-recognition during vegetative growth in fungi [19]; multilocus metabolic gene networks, such as the *GAL* network in *Saccharomyces* [20,21]; and sexual mating loci in fungi [22].

In basidiomycete fungi, there are numerous examples of balancing selection acting on loci regulating the sexual cycle [22–26]. In this phylum, the sexual cycle involves fusion (plasmogamy) of two genetically distinct monokaryotic hyphae (n or one set of chromosomes), generating a dikaryotic (n+n) hyphae [27–29]. The dikaryon is considered a more stable and long-lived state than the monokaryotic phase, but there are controversies about this assumption due to limited studies [30,31]. Due to this dikaryotic state, plasmogamy is normally separated in time from karyogamy, the fusion of both parental nuclei [32]. In basidiomycetes, karyogamy and meiosis normally occur in specialized structures, the fruit bodies [32]. Mating between two monokaryotic hyphae is determined by one or two sets of multiple allelomorphic genes in the mating (*MAT*) loci. Two different mating systems have evolved among basidiomycetes, referred to as bipolar or tetrapolar mating systems [33]. Mating-type identity in some

basidiomycetes, such as *Cryptococcus neoformans*, and members of the sister phylum Ascomycota i.e. *Saccharomyces cerevisiae*, is governed by a single *MAT* locus [34]. This case corresponds to the bipolar system, resembling the sexual system (male or female) in metazoans [35]. However, the ancestor of basidiomycetes developed an evolutionary innovation, the tetrapolar mating system, where two *MAT* loci regulate mating [36]. This new system hinders inbreeding more effectively, since only 25% of the spores from the same individual can mate, compared to 50% for the bipolar species [37]. At the same time, having multiple mating alleles in each *MAT* locus enables extremely effective outcrossing, where most monokaryotic spores or mycelia (derived from different individuals) can establish a dikaryotic mycelium when a compatible mating type partner is found [38].

In strict tetrapolar organisms, the *MATA* locus (syn. *b* or *HD*) contains a series of linked pairs of homeodomain-type transcription factor genes (*HD1-HD2*, syn. *bW-bE*), whereas the *MATB* locus (syn. *a* or *P/R*) is composed of tightly linked G-pheromone receptor genes (*STE3*, syn. *Rcb*, *pra*) and pheromone precursor genes (*Phe3*, syn. *Ph*, *mfa*) [23,39–46]. Nucleotide differences in mating-related genes, without sufficient amino acid changes in key functional domains, belong to the same allelic class [22]. Allelic classes for those genes in *MATA* and *MATB* configure the *MATA* and *MATB* type. The combination of *MATA* and *MATB* types defines mating type identity [34], which controls successful mating and completion of the life cycle [32]. When two monokaryotic (haploid) hyphae of compatible (distinct) *MATA* and *MATB* types conjugate, a structure involved in transferring one of the nuclei during cell division can be observed, called clamp connection, indicating a successful mating [47]. Proteins encoded by *MATA* genes initiate the pairing of the two parental nuclei within dikaryons, they promote clamp development, synchronize nuclear division and septum formation. Proteins encoded by *MATB* genes coordinate the completion of clamp fusion with the subapical cell after synchronized nuclear division and the release of the nucleus, which was initially trapped within the unfused clamp cell [48,49]. Once monokaryons have fused, the *MATB* proteins facilitate septum dissolution and nuclear migration [39]. Experimental crossings in various basidiomycetes, such as *Coprinopsis* and *Schizophyllum*, have been used to infer the number of *MATA* and *MATB* alleles, and results suggest that 12,800–57,600 mating types may exist [50].

However, the molecular confirmation and the knowledge of the diversity of such genomic regions are far behind, as multiple strains must be sequenced. One of the reasons to this delay, is the high nucleotide divergence among *MAT* alleles, which has complicated the study of molecular evolution of the fungal mating systems, where e.g. primer design has been a challenge. Moreover, until now, only a limited number of strains from different species have been analyzed, mainly due to sequencing costs, limiting the quantification of the strength of balancing selection, the presence of trans-species polymorphisms and the detection of mating and non-mating related genes. Due to limited availability of sequenced strains, how each gene within mating loci are involved in mating is unknown.

The type of the mating system in two non-model *Trichaptum* sister species, *Trichaptum abietinum* and *Trichaptum fuscoviolaceum* (Hymenochaetales, Basidiomycota), have been tested in the past, likely because their fruit bodies readily produce monokaryotic spores that germinates and grows *in vitro*, making it easy to conduct crossing experiments in the lab [51]. *Trichaptum abietinum* and *T. fuscoviolaceum* are wood-decay fungi with circumboreal distributions [52]. Although, we know their life cycle (Fig 1A), details about how long these organisms spend in monokaryotic or dikaryotic states are still unknown. Previous mating studies have suggested a tetrapolar mating system for *Trichaptum* with an inferred number of 385 *MATA* and 140 *MATB* alleles in *T. abietinum* [53]. The mating studies have also revealed that three intersterility groups (ISGs) occur in *T. abietinum* [50–54]. However, so far we have no information about the underlying genomic architecture and molecular divergence of *Trichaptum* mating genes.
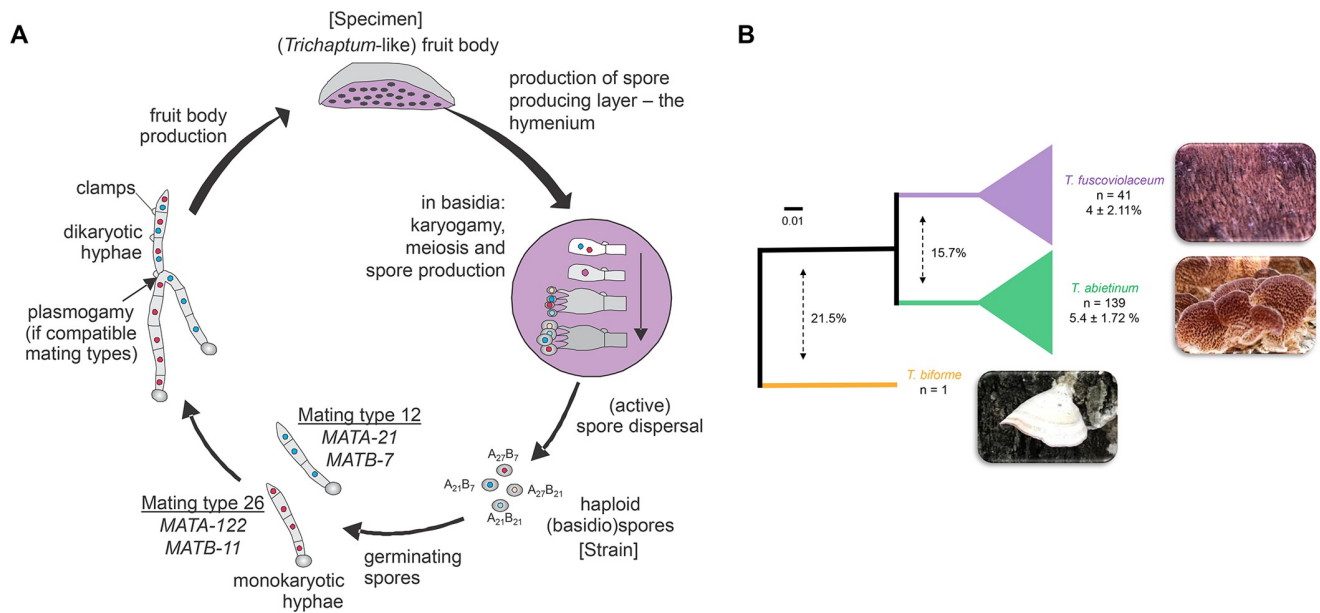
Fig 1. *Trichaptum abietinum* and *T. fuscoviolaceum* are sister-species. A) Schematic representation of the *Trichaptum* life cycle. As an example, *MATA* and *MATB* types, generating two compatible mating types are indicated. A specimen was the original dikaryotic sample, i.e. TA[Number], isolated from the wild environment and stored in a national museum or in our laboratory. Strains were isolated from fruiting bodies and due to their monokaryotic character, we added an M[Number] to the specimen name, TA[Number]M[Number]. Strains are stored in our personal collection at -80˚C. B) Schematic Neighbor-Joining (NJ) phylogenetic tree reconstructed using (100 –ANI)/100 values. ANI values go from 100% (identical genomes) to 0% (distinct genomes). In a format (100 – ANI)/100, these values represent divergence. Full NJ and ASTRAL phylogenetic trees can be found in S1 Fig and in iTOL: https://itol.embl.de/shared/Peris_D. The number of strains (n) and the average (100 –ANI)/100 within species are indicated for each species clade. The L15831 genome is included increasing the *T. abietinum* collection to 139 strains. Dashed arrows indicate the average (100 –ANI)/100 of pairwise strain comparisons for the compared species. Colors highlight the species designation after the whole genome sequencing analysis.

https://doi.org/10.1371/journal.pgen.1010097.g001

Here, we study the molecular evolution of the *MAT* genes in tetrapolar basidiomycetes, using a non-model organism. We first sequenced the full genome of a large set of new established monokaryotic cultures from sporulating fruit bodies, collected at different circumboreal locations. Then, we applied bioinformatics and *in vitro* crosses to: i) unravel the genomic location and the structure of the mating-related genes; ii) assess the allelic richness of *MAT* genes; iii) the divergence needed among the alleles in order for the fungi to recognize different mating types, then test whether the genotypic information mirrors phenotypic outcomes of *in vitro* sexual mating; iv) and reveal molecular signals of balancing selection.

Results

Mating regions are highly dynamic in Trichaptum *species*

To locate the chromosomal position of *MATA* and *MATB* and the genes delimiting the mating regions, we explored different genome assemblers using PacBio long reads and selected the best assembly (Table 1; `canu`) for one *T. abietinum* and one *T. fuscoviolaceum* strains. These two species genomes differed with an average 15.7% in a converted ANI (average nucleotide identity) value to divergence value (Fig 1B).

Both species potentially contained twelve chromosomes. The genome size of *T. abietinum* and *T. fuscoviolaceum* was 49 Mbp and 59 Mbp, respectively. Both genomes were highly syntenic with a few small inversions (S2 Fig). The *MATA* and *MATB* loci were located on chromosomes 2 and 9, respectively. *MATA* homeodomain genes were flanked by *bfg*, *GLGEN* on one end and *MIP1* coding sequences on the other (Fig 2). The *MATA* region, defined from *bfg* to

Table 1. PacBio assembly stats.

Strain name	Descended from	Species	Assembler	Before ultrascaffolding			After correction & ultrascaffolding		
				Contigs	N50 (Kb)	L50	Scaffolds	N50 (Kb)	Bases (Mb)
TA10106M1	TA-1010-6	*T. abietinum*	Canu	26	4,268.52	5	12	4,354.20	49.43
TF100210M3	TF-1002-10	*T. fuscoviolaceum*	Canu	118	2,011.66	10	12	5,547.79	59.09

https://doi.org/10.1371/journal.pgen.1010097.t001

MIP1, was 17.9 and 19.6 Kbp long in *T. abietinum* and *T. fuscoviolaceum*, respectively. Both reference genomes contained two homeodomain complexes: alpha- (*aHD*) and beta-complexes (*bHD*). In the reference *T. fuscoviolaceum MATA* region, one homeodomain pair, the *bHD1*, was lost, *bHD2* was inverted, and between the alpha and beta-complexes there was a gene encoding an *ARM*-repeat containing protein (Fig 2). *MATB* pheromone receptor and pheromone precursor genes were flanked by *PAK*, *RSM19*, *DML1*, *RIC1* and *SNF2* genes. All these genes together were defined as the *MATB* region, which was 30.3 Kbp long in both species. Four putative pheromone receptors and two pheromone precursor genes were annotated. The *MATB* region was syntenic between both species, except an inverted block containing *STE3.2* and *Phe3.2* genes in the *T. fuscoviolaceum* reference (Fig 2).
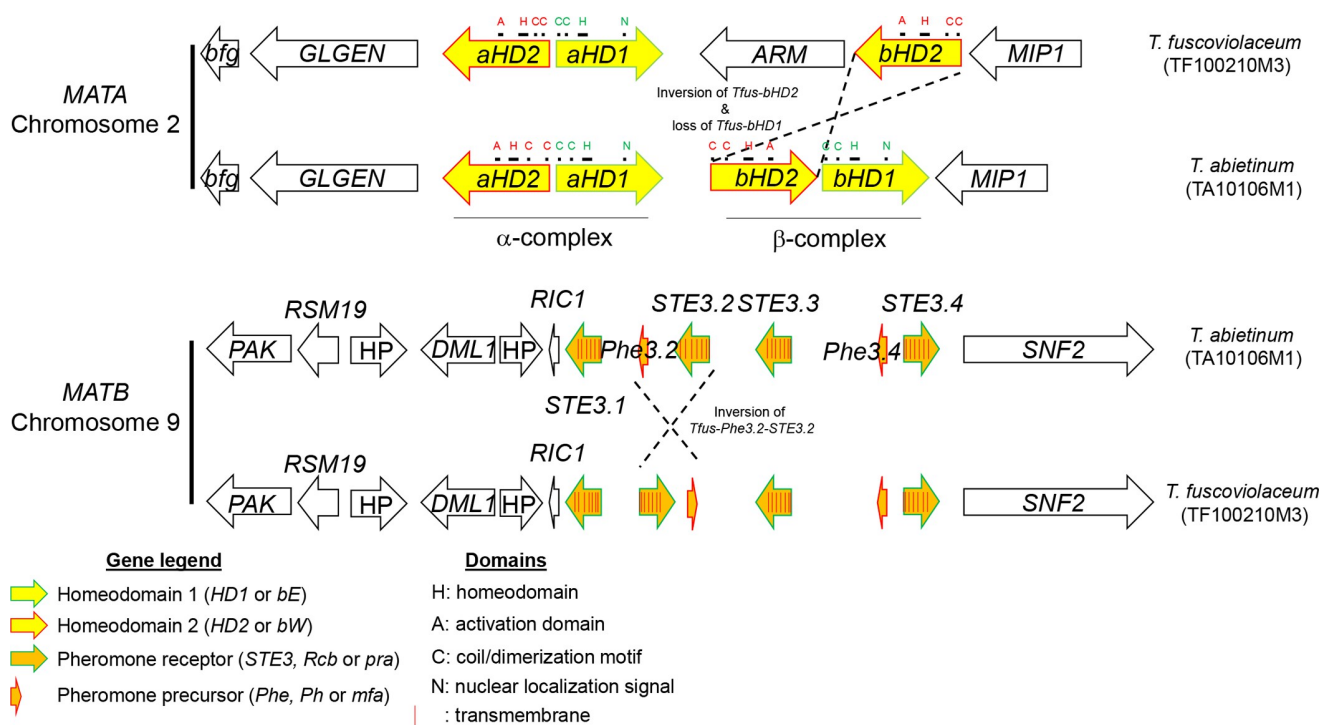
Fig 2. Two homeodomain complexes in *MATA* and four putative pheromone receptors in *MATB* were detected in *T. abietinum* and *T. fuscoviolaceum*. Schematic representation of the gene composition and direction in both reference genomes. Homeodomain, other functional domains, pheromone precursors and pheromone receptors genes are represented as indicated in the legend. The rest of the genes were colored in black, and the gene names were indicated inside the arrows. *aHD*: alpha-complex homeodomain; *ARM*: ARM-repeated containing protein; *bfg* beta-flanking gene; *bHD*: beta-complex homeodomain; *DML1*: mtDNA inheritance protein; *GLGEN*: glycogenin-1; *HP*: hypothetical protein; *MIP1*: mtDNA intermediate peptidase; *PAK*: serine/threonine protein kinase; *RSM19*: 37S ribosomal protein S19; *RIC1*: RIC1-domain containing protein; *SNF2*: Snf2 family dna-dependent ATPase; *STE3*: GPCR fungal pheromone mating factor. The Fig is not drawn to scale to facilitate visualization.

https://doi.org/10.1371/journal.pgen.1010097.g002
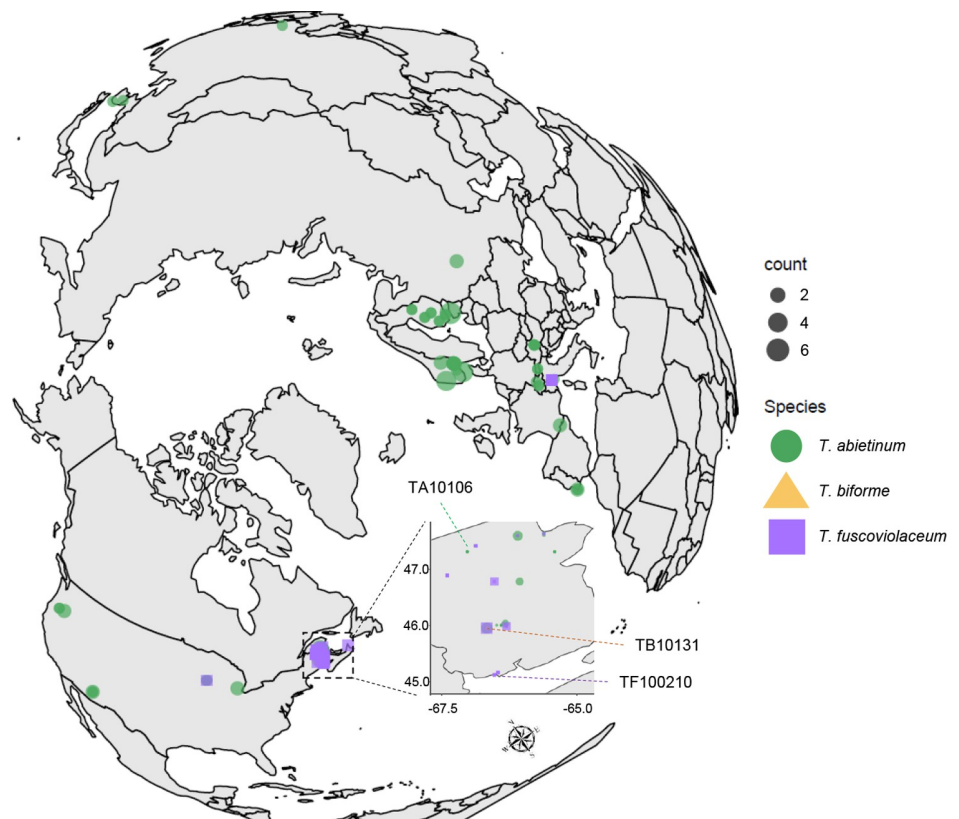
Fig 3. Circumboreal distribution of *Trichaptum* specimens. Geographic distribution of collected *Trichaptum* specimens. 77 European, 98 North American and 4 Asian *Trichaptum* specimens were collected in this study. Specimens were collected mainly from four plant hosts from the genus *Abies*, *Larix*, *Picea* and *Pinus* (Tables 2 and S1). Map was created using R and `ggplot2`.

MAT *genes displayed multiple alleles*

The annotated mating genes in the reference genomes were used to search for those genes in the 178 Illumina sequenced strains, collected at circumboreal regions (Fig 3) and a *T. abietinum* assembly downloaded from JGI (S2 Table). *Trichaptum abietinum* was the most diverse species based on this collected dataset (average converted ANI 5.4%) (Fig 1B). *MATA* genes were assembled in one contig for 75 *T. abietinum*, 25 *T. fuscoviolaceum* and 1 *T. biforme*. In the case of *MATB*, genes in that region were found in one contig for 116 *T. abietinum*, 27 *T. fuscoviolaceum* and 1 *T. biforme*. For these strains, the mating genes have potentially the same chromosomal location than in reference strains. For the rest of the sequenced strains, the mating genes were found in multiple contigs due to assembly limitations using short reads. Most of those fragmented mating regions might be organized similar to reference strains; however, we observed unexpected coding sequences for 6 strains in the *MATA* region and 2 strains in the *MATB* region, which could suggest that these regions have split and were translocated to different chromosomes or positioned in a new chromosomal location (S3 Table).

An initial analysis of nucleotide conservation of the mating regions indicated that flanking genes were conserved, as well as *STE3.1* and *STE3.3*. However, the rest of putative mating genes were highly diverse (Fig 4). Gene order comparison among strains highlighted that the most common *MATA* and *MATB* syntenic blocks were both present in *T. abietinum* and *T. fuscoviolaceum*, and the frequent *MATB* syntenic block was present in the three species
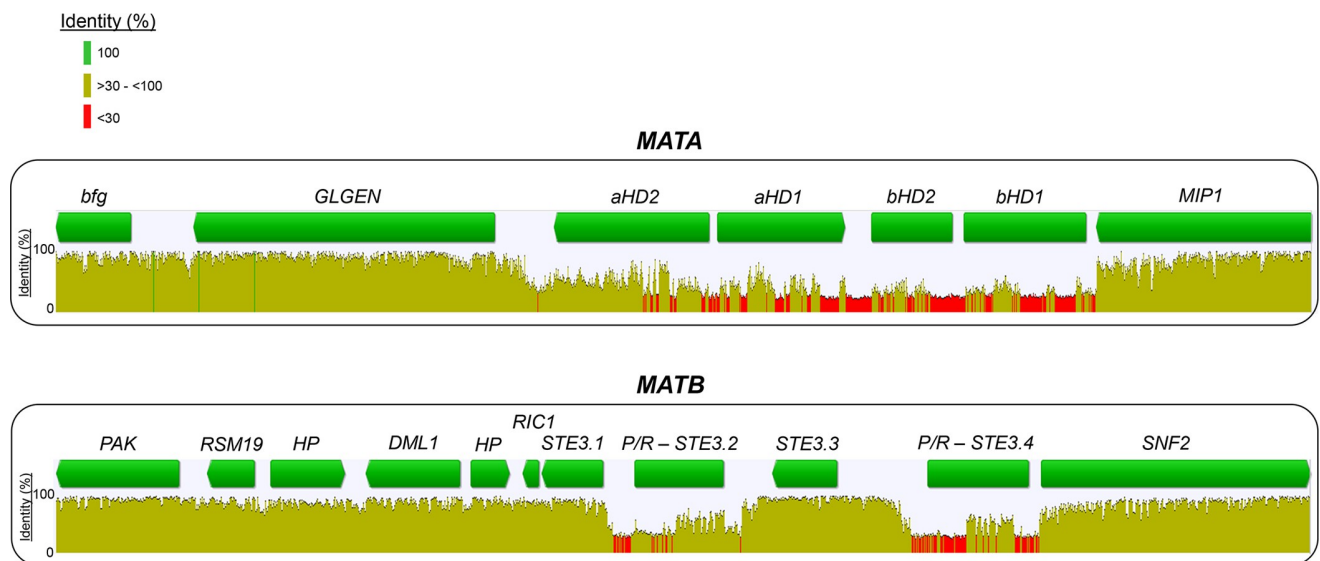
Fig 4. High nucleotide diversity among mating genes. Identity values of nucleotide alignments for *MATA* and *MATB* regions are displayed. Gene arrows indicate the coding direction; however, when gene direction differed among strains (Fig 5), we represented a green rectangle. Bar colors represented the level of identity according to the legend. `Geneious'` identity values were calculated based on each nucleotide position and represent the percentage (y-axis) of sequences with an identical nucleotide compared to the consensus sequence. The *MATA* alignment includes 175 isolates (3 species), excluding those isolates found to split the region in potentially different chromosomal locations (S3 Table). The *MATB* alignment includes 179 isolates (3 species), excluding those isolates found to split the region in potentially different chromosomal locations (S3 Table).

https://doi.org/10.1371/journal.pgen.1010097.g004

(Figs 5 and 6). *Trichaptum biforme* and five other *Trichaptum* strains, differentiated from the most frequent *MATA* configuration by the presence of a hypothetical protein (Fig 5). All this suggest that the most frequent *MATA* and *MATB* gene configurations, represented for the reference *T. abietinum* strain (Fig 2), were present in the ancestor of these three *Trichaptum* species. The gene order of *HDs* in the alpha-complex was conserved among all *Trichaptum* strains. However, frequent inversions of the *bHD2* gene and absence of one of the two *bHD* genes were detected. An interesting observation was the presence of an additional *HD2* gene (*xHD2*) upstream the alpha-complex in six *T. abietinum* strains (Fig 5). The coding sequence of *xHD2* was truncated, indicating an ongoing process of pseudogenization. In the *MATB* region, all strains contained two pheromone precursor genes, one located between *STE3.1* and *STE3.2*, and a second between *STE3.3* and *STE3.4*. The orientation of *STE3.2*, *STE3.4* and pheromone precursor genes varied among strains (Fig 6).

We were able to infer several domains and motifs in mating genes. *HD1* and *HD2* homeodomain genes contained three and four exons, respectively, whereas *STE3* genes, characterized by the presence of seven transmembrane domains, included 4 to 6 exons. Homeodomain genes were characterized by the presence of the typical homeobox domain (Fig 2). In each homeodomain protein alignment, we found conserved amino acid sequences in potentially functional homeodomains (S3 Fig), likely because they are essential for the activation of the expression of target genes. The nuclear localization signal was detected in HD1 proteins, with the presence of bipartite sequences (S3 Fig). Regions enriched in prolines are indicative of putative activation domains (AD), which were conserved in HD2 proteins (S3 Fig). It is important to note an additional conserved region at the C-terminal of HD1 proteins (S3 Fig). Coiled coils related with heterodimerization were likely located at the N-terminal (Fig 2).
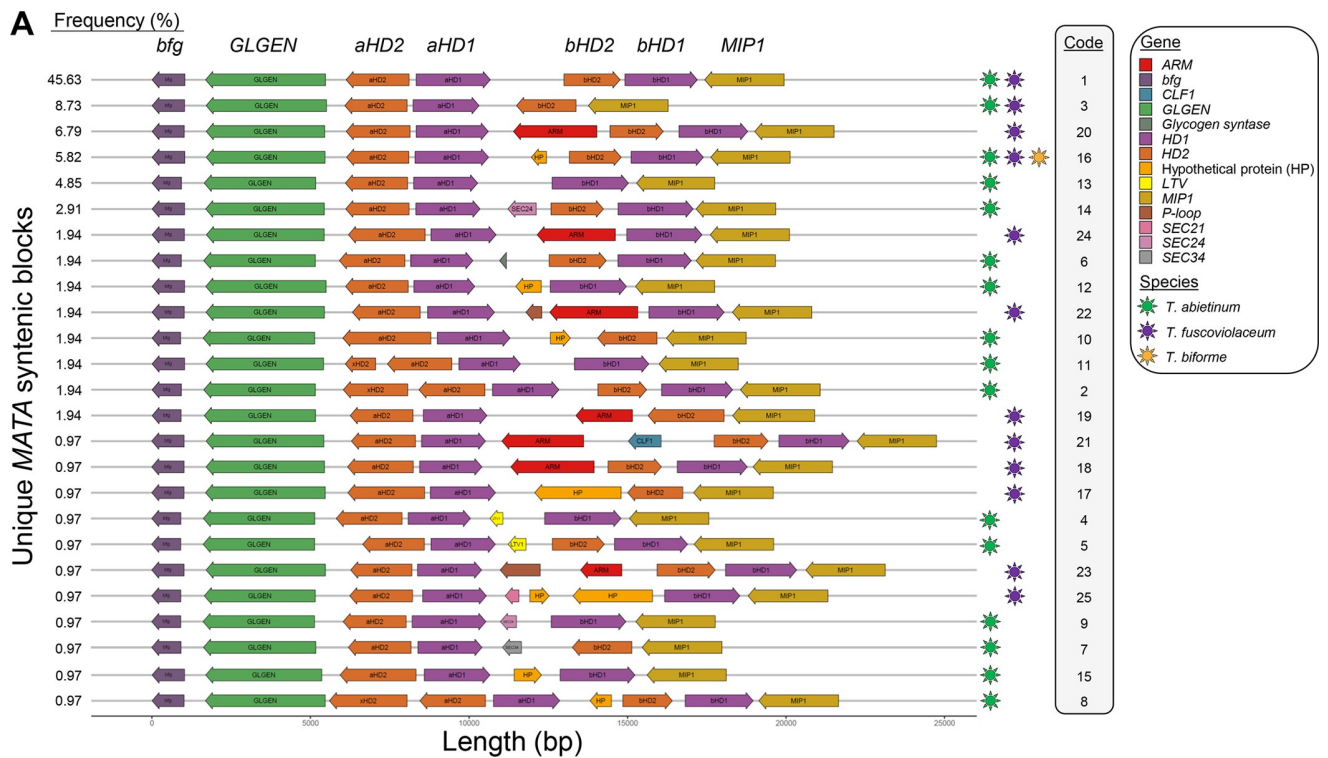
Fig 5. Mating A region is highly dynamic and show multiple rearrangements among *Trichaptum* strains. *MATA* gene order representations for *Trichaptum* strains with *MATA* genes assembled in one contig. The percentage of strains containing a specific *MAT* block order is indicated in the left. Genes were colored according to the legend on the right. Species containing a particular *MAT* block are represented by colored stars at the right of the *MAT* block and were colored according to the legend. Coding sequence direction is represented by the arrows. Code numbers link the strains in S2 Table with the displayed mating structure.

https://doi.org/10.1371/journal.pgen.1010097.g005

Using the pheromone_seeker.pl script, we were able to detect most of the pheromone precursor genes. Basically, the script searches the prenylation signal, which is important to transport the pheromone precursor peptide to the plasma membrane, where cleavage occurs in the maturation site of the precursor. Maturation will release the active pheromone, consisting of the residue from near the maturation site (E) to the C in the CaaX motif (S4 Fig). Maturation approximately generates a peptide of 10–11 amino acids [55]. However, some pheromone precursors were not detected due to unexpected amino acids in the CaaX motif (S4 Fig). We found multiple examples in both pheromone precursors (Phe3.2 and Phe3.4), where the canonical CaaX motif contained a polar (p) amino acid (threonine, T), displaying an uncommon CpaX motif. Most of the pheromones contained an aspartic amino acid following the starting methionine. The presence of both aspartic and glutamic amino acids in the maturation site was highly conserved in *Trichaptum* pheromones.

Despite the dynamic nature of both mating regions (Fig 5), where rearrangements and gene losses were frequent, and the observed high nucleotide diversity (Fig 4), the results are highlighting the effects of natural selection retaining important residues located in domains proven to be linked to the activity of mating proteins.

Distinct mating types generate compatible mating crosses within species

Mating gene combinations define mating types. To predict mating types, we first quantified the number of clades in reconstructed phylogenetic trees (Figs 7 and S5). The number of clades
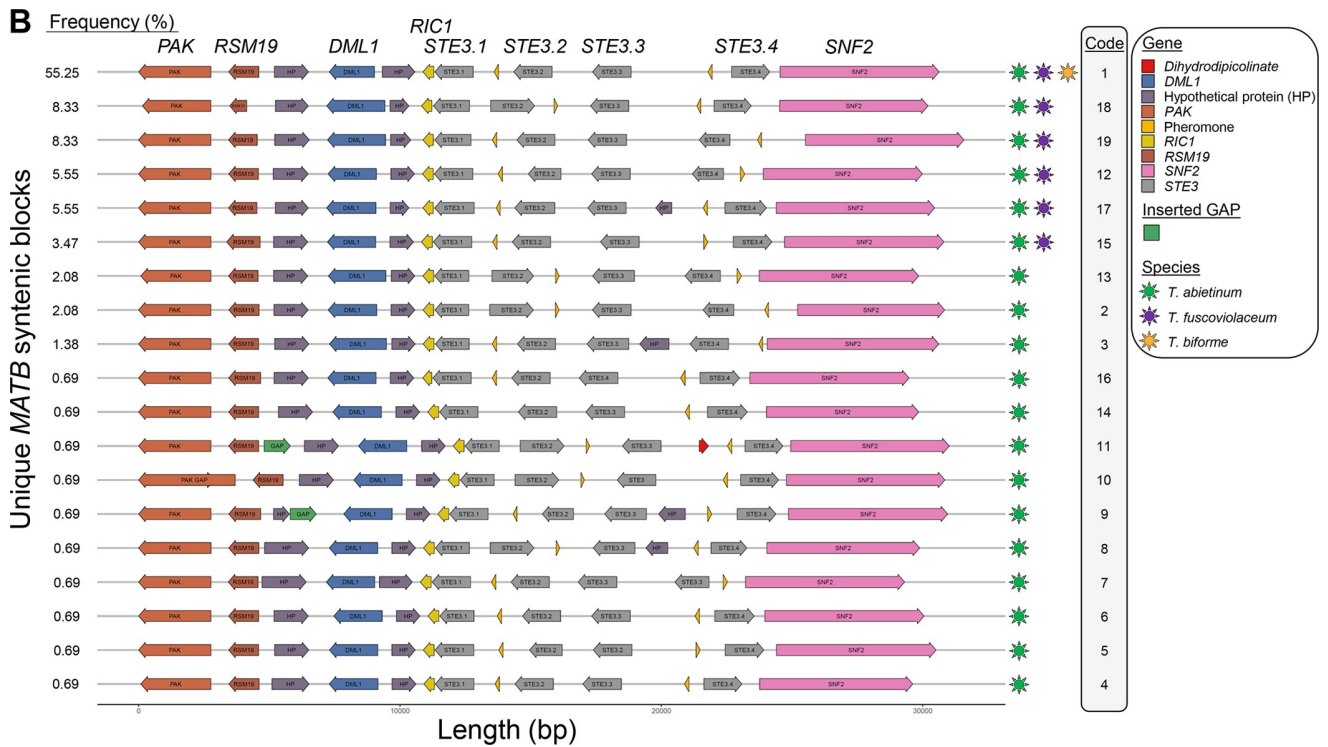
Fig 6. Mating B region is highly dynamic and show multiple rearrangements among *Trichaptum* strains. *MATB* gene order representations for *Trichaptum* strains with *MATB* genes assembled in one contig. We showed the *MATB* region for those strains where the assembly was contiguous from *RIC1* to *SNF2*. In some strains, the region from *RIC1* to *PAK* was contained in multiple contigs. Those contigs were joined (ultrascaffolding) and ordered according to the reference genomes (see Material and Methods section). To ultrascaffold, we inserted 999 Ns between joined contigs, annotated as a GAP in the legend. For that reason, GAP label is drawn. The percentage of strains containing a specific *MAT* block order is indicated in the left. Genes were colored according to the legend on the right. Species containing a particular *MAT* block are represented by colored stars at the right of the *MAT* block and were colored according to the legend. Coding sequence direction is represented by the arrows. Code numbers link the strains in S2 Table with the displayed mating structure.

https://doi.org/10.1371/journal.pgen.1010097.g006
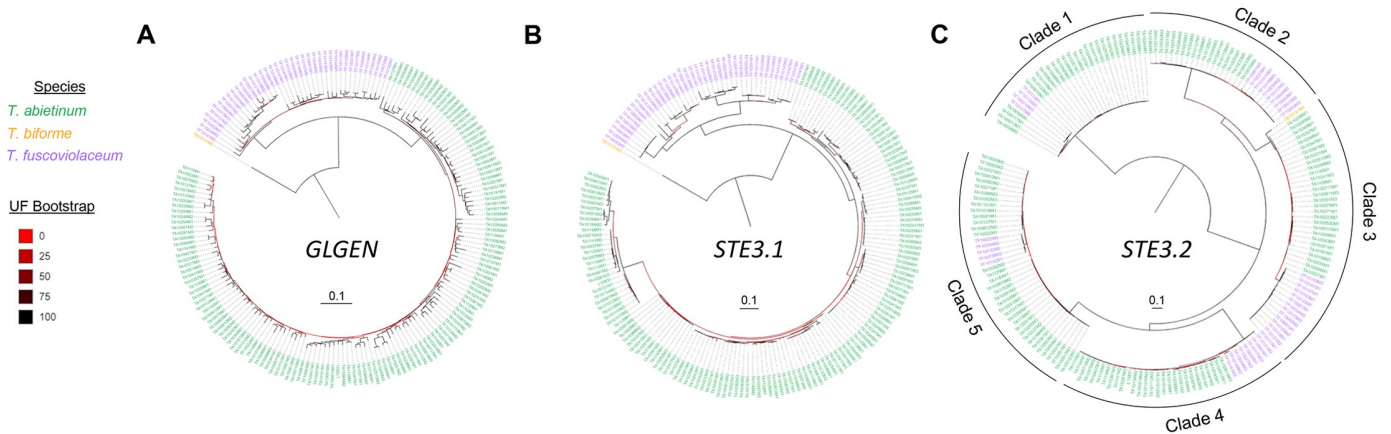
Fig 7. ML phylogenetic tree topology of mating proteins suggests balancing selection and trans-species polymorphisms. ML phylogenetic protein trees of *GLGEN* (a flanking gene), *STE3.1* (a potential non-mating pheromone receptor protein) and *STE3.2* (a mating pheromone receptor protein) are represented in panel A, B and C, respectively. Strains were colored according to the species designations as indicated in the legend. Branch support was assessed using the ultrafast bootstrap (UF bootstrap) method. UF bootstrap is indicated in each branch by a gradient color according to the legend. Scale bar is represented in number of amino acid substitutions per site. The rest of phylogenetic protein trees and more detailed trees for the represented here are found in S5 Fig.

https://doi.org/10.1371/journal.pgen.1010097.g007

in the phylogenetic trees varied from 5 to 28. Each clade was considered as a different allelic class. Sequences in the same allelic class encoded for proteins with an AAI higher than 86% (S6 Fig). The highest number of allelic classes was found among alpha-complex homeodomain genes where we detected evidence of recombination (S4 Table).

A combination of allelic classes for homeodomain gene pairs (*HD1* and *HD2*) in the alpha-complex and in the beta-complex defines the *MATA* type (S3 Table). In total, we predicted 207 *MATA* (23 alpha x 9 beta) and 189 *MATA* (21 alpha x 9 beta) types for *T. abietinum* and *T. fuscoviolaceum*, respectively. A combination of allelic classes for pheromone receptors genes *STE3.2* and *STE3.4* defined the *MATB* type (S3 Table). Predictions suggested 65 *MATB* types (5 *STE3.2* x 13 *STE3.4*) for both species. Note that *STE3.1* and *STE3.3* were not considered to be defining *MATB* types because as we describe in the next section, they were not predicted to be mating-related genes. The number of potential mating types predicted by combining *MATA* and *MATB* types is at least 13,455 (207 *MATA* x 65 *MATB*) mating types in *Trichaptum abietinum* and 12,285 (189 *MATA* x 65 *MATB*) mating types in *T. fuscoviolaceum*. Once we defined the mating types of strain samples, we calculated the AAI by pairwise comparisons of protein sequences of strains containing the same mating type. We detected high conservation within species for all proteins (AAI = 100%), and higher conservation of pheromone receptors between species (AAI > 95–98%) than for homeodomain genes (AAI > 78–83%), suggesting pheromone receptors were more constrained to accumulate non-synonymous mutations compared to homeodomains (S7 Fig).

These predicted mating types were helpful to set up mating experiments (S3 Table). We tested the outcome of crosses between selected monokaryotic strains from the same species and between species (S5 Table). We assumed a successful mating when clamp connections were formed (S8 Fig). Our expectations, based on the molecular characterization, were confirmed in all within species crosses. Crosses using strains with identical *MATA* types did not generate clamps when *MATB* types were expected to be compatible, and vice versa. These results demonstrate that identical (AAI > 86%) *MAT* allelic classes generate the first mating barrier.

We also included some strains derived from the same dikaryotic specimen (S3 Table), where most of them showed at least a pair of compatible *MATA* and/or *MATB* types. These strains helped us to unfold the original allelic class composition of the parental specimen (S3 Table). Due to the unlinked nature of *MATA* and *MATB* regions and limited number of studied strains from the same specimen, some strains had identical mating types, thus did not reveal the original mating type composition of the parental specimen.

No clamps were observed in crosses between species with compatible mating types suggesting other mechanisms are involved in the generation of pre-zygotic barriers between *Trichaptum* species.

Long-term balancing selection left footprints in the mating regions

To infer the evolutionary history of these mating genes controlling the sexual cycle, and the flanking genes, and to test whether they agree with the species tree (Figs 1B and S1), we reconstructed Maximum Likelihood (ML) individual protein trees (S5 Fig). For most proteins encoded in flanking genes and for both STE3.1 and STE3.3 proteins, phylogenetic trees clustered strain sequences according to their species designation (Figs 7A, 7B, S5A and S5B, S5H-S5L and S5N). However, phylogenetic protein trees for homeodomains (aHDs and bHDs), two pheromone receptors (STE3.2 and STE3.4), MIP1 and SNF2 disagreed with the species tree (Figs 7C and S5C–S5G, S5M, S5P, and S5O). These trees were characterized by long internal branches and a mixture of species-specific sequences in different clades. All these
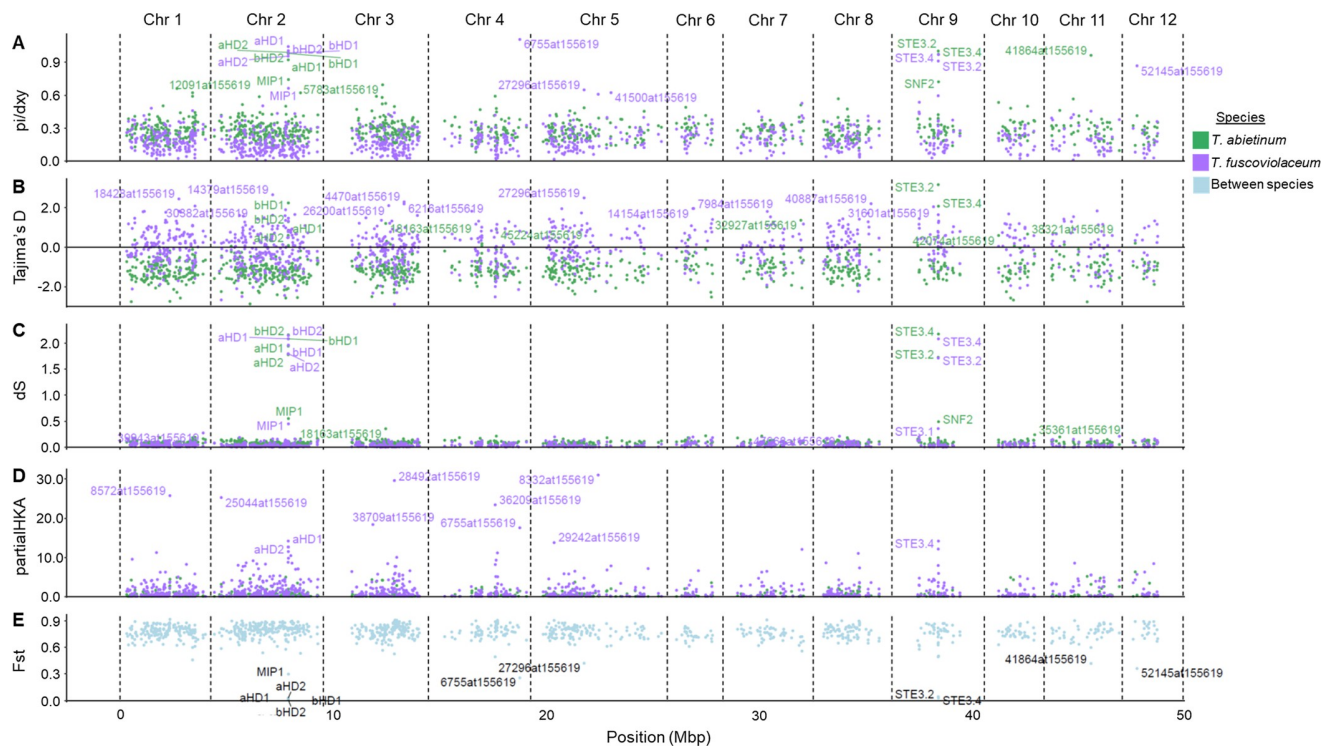
Fig 8. Multiple nucleotide statistics support long-term balancing selection in genes located in the mating region. Ratio of nucleotide diversity (Pi) and absolute divergence (dxy), Tajima's D, average number of synonymous substitutions per synonymous sites (dS), and relative divergence (Fst) values for each single-copy orthologous and mating genes are reported in panels A), B), C) and E), respectively. Gene contribution to the significance of a HKA test (partial HKA) are represented in panel D). Gene names containing 1% of the highest values (panels A, B, C, and D) or 1% of the lowest values (panels E and F) are displayed. *T. fuscoviolaceum* gene names with the highest partial HKA values are displayed due to the significant result of the HKA test (*p*-value = 3.13 x 10^{-39}). Each dot represents a gene and we used the annotation in *T. abietinum* to represent the position of each gene. Annotation file can be found in the Github page dedicated to this project. Dots were colored according to within species calculations (green or purple for *T. abietinum* and *T. fuscoviolaceum*, respectively) or between species comparison (cyan). Chr: chromosome. These analyses include all strains from both species.

https://doi.org/10.1371/journal.pgen.1010097.g008

results pointed to the presence of trans-species polymorphisms likely due to long-term balancing selection.

To further test whether long-term balancing selection is acting on the mating regions, we quantified nucleotide statistics and performed a multilocus HKA test using the mating genes and a collection of universal single-copy orthologs (BUSCO) genes. We first tested the reciprocal monophyletic nature of BUSCO gene collection. As expected from the species tree (S1B Fig), most of annotated BUSCO genes (eighty-three percent) showed reciprocal monophyly for both species, *T. abietinum* and *T. fuscoviolaceum*, and 98.64% of the rest of genes (174 genes of 1026 BUSCO genes) showed complete monophyly for one of the two species. This BUSCO dataset suggests a clear diversification of both *Trichaptum* species, and supports the utility of this dataset to set the neutral evolution values of the next analyzed nucleotide statistics.

We observed an elevated number of the average number of synonymous substitutions per synonymous sites (median dS > 1.71) and non-synonymous substitutions per non-synonymous sites (median dN > 0.22) for the mating genes compared to the flanking and BUSCO genes (Figs 8C and S9, median dS < 0.55, median dN < 0.10). dS and dN values in mating genes were more than 20x and 3x higher than values for BUSCO genes, respectively (S6 Table). This result was an additional support that balancing selection acts on the mating

regions. Moreover, similar levels of dS and dN (S9 Fig, ratio comparison of 0.95–1.03) were observed within and between species in pairwise comparisons of mating genes, indicating that these polymorphisms were not species-specific and recent introgressions were not involved in the generation of trans-species polymorphisms. This was coherent with a scenario where alleles segregated before the diversification of the species. It is important to note that dS and dN values for two putative receptors, *STE3.1* and *STE3.3*, differed from the other mating genes and that they displayed similar low values as most flanking and BUSCO genes (S9 Fig). In addition, for these two putative non-mating pheromone receptor genes, the dS and dN values were 1.41–3.17 times higher between than within species pairwise comparisons, as we would expect if most of the mutations accumulated after the speciation of *T. abietinum* and *T. fuscoviolaceum*. *MIP1* and *SNF2* dS values were slightly more elevated than BUSCO genes (S6 Table), but values from between species comparisons were more elevated than within pairwise comparisons (S9 Fig). This indicated that the elevated dS values, compared with BUSCO genes, are caused by linkage disequilibrium, where the effects of balancing selection in the closest mating gene were not completely broken by recombination.

To infer whether other nucleotide statistics supported balancing selection, we explored gene values deviating from the rest of the genome (Fig 8). Homeodomain (*HD1s* and *HD2s*) and pheromone receptor genes (*STE3.2* and *STE3.4*) deviated from the distribution of 99% of values in at least four nucleotide statistics (elevated pi/dxy ratio, high dS values, low Fst and high Tajima's D), all in agreement with a balancing selection scenario maintaining trans-species polymorphisms for multiple alleles (Fig 8). Five BUSCO genes were detected in at least two statistics, deviating from the rest of the genome (Fig 8). Those five genes were also detected to show a phylogenetic topology incongruent with a complete reciprocal monophyly, except 18163at155619 where only *T. fuscoviolaceum* sequences were monophyletic (S10 Fig). The detected genes encoded for an acetolactate synthase (27296at155619), a ribosomal protein L38e (52145at155619), a non-specific serine/threonine protein kinase (6755at155619), a protein kinase-domain-containing protein (18163at155619) and a NF-kappa-B inhibitor-like protein 1 (41864at155619).

The geographic distribution of *MATA* and *MATB* alleles did not suggest a bias towards a particular continent (S5 and S11 Figs), supporting an evolutionary scenario of long-term balancing selection for mating genes.

New mating genes generated by duplications

The diversity of allelic classes might be generated by the accumulation of point mutations or, as stated above for the alpha-complex, by recombination of existing variants. However, we detected a new HD2 (xHD2) gene in some strains (Fig 5A and S3 Table). A ML phylogenetic tree of all HD2 protein sequences (S12 Fig) clustered xHD2 proteins in two allelic classes, aHD2.8 and aHD2.10; however, the closest aHD2 in these strains were from different allelic classes: aHD2.18 and aHD2.24, respectively (S3 Table). The limited presence of *xHD2* genes in other strains and the high similarity of the proteins to two aHD2 proteins points to two recent duplications and transfers to other mating regions.

Phylogenetic analyses of homeodomain proteins with other fungal sequences indicated that the beta-complex HD proteins were much older than Hymenochaetales (S13A Fig), which was in accordance with the lower identity values observed for pairwise comparisons within bHD than within aHD (S6 Fig). Except aHD1.12, the rest of aHD proteins were identified in *Trichaptum* species. A similar result can be observed for pheromone receptors, where most *Trichaptum* pheromone receptor proteins were closely related, except two proteins, encoded in *STE3.2* and *STE3.4* genes, which were related to pheromone receptor proteins from other

fungal species (S13B Fig). It remains to be answered whether the alpha-complex was generated by a duplication from the beta-complex or a more complex scenario generated this additional homeodomain complex in *Trichaptum*.

Discussion

Mating genes diversity was maintained by balancing selection

Retaining multiple mating alleles appears to be beneficial as it promotes outcrossing [36]. The multiallelic character of mating types promotes a potential outcross event to occur in 98% of crosses [36,56]. How this mating diversity originated is not clear, but we demonstrated that some levels of recombination and duplications might play a role. Fifteen recombinant variants in the alpha-complex and two recent *aHD2* duplications were detected in *Trichaptum*. It was previously thought that recombination was suppressed or limited in the mating regions [57], and that duplication and diversification events were limited to Agaricales [42]. Recombination is suppressed by the presence of inversions and/or gene losses, which might generate hemizygous strains, observed in mating loci and genomic regions under balancing selection [58]. The rearrangements observed in *Trichaptum* beta-complex brings another layer of complexity to *MATA* region, which is comparable to the complexity previously described for *MATB* genes [36]. Rearrangements in both *MAT* loci might be an important factor suppressing recombination in these genes. On the contrary, the gene order conservation of the alpha-complex does not completely suppress recombination, in accordance with evidence of ongoing recombination between mating genes [59] and their flanking genes in other fungal organisms [60]. Our observations highlight how studying a high number of strains of the same species can unravel previously underestimated mechanisms that generate diversity in mating genes.

We have demonstrated that balancing selection is likely the main force retaining genetic diversity in the mating genes. Evidence of balancing selection has been proposed for homeodomain genes in the pathogenic root decay fungus *Heterobasidion* (Russulales) [26], as well as in pheromone receptors of *Mycrobotryum* species (Mycrobotryales) [24]. The action of balancing selection in *Trichaptum* and in other fungi appears to have occurred before the speciation event, generating multiple cases of trans-species polymorphisms [26]. The genetic signatures of balancing selection highlighted that two pheromone receptors in *Trichaptum* strains are likely non-mating genes, this could only have been unraveled by including multiple strains as we have done here. In Agaricomycotina, it is frequent to detect multiple pheromone receptors, some of them not involved in mating functions [40,42,61]. The role of these non-mating pheromone receptors will deserve further investigation.

It has long been speculated about the action of balancing selection in the *MATA* flanking gene, *MIP1* [25,60]. *MIP1* encodes a mitochondrial intermediate peptidase 1, which is a thiol-dependent metallopeptidase involved in the last step of protein maturation targeted to the mitochondria, where MIP1 cleaves off an octapeptide of immature proteins [62]. The genomic footprints detected in *MIP1* are likely due to the action of linkage disequilibrium, as *MIP1* is close to the beta-complex *HD* genes. It has been speculated that *MIP1* signals of balancing selection and trans-species polymorphisms might be due to a role in mating, such as MIP1 involvement in mitochondrial inheritance, functioning as a suppressor of selfish mtDNA [63]. However, this function is not well-supported. Other genes encoding proteins involved in mitochondrial functions have been found linked to mating genes [60]. In *T. abietinum* and *T. fuscoviolaceum*, we found *RSM19*, a 37S ribosomal protein S19, linked to *MATB*. However, we did not detect signals of balancing selection in this gene. In addition, some signals of balancing selection and trans-species polymorphisms were detected in *SNF2*, a gene located in the

MATB region, encoding a DNA-dependent ATPase protein. The analogous signals of balancing selection between *SNF2* and *MIP1* might support that the balancing selection signal in both genes is due to linkage disequilibrium, and the signal is just a consequence of the action of balancing selection in the neighbor mating genes [60].

Mating genes and organization resemble other basidiomycetes suggesting similar origin

Sampling and studying the genomes of a wide collection of *Trichaptum* strains have unraveled the dynamic nature of mating gene architectures. With two homeodomain complexes, *Trichaptum MATA* gene organization is similar to other Hymenochaetales, such as *Phellinus lamaoensis*, *Phellinus sulphurascens* (both species from the *Phyrrhoderma* genus) and *Schizopora paradoxa* [64]. The presence of homeodomain complexes with just one pair of homeodomain genes is also observed in Hymenochaetales [64]. In other Hymenochaetales species, such as *F. mediterranea* and *Porodaedalea pini*, the location of *GLGEN* gene is more distant and interrupted by multiple *ORFs* [64,65]. Notably, the *Phyrrhoderma* species and *F. mediterranea* [64,66] are bipolar, in contrasts to the tetrapolar *Trichaptum* strains. *Trichaptum* and other Hymenochaetales species, such as *Hypodontia* and *S. paradoxa*, have conserved the ancestral tetrapolar system of basidiomycetes [36].

According to mating studies, the formation of clamp connections is facilitated by the presence of at least one different allele at one of the multiple *MATA HD* complexes and one at the *MATB* P/R loci. Here, we demonstrated by mating experiments and genomic analyses that protein identity must be lower than 86% to function as different mating type, although important protein domains and motifs are conserved. We inferred that around 270 *MATA* types (30 alpha x 9 beta) and 65 *MATB* types (5 *STE3.2* x 13 *STE3.4*) are segregating in *Trichaptum* species, which indicates around 17,550 mating types. These numbers are close to the estimated number of alleles, 20,000 mating-types, in a previous study of *T. abietinum* [50], suggesting that our sequencing efforts, molecularly characterized most of the *Trichaptum* mating alleles. In other tetrapolar basidiomycete species, such as the model species *Coprinopsis cinerea* and *Schizophyllum commune*, the number of mating types is also similar, around 12,800 (160 *MATA* x 81 *MATB*) and 23,328 (288 *MATA* x 81 *MATB*), respectively [51]. We inferred that beta-complex HD alleles were segregating in other Agaricomycetes, suggesting that these HD proteins are much older than alpha HD, a result that is supported by the ongoing recombination events in the alpha HD. Moreover, we cannot discard that alpha-complex alleles may be exclusively specific of *Trichaptum*. Allele *aHD1.12* points to potential alpha-complex alleles segregating in other Hymenochaetales, but just thirteen Hymenochaetales species have been fully sequenced, and usually only one representative of each species, except for the three sequenced *Pyrrhoderma noxium* strains. Thus, there are few available genomes to compare.

A new pheromone precursor motif containing a polar amino acid in CaaX motifs was detected by this large-scale sequencing effort. We are not aware of CpaX motifs in other Basidiomycetes, although this motif was observed in pheromones of some Ascomycetes species [67,68]. The whole genome sequence of other Hymenochaetales and other fungal orders, and the increased number of strains from multiple species, will clarify the evolutionary history of the alpha-complex and protein patterns observed here.

Trichaptum—*A valuable toolset for studies of genes related with sex*

The dataset contributes with a large number of genome assemblies from two non-model species and a representative for a third species of the *Trichaptum* genus. The existence of at least two North American intersterility groups (ISGs) that are partially compatible with a third

European group in *T. abietinum* indicates three potential differentiated lineages [52–54]. Even though we did not perform a population genomic analysis in this study, multiple well-differentiated clades can be inferred by using ANI values and BUSCO phylogenetic species trees, supporting some population structure in this strain collection. The presence of ISG in *T. fuscoviolaceum* is not previously confirmed based on mating studies [52,54]. However, we hypothesize that there are at least two potential lineages due to the presence of two well-differentiated *T. fuscoviolaceum* clades, as suggested by Seierstad *et al*. [52]. ANI dissimilarity values between these lineages were nearly as high as values detected in *T. abietinum*, supporting the hypothesis about population structure in *T. fuscoviolaceum*. However, the difference in the levels of populations and the presence of clear ISG in one species and not in the other might be the reason of the differences in the distribution of Tajima's D values, with more BUSCO genes with negative Tajima's D values in *T. abietinum* than in *T. fuscoviolaceum*.

The potential number of *Trichaptum* lineages together with these new genome sequences and diversity of mating types, provide an exceptional tool for comparative genomics and functional genomics to study the evolution of sex in fungi and mechanisms involve in the sexual cycle.

Conclusion

We have demonstrated the importance of sequencing several strains of fungal species to detect mating-related genes, and to unravel the strength and footprints of long-term balancing selection in mating genes. Events previously thought of as uncommon in mating genes, such as recombination and duplications, have been detected in mating-related genes with conserved gene order. The *Trichaptum* dataset highlights how diverse and dynamic the mating loci are. These mating genes play a fundamental role in promoting outcrossing events and have consequently been targets of long-term balancing selection. The action of balancing selection leaves signatures of multiple trans-species polymorphisms beyond the genus level. Comparative genomics and phylogenomics were important tools to locate mating genes and characterize the number of alleles retained by balancing selection. Mating proteins with less than 86% identity generated compatible mating types, as we demonstrated by experimental crosses. Despite the number of alleles and the high diversity among them, important domains and motifs are still conserved due to their critical role during the life cycle. Questions regarding the effects of mutations in the interaction between homedomain proteins or receptors and pheromones, especially the presence of non-aliphatic amino acids in the CaaX motif (i.e. a CapX motif), and which role the linked mating genes, such as *MIP1*, are playing during the life cycle are exciting areas of research. This newly sequenced collection of *T. abietinum* and *T. fuscoviolaceum* makes a step-forward to re-establish these fungal organisms as a model system in evolutionary research.

Material and methods

Trichaptum *collection*

A total of 180 *Trichaptum* strains from the northern hemisphere were included in the study: 138 *T. abietinum* (67 European, 67 North American and 4 Asian), 41 *T. fuscoviolaceum* (10 European and 31 North American) and one North American *T. biforme* (S1 Table). GPS coordinate format conversion was generated with `GMScale 0.5.1` to plot the geographic distribution in R, using `ggmap 3.0.0`, `ggplot2`, `ggrepel 0.8.2`, and `mapdata 2.3.0`. *Trichaptum abietinum* were frequently isolated from *Picea* trees, whereas *T. fuscoviolaceum* was frequently associated with *Abies* (Tables 2 and S1).

Table 2. Distribution of host trees for *Trichaptum* specimens.

Genus host	*T. abietinum*	*T. biforme*	*T. fuscoviolaceum*
Abies	18	1	23
Larix	10	0	0
Picea	74	0	2
Pinus	25	0	10
Unknown	11	0	6

https://doi.org/10.1371/journal.pgen.1010097.t002

Monokaryon generation and genomic DNA isolation

To facilitate the study of highly diverse genomic regions, such as the mating loci, and to avoid heterozygosity issues in other genomic regions, we isolated single fungal spores produced by fruit bodies from dikaryon cultures (original isolated specimens, n+n), and formed monokaryotic cultures (haploid strains, n) in the lab. These monokaryotic cultures were made by hydrating dried field collected fruit bodies in the lab, and allowing the fruit bodies to eject spores onto 3% malt extract agar plates with 10 mg/L tetracyclin, 100 mg/L ampicillin, 25 mg/L streptomycin and 1 mg/L benomyl. Four germinated single spores were transferred to four new 3% malt extract agar plates with identical mixture of antibiotics and benomyl, resulting in monokaryotic cultures. All monokaryotic cultures were checked for clamp connections, which supports that only one spore was picked and mating did not already occurred among spores from the same fruit body. One of the four monokaryotic cultures were selected (except for seven specimens were more than one monokaryotic culture were included) for further analyses (S3 Table). Before DNA extraction, monokaryon cultures were grown for 2–3 weeks on nitex nylon (Sefar AG, Heiden, Switzerland) on 3% malt extract agar plates.

Two different DNA extraction protocols were used depending on the sequencing method. For Illumina sequencing, tissue from $1/4^{th}$ plate was scraped off the nylon and directly homogenized in 2 ml Lysing Matrix E tubes (MP Biomedicals, Santa Ana, CA, USA) on a FastPrep-24 (MP Biomedicals, Santa Ana, CA, USA) for 2 x 20 seconds at 4.5 m/s². Genomic DNA was extracted using the E.Z.N.A HP Fungal DNA kit (Omega Bio-Tek, Norcross, GA, USA) supplemented with 30 μl RNaseA (Qiagen, Hilden, Germany). For PacBio sequencing, tissue from 10 plates were scraped off the nylon and directly homogenized in a mortar with liquid N_2. Genomic DNA was extracted using a phenol:chloroform protocol followed by a macro (500 μg) Genomic tip (Qiagen, Hilden Germany) protocol, as described in Skrede *et al* [69].

Genome sequencing and assembly

In order to get the chromosome location and sequences of mating genes, we first Illumina sequenced the total collection of strains and provided two high-quality representative genomes for the *Trichaptum* (*T. abietinum* TA-1010-6-M1 and *T. fuscoviolaceum* TF-1002-10-M3) genus by additionally sequencing long reads (PacBio) (S2 Table).

Illumina libraries were generated by the Norwegian Sequencing Centre using the following protocol: 1 μg of genomic DNA was sheared using 96 microTUBE-50 AFA Fiber plates (Covaris Inc., Woburn, MA, USA) on a Covaris E220 system (Covaris Inc., Woburn, MA, USA). The target fragment size was 300–400 bp. gDNA samples were cleaned on a small volume Mosquito liquid handler (TTP labtech) with a 1:1 ratio of Kapa Pure beads (Roche, Basel, Switzerland) and eluted in Tris-Cl, pH 8.0. Library preparation was carried out with 500 ng sheared DNA using Kapa Hyper library prep kit (Roche, Basel, Switzerland). Barcodes were added using the Illumina UD 96 index kit (Illumina). Final libraries were PCR-amplified during 5 cycles with Kapa HIFI PCR kit (Roche, Basel, Switzerland) before standard library quality control with standard

sensitivity NGS Fragment kit (Agilent, Santa Clara, CA, USA). Quantification was performed in a qPCR with Kapa Library quantification kit (Roche, Basel, Switzerland). The first batch of library strains were sequenced with HiSeq 4000 system, and the second with NovaSeq I (S2 Table). 2x150 paired-end Illumina reads were generated by both systems. Barcodes and adapters were trimmed from final Illumina sequences using `Trim_galore 0.6.5` [70].

PacBio libraries were prepared by the Norwegian Sequencing Centre using Pacific Biosciences Express library preparation protocol (Pacific Biosciences of California, Inc, USA) without any prior fragmentation. Size selection of the final PacBio libraries was performed using BluePippin (Sage Science, Beverly, USA) and 15 Kbp cut-off. PacBio libraries were sequenced on one 1M SMRT cell using Sequel Polymerase v3.0 and sequencing chemistry v3.0. Loading was performed by diffusion and movie time was 600 min for *T. abietinum* and 900 min for both *T. fuscoviolaceum* runs.

We assembled the genome of reference *T. abietinum* using PacBio reads by different assemblers: `Flye 2.6` [71], `Canu 1.9` [72], `MECAT2` [73], `SMARTdenovo 1.0.0` [74] and `wtdbg2 2.5` [75]. Statistics of draft assemblies using these assemblers can be found in https://perisd.github.io/TriMAT/. Quality of the draft PacBio genome and percentage of consensus between draft genome and Illumina reads were quantified by `quast 5.0.2` [76] and `polca` [77], respectively. The best draft PacBio assembly based on quality statistics, canu (Table 1), was selected and Illumina-corrected using `HyPo` [78]. Scaffolds with less than 100 PacBio reads of support and less than 10 Kbp of length were removed from the final corrected genome assembly. *T. abietinum* ultrascaffolding was done using a Hymenochaetales species, *P. noxium* KPN91, which genome was assembled using PacBio reads (Accession No. GCA002287475) [79]. We first checked chromosome correspondence using `D-GENIES` [80] and manually ultrascaffolded in `Geneious 6.1.6` [81]. Chromosomes were named according to *P. noxium* chromosome similarity. We applied the same pipeline to the *T. fuscoviolaceum* reference assembly, except that ultrascaffolding was performed using `RaGOO` [82], and the *T. abietinum* genome assembly as reference. Visual inspection of syntenic comparisons were performed using `mummer 3.23` [83] and `D-GENIES`. This approach allowed us to correct the order of the ultrascaffolded chromosome 3 of *T. abietinum*, which contained 3 scaffolds. We assumed that the order of chromosome 3 must be more similar between sister-species *T. abietinum* and *T. fuscoviolaceum* than between *T. abietinum* and *P. noxium*, and the 3 scaffolds were resorted accordingly. The other ultrascaffolded *T. abietinum* chromosome 7, reminded untouched. In both *Trichaptum* assemblies, ultrascaffolded chromosomes contain artificial 10,000 Ns separating joined scaffolds. *Trichaptum fuscoviolaceum* chromosomes were composed of multiple scaffolds, except chromosome 5 that was not ultrascaffolded. Details about the ultrascaffolded canu scaffolds can be found in the GitHub page dedicated to this work. Assembly statistics of the final genomes (Table 1), such as N50, genome size, and completeness of universal single copy orthologous genes, were assessed using `quast` and `BUSCO 4.1.2` [84]. The training BUSCO database was `agaricomycetes_odb10`, which contains 2898 genes. We were able to detect 71% of the telomeric repeats (TTAGGG) [85], 20 and 14 of the 24 expected telomeric regions for each *T. abietinum* and *T. fuscoviolaceum* reference strains, respectively. For *T. abietinum* at least repeats in one telomere was detected for all chromosomes, supporting the 12 chromosome designation for this species, and suggesting that *Trichaptum* genomes were mostly telomere-telomere completed.

Genomes of the 178 strains, sequenced by the Illumina platform, were assembled with `iWGS` wrapper [86]. We selected assemblies generated by `SPAdes 3.14` [87] based on `quast` quality reports. Genome completeness was assessed with `BUSCO`. In addition, we included a DOE Joint Genome Institute (JGI) MycoCosm Illumina-sequenced and assembled *T. abietinum* strain (L15831, [88]).

Trichaptum *species classification and species tree reconstruction*

Species designation of strains was first supported based on a fast method, fastANI 1.1 [89]. With fastANI, we calculated the pairwise average nucleotide identity (ANI) among genome assemblies, whose values were then converted to a percentage dissimilarity matrix by subtracting ANI from a value of 100%. The dissimilarity data was used as distance to reconstruct a Neighbor-Joining (NJ) phylogenetic tree in MEGA v5 [90].

The utilization of gene nucleotide and amino acid sequences of universal single copy orthologs annotated with BUSCO assessed the species designation by fastANI. Individual BUSCO protein alignments were generated with MAFFT 7.455 [91]. Amino acid alignments were back translated to nucleotides using pal2nal v14 [92]. Codon columns with gaps were removed from the alignments using trimal 1.4.1 [93]. Gene sequences present in all strains that retained at least 30% of positions and with more than 300 nucleotides (100 amino acids) were selected for additional analyses. In total, 1026 BUSCO genes (35% of the genes) passed our filters. Maximum Likelihood (ML) phylogenetic trees of trimmed genes were reconstructed using IQTree 2.0.3 [94]. The best fitted evolutionary nucleotide model for each gene was estimated by ModelFinder [95] implemented in IQTree. Individual gene trees were pooled in a unique file, which was the input to reconstruct the species tree by applying a coalescent model implemented in ASTRAL 5.7.4 [96]. Species tree branch support was assessed by calculating the gene concordance factor implemented in IQTree. To assess reciprocal monophyly of BUSCO genes, ML phylogenetic trees were read in R using treeio v1.12 [97] and converted to ape v5.4 format [98]. Once species designation were associated to phylogenetic tip labels, the trees were rooted using *T. biforme* strain as an outgroup. Monophyly test was performed using spider v1.5 [99]. ML phylogenetic trees of BUSCO genes detected as top 1% in at least two nucleotide diversity statistics (see below) were drawn to a pdf using ggtree v2.2.4 [100].

Mating gene annotation, alignments and phylogenetics

Mating regions encoding the genes involved in the sexual cycle are conserved among basidiomycetes [36]. We first searched for conserved flanking genes to delimit the mating sites in these new PacBio genomes. Mating A (*MATA*) region was located using *MIP1* (mtDNA intermediate peptidase), *bfg* (beta-flanking gene) and *GLGEN* (Glycogenin-1) gene sequences. Mating B (*MATB*) region was delimited using *PAK* (syn. *CLA4*, serine/threonine protein kinase). We found both mating regions by performing a blast search in Geneious [101] using *P. noxium* flanking gene sequences as subject. Delimitation of genes and coding sequences in mating regions were performed using FGENESH and the *P. noxium* gene-finding parameters [102]. Some annotated open reading frames (ORFs) required manual curation, as the boundaries of exons vary from one strain to another stochastically. Gene designation of ORFs was assessed by BLASTing the ORF sequences, using the blastx program. An additional annotation comparison to infer the number of exons in different ORFs was done using MAKER2 [103], where we included the transcriptome dataset of L15831 *T. abietinum* as input [88].

The annotation of domains and motifs was performed using different strategies. Typical homeodomain/homeobox domains in HD proteins were annotated with CD-search using the CDD v3.18–55570 PSSMs database [104]. To differentiate *HD1* and *HD2* genes, we first screened the nuclear localization signal (NLS) domain using NLS Mapper [105]. NLS is characteristic of HD1 proteins [39,106,107]. Conserved regions enriched in proline amino acids were suggested as potential regions for activation domains (AD) for homeodomain proteins [108]. Coiled coil regions involved in the dimerization of the two homeodomain proteins were detected with Coiled coils v1.1.1 Geneious plugin. Sequence logos for each

HD protein domain were generated by using the `ggseqlogo v1.0` [109] R package after selecting representative sequences (unique sequences, `-c 1`) with `CD-HIT v4.8.1` [110]. Proteins with seven-transmembrane G protein-coupled receptor superfamily domains are usually indicative of *STE3* pheromone receptors [111]. The 7 transmembrane domains of the pheromone receptor protein were annotated with `PredictProtein` [112]. Pheromone precursor genes were screened in close proximity to the detected pheromone receptors using `pheromone_seeker.pl` script [113]. Briefly, the perl script searches common amino acid features encoded in pheromone precursor genes, such as the prenylation signal, or CaaX motif (C, cysteine; aa, two aliphatic amino acids; X is any amino acid) in the C-terminal of the pheromone precursor [40,61]. Hits with a length shorter than 100 bp or longer than 200 bp, and/or distant to *STE3* genes were considered as false positives. Consequently, we removed those hits from the annotations. Additionally, pheromone precursors in strains missing at least one hit close to *STE3.2* or *STE3.4* were searched using conserved pheromone amino acid sequences of strains in the same clade for *STE3.2* or *STE3.4* phylogenetic trees. Pheromone maturation sites were located by searching glutamic/arginine (ER) or aspartic acid/arginine (DR) amino acid motifs [39].

Once we had annotated the mating regions in the reference genomes, we were able to search for these genes in the Illumina sequenced and assembled genomes of the rest of strains. We first generated local blast databases for Illumina genomes. We BLASTed the reference flanking genes to pull out the mating regions. In case a mating region (*MATA* or *MATB*) was not contiguous (<43% and <20% of strains for *MATA* and *MATB*, respectively), but split on different contigs, we assumed those regions kept the same gene order as in the reference genomes, and we ultrascaffolded the contigs for each mating region accordingly. 999 Ns were added between joined contigs. Similar to the reference genome assemblies, we defined the mating regions to the scaffold/ultrascaffolded segment containing sequences from *bfg* to *MIP1* for *MATA* region, and from *PAK* to *SNF2* for *MATB*. Once regions were located and/or ultrascaffolded, we used the previous `FGENESH` pipeline for annotating ORFs. Gene identification was performed by BLASTing the genes from reference genomes against the mating regions. Additional identification was performed by searching family matches in the InterPro-5-RC6 database [114]. All annotations were stored in gff3 files generated by `Geneious`. Due to limitations of Illumina sequencing some genes in the mating regions were not detected probably because they were not covered by the Illumina reads (S3 Table).

For calculating the frequency of each unique gene block for each region, we followed a conservative approach. We took into account only mating regions that were assembled contiguously by `SPAdes` and did not need an ultrascaffolding step (S3 Table). The criteria apply from *bfg* to *MIP1* (*MATA*) and from *RIC1* to *SNF2* (*MATB*) genes (S3 Table). Gff3 files were the input to plot *MAT* gene order in R using `dplyr 1.0.2`, `gggenes 3.3.2`, `ggplot2 3.3.2`, and `rtracklayer 1.48.0`.

To calculate the nucleotide identity conservation of mating regions, we first aligned *MATA* and *MATB* sequence regions independently using FFT-NS-1 algorithm, 200PAM/k = 2 score matrix and default gap opening penalty and offset value with the `MAFFT 7.017` version implemented in `Geneious`. Gaps present in more than 20% of strains were removed with `trimal`. Identity plots for each region were generated in `Geneious`.

For phylogenetics, we first generated amino acid sequence alignments using `MAFFT` and back translated to nucleotides with `pal2nal`. Again, we were conservatives and codon columns with gaps were removed from the alignments using `trimal`. The trimmed alignment was converted to amino acid for ML phylogenetic tree reconstruction with `IQTree`. An evolutionary protein model for each protein was estimated by `ModelFinder`. Homeodomain and pheromone receptors were classified in clades/allelic classes according to visual inspection

of ML phylogenetic trees and pairwise amino acid identity percentages calculated in `Geneious`. Note here that allelic classes refer to similar protein sequences enclosed in a clade and not to haplotype sequences.

Mating genes, flanking genes and the species tree were plotted with `iTOL 6.5` [115]. *T. biforme* was used as the outgroup to root the trees when possible. To detect whether a mating related gene was segregating before the speciation event, we selected a random protein sequence of each allelic class to infer the phylogenetic relationship with proteins from other Hymenochaetales species, two reference species of Agaricales and one species from Polyporales.

Nucleotide statistics, tests to detect balancing selection and recombination

Trimmed codon-based sequence alignments of mating genes, their flanking genes and BUSCO genes were the input for the calculation of nucleotide statistics. Pairwise sequence estimation of synonymous and nonsynonymous substitution rates were calculated using the model of Yang and Nielsen [116] implemented in the `yn00` program of `PAML 4.9` [117]. We calculated nucleotide statistics, absolute nucleotide divergence (dxy) and relative divergence (Fst) using the `PopGenome 2.7.5` package in `R 4.0.2` [118]. Sequences were split in different alignments based on the species designation inferred from the species tree phylogeny. Each species-specific alignment was the input to calculate nucleotide diversity (π, Pi) and Tajima's D using `PopGenome`. A multilocus test for detecting balancing selection was performed with `HKAdirect 0.70b` [13]. We generated species-specific input tables for `HKAdirect` using `PopGenome`. The input tables consisted of the number of samples (nsam), segregating sites (S), absolute divergence (Divergence) and length for each species-specific gene (length_-pol and length_div). We set factor_chrm to 1 because genes are encoded in the nuclear genome. The input tables were necessary to run the multilocus test.

dS and dN boxplots, and genome-wide gene nucleotide statistic plots were generated in `R` using `cowplot 1.0.0, dplyr, ggplot2, ggrepel, PopGenome, reshape2 1.4.4,` and `rtracklayer`.

To detect evidence of recombination, homeodomain and pheromone receptor individual nucleotide alignments were analyzed in `RDPv4` [119]. Recombination events significantly detected by all seven methods (`RDP, GENECONV, Bootscan, Maxchi, Chimaera, SiSscan` and `3Seq`) were reported.

Crosses of monokaryotic strains

To test the compatibility of the inferred mating types, we designed putative compatible and incompatible crosses (S5 Table). Mating types were defined according to S3 Table and based on the phylogenetic analyses and AAI. For example, mating type 158 (A_1B_{56}) is defined by the presence of *MATA*-1 and *MATB*-56 (S3 Table). *MATA*-1 is the combination of aHD.1 (aHD2 allelic class 1 plus aHD1 allelic class 9, S5 Fig) and bHD.2 (bHD2 allelic class 2 plus bHD1 allelic class 4). And *MATB*-56 is composed by *STE3.2* allelic class 5 plus *STE3.4* allelic class 10 (S5 Fig). The mating classification was arbitrary. For that reason, for simplicity, selected candidates were described as having or not having a compatible alpha-/beta-complex and *STE3.2/ STE3.4* in S5 Table. We expected a compatible cross when one of the *MATA* complexes (aHD or bHD) and one of the pheromone receptors (STE3.2 or STE3.4) were distinct among the selected strains.

A total of 21 and 10 crosses were designed for crosses within *T. abietinum* and *T. fuscoviolaceum*, respectively, and 10 crosses between both species. Crosses were performed by plating monokaryons on 3% malt extract agar plates at 4 cm distance between the two monokaryons.

After 2–4 weeks, hyphal growth generated contact zones between both monokaryons. Then, a small piece from the middle area of the contact zone was extracted and re-plated on a new 3% malt extract agar plate. After one week of growth, we examined clamp connections by placing a sample of the culture on a slide under a Nikon Eclipse 50i (Nikon Instruments Europe BV, Amsterdam Netherlands). Images of the microscopic slides were acquired under a Zeiss Axioplan-2 imaging with Axiocam HRc microscope camera (Zeiss, Oberkochen Germany). All crosses were performed in triplicates.

Bioinformatic tools

All bioinformatic tools, programs and most scripts were implemented in UNINETT Sigma2 SAGA High-Performance Computing system (technical details here: https://bit.ly/2VkIXM2), except most R steps. R analyses were performed in Windows 10 operative system, implemented in RStudio 1.3.1073 with an R version 4.0.2. Bioinformatic tools were installed through conda [120] under the SAGA module Anaconda2/2019.03. Non-computational demanding and/or simple python steps were implemented in Jupyter notebooks using python modules installed through conda under Windows 10 Anaconda 1.9.12 version.

Dryad DOI

https://doi.org/10.5061/dryad.fxpnvx0t4 [122]

Supporting information

S1 Fig. Phylogenetic trees suggest some population structure in *Trichaptum* species. A) Neighbor-Joining tree using the (100 –ANI)/100 values as distances to reconstruct the tree. Scale bar represents (100 –ANI) / 100. B) Coalescent species tree using 1026 BUSCO ML phylogenetic trees. Scale bar represents coalescent units. Bar colors represent the species designation according to the legend. Circles in branches represent the concordance factor support (0: none ML tree agrees– 100: all 1028 ML trees agree). More detailed phylogenetic trees can be found in iTOL: https://itol.embl.de/shared/Peris_D.
(PDF)

S2 Fig. Genomes of *T. abietinum* and *T. fuscoviolaceum* are mostly syntenic. D-GENIES dot-plot of our two reference genomes. Alignment matches are represented by dots and the identity values are colored according to the legend. *MAT* region locations are indicated. Dot identity values are defined as: (number of residue matches for a segment / alignment segment length) * 100. These identity values are calculated from column 10 and 11 in PAF (Pairwise mApping Format) files generated by minimap2 [121], program implemented in D-GENIES.
(PDF)

S3 Fig. Amino acid sequence conservation in homeodomain proteins. Sequence logo plots of protein domains involve in the function of homeodomain proteins. Although the C-terminal domain was not related to a function, it was displayed due to its high conservation in protein sequences. Amino acids are colored according to chemistry as indicated in to the legend.
(PDF)

S4 Fig. Non-common CpaX motifs were detected in *Trichaptum* pheromone precursor proteins. Phe3.2 and Phe3.4 sequence alignments of unique pheromone precursor proteins are represented in panels A) and B). Sequence logo, generated by Geneious R6, is

represented at the top of each alignment to highlight conserved amino acids. Polar amino acids in the CaaX motif are squared in red. Red lines split the pheromone precursor sequences according to the allelic class of the closest mating-related pheromone receptor gene, as indicated in the sequence names on the left (PheX.X.Y, where Y is the allelic class). Letters in sequence names (i.e. -A, -B, etc) indicate unique sequences.
(PDF)

S5 Fig. ML protein phylogenetic trees show signals of balancing selection in mating genes and linked genes. ML phylogenetic trees of individual proteins from the *MATA* and *MATB* regions are represented. Species designation and continental isolation are indicated by colored bars according to the legend. Branch support was assessed using the ultrafast bootstrap (UF bootstrap) method. UF bootstrap is indicated in each branch by a gradient color according to the legend. Scale bar is represented in number of amino acid substitutions per site.
(PDF)

S6 Fig. Pairwise amino acid identity within mating proteins. Pairwise amino acid identity was calculated for protein sequences within an allelic class and between protein sequences from different allelic classes. Dots represent the average value for within or between pairwise comparisons. Median values for all proteins are represented by horizontal lines inside the boxes, and the upper and lower whiskers represent the highest and lowest values of the 1.5 * IQR (inter-quartile range), respectively. Box plots and dots were colored according to the species where the pairwise comparison was performed. Horizontal dashed line represents the maximum value of 100 - % amino acid identity. We considered 86% amino acid identity a threshold to classify sequences in an allelic class.
(PDF)

S7 Fig. Pairwise amino acid identity of mating proteins from strains with identical mating types. Pairwise amino acid identity was calculated for protein sequences within an allelic class of the same species (2 pairwise comparisons for *T. fuscoviolaceum*) and between species (2 pairwise comparisons between 2 *T. abietinum* and 2 *T. fuscoviolaceum*). Dots represent the average value for within or between pairwise comparisons. Horizontal dashed line represents the 86% amino acid identity threshold detected in S5 Fig.
(PDF)

S8 Fig. Experimental crosses support predicted compatible and incompatible mating types. Example plate and microscope pictures of the strain cross experiments are displayed on the left and on the right, respectively. Codes on the right, such as TFx1, indicate the type of cross (S5 Table). Pictures of additional crosses are indicated in S5 Table and they can be found in https://perisd.github.io/TriMAT/. When types were distinct in both mating loci clamp connections (red arrows) are observed in septae. Strain names and the inferred allelic classes for each mating gene (S2 Table) are displayed. Compatible *MATA* complexes or pheromone receptors are highlighted in green in each strain. *Tabi*, *Trichaptum abietinum*; *Tfus*, *Trichaptum fuscoviolaceum*.
(PDF)

S9 Fig. dS and dN values for mating, flanking and BUSCO genes supports balancing selection in mating genes. Panels A) and C) report the pairwise dS within each species (colored according to the legend) or between species (black) for each gene in the *MATA* and *MATB* regions, respectively. Similarly, panels B) and D) report the pairwise dN. Median values for all genes are represented by horizontal lines inside the boxes, and the upper and lower whiskers represent the highest and lowest values of the 1.5 * IQR (inter-quartile range), respectively.

Median values for BUSCO genes are represented by horizontal dashed lines and they are colored according to the legend, green and purple for within *T. abietinum* and *T. fuscoviolaceum* comparisons, respectively, and black between species comparisons.
(PDF)

S10 Fig. Detected BUSCO genes are shown to have some signal of non-reciprocal monophyly. Maximum-Likelihood phylogenetic trees of five detected BUSCO genes based on nucleotide statistics (Fig 7) are represented. Scale bar is represented in number of nucleotide substitutions per site.
(PDF)

S11 Fig. Geographic distribution of mating alleles supports long-term segregation. Stacked bar plots are represented for each mating gene. For each allelic class a bar colored according to the geographic location is drawn.
(PDF)

S12 Fig. Two recent duplications of *aHD2* genes generated xHD2 proteins. ML phylogenetic trees of a protein sequence alignment containing xHD2, aHD2 and bHD2. xHD2 sequences are highlighted with red arrows. Branch support was assessed using the ultrafast bootstrap (UF bootstrap) method. UF bootstrap is indicated in each branch by a gradient color according to the legend. Scale bar is represented in number of amino acid substitutions per site.
(PDF)

S13 Fig. Some mating alleles are older than *Trichaptum* genus. Selected regions of ML phylogenetic trees of trimmed (`trimal-gt 0.8`) protein sequence alignments containing HD2-HD1 and STE3 are displayed in panels A) and B), respectively. Branch support was assessed using the ultrafast bootstrap (UF bootstrap) method. UF bootstrap is indicated in each branch by a gradient color according to the legend. Scale bar is represented in number of amino acid substitutions per site. *Trichaptum* proteins are highlighted by red arrows or enclosed in a red bar. Allelic classes are indicated in the protein name (i.e. aHDX.Y, where Y is the allelic class). Protein sequences were retrieved from DOE-JGI MycoCosm and download from NCBI as indicated: 1. Hymneochaetales JGI protein list: Fomme: *Fomitiporia mediterranea* (MF3/22), Onnsc: *Onnia scaura* (P-53A), Phefer: *Phellinidium ferrugineofuscum* (SpK3Phefer14), Pheign: *Phellinus ignarius* (CCBS575), Phevit: *Phellinus viticola* (PhevitSig-SM15), Pheni: *Phellopilus* (*Phellinus*) *nigrolimitatus* (SigPhenig9), Porchr: *Porodaedalea chrysoloma* (FP-135951), Pornie: *Porodaedalea niemelaei* (PN71-100-IP13), Resbic: *Resinicium bicolor* (OMC78), Ricfib: *Rickenella fibula* (HBK330-10), Ricmel: *Rickenella mellea* (SZMC22713), Schpa: *Schizopora paradoxa* (KUC8140), Sidvul: *Sidera vulgaris* (OMC1730). 2. Downloaded from NCBI: [HYMENOCHAETALES] *Fomitiporia mediterranea* (MF3/22), *Pyrrhoderma noxium* (KPN91), *Shanghuangporus baumii* (Bpt 821), *Rickenella mellea* (SZMC22713); [AGARICALES] *Laccaria bicolor* (S238N-H82), *Coprinopsis cinerea* (Okayama7#130); [POLYPORALES] *Rhodonia* (*Postia*) *placenta* (Mad-698-R). To remove protein redundancy in protein collection of species retrieved from JGI, a blastp using the downloaded NCBI protein sequences and HDs and STE3s protein representatives of each allelic class was performed. For each input sequence two hits were used for sequence alignments, a protein sequence with the lowest e-value and the protein sequence with the highest coverage value. Complete ML phylogenetic trees are deposited in a shared iTOL folder: https://itol.embl.de/shared/Peris_D.
(PDF)

S1 Table. Strains used in this study. Geographical and source of isolation.
(XLSX)

S2 Table. Whole genome sequencing statistics.
(XLSX)

S3 Table. Mating gene allelic designation and predicted mating types.
(XLSX)

S4 Table. Sequences detected as recombinant.
(XLSX)

S5 Table. dS and dN average pairwise comparisons between mating genes and BUSCO genes (dX Gene/dX BUSCO).
(XLSX)

S6 Table. Experimental mating crosses.
(XLSX)

Author Contributions

Conceptualization: David Peris, Dabao Sun Lu, Håvard Kauserud, Inger Skrede.

Data curation: David Peris, Dabao Sun Lu, Vilde Bruhn Kinneberg, Ine-Susanne Methlie, Malin Stapnes Dahl, Inger Skrede.

Formal analysis: David Peris.

Funding acquisition: David Peris, Inger Skrede.

Investigation: David Peris, Dabao Sun Lu, Vilde Bruhn Kinneberg, Inger Skrede.

Methodology: David Peris, Dabao Sun Lu, Inger Skrede.

Project administration: Inger Skrede.

Resources: David Peris, Dabao Sun Lu, Vilde Bruhn Kinneberg, Ine-Susanne Methlie, Malin Stapnes Dahl, Inger Skrede.

Software: David Peris.

Supervision: David Peris, Håvard Kauserud, Inger Skrede.

Validation: David Peris, Dabao Sun Lu, Vilde Bruhn Kinneberg, Timothy Y. James, Håvard Kauserud, Inger Skrede.

Visualization: David Peris.

Writing – original draft: David Peris, Dabao Sun Lu, Inger Skrede.

Writing – review & editing: David Peris, Dabao Sun Lu, Vilde Bruhn Kinneberg, Ine-Susanne Methlie, Timothy Y. James, Håvard Kauserud, Inger Skrede.

References

1. Charlesworth D (2006) Balancing Selection and its effects on sequences in nearby genome regions. Plos Genetics 2: e64. https://doi.org/10.1371/journal.pgen.0020064 PMID: 16683038

2. Dobzhansky T (1951) Genetics and the Origin of Species. New York: Columbia University Press.

3. Johnston SE, Gratten J, Berenos C, Pilkington JG, Clutton-Brock TH, et al. (2013) Life history trade-offs at a single locus maintain sexually selected genetic variation. Nature 502: 93–95. https://doi.org/10.1038/nature12489 PMID: 23965625

4. Mitchell-Olds T, Willis JH, Goldstein DB (2007) Which evolutionary processes influence natural genetic variation for phenotypic traits? Nature Reviews Genetics 8: 845–856. https://doi.org/10.1038/nrg2207 PMID: 17943192

5. Bergland AO, Behrman EL, O'Brien KR, Schmidt PS, Petrov DA (2014) Genomic evidence of rapid and stable adaptive oscillations over seasonal time scales in *Drosophila*. Plos Genetics 10: e1004775. https://doi.org/10.1371/journal.pgen.1004775 PMID: 25375361

6. Úbeda F, Haig D (2004) Sex-specific meiotic drive and selection at an imprinted locus. Genetics 167: 2083–2095. https://doi.org/10.1534/genetics.103.021303 PMID: 15342542

7. Charlesworth B, Charlesworth D (2010) Elements of evolutionary genetics. Roberts and Co. Publishers. Available: http://books.google.com/books?id=dgNFAQAAIAAJ.

8. Klein J (1980) Generation of diversity at *MHC* loci: implications for T-cell receptor repertoires. In: Fougereau M, Dausset J, editors. Immunology. London, UK: Academic Press. pp. 239.

9. Hein J, Schierup M, and Wiuf C (2005) Gene Genealogies, Variation and Evolution: A Primer in Coalescent Theory. New York: Oxford University Press.

10. Richman A (2000) Evolution of balanced genetic polymorphism. Mol Ecol 9: 1953–1963. https://doi.org/10.1046/j.1365-294x.2000.01125.x PMID: 11123608

11. Hudson RR, Kreitman M, Aguadé M (1987) A Test of neutral molecular evolution based on nucleotide data. Genetics 116: 153. https://doi.org/10.1093/genetics/116.1.153 PMID: 3110004

12. Tajima F (1989) Statistical method for testing the neutral mutation hypothesis by DNA polymorphism. Genetics 123: 585–595. https://doi.org/10.1093/genetics/123.3.585 PMID: 2513255

13. Esteve-Codina A, Paudel Y, Ferretti L, Raineri E, Megens HJ, et al. (2013) Dissecting structural and nucleotide genome-wide variation in inbred Iberian pigs. BMC Genomics 14: 148. https://doi.org/10.1186/1471-2164-14-148 PMID: 23497037

14. Ségurel L, Thompson EE, Flutre T, Lovstad J, Venkat A, et al. (2012) The ABO blood group is a trans-species polymorphism in primates. Proc Natl Acad Sci U S A 201210603. https://doi.org/10.1073/pnas.1210603109 PMID: 23091028

15. Cagliani R, Fumagalli M, Biasin M, Piacentini L, Riva S, et al. (2010) Long-term balancing selection maintains trans-specific polymorphisms in the human *TRIM5* gene. Human Genetics 128: 577–588. https://doi.org/10.1007/s00439-010-0884-6 PMID: 20811909

16. Cagliani R, Guerini FR, Fumagalli M, Riva S, Agliardi C, et al. (2012) A trans-specific polymorphism in *ZC3HAV1* is maintained by long-standing balancing selection and may confer susceptibility to multiple sclerosis. Mol Biol Evol 29: 1599–1613. https://doi.org/10.1093/molbev/mss002 PMID: 22319148

17. Castric V, Vekemans X (2004) Plant self-incompatibility in natural populations: a critical assessment of recent theoretical and empirical advances. Mol Ecol 13: 2873–2889. https://doi.org/10.1111/j.1365-294X.2004.02267.x PMID: 15367105

18. Wright S (1939) The distribution of self-sterility alleles in populations. Genetics 24: 538. https://doi.org/10.1093/genetics/24.4.538 PMID: 17246937

19. Wu J, Saupe SJ, Glass NL (1998) Evidence for balancing selection operating at the *het-c* heterokaryon incompatibility locus in a group of filamentous fungi. Proc Natl Acad Sci U S A 95: 12398. https://doi.org/10.1073/pnas.95.21.12398 PMID: 9770498

20. Hittinger CT, Gonçalves P, Sampaio JP, Dover J, Johnston M, Rokas A (2010) Remarkably ancient balanced polymorphisms in a multi-locus gene network. Nature 464: 54–58. https://doi.org/10.1038/nature08791 PMID: 20164837

21. Boocock J, Sadhu MJ, Bloom JS, Kruglyak L (2021) Ancient balancing selection maintains incompatible versions of the galactose pathway in yeast. Science 371: 415–419. https://doi.org/10.1126/science.aba0542 PMID: 33479156

22. May G, Shaw F, Badrane H, Vekemans X (1999) The signature of balancing selection: Fungal mating compatibility gene evolution. Proc Natl Acad Sci U S A 96: 9172. https://doi.org/10.1073/pnas.96.16.9172 PMID: 10430915

23. James TY, Liou SR, Vilgalys R (2004) The genetic structure and diversity of the A and B mating-type genes from the tropical oyster mushroom, *Pleurotus djamor*. Fungal Genet Biol 41: 813–825. https://doi.org/10.1016/j.fgb.2004.04.005 PMID: 15219565

24. Devier B, Aguileta G, Hood ME, Giraud T (2009) Ancient *trans*-specific polymorphism at pheromone receptor genes in Basidiomycetes. Genetics 181: 209. https://doi.org/10.1534/genetics.108.093708 PMID: 19001292

25. Engh IB, Skrede I, Sætre GP, Kauserud H (2010) High variability in a mating type linked region in the dry rot fungus *Serpula lacrymans* caused by frequency-dependent selection? BMC Genetics 11: 64. https://doi.org/10.1186/1471-2156-11-64 PMID: 20624315

26. van Diepen LTA, Olson Å, Ihrmark K, Stenlid J, James TY (2013) Extensive trans-specific polymorphism at the Mating type locus of the root decay fungus *Heterobasidion*. Mol Biol Evol 30: 2286–2301. https://doi.org/10.1093/molbev/mst126 PMID: 23864721

27. Bensaude, M (1918) Recherches sur le cycle évolutif et la sexualité chez lesBasidiomycètes. [dissertation]. Faculté des Sciences de Paris, Imprimerie Nemourienne, Henri Bouloy, Nemours, France.

28. Casselton LA, Econoumou A (1985) Dikaryon formation. In: Moore D, Casselton LA, Wood DA, Frankland JC, editors. Developmental biology of higher fungi. Cambridge, United Kingdom: Cambridge University Press. pp. 213–229.

29. Kemp RFO (1977) Oidial homing and the taxonomy and speciation of basidiomycetes with special reference to the genus *Coprinus*. In: Clémençon H, editors. The species concept in hymenomycetes. Vaduz, Switzerland: Cramer. pp. 259–276.

30. Crockatt ME, Pierce GI, Camden RA, Newell PM, Boddy L (2008) Homokaryons are more combative than heterokaryons of *Hericium coralloides*. Fungal Ecology 1: 40–48.

31. Hiscox J, Hibbert C, Rogers HJ, Boddy L (2010) Monokaryons and dikaryons of *Trametes versicolor* have similar combative, enzyme and decay ability. Fungal Ecology 3: 347–356.

32. Kües U (2000) Life history and developmental processes in the Basidiomycete *Coprinus cinereus*. Microbiol Mol Biol R 64: 316. https://doi.org/10.1128/MMBR.64.2.316-353.2000 PMID: 10839819

33. Whitehouse HLK (1949) Multiple-allelomorph heterothallism in the fungi. New Phytol 48: 212–244. https://doi.org/10.1111/j.1469-185x.1949.tb00582.x PMID: 24536314

34. Bennett Richard J. and Turgeon B. Gillian (2017) Fungal Sex: The Ascomycota. American Society of Microbiology. https://doi.org/10.1099/mic.0.000401 PMID: 27902434

35. Heitman J, Sun S, James TY (2013) Evolution of fungal sexual reproduction. Mycologia 105: 1–27. https://doi.org/10.3852/12-253 PMID: 23099518

36. Coelho MA, Bakkeren G, Sun S, Hood ME, and Giraud T (2017) Fungal Sex: The Basidiomycota. American Society of Microbiology. https://doi.org/10.1099/mic.0.000401 PMID: 27902434

37. Kües U, Casselton LA (1993) The origin of multiple mating types in mushrooms. J Cell Sci 104: 227.

38. Hiscock SJ, Kües U (1999) Cellular and molecular mechanisms of sexual incompatibility in plants and fungi. In: Jeon KW, editors. International Review of Cytology. Academic Press. pp. 165–295. https://doi.org/10.1016/s0074-7696(08)61781-7 PMID: 10494623

39. Casselton LA, Olesnicky NS (1998) Molecular genetics of mating recognition in Basidiomycete fungi. Microbiol Mol Biol R 62: 55. https://doi.org/10.1128/MMBR.62.1.55-70.1998 PMID: 9529887

40. Niculita-Hirzel H, Labbé J, Kohler A, Le Tacon F, Martin F, Sanders IR, Kües U (2008) Gene organization of the mating type regions in the ectomycorrhizal fungus *Laccaria bicolor* reveals distinct evolution between the two mating type loci. New Phytol 180: 329–342. https://doi.org/10.1111/j.1469-8137.2008.02525.x PMID: 18557817

41. Lee SC, Ni M, Li W, Shertz C, Heitman J (2010) The evolution of sex: a perspective from the fungal kingdom. Microbiol Mol Biol R 74: 298–340. https://doi.org/10.1128/MMBR.00005-10 PMID: 20508251

42. Kües U, James TY, Heitman J (2011) Mating type in Basidiomycetes: unipolar, bipolar, and tetrapolar patterns of sexuality. In: Pöggeler S, Wöstemeyer J, editors. Evolution of Fungi and Fungal-Like Organisms. Berlin, Heidelberg: Springer Berlin Heidelberg. pp. 97–160.

43. Skrede I, Maurice S, Kauserud H (2013) Molecular characterization of sexual diversity in a population of *Serpula lacrymans* a tetrapolar Basidiomycete. G3 3: 145. https://doi.org/10.1534/g3.112.003731 PMID: 23390592

44. Maia TM, Lopes ST, Almeida JMGC, Rosa LH, Sampaio JP, Gonçalves P, Coelho MA (2015) Evolution of mating systems in Basidiomycetes and the genetic architecture underlying mating-type determination in the yeast *Leucosporidium scottii*. Genetics 201: 75. https://doi.org/10.1534/genetics.115.177717 PMID: 26178967

45. Wang W, Lian L, Xu P, Chou T, Mukhtar I, Osakina A, et al. (2016) Advances in understanding mating type gene organization in the mushroom-forming fungus *Flammulina velutipes*. G3 Genes|Genomes|Genetics 6: 3635–3645. https://doi.org/10.1534/g3.116.034637 PMID: 27621376

46. Sipos G, Prasanna AN, Walter MC, O'Connor E, Bálint B, et al. (2017) Genome expansion and lineage-specific genetic innovations in the forest pathogenic fungi Armillaria. Nature Ecology & Evolution 1: 1931–1941.

47. Furtado JS (1966) Significance of the clamp-connection in the Basidiomycetes. Persoonia—Molecular Phylogeny and Evolution of Fungi 4: 125–144.

48. Swiezynski KM, Day PR (1960) Heterokaryon formation in *Coprinus lagopus*. Genetics Research 1: 114–128.

49. Swiezynski KM, Day PR (1960) Migration of nuclei in *Coprinus lagopus*. Genetics Research 1: 129–139.

50. Burnett JH (1965) The natural history of recombination systems. In: Esser K, Raper JR, editors. Incompatibility in Fungi: A Symposium held at the 10th International Congress of Botany at Edinburgh, August 1964. Berlin, Heidelberg: Springer Berlin Heidelberg. pp. 98–113.

51. Raper, John R (3-1-1966) Genetics of sexuality in higher fungi. Ronal Press Company.

52. Seierstad KS, Fossdal R, Miettinen O, Carlsen T, Skrede I, Kauserud H (2021) Contrasting genetic structuring in the closely related basidiomycetes *Trichaptum abietinum* and *T. fuscoviolaceum* (Hymenochaetales). Fungal Biology 125: 269–275. https://doi.org/10.1016/j.funbio.2020.11.001 PMID: 33766305

53. Magasi LP (1976) Incompatibility factors in *Polyporus abietinus*, their numbers and distribution. Memoirs of the New York botanical garden 28: 163–173.

54. Macrae R (1966) Pairing incompatibility and other distinctions among *Hirschioporus* [*Polyporus*] *abietinus*, *H. fusco-violaceus*, and *H. laricinus*. Canadian Journal of Botany 45: 1371–1398.

55. Chen P, Sapperstein SK, Choi JD, Michaelis S (1997) Biogenesis of the *Saccharomyces cerevisiae* mating pheromone a-factor. Journal of Cell Biology 136: 251–269. https://doi.org/10.1083/jcb.136.2.251 PMID: 9015298

56. Raper JR, Krongelb GS, Baxter MG (1958) The number and distribution of incompatibility factors in *Schizophyllum*. JSTOR 92: 221–232.

57. Idnurm A, Hood ME, Johannesson H, Giraud T (2015) Contrasted patterns in mating-type chromosomes in fungi: Hotspots versus coldspots of recombination. Fungal Biology Reviews 29: 220–229. https://doi.org/10.1016/j.fbr.2015.06.001 PMID: 26688691

58. Gutiérrez-Valencia J, Hughes PW, Berdan EL, Slotte T (2021) The genomic architecture and evolutionary fates of supergenes. Genome Biol Evol 13. https://doi.org/10.1093/gbe/evab057 PMID: 33739390

59. Lukens L, Yicun H, May G (1996) Correlation of genetic and physical maps at the A mating-type locus of *Coprinus cinereus*. Genetics 144: 1471–1477. https://doi.org/10.1093/genetics/144.4.1471 PMID: 8978036

60. James TY (2007) Analysis of mating-type locus organization and synteny in mushroom fungi: beyond model species. In: Heitman J, editors. Sex in Fungi: Molecular Determinationand Evolutionary Implications. Washington, D.C.: ASM Press. pp. 317–331.

61. Raudaskoski M, Kothe E (2010) Basidiomycete mating type genes and pheromone signaling. Eukaryot Cell 9: 847. https://doi.org/10.1128/EC.00319-09 PMID: 20190072

62. Isaya G (2004) 90—Mitochondrial intermediate peptidase. In: Barrett AJ, Rawlings ND, Woessner JF, editors. Handbook of Proteolytic Enzymes (Second Edition). London: Academic Press. pp. 366–369.

63. Röhr H, Kües U, Stahl U (1999) Recombination:organelle DNA of plants and fungi: inheritance and recombination. In: Esser K, Kadereit JW, Lüttge U, Runge M, editors. Progress in Botany: Genetics Cell Biology and Physiology Systematics and Comparative Morphology Ecology and Vegetation Science. Berlin, Heidelberg: Springer Berlin Heidelberg. pp. 39–87.

64. Chung CL, Lee TJ, Akiba M, Lee HH, Kuo TH, et al. (2017) Comparative and population genomic landscape of *Phellinus noxius*: A hypervariable fungus causing root rot in trees. Mol Ecol 26: 6301–6316. https://doi.org/10.1111/mec.14359 PMID: 28926153

65. James TY, Sun S, Li W, Heitman J, Kuo HC, et al. (2013) Polyporales genomes reveal the genetic architecture underlying tetrapolar and bipolar mating systems. Mycologia 105: 1374–1390. https://doi.org/10.3852/13-162 PMID: 23928418

66. Fischer M (2002) A new wood-decaying basidiomycete species associated with esca of grapevine: *Fomitiporia mediterranea* (Hymenochaetales). Mycological Progress 1: 315–324.

67. Schmoll M, Seibel C, Tisch D, Dorrer M, Kubicek CP (2010) A novel class of peptide pheromone precursors in ascomycetous fungi. Mol Microbiol 77: 1483–1501.

68. Martin SH, Wingfield BD, Wingfield MJ, Steenkamp ET (2011) Causes and consequences of variability in peptide mating pheromones of Ascomycete fungi. Mol Biol Evol 28: 1987–2003. https://doi.org/10.1093/molbev/msr022 PMID: 21252281

69. Skrede I, Murat C, Hess J, Maurice S, Stønstebø JH, et al. (2021) Contrasting demographic histories revealed in two invasive populations of the dry rot fungus *Serpula lacrymans*. Mol Ecol. https://doi.org/10.1111/mec.15934 PMID: 33955084

70. Krueger F (2019) Trim Galore! https://github.com/FelixKrueger/TrimGalore

71. Kolmogorov M, Yuan J, Lin Y, Pevzner PA (2019) Assembly of long, error-prone reads using repeat graphs. Nature biotechnology 37: 540–546. https://doi.org/10.1038/s41587-019-0072-8 PMID: 30936562

72. Koren S, Walenz BP, Berlin K, Miller JR, Bergman NH, Phillippy AM (2017) Canu: scalable and accurate long-read assembly via adaptive k-mer weighting and repeat separation. Genome Res 27: 722–736. https://doi.org/10.1101/gr.215087.116 PMID: 28298431

73. Xiao CL, Chen Y, Xie SQ, Chen KN, Wang Y, et al. (2017) MECAT: fast mapping, error correction, and de novo assembly for single-molecule sequencing reads. Nature Methods 14: 1072–1074. https://doi.org/10.1038/nmeth.4432 PMID: 28945707

74. Liu H, Wu S, Li A, Ruan J (2020) SMARTdenovo: A de novo assembler using long noisy reads. Preprints.

75. Ruan J, Li H (2020) Fast and accurate long-read assembly with wtdbg2. Nature Methods 17: 158. https://doi.org/10.1038/s41592-019-0669-3 PMID: 31819265

76. Gurevich A, Saveliev V, Vyahhi N, Tesler G (2013) QUAST: quality assessment tool for genome assemblies. Bioinformatics 29: 1072–1075. https://doi.org/10.1093/bioinformatics/btt086 PMID: 23422339

77. Zimin AV, Salzberg SL (2020) The genome polishing tool POLCA makes fast and accurate corrections in genome assemblies. PLoS Computational Biology 16: e1007981. https://doi.org/10.1371/journal.pcbi.1007981 PMID: 32589667

78. Kundu R, Casey J, Sung WK (2019) HyPo: super fast & accurate polisher for long read genome assemblies. bioRxiv 2019.

79. Lee HH, Ke HM, Lin CYI, Lee TJ, Chung CL, Tsai IJ (2019) Evidence of extensive intraspecific noncoding reshuffling in a 169-kb mitochondrial genome of a Basidiomycetous fungus. Genome Biol Evol 11: 2774–2788. https://doi.org/10.1093/gbe/evz181 PMID: 31418013

80. Cabanettes F, Klopp C (2018) D-GENIES: dot plot large genomes in an interactive, efficient and simple way. PeerJ 6: e4958. https://doi.org/10.7717/peerj.4958 PMID: 29888139

81. Kearse M, Moir R, Wilson A, Stones-Havas S, Cheung M, et al. (2012) Geneious Basic: An integrated and extendable desktop software platform for the organization and analysis of sequence data. Bioinformatics 28: 1647–1649. https://doi.org/10.1093/bioinformatics/bts199 PMID: 22543367

82. Alonge M, Soyk S, Ramakrishnan S, Wang X, Goodwin S, et al. (2019) RaGOO: fast and accurate reference-guided scaffolding of draft genomes. Genome Biol 20: 224. https://doi.org/10.1186/s13059-019-1829-6 PMID: 31661016

83. Kurtz S, Phillippy A, Delcher A, Smoot M, Shumway M, et al. (2004) Versatile and open software for comparing large genomes. Genome Biol 5: R12. https://doi.org/10.1186/gb-2004-5-2-r12 PMID: 14759262

84. Waterhouse RM, Seppey M, Simão FA, Manni M, Ioannidis P, et al. (2018) BUSCO applications from quality assessments to gene prediction and phylogenomics. Mol Biol Evol 35: 543–548. https://doi.org/10.1093/molbev/msx319 PMID: 29220515

85. Cervenák F, Sepšiová R, Nosek J, Tomáška L (2021) Step-by-step evolution of telomeres: lessons from yeasts. Genome Biol Evol 13. https://doi.org/10.1093/gbe/evaa268 PMID: 33537752

86. Zhou X, Peris D, Kominek J, Kurtzman CP, Hittinger CT, Rokas A (2016) *in silico* Whole Genome Sequencer & Analyzer (iWGS): a computational pipeline to guide the design and analysis of de novo genome sequencing studies. G3 (Bethesda) 6: 3655–3670.

87. Bankevich A, Nurk S, Antipov D, Gurevich AA, Dvorkin M, et al. (2012) SPAdes: a new genome assembly algorithm and its applications to single-cell sequencing. Journal of Computational Biology 19: 455–477. https://doi.org/10.1089/cmb.2012.0021 PMID: 22506599

88. Varga T, Krizsán K, Földi C, Dima B, Sánchez-García M, et al. (2019) Megaphylogeny resolves global patterns of mushroom evolution. Nature Ecology & Evolution 3: 668–678.

89. Jain C, Rodriguez R, Phillippy AM, Konstantinidis KT, Aluru S (2018) High throughput ANI analysis of 90K prokaryotic genomes reveals clear species boundaries. Nature Communications 9: 5114. https://doi.org/10.1038/s41467-018-07641-9 PMID: 30504855

90. Tamura K, Peterson D, Peterson N, Stecher G, Nei M, Kumar S (2011) MEGA5: Molecular Evolutionary Genetics Analysis using Maximum Likelihood, evolutionary distance, and Maximum Parsimony methods. Mol Biol Evol 28: 2731–2739. https://doi.org/10.1093/molbev/msr121 PMID: 21546353

91. Katoh K, Standley DM (2013) MAFFT multiple sequence alignment software version 7: improvements in performance and usability. Mol Biol Evol 30: 772–780. https://doi.org/10.1093/molbev/mst010 PMID: 23329690

92. Suyama M, Torrents D, Bork P (2006) PAL2NAL: robust conversion of protein sequence alignments into the corresponding codon alignments. Nucl Acids Res 34: W609–W612. https://doi.org/10.1093/nar/gkl315 PMID: 16845082

93. Capella-Gutiérrez S, Silla-Martínez JM, Gabaldón T (2009) trimAl: a tool for automated alignment trimming in large-scale phylogenetic analyses. Bioinformatics 25: 1972–1973. https://doi.org/10.1093/bioinformatics/btp348 PMID: 19505945

94. Minh BQ, Schmidt HA, Chernomor O, Schrempf D, Woodhams MD, et al. (2020) IQ-TREE 2: new models and efficient methods for phylogenetic inference in the genomic era. Mol Biol Evol 37: 1530–1534. https://doi.org/10.1093/molbev/msaa015 PMID: 32011700

95. Kalyaanamoorthy S, Minh BQ, Wong TKF, von Haeseler A, Jermiin LS (2017) ModelFinder: fast model selection for accurate phylogenetic estimates. Nat Meth 14: 587–589. https://doi.org/10.1038/nmeth.4285 PMID: 28481363

96. Yin J, Zhang C, Mirarab S (2019) ASTRAL-MP: scaling ASTRAL to very large datasets using randomization and parallelization. Bioinformatics 35: 3961–3969. https://doi.org/10.1093/bioinformatics/btz211 PMID: 30903685

97. Wang LG, Lam TT-Y, Xu S, Dai Z, Zhou L, et al. (2019) Treeio: an R package for phylogenetic tree input and output with richly annotated and associated data. Mol Biol Evol 37: 599–603. https://doi.org/10.1093/molbev/msz240 PMID: 31633786.

98. Paradis E, Schliep K (2018) ape 5.0: an environment for modern phylogenetics and evolutionary analyses in R. Bioinformatics 35: 526–528.

99. Brown SDJ, Collins RA, Boyer S, Lefort MC, Malumbres-Olarte J, et al. (2012) Spider: An R package for the analysis of species identity and evolution, with particular reference to DNA barcoding. Mol Ecol Resour 12: 562–565. https://doi.org/10.1111/j.1755-0998.2011.03108.x PMID: 22243808

100. Yu G, Smith DK, Zhu H, Guan Y, Lam TT-Y (2017) ggtree: an R package for visualization and annotation of phylogenetic trees with their covariates and other associated data. Methods Ecol Evol 8: 28–36. https://doi.org/10.1111/2041-210X.12628

101. Altschul S, Gish W, Miller W, Myers E, Lipman D (1990) Basic local alignment search tool. J Mol Biol 215: 403–410. https://doi.org/10.1016/S0022-2836(05)80360-2 PMID: 2231712

102. Solovyev V, Kosarev P, Seledsov I, Vorobyev D (2006) Automatic annotation of eukaryotic genes, pseudogenes and promoters. Genome Biol 7: S10. https://doi.org/10.1186/gb-2006-7-s1-s10 PMID: 16925832

103. Holt C, Yandell M (2011) MAKER2: an annotation pipeline and genome-database management tool for second-generation genome projects. BMC Bioinformatics 12: 491. https://doi.org/10.1186/1471-2105-12-491 PMID: 22192575

104. Marchler-Bauer A, Bryant SH (2004) CD-Search: protein domain annotations on the fly. Nucl Acids Res 32: W327–W331. https://doi.org/10.1093/nar/gkh454 PMID: 15215404

105. Kosugi S, Hasebe M, Tomita M, Yanagawa H (2009) Systematic identification of cell cycle-dependent yeast nucleocytoplasmic shuttling proteins by prediction of composite motifs. Proc Natl Acad Sci U S A 106: 10171. https://doi.org/10.1073/pnas.0900604106 PMID: 19520826

106. Kronstad JW, Leong SA (1990) The b mating-type locus of *Ustilago maydis* contains variable and constant regions. Gene Dev 4: 1384–1395. https://doi.org/10.1101/gad.4.8.1384 PMID: 2227416

107. Spit A, Hyland RH, Mellor EJ, Casselton LA (1998) A role for heterodimerization in nuclear localization of a homeodomain protein. Proc Natl Acad Sci U S A 95: 6228. https://doi.org/10.1073/pnas.95.11.6228 PMID: 9600947

108. James TY, Lee M, van Diepen LTA (2011) A single mating-type locus composed of homeodomain genes promotes nuclear migration and heterokaryosis in the white-rot fungus *Phanerochaete chrysosporium*. Eukaryot Cell 10: 249. https://doi.org/10.1128/EC.00212-10 PMID: 21131435

109. Wagih O (2017) ggseqlogo: a versatile R package for drawing sequence logos. Bioinformatics 33: 3645–3647. https://doi.org/10.1093/bioinformatics/btx469 PMID: 29036507

110. Fu L, Niu B, Zhu Z, Wu S, Li W (2012) CD-HIT: accelerated for clustering the next-generation sequencing data. Bioinformatics 28: 3150–3152. https://doi.org/10.1093/bioinformatics/bts565 PMID: 23060610

111. Riquelme M, Challen MP, Casselton LA, Brown AJ (2005) The origin of multiple *B* mating specificities in *Coprinus cinereus*. Genetics 170: 1105. https://doi.org/10.1534/genetics.105.040774 PMID: 15879506

112. Bernhofer M, Dallago C, Karl T, Satagopam V, Heinzinger M, et al. (2021) PredictProtein—predicting protein structure and function for 29 years. Nuc Acids Res 49:W535–W540. https://doi.org/10.1093/nar/gkab354 PMID: 33999203

113. Mujic AB, Kuo A, Tritt A, Lipzen A, Chen C, et al. (2017) Comparative genomics of the ectomycorrhizal sister species *Rhizopogon vinicolor* and *Rhizopogon vesiculosus* (Basidiomycota: Boletales) reveals a divergence of the mating type B locus. G3 Genes|Genomes|Genetics 7: 1775–1789. https://doi.org/10.1534/g3.117.039396 PMID: 28450370

114. Mitchell AL, Attwood TK, Babbitt PC, Blum M, Bork P, et al. (2019) InterPro in 2019: improving coverage, classification and access to protein sequence annotations. Nucl Acids Res 47: D351–D360. https://doi.org/10.1093/nar/gky1100 PMID: 30398656

115. Letunic I, Bork P (2019) Interactive Tree Of Life (iTOL) v4: recent updates and new developments. Nucl Acids Res 47: W256–W259. https://doi.org/10.1093/nar/gkz239 PMID: 30931475

116. Yang Z, Nielsen R (2000) Estimating synonymous and nonsynonymous substitution rates under realistic evolutionary models. Mol Biol Evol 17: 32–43. https://doi.org/10.1093/oxfordjournals.molbev.a026236 PMID: 10666704

117. Yang Z (2007) PAML 4: Phylogenetic Analysis by Maximum Likelihood. Mol Biol Evol 24: 1586–1591. https://doi.org/10.1093/molbev/msm088 PMID: 17483113

118. R Development Core Team (2010) R: a language and environment for statistical computing. Vienna, Austria: R Foundation for Statistical Computing. Available.

119. Martin DP, Murrell B, Golden M, Khoosal A, Muhire B (2015) RDP4: Detection and analysis of recombination patterns in virus genomes. Virus Evol 1. https://doi.org/10.1093/ve/vev003 PMID: 27774277

120. Anaconda Software Distribution (2016) Conda, version 4.8.3.

121. Li H (2018) Minimap2: pairwise alignment for nucleotide sequences. Bioinformatics 34: 3094–3100. https://doi.org/10.1093/bioinformatics/bty191 PMID: 29750242

122. Peris D, Lu DS, Kinneberg VB, Methlie I-S, Dahl MS, James TY, Kauserud H, Skrede I (2022) Large-scale fungal strain sequencing unravels the molecular diversity in mating loci maintained by long-term balancing selection. https://doi.org/10.5061/dryad.fxpnvx0t4

Paper II

Introgression between highly divergent fungal sister species

Vilde Bruhn Kinneberg[1,4*], Dabao Sun Lü[1], David Peris[1,2], Mark Ravinet[3] and Inger Skrede[1]

[1]Section for Genetics and Evolutionary Biology, Department of Biosciences, University of Oslo, 0316 Oslo, Norway

[2]Department of Food Biotechnology, Institute of Agrochemistry and Food Technology (IATA), CSIC, Valencia Spain.

[3]School of Life Sciences, University of Nottingham, Nottingham, NG7 2RD United Kingdom

[4]Evolution and Paleobiology, Natural History Museum, University of Oslo, 0318 Oslo, Norway

*Corresponding author: v.b.kinneberg@nhm.uio.no

Abstract

To understand how species evolve and adapt to changing environments, it is important to study gene flow and introgression due to their influence on speciation and radiation events. Here, we apply a novel experimental system for investigating these mechanisms using natural populations. The system is based on two fungal sister species with morphological and ecological similarities occurring in overlapping habitats. We examined introgression between these species by conducting whole genome sequencing of individuals from populations in North America and Europe. We assessed genome wide nucleotide divergence and performed crossing experiments to study reproductive barriers. We further used ABBA-BABA statistics together with a network analysis to investigate introgression, and conducted demographic modelling to gain insight into divergence times and introgression events. The results revealed that the species are highly divergent and incompatible in vitro. Despite this, small regions of introgression were scattered throughout the genomes and one introgression event likely involves a ghost population (extant or extinct). This study demonstrates that introgression can be found among divergent species and that population histories can be studied without collections of all the populations involved. Moreover, the experimental system is shown to be a useful tool for research on reproductive isolation in natural populations.

Introduction

Speciation can occur rapidly, changing the course of evolution in a single event, or over a long period of time with gradual shifts from semi-compatible populations to complete divergence (Nosil et al. 2017). When speciation occurs gradually, barriers to gene exchange do not arise immediately and gene flow between diverging populations can be maintained. Consequently, hybrid individuals may form (Harrison and Larson 2014; Ravinet et al. 2018). If these hybrids backcross into one of the parental species, a scenario termed introgression (Anderson and Hubricht 1938; Aguillon 2022), it can result in unique genetic combinations (Stukenbrock 2016). The amalgamation of genes across lineages can also contribute to the strengthening of barriers to gene exchange. These barriers can arise when selection increases reproductive isolation, a process known as reinforcement (Butlin 1987). Hybrids can contribute to reinforcement because they often have detrimental gene combinations, resulting in poorer fitness compared to the parental species (Abbott et al. 2013). Consequently, hybridization can both give rise to beneficial gene combinations which selection can act upon, and at the same time accelerate the divergence process by contributing to reproductive barriers (Abbott et al. 2010).

Hybridization is established as a common event in nature and can lead to the formation of hybrid species (Mallet et al. 2016; Ackermann et al. 2019; Eberlein et al. 2019; Grant and Grant 2019), as seen in several taxa including plants (e.g., *Senecio* spp.; Hegarty and Hiscock 2005; Wood et al. 2009), animals (e.g., *Heliconius* butterflies; Mavárez et al. 2006 and *Passer italiae*; Hermansen et al. 2011), and certain fungal groups (e.g., *Saccharomyces* spp.; Langdon et al. 2019 and *Zymoseptoria pseudotritici*; Stukenbrock et al. 2012). Even though hybridization seems to be common in contemporary populations, we do not know the impact it will have on future populations. By studying the genomes of contemporary species with a history of hybridization and introgression, it may be possible to understand how previous interspecific gene exchange have influenced the populations we observe today and infer the future effect of current events.

When taxa have diverged over a long period of time, it can be difficult to discover ancient admixture as genomic signals of introgression can be blurred over macroevolutionary time. Moreover, detection might be difficult due to the deleterious nature of most introgressed genes between divergent species, or lack of time for introgressed regions to spread in the population if the gene flow is recent (Maxwell et al. 2018). Hence, the evolutionary history of a genus can be complex even though current investigations recover clear and resolved phylogenies (Keuler et al. 2020).

Signs of ancient or low frequency introgression have been possible to detect using high-throughput sequencing and statistical models (e.g., Crowl et al. 2019; Ravinet et al. 2018). Regions of introgression might constitute small parts of otherwise divergent genomes due to erosion of linkage by recombination coupled with a long period of mostly independent evolution or purifying selection (Maxwell et al. 2018; Ravinet et al. 2018; Schumer et al. 2018; Martin et al. 2019; Cuevas et al. 2022). The retention of specific introgressed regions can for example represent adaptational benefits (Racimo et al. 2015), regions of high recombination rate (Nachman and Payseur 2012; Ravinet et al. 2018; Schumer et al. 2018), and regions under lower constraint or less purifying selection (Schumer et al. 2016). However, the patterns of introgression can also be difficult to distinguish from mechanisms such as incomplete lineage sorting (i.e., preservation of ancestral polymorphisms; Platt et al. 2019). Methods have been developed to circumvent confounding signals (e.g., ABBA-BABA statistics; Green et al. 2010; Durand et al. 2011) and in principle it is possible to separate introgression from other evolutionary processes (Martin et al. 2014). Research on introgression between divergent species can reveal important contributions to the evolutionary history of the taxa involved (e.g., ecological adaptations; Nelson et al. 2021) and increase our understanding of how such mechanisms can affect contemporary populations in the future and how robust reproductive barriers are against gene flow between divergent species.

There is currently a need for tractable experimental systems to study reproductive isolation in natural populations (White et al. 2019; Stankowski and Ravinet 2021). An interesting experimental system for investigating such processes appear in fungal species complexes in the Agaricomycotina. The subphylum Agaricomycotina is a diverse taxon, with about 20,000 species described worldwide and a crown age estimate of around 429 million years (Floudas et al. 2012). Research on hybridization and introgression among species of the Agaricomycotina (mushroom-forming fungi) is limited, but there are some examples from genera including *Pleurotus* (Bresinsky et al. 1987), *Heterobasidion* (Garbelotto et al. 1996; Stenlid and Karlsson 1991; Giordano et al. 2018), and *Armillaria* (Baumgartner et al. 2012), indicating that hybridization may be a common but understudied mechanism of speciation and gene exchange among taxa in this branch of the tree of life. Moreover, recent research shows that the reproductive barrier in fungi can be permeable despite high divergence between species (Maxwell et al. 2018), making the fungal reproductive system an interesting case study for expanding our knowledge on reproductive isolation and the speciation continuum (Maxwell et al. 2018).

In this study we use natural populations of the sister species *Trichaptum fuscoviolaceum* (Ehrenb.) Ryvarden and *Trichaptum abietinum* (Dicks.) Ryvarden (pictured in Figure 2) as our model organisms. These fungi are saprotrophic white rot fungi growing on conifers across the northern hemisphere. The two species are broadly sympatric and can grow on the same host, and sometimes they are found together on the same substrate. They are in general phylogenetically well separated species (Kauserud and Schumacher 2003; Seierstad et al. 2020; Peris et al. 2022), but some individuals clustered incongruently for different loci in a previous study including only a few molecular markers (Seierstad et al. 2020). Here, the authors suggested introgression or incomplete lineage sorting as possible explanations for the conflicting phylogenetic signals (Seierstad et al. 2020). Reproductive barriers between the two species have been documented by in vitro crossing experiments (Macrae 1967).

Population structure has been found within *T. abietinum* (Seierstad et al. 2020; Peris et al. 2022). In North America, two populations referred to as the North American A and the North American B population occur in sympatry and are reproductively isolated (i.e., form intersterile groups; Macrae 1967; Magasi 1976; Peris et al. 2022). Such intersterile groups have not been detected among populations of *T. fuscoviolaceum* (Macrae 1967; Peris et al. 2022), but some genetic structuring of populations has been observed (Seierstad et al. 2020; Peris et al. 2022).

Through this study, we aimed at exploring how signs of introgression can be discovered in genomes of extant species and how regions retained from past or current introgression might influence the evolution of contemporary populations. We used the experimental system to investigate potential introgression by conducting whole genome sequencing of individuals belonging to different populations of *T. abietinum* and *T. fuscoviolaceum*. Further, we assessed the possibility of current gene flow by testing compatibility across species through in vitro crossing experiments. We hypothesised that *T. fuscoviolaceum* and *T. abietinum* populations do not hybridize frequently due to a well resolved phylogeny and earlier crossing experiments revealing the species to be intersterile (Macrae 1967; Seierstad et al. 2020; Peris et al 2022). However, due to their overlapping habitat, ecology, and morphology, we hypothesised that the sister species might have a shared history that involves introgression, possibly having occurred among ancient populations.

Materials and Methods

Sampling

Individuals of *T. abietinum* and *T. fuscoviolaceum* were collected in New Brunswick, Canada and Pavia, Italy during the autumn of 2018. One individual of *T. biforme* was collected in New

Brunswick, Canada, and included as an outgroup. For all collection sites, ten individuals were sampled from separate logs, or two meters apart on the same log, within one square kilometre. For every individual, a cluster of sporocarps (covering no more than 3 x 3 cm) was collected and placed in separate paper bags. Notes on host substrate, GPS coordinates and locality were made for all individuals, and they were given a collection ID according to collection site and species identification based on morphology. The sporocarps were dried at room temperature for 2 – 3 days, or in a dehydrator at 30 °C, and later stored at room temperature in paper bags. Individuals included in this study are presented in Table S1.

Culturing

Since haploid sequences are bioinformatically convenient to work with, we isolated monokaryotic mycelia for sequencing by culturing collected *Trichaptum* individuals as follows (illustrated in Figure 1): (A) To revive dried individuals for spore shooting, sporocarps were placed in a moist paper towel and left in the fridge until soaked through (about 3 hours). (B) While working in a safety cabinet (Labculture® ESCO Class II Type A2 BSC, Esco Micro Pte. Ltd., Singapore), sporocarps were attached, hymenium facing media, with silicon grease from Merck Millipore (Darmstadt, Germany) to the lid of a petri dish containing 3% malt extract agar (MEA), with antibiotics and fungicides (10 mg/l Tetracycline, 100 mg/l Ampicillin, 25 mg/l Streptomycin and 1 mg/l Benomyl) to avoid contamination. The sporocarps were left for a minimum of one hour for spores to shoot onto the MEA plates. Subsequently, the sporocarps were removed to minimize spore shooting and the petri dish was sealed off with Parafilm M® (Neenah, WI, USA). The cultures were left for approximately one week, or until hyphal patches were visible, at 20 °C in a dark incubator (Termaks AS KB8400/KB8400L, Bergen, Norway). (C) Working in a safety cabinet, single, germinated spores were picked with a sterile scalpel and placed onto new MEA dishes with antibiotics and fungicides. The new cultures were left in a dark incubator at 20 °C for a few days until a mycelial patch could be observed. (D) The hyphae were checked for clamp connections in a Nikon Eclipse 50i light microscope (Tokyo, Japan) using 0.1% Cotton Blue to accentuate cells (examples in Figure S5). Clamps indicate a dikaryotic hyphae and we proceeded with the cultures lacking clamps (i.e., monokaryotic hyphae). (E) Monokaryotic cultures were replated onto new MEA dishes without antibiotics and fungicides (the mycelia grow better without these substances and the cultures were now free from contaminants) and placed in an incubator at 20 °C prior to sequencing and experiments.
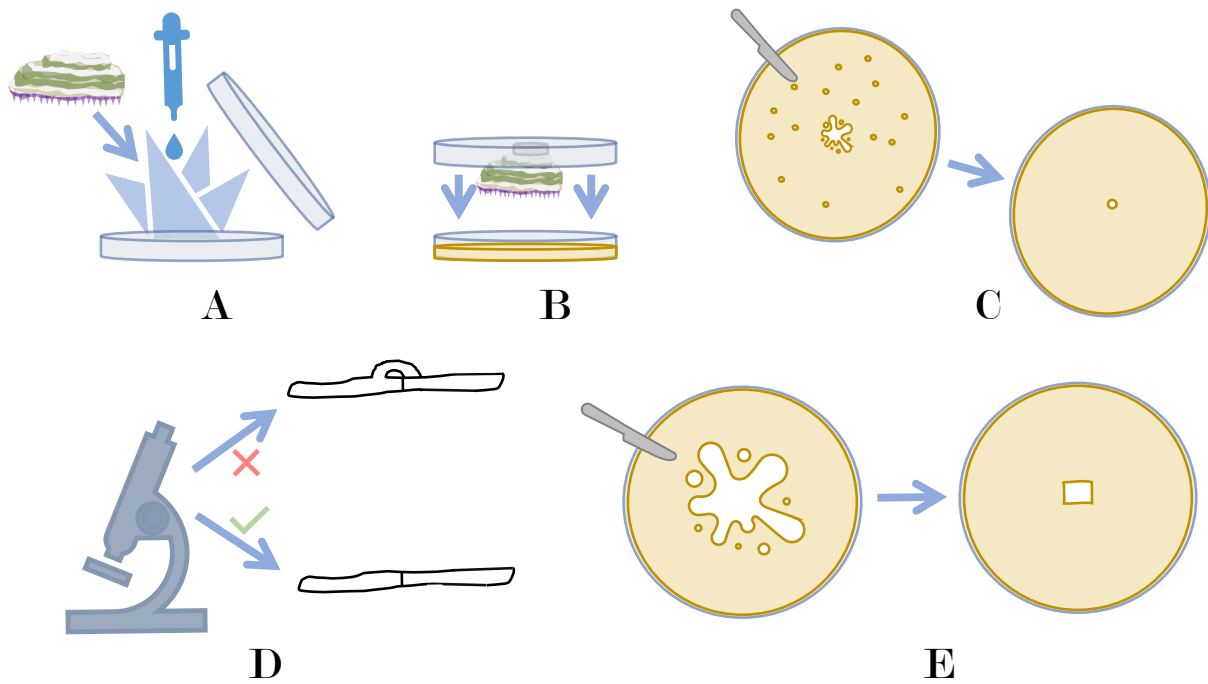
Figure 1. Procedure for culturing monokaryotic fungal individuals. (A) A sporocarp is placed in a wet paper towel onto a petri dish. (B) The sporocarp is glued to the lid of the petri dish to allow for spore shooting onto agar. (C) Hyphae from single spores are picked with a scalpel and placed onto new agar. (D) Microscopy of hyphae to confirm monokaryotic cultures. Hypha with clamp connection is indicated with a red cross and hypha without clamp connection is indicated with a green check symbol. (E) Mycelium from the new culture made in (C) is cut out with a scalpel and placed onto new agar.

PCR and Sanger sequencing

We Sanger sequenced the internal transcribed spacer (ITS; the fungal barcode), to confirm correct species designation of the cultures. The ITS region was amplified using the ITS1 (5' – TCCGTAGGTGAACCTGCGG – 3'; White et al. 1990) and ITS4 (5' – TCCTCCGCTTATTGATATGC – 3'; White et al. 1990) primers and the Thermo ScientificTM Phire Plant Direct PCR Kit (Waltham, USA) according to the manufacturer's protocol (using a small piece of mycelia instead of plant tissue). The following PCR program was used: 4 min at 95 °C, followed by 40 cycles of 25 sec at 95 °C, 30 sec at 53 °C and 60 sec at 72 °C, followed by a 10 min extension at 72 °C and an indefinite hold at 10 °C. PCR products were purified using 0.2 µl ExoProStar 1-Step (GE Healthcare, Chicago, USA), 1.8 µl H_2O and 8 µl PCR product. The samples were Sanger sequenced by Eurofins Scientific (Hamburg, Germany).

We assessed, trimmed, and aligned the resulting forward and reverse sequences to consensus sequences using Geneious Prime v2020.1.2 (https://www.geneious.com). To verify species designation of cultures, the consensus sequences were checked with the Basic local alignment search tool (BLAST; Altschul et al. 1990) against the National Centre of

Biotechnology Information (NCBI) database (U.S. National Library of Medicine, Bethesda, MD, USA). We kept cultures identified as *T. fuscoviolaceum* or *T. abietinum* and updated the *Trichaptum* cultures with incorrect initial species designation (based on sporocarp morphology).

DNA extraction and Illumina sequencing

Where possible, we chose approximately five individuals of *T. fuscoviolaceum* from each collection site and five individuals of *T. abietinum* corresponding to the same sites for DNA extraction and Illumina sequencing, including one individual of *T. biforme* as an outgroup.

Mycelia from fresh cultures, grown on 3% MEA with a nylon sheet between the MEA and the mycelia, were scraped off the plate and DNA was extracted using the E.Z.N.A.® Fungal HP DNA Kit (Omega Bio-Tek, Norcross, GA, USA) according to the DNA extraction procedure explained in Peris et al. (2022). Illumina libraries were prepared by the Norwegian Sequencing Center (NSC) as explained in Peris et al. (2022). Samples were sequenced at NSC on either the Illumina Hiseq 4000 or the Illumina Novaseq I. The samples where distributed across sequencing runs (3 runs).

Crossing experiments

To assess mating compatibility between species (i.e., between individuals of *T. abietinum* and *T. fuscoviolaceum*), we performed crossing experiments. As a positive control, we also crossed individuals of *T. fuscoviolaceum*. Some of the crosses include European *T. abietinum* that are not sequenced in the present study, but included in Peris et al. (2022). The crossing set-ups were planned according to a mating compatibility scheme based on mating loci (*MAT*) predicted in Peris et al. (2022). We crossed individuals that were both expected and not expected to mate based on their predicted mating type (i.e., dissimilar or similar allelic classes on both *MAT* loci). Individuals used for the experiment are presented in Table S2 (see Peris et al. (2022) for further details on the European *T. abietinum* individuals). Three replicates were made for all crosses to strengthen the confidence in the observations.

Pairs of monokaryotic individuals (circular 0.8 cm in diameter plugs) were plated 4 cm apart on petri dishes containing 3% MEA. The petri dishes were placed in a dark incubator at 19 °C until the two mycelia had grown together (about 2 weeks). The cultures were photographed using a Nikon D600 Digital Camera (Tokyo, Japan). To investigate if the crossing experiments were successful, we assessed the presence or absence of clamp connections using a Zeiss Axioplan 2 imaging light microscope (Güttingen, Germany) with

Zeiss AxioCam HRc (Güttingen, Germany). The process is similar to the description in Figure 1D. Microscopic photographs of hyphae were taken at 400 and 630 × magnification.

Reference genomes

We used the two genomes of *T. abietinum* (strain TA10106M1) and *T. fuscoviolaceum* (strain TF100210M3) as reference genomes (Bioproject PRJNA679164; https://doi.org/10.5061/dryad.fxpnvx0t4; Peris et al. 2022). In addition, we made a combined reference genome by merging the *T. abietinum* (acc. no. GCA 910574555) and *T. fuscoviolaceum* (acc. no. GCA 910574455) reference genomes with *sppIDer* (Langdon et al. 2018).

Preparation and initial mapping of whole genome data

Illumina raw sequences were quality filtered, removing sequences with a Phred quality score less than 30, using *Trim Galore! v0.6.2* (https://www.bioinformatics.babraham.ac.uk/projects/trim_galore/; Krueger 2015), and assessed using *FastQC* (Andrews 2010) and *MultiQC* (Ewels et al. 2016). After pre-processing, we used *BWA v0.7.17* (Li and Durbin 2009) to search for recent hybrids by mapping Illumina reads from *T. fuscoviolaceum* to a combined reference genome of *T. fuscoviolaceum* and *T. abietinum* using the wrapper *sppIDer* (Langdon et al. 2018). The wrapper generates a reference genome with chromosomes from both *T. fuscoviolaceum* and *T. abietinum*. The reads from a strain from one species can then be mapped to the combined genome. If the reads of an individual map equally well to chromosomes of both species, it indicates that the strain is a hybrid. No hybrids were revealed among the *T. fuscoviolaceum* individuals in the *sppIDer* analysis (Figure S1). All *T. fuscoviolaceum* individuals mapped with greater depth to the *T. fuscoviolaceum* part of the combined reference genome than the *T. abietinum* part. Since there were no recent hybrid individuals and we could only use one reference for further analyses, we chose to continue with the *T. fuscoviolaceum* reference genome.

Re-mapping with Stampy

To improve mapping of *T. abietinum*, *T. biforme* and the Italian *T. fuscoviolaceum* to the reference genome (based on a Canadian *T. fuscoviolaceum* individual), the raw sequences were mapped with *Stampy v1.0.32* (Lunter and Goodson 2011), which is designed to be more sensitive to divergent sequences (Lunter and Goodson 2011), before continuing with further analyses. Based on nucleotide divergence estimates found in Peris et al. (2022) by conversion

of average nucleotide identity using *FastANI* (Jain et al. 2018), the substitution rate flag was set to 0.23 for *T. biforme*, 0.067 for the Italian *T. fuscoviolaceum*, and 0.157 for *T. abietinum* when mapping each to the reference. The raw sequences were not trimmed before mapping due to limitations on hard clipping in *Stampy* (i.e., sequences are sometimes too short for *Stampy*), but poor sequences were filtered away at a later stage (see below).

SNP calling and filtering

To obtain a dataset with single nucleotide polymorphisms (SNPs), we first used *GATK HaplotypeCaller v4.1.4.* (McKenna et al. 2010). To create the dictionary files and regroup the mapped files before SNP calling, we used *Picard v2.21.1* (https://broadinstitute.github.io/picard/) and reference index files were made using *SAMtools* faidx (Li et al. 2009). We ran *HaplotypeCaller* in haploid mode with otherwise default settings. Subsequently, we used the resulting Variant Call Format (VCF) files in *GATK GenomicsDBImport* (McKenna et al. 2010) to create a database used as input for *GATK GenotypeGVCF* (McKenna et al. 2010), which creates a VCF file containing SNPs for all individuals. *GenomicsDBImport* was used with default settings together with the java options ('--java-options') '-Xmx4g' and '-Xms4g' and an interval text file ('--intervals') containing names of the different scaffolds. *GenotypeGVCF* was used with default settings. To remove indels, bad SNPs, and individuals with high missingness, we filtered the resulting VCF file with *GATK VariantFiltration* (McKenna et al. 2010) and *BCFtools v1.9* filter (Danecek et al. 2021). We used GATK's hard filtering recommendations (https://gatk.broadinstitute.org/hc/en-us/articles/360035890471-Hard-filtering-germline-short-variants) together with the Phred quality score option of removing SNPs with a score less than 30.0 ('QUAL < 30.0'). With *BCFtools* filter, we removed indels and poor SNPs using these options: minimum read depth (DP) < 3, genotype quality (GP) < 3 and '-v snps'. We also used *BCFtools* filter to remove multiallelic SNPs ('view -M2'), SNPs close to indels ('--SnpGap 10'), variants with a high number of missing genotypes ('-e 'F_MISSING > 0.2''), minimum allele frequency ('MAF <= 0.05'), and invariant sites and monomorphic SNPs ('-e 'AC==0 || AC==AN''). We made one dataset where monomorphic SNPs were removed and the MAF filter was applied (Dataset 1 with 2 040 885 SNPs) and two datasets, one with the outgroup and one without, not applying these filters (Dataset 2 with 3 065 109 SNPs; Dataset-O 2 with 3 118 957 SNPs, where O = outgroup), because monomorphic sites were required to calculate some divergence statistics. After filtering, individuals with high missingness or high heterozygosity (i.e., dikaryons) were removed. The final datasets consisted of 32 individuals from the Canadian *T. fuscoviolaceum*

population, 9 individuals from the Italian *T. fuscoviolaceum* population, 30 individuals from the North American B *T. abietinum* population, and 6 individuals from the North American A *T. abietinum* population.

Phylogenetic tree analysis

To confirm the phylogenetic relationship between the different populations of *T. abietinum* and *T. fuscoviolaceum*, we performed a maximum likelihood phylogenetic tree analysis using *IQ-TREE 2* (Minh et al. 2020). The VCF-file from Dataset-O 2 was converted into a PHYLIP file by using Edgardo M. Ortiz's script *vcf2phylip.py* (https://raw.githubusercontent.com/edgardomortiz/vcf2phylip/master/vcf2phylip.py). The IQ-TREE analysis was run on the PHYLIP file using the flags '-T auto', '-m GTR+ASC', '-alrt 1000' and '-B 1000'. GTR+ASC is a standard model.

Principal component and divergence analyses

To explore the data and investigate population groupings, we performed a principal component analysis (PCA) with *PLINK v2.00-alpha* (www.cog-genomics.org/plink/2.0/; Chang et al. 2015). To prepare the input file, we linkage pruned Dataset 1 in *PLINK*, using the flags '--vcf $vcf_file', '--double-id', '--allow-extra-chr', '--set-missing-var-ids @:#', --out $out_file' and '--indep-pairwise 50 10 0.1', retaining 56 046 SNPs. The '--indep-pairwise' flag performs the linkage pruning, where '50' denotes a 50 Kb window, '10' sets the window step size to 10 bp, and '0.1' denotes the r^2 (or linkage) threshold. A PCA was subsequently performed on the pruned VCF file, using the flags, '--vcf $vcf_file', '--double-id', '--allow-extra-chr', '--set-missing-var-ids @:#', '--extract $prune.in_file', '--make-bed', '--pca', and '--out $out_file' (both linkage pruning and PCA flags were based on the Physalia tutorial https://speciationgenomics.github.io/pca/).

To investigate the divergence between populations, we applied a sliding window approach on Dataset 2 to calculate the fixation index (F_{ST}) and absolute divergence (d_{XY}) along the genome. We also performed a sliding window analysis to calculate within population divergence (π). The analyses were performed using Simon Martin's script *popgenWindows.py* (https://github.com/simonhmartin/genomics_general/blob/master/popgenWindows.py) with *Python v3.8* (Van Rossum and Drake 2009). We set the window size to 20 000 bp ('-w 20000'), step to 10 000 bp ('-s 10000') and the minimum number of SNPs in each window to 10 ('-m 10').

Introgression analyses with D-statistics

To investigate introgression between populations, we used the *R* (R Core Team 2020) package *admixr* (Petr et al. 2019) and Dataset-O 2 to calculate the *D* (Green et al. 2010; Durand et al. 2011), outgroup f_3 (Raghavan et al. 2014) and f_4-ratio (Reich et al. 2009; 2011; Patterson et al. 2012) statistics between different populations (based on recommendations from the Physalia tutorial https://speciationgenomics.github.io/ADMIXTOOLS_admixr/, and the *admixr* tutorial https://bodkan.net/admixr/articles/tutorial.html#f4-ratio-statistic-1). To prepare the input file from VCF to Eigenstrat format, we used the conversion script *convertVCFtoEigenstrat.sh* by Joana Meier (https://github.com/speciationgenomics/scripts), which utilizes *VCFtools v0.1.16* (Danecek et al. 2011) and *EIGENSOFT v7.2.1* (Patterson 2006; Price 2006). The script has a default recombination rate of 2.0 cM/Mb, which we changed to 2.5 cM/Mb based on earlier findings in the class Agaricomycetes, where *Trichaptum* belongs (Heinzelmann et al. 2020). We further used another *Python* script developed by Simon Martin, *ABBABABAwindows.py* (https://github.com/simonhmartin/genomics_general/blob/master/ABBABABAwindows.py), for a sliding window ABBA-BABA analysis on Dataset-O 2 to calculate the proportion of introgression (f_{dM}; Malinsky et al. 2015). The window size was set to 20 000 ('-w 20000'), step size to zero, and minimum number of SNPs per window to 100 ('-m 100'), together with '--minData 0.5' to specify that at least 50% of the individuals in each population must have data for a site to be included (based on recommendations from the Physalia tutorial https://speciationgenomics.github.io/sliding_windows/). We used *T. biforme* as outgroup and tested introgression between the Canadian *T. fuscoviolaceum* and the *T. abietinum* populations in addition to the Italian *T. fuscoviolaceum* and the *T. abietinum* populations (the phylogenetic topology was based on results from the f_3 analysis). Outlier windows were extracted from the results with a Hidden Markov-model approach using the *R* package *HiddenMarkov* (Harte 2021) following Ravinet et al. (2018). Since the HMM approach cannot analyse negative values, the f_{dM} distribution was rescaled by adding 2 to all values. Annotated genes in these outlier windows were retrieved from the annotated *T. fuscoviolaceum* reference genome. The reference genome was annotated using *RepeatModeler* (Flynn et al. 2020), *RepeatMasker* (Smit et al. 2013-2015) and *MAKER2* (Holt and Yandell 2011). Functional annotation and protein domain annotations of detected coding sequences and the encoded proteins were performed using *blastp* (Altschul et al. 1990) against a local UniProt database and InterProScan (Jones et al. 2014), respectively. All annotations were encoded in a General Feature Format (GFF) file, which was used to match the significant windows and extract the genes. Gene ontology terms annotated in the GFF file were extracted using the package *rtracklayer v1.48* (Lawrence et al.

2009) in *R*. Gene ontology (GO) enrichment analysis was performed in *R* using the *TopGO v2.40* package (Alexa and Rahnenfuhrer 2021). Lastly, a false discovery rate (FDR) analysis was performed on the resulting raw p-values.

Network analysis with TreeMix

To further explore possible introgression events and direction of introgression, we applied a network analysis with *TreeMix* (Pickrell and Pritchard 2012). To prepare the VCF-file (Dataset-O 2) for analysis, we removed sites with missing data using *VCFtools* (with '--max-missing 1') and linkage pruned the data using *PLINK*. Linkage pruning was performed in the same way as with the PCA, except the file was recoded into a new VCF file using the flags '--bfile' and '--recode vcf' after pruning. To convert the data to *TreeMix* format, we ran Joana Meier's script *vcf2treemix.sh* (https://github.com/speciationgenomics/scripts/blob/master/vcf2treemix.sh). The script *vcf2treemix.sh* also requires the script *plink2treemix.py* (https://bitbucket.org/nygcresearch/treemix/downloads/plink2treemix.py).

TreeMix was run using the options '-global' and '-root *T_biforme*'. The option for number of edges ('-m') was analysed from 1-10 and the option for block size ('-k') was varied between 300 and 800 to avoid identical likelihoods. For each block size, the analysis was repeated three times for 'm' 1-10 edges. To find the optimal number of edges, we ran *OptM* (Fitak 2021) in *R*. The residuals and network with different edges were plotted in *R* using the functions provided by *TreeMix* (*plotting_funcs.R*) together with the packages *RColorBrewer v1.1-2* (Neuwirth 2014) and *R.utils* (Bengtsson 2021).

Demographic modelling

To explore divergence and introgression, we applied demographic modelling with *fastsimcoal2* (Excoffier et al. 2021). To prepare the data (Dataset 2) for analysis, frequency spectrum files were created using *easySFS* (https://github.com/isaacovercast/easySFS). Subsequently, *fastsimcoal2* was run using the output files from *easySFS* together with a template file defining the demographic model and a parameter estimation file. The analyses were run with the flags '-m', '-0', '-n 200000', '-L 50', '-s 0' and '-M'. Each model was run 100 times and the run with the best likelihood was extracted using Joana Meier's script *fsc-selectbestrun.sh* (https://raw.githubusercontent.com/speciationgenomics/scripts/master/fsc-selectbestrun.sh). We selected several different models based on likely events inferred from the results of the ML tree, *D* and *f* statistics, and *TreeMix* analyses to test for different scenarios of introgression, both with and without one or two ghost populations (Figure S7). To compare different models, an

AIC value was calculated from the run with the best likelihood for each model using the *R* script based on code by Vitor Sousa, *calculateAIC.sh* (https://github.com/speciationgenomics/scripts/blob/master/calculateAIC.sh). Since AIC can overestimate support for the best model when SNPs are in linkage, we also calculated the likelihood distributions using the best run from all the models. The models were run with the best parameter values ({PREFIX}_maxL.par output from the first run) 100 times in *fastimcoal2* with the options '-n 1000000', '-m', '-q', and '-0'. The likelihood values were collected and plotted in *R* for comparison between models (i.e., look for overlapping distributions).

Results

Crossing experiments confirm incompatibility between species

To ensure that the previous results of intersterility between *T. abietinum* and *T. fuscoviolaceum* were also the case for our collections, we performed new crossing experiments with our individuals. We did not observe clamp connections between crosses of *T. fuscoviolaceum* and *T. abietinum* individuals. This was the case for mate pairs that were predicted to mate based on mating type alleles and for those predicted not to (Table S2; Figure S2 and S4; mating type alleles were annotated in Peris et al. 2022). It was difficult to observe compatible crosses by investigating the cultures macroscopically, but there was often a sharper line between individuals on the petri dish when the crosses were incompatible (Figure S2). The *T. fuscoviolaceum* individuals mated as expected (i.e., those that were predicted to be incompatible due to identical mating types showed no clamp connections and those that were predicted to be compatible had clamp connections; Table S2; Figure S3 and S5).

Phylogeny, principal component and divergence analyses reveal high divergence between species

After confirming mating incompatibility, we continued with assessing the nucleotide divergence between the species. The maximum likelihood (ML) phylogeny clustered the species and populations into well-defined clades with high support (Figure 2).

The PCA also indicated clear groupings of species and populations of *T. fuscoviolaceum* and *T. abietinum*, with PC1 and PC2 explaining 57.5% and 12.2% of the observed variation, respectively (Figure 3). The Italian and Canadian *T. fuscoviolaceum* populations were closer to each other than either was to the North American A and the North American B *T. abietinum* population along PC1. PC2 positioned the Italian *T. fuscoviolaceum* population and the North American A *T. abietinum* population at opposite ends of the axis, while the Canadian *T.*

fuscoviolaceum and the North American B *T. abietinum* population were placed closer in the middle of the axis.
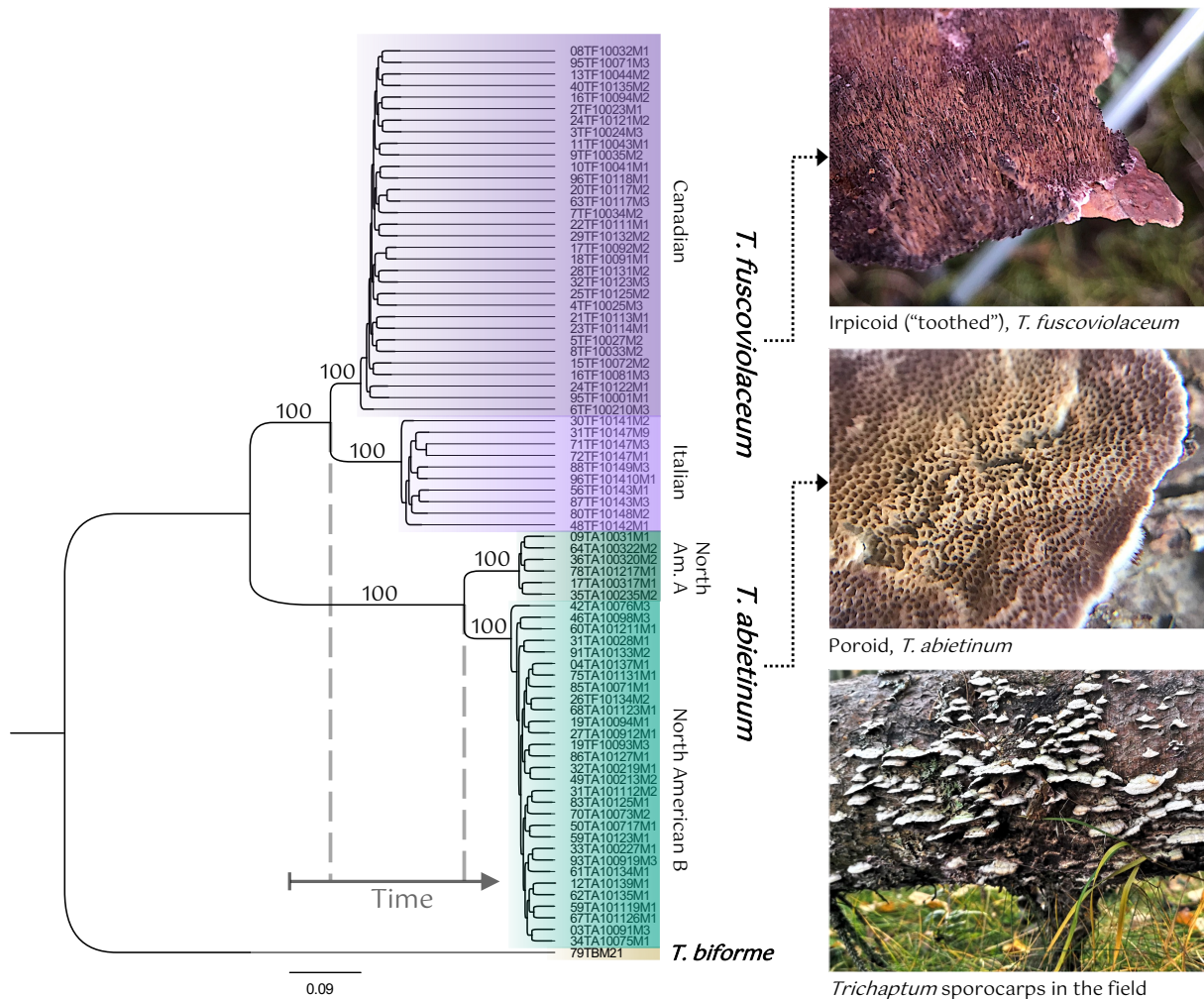
Figure 2. Clear population structure in the phylogenetic tree analysis. The analysis is based on a single nucleotide polymorphism (SNP) dataset of 3 118 957 SNPs. The tree is constructed using *IQ-TREE 2* (Minh et al. 2020) with the model GTR+ASC. The numbers on the branches represent bootstrap branch support. Populations of *Trichaptum fuscoviolaceum* (TF) are coloured in shades of purple and populations of *T. abietinum* (TA) are coloured in shades of green. The outgroup, *T. biforme*, is coloured in brown. The scale bar on the bottom is the number of substitutions per site. The time axis illustrates relative split of the TF and TA populations (see Figure S6). The shade of purple of the hymenium (spore producing layer) can vary. The two TF individuals in the North American B population are confirmed as North American B TA individuals after genomic analyses (wrongly assigned in the field). Photographs of hymenia by Inger Skrede and photograph of sporocarps by Malin Stapnes Dahl.
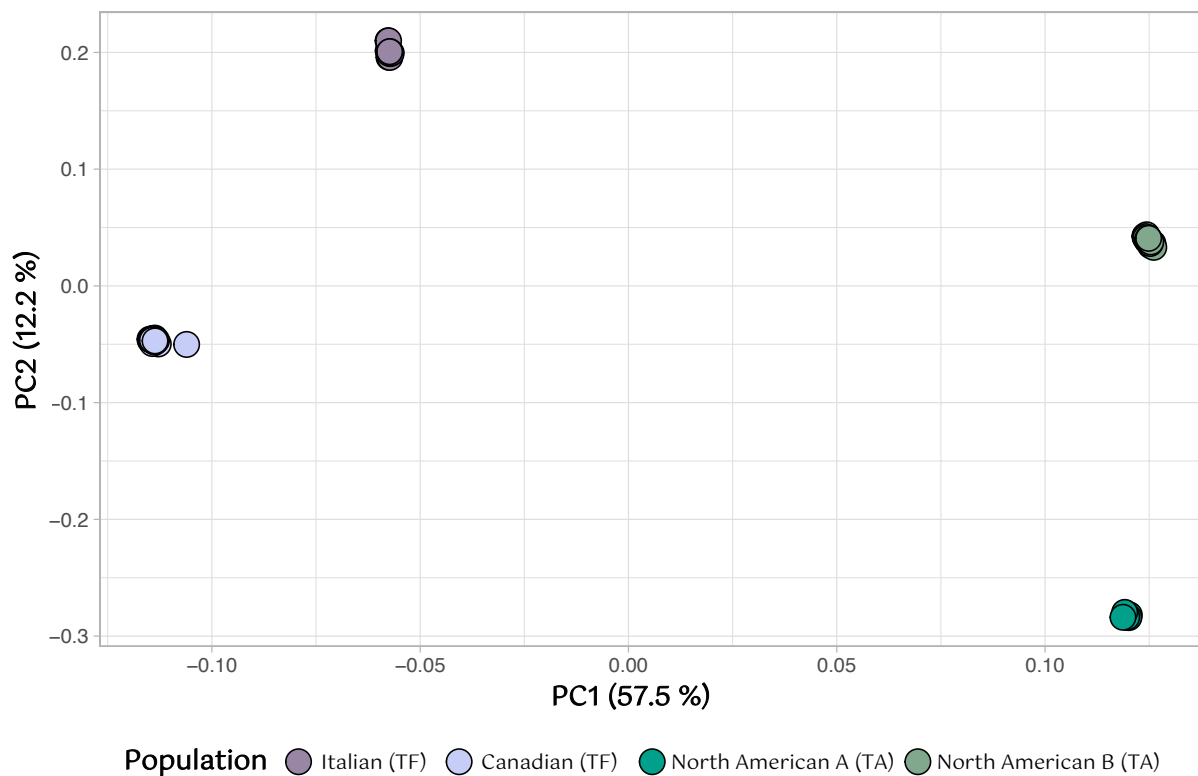
Figure 3. Clear groupings according to species and populations in the principal component analysis (PCA). The PCA is based on a single nucleotide polymorphism (SNP) dataset of 2 040 885 SNPs linkage pruned to 56 046 SNPs. The x and y-axes represent PC1 and PC2, respectively, with percentage of variance explained in parentheses. Points are individuals colored by population as indicated in the legend. TF = *Trichaptum fuscoviolaceum* and TA = *T. abietinum*. The figure is made in *R v4.0.2* using the packages *ggplot2* (Wickham 2016) and *wesanderson* (Ram and Wickham 2018).

The clear distinction spotted in the PCA was corroborated by the fixation index (F_{ST}), which showed a high degree of divergence both between populations of different species and between populations of same species. The F_{ST} means across the genome for between species comparisons were ranging from 0.6 – 0.8. The within-species comparisons showed higher differentiation between the two *T. fuscoviolaceum* populations than between the *T. abietinum* populations (mean F_{ST} between *T. fuscoviolaceum* populations was 0.46, while mean F_{ST} between *T. abietinum* populations was 0.33).

The absolute between populations divergence (d_{XY}) echoed the patterns of the PCA and F_{ST} scan, with generally high divergence both between populations of different species and between populations within species. The mean d_{XY} values between populations of different species were about 0.4, while the mean values between populations of same species were slightly less than 0.2 for both comparisons.

The within population variation calculated by the nucleotide diversity, π, had a mean value of about 0.05 for all populations.

Introgression analyses indicate a complex evolutionary history

The divergence analyses suggested that the species had diverged for a long time. Thus, we wanted to explore signs of ancestral introgression not revealed by assessing current interbreeding with crossing experiments. The *D* statistic, used to detect signs of introgression across the genome, gave significant *D* values (|z-score| > 3) between the Italian *T. fuscoviolaceum* and both the *T. abietinum* populations (Table 1), indicating introgression between the Italian *T. fuscoviolaceum* and the *T. abietinum* populations. The test of introgression between the *T. abietinum* populations and either of the *T. fuscoviolaceum* populations did not reveal significant positive or negative *D*-values. There was also a larger discrepancy between ABBA and BABA sites in the significant topologies (Table 1, row three and four), than in the nonsignificant topologies (Table 1, row one and two).

Table 1. The *D* statistic indicates introgression. The analysis is performed on a single nucleotide polymorphism (SNP) dataset of 3 118 957 SNPs. The *D* statistic is based on a phylogenetic tree hypothesis of (((W, X), Y), Z) and tests introgression between Y and X (negative D) and Y and W (positive D). Z is the outgroup. The table includes the *D* value (*D*), standard error (std error), significance of the *D* values (z-score; an absolute z-score larger than 3 is considered significant), the number of SNPs shared between Y and W (BABA), the number of SNPs shared between Y and X (ABBA), and the number of SNPs used for the comparison (n SNPs). Significant introgression between populations is highlighted in bold.

D statistics									
W	**X**	**Y**	**Z**	**D**	**std error**	**z-score**	**BABA**	**ABBA**	**n SNPs**
NAmB TA	NamA TA	Can TF	TB	-0.0041	0.004744	-0.868	9142	9217	662894
NamA TA	NamB TA	It TF	TB	0.0060	0.005929	1.018	9706	9590	662712
Can TF	**It TF**	**NamA TA**	TB	-0.2257	0.008558	-26.376	8072	12776	662712
It TF	Can TF	**NamB TA**	TB	0.2287	0.009069	25.221	12810	8041	664963

Can TF = Canadian *Trichaptum fuscoviolaceum*, It TF = Italian *T. fuscoviolaceum*, NamB TA = North American B *T. abietinum*, NamA TA = North American A *T. abietinum*, TB = *T. biforme*

The four-population *f* statistic (*f₄* ratio), used to test proportion of introgression, resulted in a violation of the statistical model (i.e., negative alpha values; valid values are proportions between 0 and 1) when placing *T. abietinum* and *T. fuscoviolaceum* as sister groups with the Canadian *T. fuscoviolaceum* or the North American B *T. abietinum* at the X position (Table 2). Reversing the positions of the Canadian and Italian *T. fuscoviolaceum* or the two *T. abietinum* populations at X and C resulted in a positive alpha value, which did not violate the model (Table

2). The alpha value indicated about 5.7% shared ancestry between the *T. abietinum* populations and the Italian *T. fuscoviolaceum* population (Table 2, row three and four). The small amount of shared ancestry (0.1 – 0.2%) between the North American A *T. abietinum* and the two *T. fuscoviolaceum* populations did not show a significant z-score (< 3; Table 2, row seven and eight).

Table 2. The four-population *f* statistic (*f₄*) show further signs of introgression. The analysis is based on a single nucleotide polymorphism (SNP) dataset of 3 118 957 SNPs. The table shows the different configurations tested from a hypothesis of the phylogenetic relationship presented as (((A, B), (X, C)), O), where X is the introgressed population and C its sister population, with the B population as the source of introgression and A as its sister population. O is the outgroup (*Trichaptum biforme*). The alpha value indicates proportion of gene flow with standard error (std error) and significance (z-score; considered significant when larger than 3). Negative alpha values are due to violation of the statistical model. Significant introgression between populations is highlighted in bold.

F_4 ratio							
A	**B**	**X**	**C**	**O**	**alpha**	**std error**	**z-score**
NamA TA	NamB TA	Can TF	It TF	*T. biforme*	-0.061072	0.003619	-16.878
NamB TA	NamA TA	Can TF	It TF	*T. biforme*	-0.060432	0.003553	-17.010
NamB TA	**NamA TA**	**It TF**	Can TF	*T. biforme*	0.057001	0.003160	18.038
NamA TA	**NamB TA**	**It TF**	Can TF	*T. biforme*	0.057569	0.003215	17.906
It TF	Can TF	NamB TA	NamA TA	*T. biforme*	-0.002591	0.002536	-1.022
Can TF	It TF	NamB TA	NamA TA	*T. biforme*	-0.001535	0.001703	-0.902
Can TF	It TF	NamA TA	NamB TA	*T. biforme*	0.001536	0.001697	0.905
It TF	Can TF	NamA TA	NamB TA	*T. biforme*	0.002592	0.002522	1.028

NamA = North American A, NamB = North American B, Can = Canadian, It = Italian, TF = *T. fuscoviolaceum*, TA = *T. abietinum*

Further investigation of introgression with the three-population outgroup *f* statistic (*f₃*), which estimates shared genetic drift (or branch length), revealed that the *T. abietinum* populations split later (share more genetic drift) than the *T. fuscoviolaceum* populations (Figure S6; Figure 2). As with the *f₄* ratio analysis, the Italian *T. fuscoviolaceum* population exhibited slightly more shared genetic drift with the *T. abietinum* populations than the Canadian *T. fuscoviolaceum* population. Nevertheless, the difference between the two *T. fuscoviolaceum* populations was miniscule. The *f₃* analysis indicated a phylogenetic topology where the *T. fuscoviolaceum* populations diverged earlier than the *T. abietinum* populations. A reasonable next step was therefore to test introgression between the *T. abietinum* populations and each of the *T. fuscoviolaceum* populations in subsequent sliding window introgression analyses (i.e., a

(((North American B *T. abietinum*, North American A *T. abietinum*), *T. fuscoviolaceum* population), *T. biforme*) phylogenetic topology).

The network analysis conducted in *TreeMix,* based on the model with the most optimal number of edges, supported introgression from the Italian *T. fuscoviolaceum* into the North American B *T. abietinum* population (Figure 4). From the residual plot (Figure S8, 1 edge), it was clear that a large proportion of the residuals were not accounted for between the Italian *T. fuscoviolaceum* and the North American B *T. abietinum*. When testing which number of edges was the most optimal, the model with 1 migration edge got the best support (Figure S9).

The sliding window proportion of introgression (f_{dM}) calculated across the genome, which was set up based on the results from the f_3 analysis, revealed small regions of possible introgression (Figure 5; Figure S10). There were several windows of significant positive f_{dM} values (e.g., more shared derived polymorphisms than expected between the *T. fuscoviolaceum* populations and the North American A *T. abietinum* population), which suggests regions of introgressed genes (HMM outliers are marked in Figure 5 and S10, and presented in Table S3).
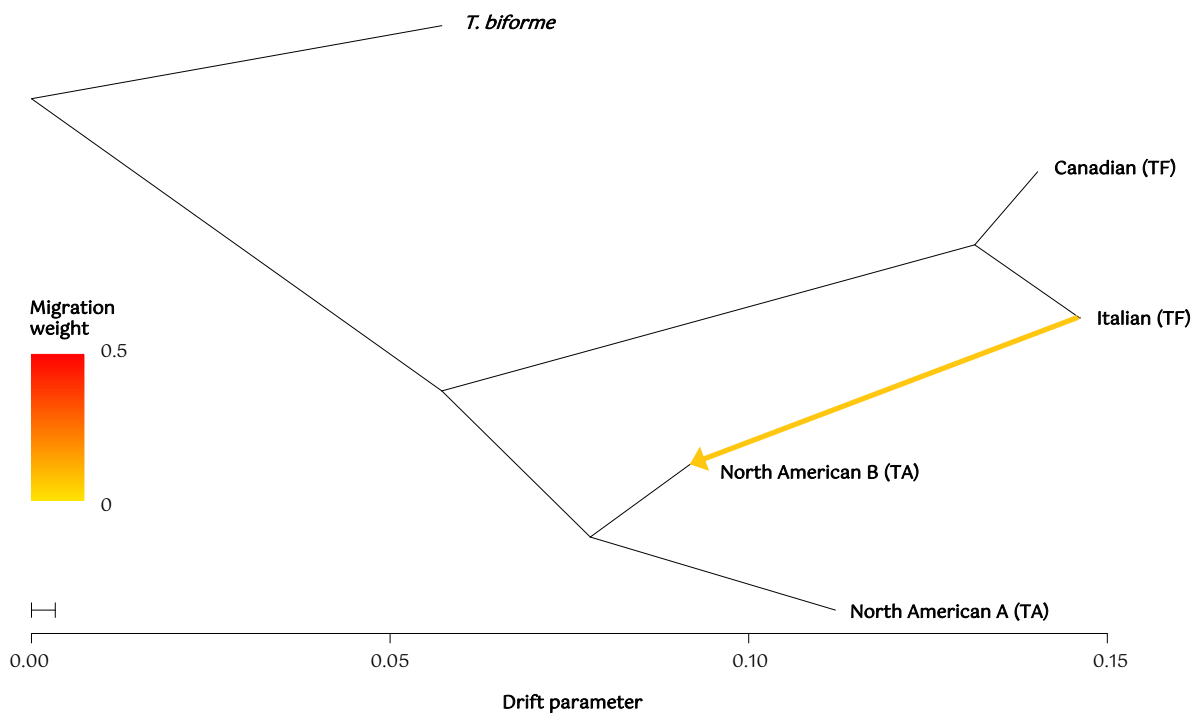
Figure 4. Introgression from the Italian *Trichaptum fuscoviolaceum* into the North American B *T. abietinum*. The plot is based on results from the *TreeMix* (Fitak 2021) analysis with a block size (-k) of 700 and 1 migration edge (-m 1). The yellow arrow shows the direction of migration (introgression). The bar on the left depicts the migration weight (proportion of admixture). The bottom scale bar shows the drift parameter (amount of genetic drift along each population; Wang et al. 2016). TF = *T. fuscoviolaceum* and TA = *T. abietinum*. The outgroup is *T. biforme.*

18

The genes in outlier windows coded for many unknown proteins, but also proteins similar to those found in common model organisms such as *Saccharomyces* spp. and *Arabidopsis thaliana*. The genes with similarity to other organisms are annotated to many different functions (i.e., there are genes involved in oxidoreductases, hydrolases, and transport, among others; The UniProt Consortium 2021). The GO enrichment analysis found some of the HMM outlier genes to be [...] abo[...] and c[...] ort, [...] (Ta Howe[...] g Fl[...] he r[...] he o[...] ter signif[...]

((((North American B TA, North American A TA), Canadian TF), *T. biforme*)

phylogenetic hypothesis, ((((P1, P2), P3), O), where P1, P2 and P3 are populations investigated for introgression and O is the outgroup. A positive value indicates more shared derived polymorphisms than

Trichaptum abietinum and TF = *T. fuscoviolaceum*. The figure is made in *R v4.0.2* using the packages *tidyverse* (Wickham et al. 2019) and *wesanderson* (Ram and Wickham, 2018).

Demographic modelling indicates involvement of a ghost population.

Lastly, we conducted demographic modelling to gain insight into divergence times and introgression events. We were also able to include a ghost population (Beerli et al. 2004) to test for introgression from unsampled or extinct populations. The best model, supported by both AIC and likelihood distribution comparison, showed introgression occurring twice; first between an ancestral *T. fuscoviolaceum* population and an ancestral *T. abietinum* population and later between a more recent ghost population related to the Italian *T. fuscoviolaceum* and the North American B *T. abietinum* (Figure 6; Figure S7, Ghost migration 9; Figure S11). These results were partly congruent with the network analysis where introgression was inferred between the Italian *T. fuscoviolaceum* and the North American B *T. abietinum* (Figure 4), but in the network analysis a ghost population could not be included. The best supported model further indicated that the sister species split 524 473 generations ago, while the *T. fuscoviolaceum* populations split 99 101 generations ago and the *T. abietinum* populations 92 260 generations ago. The estimated split of the ghost population from the Italian *T. fuscoviolaceum* populations was 94 536 generations ago (Figure 6; Table S5). The analysis further showed that there were few migrants between the ancestral populations and a little more between the ghost population and the North American B *T. abietinum* population, but mostly from the ghost population into the North American B (migration values; Figure 6).

Discussion

Divergent sister species show signs of introgression

Population genomic and introgression analyses present a window into exploring the dynamics of populations, their genomes and how they evolve. By searching beyond current phylogenies and population structures, intricate evolutionary histories can be revealed. Our study presents several results indicating that high divergence between *T. abietinum* and *T. fuscoviolaceum* does not exclude the possibility of admixture.

Firstly, high divergence values are prevalent throughout the genomes. The large genetic differences between the species can be a result of mechanisms such as reproductive isolation (Nei et al. 1983) and random events over time (i.e., genetic drift; Watterson 1985). The fungi make up an ancient and diverse kingdom that originated over a billion years ago (Berbee et al. 2020), with the oldest fungal-like fossil dating back to 2.4 billion years ago (Bengtson et al.
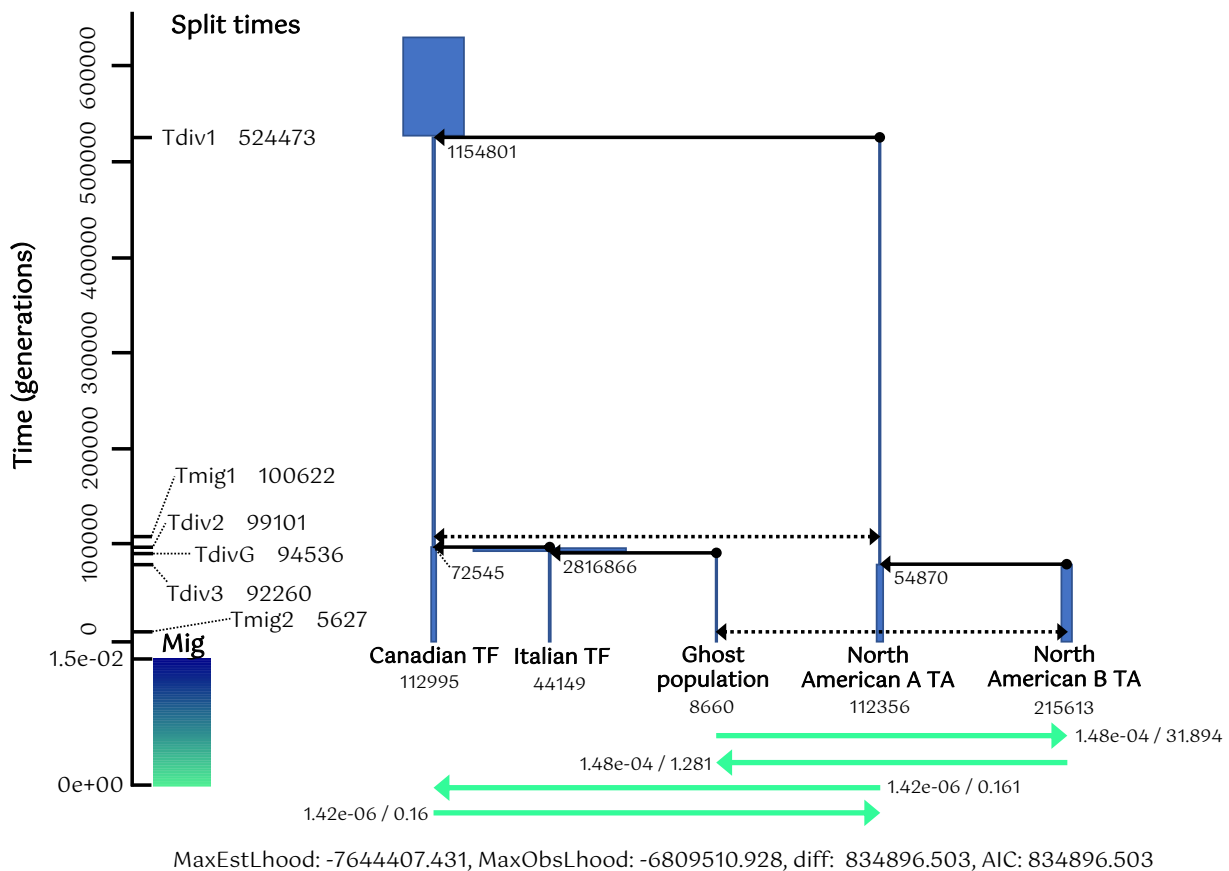
Figure 6. Demographic modelling indicates both ancient introgression and introgression from a recent ghost population. The analysis is performed in *fastsimcoal2* (Excoffier et al. 2021) using a dataset of 3 065 109 SNPs. The figure shows the best supported model. The right side of the axis indicate time in generations, with the lower part (below 0; colour bar) showing proportion of migration. The split times (divergence times) and migration times are presented on the right side of the axis. The blue bars and fully drawn black arrows show the estimated effective population sizes and splits, respectively. Effective population sizes are also written next to the splits and below the current populations. The dotted black arrows are the introgression events. The green arrows display the amount of migration (related to the colour bar) with the proportion of introgression estimate / the calculation to the number from the effective population size (current populations) that migrated. Tdiv = time of divergence, Tmig = time of migration, mig = migration, MaxEstLhood = maximum estimated likelihood, MaxObsLhood = maximum observed likelihood, diff. = difference between MaxEstLhood and MaxObsLhood, TF = *Trichaptum fuscoviolaceum* and TA = *T. abietinum*. Note: All parameters (including migration) are plotted backward in time and in haploid numbers.

2017) and an estimate of 2 – 5 million extant species (Li et al. 2021). The divergence of the order Hymenochaetales, which *Trichaptum* belongs to, dates to the Jurassic, about 167 million years ago (Varga et al. 2019). There are 37 accepted species in the genus (Index Fungorum 2021), but no estimates of the age of *Trichaptum*. Seeing the old age of Hymenochaetales and assuming the genus *Trichaptum* is old, time is a likely explanation for the genome wide high

divergence between *T. fuscoviolaceum* and *T. abietinum*. The demographic modelling also suggests that the two species split quite some time ago (524 473 generations).

Today, the sister species occur in the same habitat, with similar morphology and ecology, acting as early saprotrophs on newly deceased conifers in the northern hemisphere (Kauserud and Schumacher 2003). As mentioned, the two species can grow on the same host and are sometimes found on exactly the same substrate. However, *T. fuscoviolaceum* is usually found on pine (*Pinus*) and fir (*Abies*; most individuals in this study were collected on balsam fir; *A. balsamea*), while *T. abietinum* is more common on spruce (*Picea*) and larch (*Larix*; Macrae 1967; Peris et al. 2022). Even though habitats overlap, the crossing experiments corroborate previous results (Macrae 1967) in that the sister species do not hybridize in vitro, and the genomic analyses suggest that this does not happen between contemporary populations in the wild either. However, the detection of gene flow between more recent populations in the demographic modelling does imply that mating between species can occur occasionally in nature.

The crossing experiments between *T. fuscoviolaceum* individuals of different populations demonstrate that individuals can still mate successfully even though the divergence analyses exhibit high F_{ST} and d_{XY} values. The high divergence could be due to geographic separation of the Italian and Canadian population, reducing gene flow between these populations. Compatibility is not observed among all populations of *T. abietinum*, where intersterility is detected between some populations that occur in sympatry (Macrae 1967; Magasi 1976). The genus *Trichaptum* consists of tetrapolar fungi, which means individuals are compatible only when they have different alleles on both of the two mating loci (*MATA* and *MATB*; Fraser et al. 2007; Peris et al. 2022). Previous studies have shown that fungal mating loci are diverse and maintained by balancing selection (May et al. 1999; James et al. 2004), which was recently demonstrated in *Trichaptum* as well (Peris et al. 2022). In *T. abietinum*, additional reproductive barriers other than incompatible mating loci are at play, causing the formation of intersterility groups. However, such barriers can remain incomplete. If reproductive barriers were incomplete during the divergence of *T. abietinum* and *T. fuscoviolaceum*, and they diverged mostly due to genetic drift in geographic isolation, conserved diversity on the mating loci (as observed in Peris et al. 2022) over time can have allowed for introgression by maintaining reproductive compatibility across species. The demographic modelling does suggest that introgression happened quite some time after the split between the species. This again supports allopatric divergence, making it possible for the species to reproduce at a later stage due to the possible lack of reproductive barriers (no

reinforcement). Gene flow between divergent species of fungi has been detected before (e.g., Maxwell et al. 2018). This could be facilitated by the flexible developmental biology of some fungi, with the capability of tolerating developmental imprecision and distortion of their genetic makeup and still be able to grow and reproduce (Moore et al. 2011; Stukenbrock 2016).

The specific mechanisms behind how the two species diverged are difficult to untangle based on our results. However, the *D* and *f* statistics, together with the network analysis and demographic modelling, show signs of introgression between *T. abietinum* and *T. fuscoviolaceum*. Based on the *D* statistic, the ancestor of the Italian *T. fuscoviolaceum* population appears to have admixed with the *T. abietinum* populations. This might be somewhat counterintuitive, as it is the Canadian *T. fuscoviolaceum* population that currently occurs in sympatry with the collected *T. abietinum* populations. However, reproductive barriers can be produced between species in sympatry due to reinforcement (Abbott et al. 2013). When an allopatric lineage, such as the Italian *T. fuscoviolaceum* or a closely related ghost population, is encountered, reproductive barriers may not be in place and gene flow can occur. The f_4 ratio test further corroborates these results, indicating that the Italian *T. fuscoviolaceum* shares a larger proportion of the genome with the *T. abietinum* populations than the Canadian *T. fuscoviolaceum*. The violation of the statistical model for some topologies with *T. abietinum* and *T. fuscoviolaceum* populations as sister species in the f_4 ratio test can be due to lack of data from populations not sampled (extinct and extant; i.e., ghost populations; Beerli 2004), suggesting a more complex evolutionary history of *Trichaptum* than the collected data can disclose. This is corroborated by the demographic modelling, where the best model includes introgression between a ghost population related to the Italian *T. fuscoviolaceum* and the North American B *T. abietinum* population. According to the network analysis, introgression has occurred from the Italian *T. fuscoviolaceum* population into the North American B *T. abietinum* population. However, *TreeMix* is not always able to reveal the true introgression scenario when the actual admixed populations are related to the populations used in the analyses (Fitak 2021). Therefore, introgression has not necessarily occurred between these two populations but most likely between the ghost population incorporated in the demographic modelling and the North American B *T. abietinum* population. This is similar to introgression inferred between archaic hominins, such as Denisovans and Neanderthals, and present-day humans (Durvasula and Sankararaman 2019). A wider collection, including more populations across the northern hemisphere (e.g., Asia and throughout Europe and North America), could capture the ghost population (if extant) and help untangle the shared evolutionary history of *T. abietinum* and *T. fuscoviolaceum*.

The f_{dM} analysis shows only slightly more significantly introgressed regions between the Italian *T. fuscoviolaceum* population tested against the *T. abietinum* populations than the Canadian *T. fuscoviolaceum* (14 vs. 11). All the introgressed regions occur between the *T. fuscoviolaceum* populations and the North American A population (not including B as in the other analyses). The demographic modelling did detect introgression between an ancestral *T. fuscoviolaceum* population and a *T. abietinum* population leading up to the current North American A population. Since this event is ancient, introgressed genes have had time to spread and become fixed in the genomes of current populations (in this case the North American A population and the *T. fuscoviolaceum* populations), which could explain why significant regions are only detected between the North American A population and the *T. fuscoviolaceum* populations.

Many of the genes are also found in the same regions across the genome for both comparisons. This further suggests that the f_{dM} analysis is detecting ancestral and not recent gene flow because the regions are conserved through the population splits (i.e., the introgression happened before the current populations diverged). The small regions of scattered introgression in the f_{dM} analysis also imply more ancient introgression. This follows similar patterns with highly divergent genomes and localized regions of introgression as found in analyses of three-spined stickleback species pairs in the Japanese archipelago (Ravinet et al. 2018) and *Heliconius* butterflies in Brazil (Zhang et al. 2016).

Population histories, introgression and its implications

In this study, we only have four populations of two widespread species, thus we are not covering the full diversity of the species. Still, the analyses are able to detect intricate population histories including both a population that is not sampled and ancestral populations leading up to the current ones (i.e., ghost populations). The chance of being able to sample all populations or have no ghost populations in an evolutionary study system consisting of natural populations is minor. It is therefore promising that we are able to extract interesting results based on a relatively small sample. The method is also useful for organismal groups such as fungi, where ancient genomes cannot be retrieved due to poor fossilization (Berbee et al. 2020; ancestral populations have for example been detected through genomic sequencing of subfossil in a study of the giant panda; Sheng et al. 2019). Including ghost populations in modelling can improve the estimate of migration rates (Beerli et al. 2004; Slatkin 2004). To increase our understanding of how introgression and gene flow affects the speciation continuum, it requires that researchers

account for scenarios such as extinct lineages and ghost populations when performing model testing.

The signs of introgression observed in the oldest migration in the demographic modelling and in the *D* and *f* statistics are likely a case of introgression from extinct lineages. Ancient introgression has previously been detected from extinct cave bears in the genomes of brown bears (*Ursus arctus*; Barlow et al. 2018), through phenotype analyses of beak sizes in one of Darwin's finches (*Geospiza fortis*; Grant and Grant 2021), and in the mitochondrial genome of the intermediate horseshoe bat (*Rhinolophus affinis*; Mao et al. 2012), to name a few. Genes or alleles from extinct lineages can therefore persist in extant species and might impose adaptive benefits (The *Heliconius* Genome Consortium et al. 2013; Racimo et al. 2015). It is difficult to say if this is the case with *T. abietinum* and *T. fuscoviolaceum*, but genes found in HMM outlier windows of the f_{dM} analysis may represent putatively adaptive genes with an ancient introgression origin. Many of the genes code for proteins of unknown function, which is common in non-model organisms due to limited research. However, the genes with similarity to other functional annotated genes are involved in several different functions in organisms. For example, oxidoreductases and hydrolases partake in numerous enzymatic reactions and are known to be important for wood decaying fungi to depolymerize the recalcitrant woody substrate (Floudas et al. 2012; Presley and Schilling 2017). Nevertheless, whether any of these genes are involved in adaptive introgression cannot be concluded based on the f_{dM} analysis alone. To extrapolate any adaptive implications from the GO enrichment analysis would not be appropriate, based on the lack of significance. It is also possible that the signs of introgression observed in the f_{dM} analysis are due to non-adaptive factors. For example, parts of the genome stemming from ancient introgression can persist due to recombination and constraint (e.g., Schumer et al. 2016; 2018). Thus, this question remains inconclusive until further analyses are conducted (e.g., recombination rates along the genome). However, the retention of the introgressed regions in the genome is still interesting and acts as a detection marker for the evolutionary history of these species not revealed by examining compatibility in the current populations.

Genes transferred through introgression can lead to an expansion of a species' distribution range, as for example seen for habitat and climate adaptation in cypress species (*Cupressus* spp.; Ma et al. 2019). The divergence of many of the taxa in the family Pinaceae, which includes the current host species of *T. fuscoviolaceum* and *T. abietinum*, is dated to the Jurassic (< ~185 million years ago; Ran et al. 2018), the same period as the divergence of Hymenochaetales (Varga et al. 2019). There are several examples where research on cryptic

diversity in fungi has revealed high divergence and old divergence times when species initially were thought to be closely related (summarized in Skrede 2021), which may also be the case for *T. fuscoviolaceum* and *T. abietinum*. Our results do not conclusively show adaptive introgression but based on the large nucleotide discrepancies and most likely old divergence, one could speculate that the introgression from ancestral populations has facilitated adaption to a larger host range of *T. fuscoviolaceum* and *T. abietinum* as conifers diverged and expanded across the northern hemisphere. However, additional research (e.g., protein function analysis) is needed to say anything certain about the implications of introgression between the sister species.

Since introgression can have impacts on the evolutionary trajectory of a species, it is an important mechanism to consider when investigating the evolutionary history of taxa. Introgression is not well examined in fungi or within an experimental system based on natural populations, and historically most introgression studies have been carried out on mammals or plants (Dagilis et al. 2021). Our results indicate that ancient introgression can be detected also among divergent species. Even though the phylogenetic relationship between *T. fuscoviolaceum* and *T. abietinum* is well-defined (Seierstad et al. 2020; Peris et al. 2022), signals of introgression lingering in their genomes suggests that the evolutionary history of these species is more complex than the current phylogenies can reveal.

Conclusion

Our study corroborates earlier findings, indicating that *T. abietinum* and *T. fuscoviolaceum* do not hybridize in vitro. Our results show that the sister species are highly divergent, exhibiting large genetic differences and are reproductively isolated. Nevertheless, introgression analyses display admixture, with small regions of introgression occurring throughout the genomes. These signs point to cases of both ancient and recent introgression between ancestral and current populations of *T. abietinum* and *T. fuscoviolaceum*, including a ghost population of a non-sampled or extinct population. Regardless of a well-resolved phylogeny, the evolutionary history of these species is intricate, including transfer of genes across lineages with unknown implications. This study builds on a small collection of studies detecting introgression between highly divergent species, expanding our knowledge on speciation and the permeability of reproductive barriers. The study also presents a novel system including natural populations and in vitro experiments, which is much needed for understanding the speciation continuum. The ceaselessness of speciation will naturally leave traces of historical events in the genomes of extant organisms. Accounting for these events when investigating speciation and adaptation

can give insight into how evolution proceeds and shapes the diversity we observe today, as well as how populations are affected in the future. It will be interesting to use this fungal experimental system applying other approaches, including protein function analysis, to link introgression to historical events (e.g., host shifts) and increase insight into the mechanisms governing divergence and adaptation.

Literature cited

Abbott, R., D. Albach, S. Ansell, J. W. Arntzen, S. J. E. Baird, N. Bierne, J. Boughman, A. Brelsford, C. A. Buerkle, R. Buggs, R. K. Butlin, U. Dieckmann, F. Eroukhmanoff, A. Grill, S.H. Cahan, J. S. Hermansen, G. Hewitt, A. G. Hudson, C. Jiggins, J. Jones, B. Keller, T. Marczewski, J. Mallet, P. Martinez-Rodriguez, M. Möst, S. Mullen, R. Nichols, A. W. Nolte, C. Parisod, K. Pfennig, A. M. Rice, M. G. Ritchie, B. Seifert, C. M. Smadja, R. Stelkens, J. M. Szymura, R. Väinölä, J. B. W. Wolf, and D. Zinner. 2013. Hybridization and speciation. J. Evol. Biol. 26:229-246.

Abbott, R. J., M. J. Hegarty, S. J. Hiscock, and A. C. Brennan. 2010. Homoploid hybrid speciation in action. Taxon. 59:1375-1386.

Ackermann, R. R., M. L. Arnold, M. D. Baiz, J. A. Cahill, L. Cortés-Ortiz, B. J. Evans, B. R. Grant, P. R. Grant, B. Hallgrimsson, R. A. Humphreys, C. J. Jolly, J. Malukiewicz, C. J. Percival, T. B. Ritzman, C. Roos, C. C. Roseman, L. Schroeder, F. H. Smith, K. A. Warren, R. K. Wayne, and D. Zinner. 2019. Hybridization in human evolution: Insights from other organisms. Evol. Anthropol. 28:189-209.

Aguillon, S. M., T. O. Dodge, G. A. Preising, and M. Schumer. 2022. Introgression. Curr. Biol. 32:R865-R868.

Alexa, A. and J. Rahnenfuhrer. 2021. topGO: Enrichment Analysis for Gene Ontology. R package version 2.40. Available at: https://bioconductor.org/packages/release/bioc/html/topGO.html. Accessed March 2, 2022.

Altschul, S. F., W. Gish, W. Miller, E. W. Meyers, and D. J. Lipman. 1990. Basic local alignment search tool. J. Mol. Biol. 215:403-410.

Anderson, E. and L. Hubricht. 1938. Hybridization in *Tradescantia*. III. The evidence for introgressive hybridization. Botany. 25:396-402.

Andrews, S. 2010. FastQC: A Quality Control Tool for High Throughput Sequence Data. Available at http://www.bioinformatics.babraham.ac.uk/projects/fastqc/. Accessed March 2, 2022.

Barlow, A., J. A. Cahill, S. Hartmann, C. Theunert, G. Xenikoudakis, G. G. Fortes, J. L. A. Paijmans, G. Rabeder, C. Frischauf, A. Grandal-d'Anglade, A. García-Vázquez, M. Murtskhvaladze, U. Saarma, P. Anijalg, T. Skrbinšek, G. Bertorelle, B. Gasparian, G. Bar-Oz, R. Pinhasi,M.

Slatkin, L. Dalén, B. Shapiro, and M. Hofreiter. 2018. Partial genomic survival of cave bears in living brown bears. Nat. Ecol. Evol. 2:1563-1570.

Baumgartner, K., B. R. Baker, K. Korhonen, J. Zhao, K. W. Hughes, J. Bruhn, T. S. Bowman, and S. E. Bergemann. 2012. Evidence of natural hybridization among homothallic members of the basidiomycete *Armillaria mellea* sensu stricto. Fungal Biol. 116:677-691.

Beerli, P. 2004. Effect of unsampled populations on the estimation of population sizes and migration rates between sampled populations. Mol. Ecol. 13:827-836.

Bengtson, S., B. Rasmussen, M. Ivarsson, J. Muhling, C. Broman, F. Marone, M. Stampanoni, and A. Bekker. 2017. Fungus-like mycelial fossils in 2.4-billion-year-old vesicular basalt. Nat. Ecol. Evol. 1:0141.

Bengtsson, H. 2021. R.utils: Various Programming Utilities. R package version 2.11.0. Available at: https://CRAN.R-project.org/package=R.utils. Accessed June 27, 2022.

Berbee, M. L., C. Strullu-Derrien, P.-M. Delaux, P. K. Strother, P. Kenrick, M.-A. Selosse, and J. W. Taylor. 2020. Genomic and fossil windows into the secret lives of the most ancient fungi. Nat. Rev. Microbiol. 18:717–730.

Bresinsky, A., M. Fischer, B. Meixner, and W. Paulus. 1987. Speciation in *Pleurotus*. Mycologia. 79:234-245.

Butlin, R. (1987). Speciation by reinforcement. Trends Ecol. Evol. 2:8-13.

Chang, C. C., C. C. Chow, L. C. Tellier, S. Vattikuti, S. M. Purcell, and J. J. Lee. 2015. Second generation PLINK: rising to the challenge of larger and richer datasets. GigaScience. 4: s13742-015-0047-8.

Crowl, A. A., P. S. Manos, J. D. McVay, A. R. Lemmon, E. M. Lemmon, and A. L. Hipp. 2019. Uncovering the genomic signature of ancient introgression between white oak lineages (*Quercus*). New Phytol. 226:1158-1170.

Cuevas, A., F. Eroukhmanoff, M. Ravinet, G.-L. Sætre, and A. Runemark. 2022. Predictors of genomic differentiation within a hybrid taxon. PLoS Genet. 18:e1010027.

Dagilis, A. J., D. Peede, J. M. Coughlan, G. I. Jofre, E. R. R. D'Agostino, H. Mavengere, A. D. Tate, and D. R. Matute. 2021 (Prepr.). 15 years of introgression studies: quantifying gene flow across Eukaryotes. bioRxiv. Available at: https://doi.org/10.1101/2021.06.15.448399.

Danecek, P., A. Auton, G. Abecasis, C. A. Albers, E. Banks, M. A. DePristo, R. Handsaker, G. Lunter, G. Marth, S. T. Sherry, G. McVean, R. Durbin, and 1000 Genomes Project Analysis Group. 2011. The variant call format and VCFtools. Bioinformatics. 27:2156-2158.

Danecek, P., J. K. Bonfield, J. Liddle, J. Marshall, V. Ohan, M. O. Pollard, A. Whitwham, T. Keane, S. A. McCarthy, R. M. Davies, and H. Li. 2021. Twelve years of SAMtools and BCFtools. GigaScience. 10:giab008.

Durand, E. Y., N. Patterson, D. Reich, and M. Slatkin. 2011. Testing for Ancient Admixture between Closely Related Populations, Mol. Biol. Evol. 28:2239-2252.

Durvasula, A. and S. Sankararaman. 2019. A statistical model for reference-free inference of archaic local ancestry. PLoS Genet. 15: e1008175.

Eberlein, C., M. Hénault, A. Fijarczyk, G. Charron, M. Bouvier, L. M. Kohn, J. B. Anderson, and C. R. Landry. 2019. Hybridization is a recurrent evolutionary stimulus in wild yeast speciation. Nat.Commun. 10:923.

Ewels, P., M. Magnusson, S. Lundin, and M. Käller. 2016. MultiQC: Summarize analysis results for multiple tools and samples in a single report. Bioinformatics. 32:3047-3048.

Excoffier, L., N. Marchi, D. A. Marques, R. Matthey-Doret, A. Gouy, and V. C. Sousa. 2021. *fastsimcoal2*: demographic inference under complex evolutionary scenarios. Bioinformatics. 37:4882-4885.

Fitak, R. R. 2021. OptM: estimating the optimal number of migration edges on population trees using Treemix. Biol. Methods Protoc. 6:bpab017.

Floudas, D., M. Binder, R. Riley, K. Barry, R. A. Blanchette, B. Henrissat, A. T. Martínez, R. Otillar, J. W. Spatafora, J. S. Yadav, A. Aerts, I. Benoit, A. Boyd, A. Carlson, A. Copeland, P. M. Coutinho, R. P. de Vries, P. Ferreira, K. Findley, B. Foster, J. Gaskell, D. Glotzer, P. Górecki, J. Heitman, C. Hesse, C. Hori, K. Igarashi, J. A. Jurgens, N. Kallen, P. Kersten, A. Kohler, U. Kües, T. K. Kumar, A. Kuo, K. LaButti, L. F. Larrondo, E. Lindquist, A. Ling, V. Lombard, S. Lucas, T. Lundell, R. Martin, D. J. McLaughlin, I. Morgenstern, E. Morin, C. Murat, L. G. Nagy, M. Nolan, R. A. Ohm, A. Patyshakuliyeva, A. Rokas, F. J. Ruiz-Dueñas, G. Sabat, A. Salamov, M. Samejima, J. Schmutz, J. C. Slot, F. St John, J. Stenlid, H. Sun, S. Sun, K. Syed, A. Tsang, A. Wiebenga, D. Young, A. Pisabarro, D. C. Eastwood, F. Martin, D. Cullen, I. V. Grigoriev, and D. S. Hibbett. 2012. The Paleozoic origin of enzymatic lignin decomposition reconstructed from 31 fungal genomes. Science. 336:1715–1719.

Flynn, J. M., R. Hubley, C. Goubert, J. Rosen, A. G. Clark, C. Feschotte, and A. F. Smit. 2020. RepeatModeler2 for automated genomic discovery of transposable element families. Proc. Natl. Acad. Sci. U. S. A. 117:9451-9457.

Fraser, J. A., Y.-P. Hsueh, K. M. Findley, and J. Heitman. 2007. Evolution of the Mating-Type Locus: The Basidiomycetes. Pp. 19-34 *in* J. Heitman, J. W. Kronstad, J. W. Taylor and L. A. Casselton, eds. Sex in fungi. ASM Press, Washington, D.C

Garbelotto, M., A. Ratcliff, T. D. Bruns, F. W. Cobb, and W. J. Otrosina. 1996. Use of taxon-specific competitive-priming PCR to study host specificity, hybridization, and intergroup gene flow in intersterility groups of *Heterobasidion annosum*. Phytopathology. 86:543-551.

Giordano, L., F. Sillo, M. Garbelotto, and P. Gonthier. 2018. Mitonuclear interactions may contribute to fitness of fungal hybrids. Sci. Rep. 8:1706.

Grant, P. R. and B. R. Grant. 2021. Morphological ghosts of introgression in Darwin's finch populations. Proc. Natl. Acad. Sci. U.S.A. 118:e2107434118.

Grant, P. R. and B. R. Grant. 2019. Hybridization increases population variation during adaptive radiation. Proc. Natl. Acad. Sci. U. S. A. 116: 23216-23224.

Green, R. E., J. Krause, A. W. Briggs, T. Maricic, U. Stenzel, M. Kircher, N. Patterson, H. Li, W. W. Zhai, M. H. Y. Fritz, N. F. Hansen, E. Y. Durand, A. S. Malaspinas, J. D. Jensen, T. Marques-Bonet, C. Alkan, K. Prufer, M. Meyer, H. A. Burbano, J. M. Good, R. Schultz, A. Aximu-Petri, A. Butthof, B. Höber, B. Höffner, M. Siegemund, A. Weihmann, C. Nusbaum, E. S. Lander, C. Russ, N. Novod, J. Affourtit, M. Egholm, C. Verna, P. Rudan, D. Brajkovic, Ž. Kucan, I. Gušic, V. B. Doronichev, L. V. Golovanova, C. Lalueza-Fox, M. de la Rasilla, J. Fortea, A. Rosas, R. W. Schmitz, P. L. F. Johnson, E. E. Eichler, D. Falush, E. Birney, J. C. Mullikin, M. Slatkin, R. Nielsen, J. Kelso, M. Lachmann, D. Reich, and S. Pääbo. 2010. A draft sequence of the Neandertal genome. Science. 328:710-722.

Harrison, R. G. and E. L. Larson. 2014. Hybridization, introgression, and the nature of species boundaries. J. Hered. 105:795–809.

Harte, D. 2021. HiddenMarkov: Hidden Markov Models. R package version 1.8-13. Wellington: Statistics Research Associates. Available at https://cran.r-project.org/web/packages/HiddenMarkov/index.html. Accessed March 2, 2022.

Hegarty, M. J. S. J. and Hiscock. 2005. Hybrid speciation in plants: new insights from molecular studies. New Phytol. 165:411-423.

Heinzelmann, R., D. Rigling, G. Sipos, M. Münsterkötter, and D. Croll. 2020. Chromosomal assembly and analyses of genome-wide recombination rates in the forest pathogenic fungus *Armillaria ostoyae*. Heredity. 124:699-713.

The Heliconius Genome Consortium, K. K. Dasmahapatra, J. Walters, A. D. Briscoe, J. W. Davey, A. Whibley, N. J. Nadeau, A. V. Zimin, D. S. T. Hughes, L. C. Ferguson, S. H. Martin, C. Salazar, J. J. Lewis, S. Adler, S.-J. Ahn, D. A. Baker, S. W. Baxter, N. L. Chamberlain, R. Chauhan, B. A. Counterman, T. Dalmay, L. E. Gilbert, K. Gordon, D. G. Heckel, H. M. Hines, K. J. Hoff, P. W. H. Holland, E. Jacquin-Joly, F. M. Jiggins, R. T. Jones, D. D. Kapan, P. Kersey, G. Lamas, D. Lawson, D. Mapleson, L. S. Maroja, A. Martin, S. Moxon, W. J. Palmer, R. Papa, A. Papanicolaou, Y. Pauchet, D. A. Ray, N. Rosser, S. L. Salzberg, M. A. Supple, A. Surridge, A. T. Trolander, H. Vogel, P. A. Wilkinson, D. Wilson, J. A. Yorke, F. Yuan, A. L. Balmuth, C. Eland, K. Gharbi, M. Thomson, R. A. Gibbs, Y. Han, J. C. Jayaseelan, C. Kovar, T. Mathew, D. M. Muzny, F. Ongeri, L.-L. Pu, J. Qu, R. L. Thornton, K. C. Worley, Y.-Q. Wu, M. Linares, M. L. Blaxter, R. H. F. Constant, M. Joron, M. R. Kronforst, S. P. Mullen, R. D. Reed, S. E. Scherer, S. Richards, J. Mallet, W. O. McMillan, and C. D. Jiggins. 2012. Butterfly genome reveals promiscuous exchange of mimicry adaptations among species. Nature. 487:94-98.

Hermansen, J. S., S. A. Sæther, T. O. Elgvin, T. Borge, E. Hjelle, and G.-P. Sætre. 2011. Hybrid speciation in sparrows I: phenotypic intermediacy, genetic admixture and barriers to gene flow. Mol. Ecol. 20:3812-3822.

Holt, C. and M. Yandell. 2011. MAKER2: an annotation pipeline and genome-database management tool for second-generation genome projects. BMC Bioinformatics. 12:491.

Index Fungorum. 2021. Index Fungorum, The Royal Botanic Gardens Kew. Available at www.indexfungorum.org. Accessed March 2, 2022.

Jain, C., L. M. Rodriguez-R, A. M. Phillippy, K. T. Konstantinidis, and S. Aluru. 2018. High throughput ANI analysis of 90K prokaryotic genomes reveals clear species boundaries. Nat. Commun. 9:5114.

James, T. Y., U. Kües, S. A. Rehnerc, and R. Vilgalysa. 2004. Evolution of the gene encoding mitochondrial intermediate peptidase and its cosegregation with the A mating-type locus of mushroom fungi. Fungal Genet. Biol. 41:381-390.

Jones, P., D. Binns, H.-Y. Chang, M. Fraser, W. Li, C. McAnulla, H. McWilliam, J. Maslen, A. Mitchell, G. Nuka, S. Pesseat, A. F. Quinn, A. Sangrador-Vegas, M. Scheremetjew, S.-Y. Yong, R. Lopez, and S. Hunter. 2014. InterProScan 5: genome-scale protein function classification. Bioinformatics. 30:1236-40.

Kauserud, H. and T. Schumacher. 2003. Ribosomal DNA variation, recombination and inheritance in the basidiomycete *Trichaptum abietinum*: implications for reticulate evolution. Heredity. 91:163-172.

Keuler, R., A. Garretson, T. Saunders, R. J. Erickson, N. St. Andre, F. Grewe, H. Smith, H. T. Lumbsch, J.-P. Huang, L. L. St. Clair, and S. D. Leavitt. 2020. Genome-scale data reveal the role of hybridization in lichen-forming fungi. Sci. Rep. 10:1497.

Krueger, F. 2015. Trim Galore!: A wrapper tool around Cutadapt and FastQC to consistently apply quality and adapter trimming to FastQ files. Available at http://www.bioinformatics.babraham.ac.uk/projects/trim_galore/. Accessed March 2, 2022.

Langdon, Q. K., D. Peris, E. C. P. Baker, D. A. Opulente, H.-V. Nguyen, U. Bond, P. Gonçalves, J. P. Sampaio, D. Libkind, and C. T. Hittinger. 2019. Fermentation innovation through complex hybridization of wild and domesticated yeasts. Nat. Ecol. Evol. 3:1576 1586.

Langdon, Q. K., D. Peris, B. Kyle, and C. T. Hittinger. 2018. sppIDer: A species identification tool to investigate hybrid genomes with high-throughput sequencing. Mol. Biol. Evol. 35:2835-2849.

Lawrence, M., R. Gentleman, and V. Carey. 2009. rtracklayer: an R package for interfacing with genome browsers. Bioinformatics. 25:1841-1842.

Li, H. and R. Durbin. 2009. Fast and accurate short read alignment with Burrows-Wheeler Transform. Bioinformatics. 25:1754-60.

Li, H., B. Handsaker, A. Wysoker, T. Fennell, J. Ruan, N. Homer, G. Marth, G. Abecasis, R. Durbin, and 1000 Genome Project Data Processing Subgroup. 2009. The Sequence alignment/map (SAM) format and SAMtools. Bioinformatics. 25:2078-2079.

Li, Y., J. L. Steenwyk, Y. Chang, Y. Wang, T. Y. James, J. E. Stajich, J. W. Spatafora, M. Groenewald, C. W. Dunn, C. T. Hittinger, X.-X. Shen, and A. Rokas. 2021. A genome scale phylogeny of the kingdom Fungi. Curr. Biol. 31:1653-1665.e5

Lunter, G. and M. Goodson. 2011. Stampy: A statistical algorithm for sensitive and fast mapping of Illumina sequence reads. Genome Res. 21:961-973.

Ma, Y., J. Wang, Q. Hu, J. Li, Y. Sun, L. Zhang, R. J. Abbott, J. Liu, and L. Mao. 2019. Ancient introgression drives adaptation to cooler and drier mountain habitats in a cypress species complex. Commun. Biol., 2:213.

Macrae, R. 1967. Pairing incompatibility and other distinctions among *Hirchiporus abietinus*, *H. fusco-violaceus*, and *H. laricinus*. Can. J. Bot. 45:1371-1399

Magasi, L.P. 1976. Incompatibility factors in *Polyporus abietinus*, their numbers and distribution. Mem. NY. Bot. Gard. 28:163–173.

Malinsky, M., R. J. Challis, M. T. Alexandra, S. Schiffels, Y. Terai, P. B. Ngatunga, E. A. Miska, R. Durbin, M. J. Genner, and G. F. Turner. 2015. Genomic islands of speciation separate cichlid ecomorphs in an East African crater lake. Science. 350:1493-1498.

Mallet, J., N. Besansky, and M. W. Hahn. 2016. How reticulated are species? BioEssays. 38: 140-149.

Mao, X., G. He, P. Hua, G. Jones, S. Zhang, S. J. Rossiter. 2012. Historical introgression and the persistence of ghost alleles in the intermediate horseshoe bat (*Rhinolophus affinis*). Mol. Ecol. 22:1035-1050.

Martin, S. H., J. W. Davey, and C. D. Jiggins. 2014. Evaluating the use of ABBA–BABA statistics to locate introgressed loci. Mol. Biol. Evol. 32:244–257.

Martin, S. H., J. W. Davey, C. Salazar, and C. D. Jiggins. 2019. Recombination rate variation shapes barriers to introgression across butterfly genomes. PLoS Biol. 17:e2006288.

Mavárez, J., C. A. Salazar, E. Bermingham, C. Salcedo, C. D. Jiggins, and M. Linares. 2006. Speciation by hybridization in *Heliconius* butterflies. Nature. 441:868-871.

May, G., F. Shaw, H. Badrane, and X. Vekemans. 1999. The signature of balancing selection: Fungal mating compatibility gene evolution. Proc. Natl. Acad. Sci. U. S. A. 96:9172–9177.

McKenna, A., H. Hanna, E. Banks, A. Sivachenko, K. Cibulskis, A. Kernytsky, K. Garimella, D. Altshuler, S. Gabriel, M. Daly, and M. A. DePristo. 2010. The Genome Analysis Toolkit: A MapReduce framework for analyzing next-generation DNA sequencing data. Genome Res. 20:1297-1303.

Minh, B. Q., H. A. Schmidt, O. Chernomor, D. Schrempf, M. D. Woodhams, A. von Haeseler, R., and Lanfear. 2020. IQ-TREE 2: New Models and Efficient Methods for Phylogenetic Inference in the Genomic Era. Mol. Biol. Evol. 37:1530–1534.

Moore, D., G. D. Robson, and A. P. J. Trinci. 2011. Chapter 12.14 – Classic genetic approaches to study development and the impact of genomic data mining. Pp. 311-315 *in* D. Moore, G. D. Robson, and A. P. J. Trinci, eds. 21st century guidebook to fungi. Cambridge University Press, New York.

Nachman, M. W. and B. A. Payseur. 2012. Recombination rate variation and speciation: theoretical predictions and empirical results from rabbits and mice. Philos. Trans. R. Soc. Lond., B, Biol. Sci. 367: 409–421.

Nei, M., T. Maruyama, and C.-I. Wu. 1983. Models of evolution of reproductive isolation. Genetics 103:557-579.

Nelson, T. C., A. M. Stathos, D. D. Vanderpool, F. R. Finseth, Y-w. Yuan, and L. Fishman. 2021. Ancient and recent introgression shape the evolutionary history of pollinator adaptation and speciation in a model monkeyflower radiation (Mimulus section Erythranthe). PLoS Genet. 17:e1009095.

Neuwirth, E. 2014. RColorBrewer: ColorBrewer Palettes. R package version 1.1-2. Available at: https://CRAN.R-project.org/package=RColorBrewer. Accessed June 27, 2022.

Nosil, P., J. L. Feder, S. M. Flaxman, and Z. Gompert. 2017. Tipping points in the dynamics of speciation. Nat. Ecol. Evol. 1:0001.

Patterson, N., P. Moorjani, Y. Luo, S. Mallick, N. Rohland, Y. Zhan, T. Genschoreck, T. Webster, and D. Reich. 2012. Ancient admixture in human history. Genetics. 192:1065–1093.

Patterson, N., A. L. Price, and D. Reich. 2006. Population structure and eigenanalysis. PLoS Genet. 2:2074-2093.

Peris, D., D. Lu, V. B. Kinneberg, I.-S. Methlie, M. S. Dahl, T. Y. James, H. Kauserud, and I. Skrede. 2022. Large-scale fungal strain sequencing unravels the molecular diversity in mating loci maintained by long-term balancing selection. PLoS Genet. 18: e1010097.

Petr, M., B. Vernot, and J. Kelso. 2019. admixr—R package for reproducible analyses using ADMIXTOOLS. Bioinformatics. 35:3194–3195.

Pickrell, J. K. and J. K. Pritchard. 2012. Inference of population splits and mixtures from genome wide allele frequency Data. PLoS Genet. 8: e1002967.

Platt R. N., II, M. McDew-White, W. L. Clec'h, F. D. Chevalier, F. Allan, A. M. Emery, A. Garba, A. A. Hamidou, S. M. Ame, J. P. Webster, D. Rollinson, B. L. Webster, and T. J. C. Anderson. 2019. Ancient hybridization and adaptive introgression of an invadolysin gene in schistosome parasites. Mol. Biol. Evol. 36:2127–2142.

Presley, G. N. and J. S. Schilling. 2017. Distinct Growth and Secretome Strategies for Two Taxonomically Divergent Brown Rot Fungi. Appl. Environ. Microbiol. 83:e02987-16.

Price, A. L., N. J. Patterson, R. M. Plenge, M. E. Weinblatt, N. A. Shadick, and D. Reich. 2006. Principal components analysis corrects for stratification in genome-wide association studies. Nat. Genet. 38:904-909.

R Core Team. 2020. R: A language and environment for statistical computing, R Foundation for Statistical Computing. Vienna, Austria. Available at https://www.R-project.org/. Accessed March 2, 2022.

Racimo, F., S. Sankararaman, R. Nielsen, and E. Huerta-Sánchez. 2015. Evidence for archaic adaptive introgression in humans. Nat. Rev. Genet. 16:359–371.

Raghavan, M., P. Skoglund, K. Graf, M. Metspalu, A. Albrechtsen, I. Moltke, S. Rasmussen, T. W. Stafford Jr, L. Orlando, E. Metspalu, M. Karmin, K. Tambets, S. Rootsi, R. Mägi, P. F. Campos, E. Balanovska, O. Balanovsky, E. Khusnutdinova, S. Litvinov, L. P. Osipova, S. A. Fedorova, M. I. Voevoda, M. DeGiorgio, T. Sicheritz-Ponten, S. Brunak, S. Demeshchenko, T. Kivisild, R. Villems, R. Nielsen, M. Jakobsson, and E. Willerslev. 2014. Upper Palaeolithic Siberian genome reveals dual ancestry of Native Americans. Nature. 505:87–91.

Ram, K. and H. Wickham. 2018. wesanderson: A Wes Anderson Palette Generator. R package version 0.3.6. Available at https://CRAN.R-project.org/package=wesanderson. Accessed March 2, 2022.

Ran, J.-H., T.-T. Shena, H. Wu, X. Gong, and X.-Q. Wang. 2018. Phylogeny and evolutionary history of Pinaceae updated by transcriptomic analysis. Mol. Phylogenet. Evol. 129:106-116.

Ravinet, M., K. Yoshida, S. Shigenobu, A. Toyoda, A. Fujiyama, and J. Kitano. 2018. The genomic landscape at a late stage of stickleback speciation: High genomic divergence interspersed by small localized regions of introgression. PLoS Genet. 14:e1007358.

Reich, D. N. Patterson, M. Kircher, F. Delfin, M. R. Nandineni, I. Pugach, A. M.-S. Ko, Y. C. Ko, T. A. Jinam, M. E. Phipps, N. Saitou, A. Wollstein, M. Kayser, S. Pääbo, and M. Stoneking. 2011. Denisova admixture and the first modern human dispersals into southeast Asia and Oceania. Am. J. Hum. Genet. 89:516-528.

Reich, D., K. Thangaraj, N. Patterson, A. L. Price, and L. Singh. 2009. Reconstructing Indian population history. Nature. 461:489-494.

Schumer, M., C. Xu, D. L. Powell, A. Durvasula, L. Skov, C. Holland, J. C. Blazier, S. Sankararaman, P. Andolfatto, G. G. Rosenthal, and M. Przeworski. 2018. Natural selection interacts with recombination to shape the evolution of hybrid genomes. Science. 360:656-660.

Schumer, M., R. Cui, D. L. Powell, G. G. Rosenthal, and P. Andolfatto. 2016. Ancient hybridization and genomic stabilization in a swordtail fish. Mol. Ecol. 25: 2661-2679.

Seierstad, K. S., R. Fossdal, O. Miettinen, T. Carlsen, I. Skrede, and H. Kauserud. 2020. Contrasting genetic structuring in the closely related basidiomycetes *Trichaptum abietinum* and *Trichaptum fuscoviolaceum* (Hymenochaetales). Fungal Biol. 124:269-275.

Skrede, I. 2021. Chapter one - Diversity and distribution of ligninolytic fungi. Pp. 1-36 *in* M. Morel-Rouhier and R. Sormani, eds. Advances in botanical research, vol 99: Wood degradation and ligninolytic fungi. Academic Press, Cambridge.

Slatkin, M. 2004. Seeing ghosts: the effect of unsampled populations on migration rates estimated for sampled populations. Mol. Ecol. 14:67-73.

Smit, A. F. A., R. Hubley, and P. Green. 2013-2015. RepeatMasker Open-4.0. Available at http://www.repeatmasker.org. Accessed March 2, 2022.

Stankowski, S. and M. Ravinet. 2021. Defining the speciation continuum. Evolution. 75: 1256-1273.

Stenlid, J. and J.-O. Karlsson. 1991. Partial intersterility in *Heterobasidion annosum*. Mycol. Res. 95:1153-1159.

Stukenbrock, E. H. 2016. The role of hybridization in the evolution and emergence of new fungal plant pathogens. Phytopathology. 106:104-112.

Stukenbrock, E. H., F. B. Christiansen, T. T. Hansen, J. Y. Dutheil, and M. H. Schierup. 2012. Fusion of two divergent fungal individuals led to the recent emergence of a unique widespread pathogen species. Proc. Natl. Acad. Sci. U. S. A. 109:10954-10959.

The UniProt Consortium. 2021. UniProt: the universal protein knowledgebase in 2021. Nucleic Acids Res. 49:D480-D489.

Van Rossum, G. and F. L. Drake. 2009. Python 3 Reference Manual. CreateSpace, Scotts Valley, CA.

Varga, T., K. Krizsán, C. Földi, B. Dima, M. Sánchez-García, S. Sánchez-Ramírez, G. J. Szöllősi, J. G. Szarkándi, V. Papp, L. Albert, W. Andreopoulos, C. Angelini, V. Antonín, K. W. Barry, N. L. Bougher, P. Buchanan, B. Buyck, V. Bense, P. Catcheside, M. Chovatia, J. Cooper, W. Dämon, D. Desjardin, P. Finy, J. Geml, S. Haridas, K. Hughes, A. Justo, D. Karasiński, I. Kautmanova, B. Kiss, S. Kocsubé, H. Kotiranta, K. M. LaButti, B. E. Lechner, K. Liimatainen, A. Lipzen, Z. Lukács, S. Mihaltcheva, L. Morgado, T. Niskanen, M. E. Noordeloos, R. A. Ohm, B. Ortiz-Santana, C. Ovrebo, N. Rácz, R. Riley, A. Savchenko, A. Shiryaev, K. Soop, V. Spirin, C. Szebenyi, M. Tomšovský, R. E. Tulloss, J. Uehling, I. V. Grigoriev, C. Vágvölgyi, T. Papp, F. M. Martin, O. Miettinen, D. S. Hibbett, and L. G. Nagy. 2019. Megaphylogeny resolves global patterns of mushroom evolution. Nat. Ecol. Evol. 3:668-678.

Wang, G. D., W. Zhai, H. C. Yang, L. Wang, L. Zhong, Y. H. Liu, R. X. Fan, T. T. Yin, C. L. Zhu, A. D. Poyarkov, D. M. Irwin, M. K. Hytönen, H. Lohi, C. I. Wu, P. Savolaine, and Y. P. Zhang. 2016. Out of southern East Asia: the natural history of domestic dogs across the world. Cell. Res. 26:21-33.

Watterson, G. A. 1985. The genetic divergence of two populations. Theor. Popul. Biol. 27:298-317.

White, N. J., R. R. Snook, and I. Eyres. 2019. The past and future of experimental speciation. Trends Ecol. Evol. 35:10-21.

White, T. J., T. Bruns, S. Lee, and J. W. Taylor. 1990. Amplification and direct sequencing of fungal ribosomal RNA genes for phylogenetics. Pp. 315-322 *in* M. A. Innis, D. H. Gelfand, J. J. Sninsky and T. J. White, eds. PCR protocols: a guide to methods and applications. Academic Press, Inc., New York.

Wickham, H. 2016. ggplot2: Elegant Graphics for Data Analysis. 2nd ed. Springer, New York, NY.

Wickham, H., M. Averick, J. Bryan, W. Chang, L. D. McGowan, R. François, G. Grolemund, A. Hayes, L. Henry, J. Hester, M. Kuhn, T. L. Pedersen, E. Miller, S. M. Bache, K. Müller, J. Ooms, D. Robinson, D. P. Seidel, V. Spinu, K. Takahashi, D. Vaughan, C. Wilke, K. Woo, and H. Yutani. 2019. Welcome to the tidyverse. J. Open Source Softw. 4:1686.

Wood, T. E., N. Takebayashi, M. S. Barker, I. Mayrose, P. B. Greenspoon, and L. H. Rieseberg. 2009. The frequency of polyploid speciation in vascular plants. Proc. Natl. Acad. Sci. U. S. A. 106:13875-13879.

Zhang, W., K. K. Dasmahapatra, J. Mallet, G. R. P. Moreira, and M. R. Kronforst. 2016. Genome wide introgression among distantly related *Heliconius* butterfly species. Genome Biol. 17:25.

Supporting information:
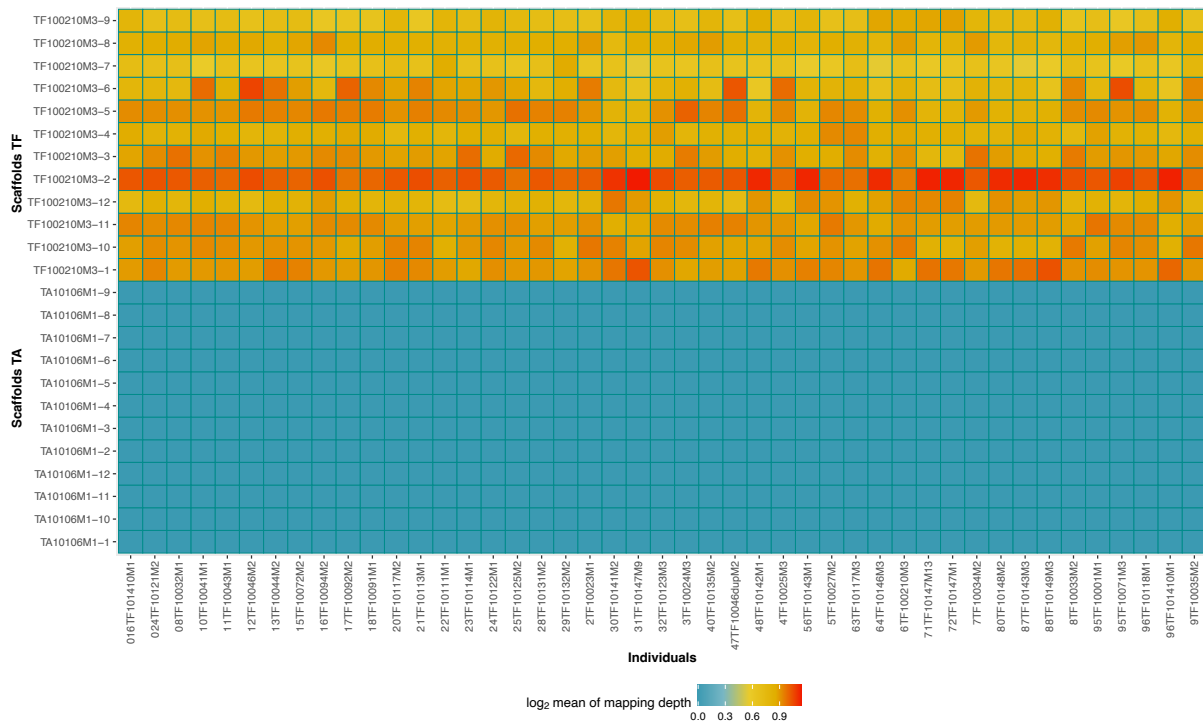
Figure S1. **The *sppIDer* analysis did not detect hybrid individuals.** On the y-axis are scaffolds (chromosomes) of the *Trichaptum fuscoviolaceum* (TF) and *T. abietinum* (TA) combined reference genome. The x-axis shows *T. fuscoviolaceum* individuals mapped to the combined reference genome. The legend on the bottom shows a colour gradient for the $\log_2$ mean of the mapping depth. Cooler colours indicate poorer mapping, while warmer colours indicate better mapping. The figure is made in *R v4.0.2* (R Core Team) using the packages *ggplot2* (Wickham, 2016), *wesanderson* (Ram and Wickham, 2018), *viridis* (Garnier, 2018), *readtext* (Benoit and Obeng, 2020), *data.table* (Dowle and Srinivasan, 2020) and *hrbrthemes* (Rudis, 2020).
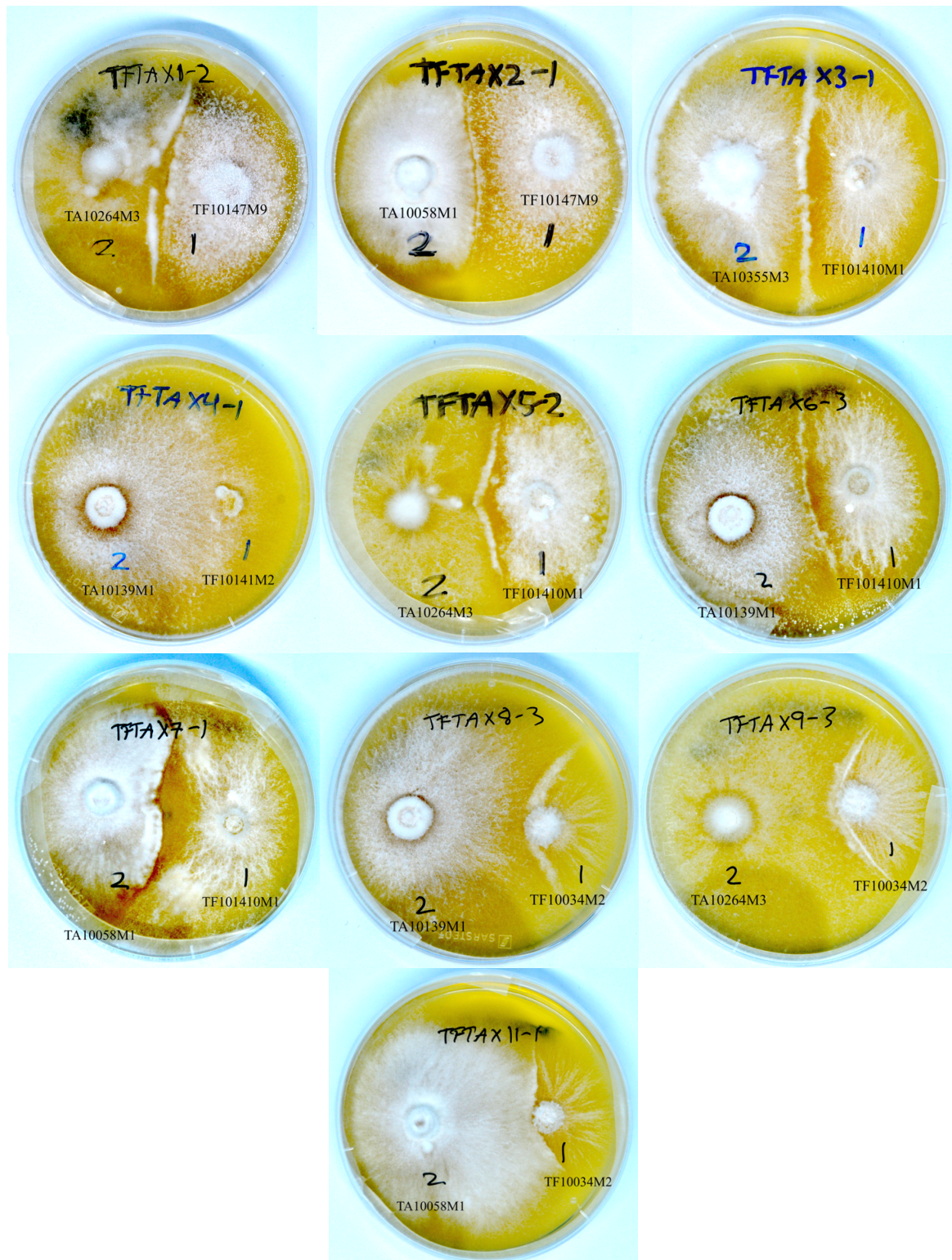
Figure S2. No successful crossings between *Trichaptum abietinum* and *T. fuscoviolaceum.* Photographs of cultures from the crossing experiments (one of the three replicates for each cross). None of the crossings are successful. The cross name is indicated at the top (number after the dashed line indicates replicate number), and the individuals are noted on the bottom. TA = *T. abietinum* and TF = *T. fuscoviolaceum*. Photographs were taken with a Nikon D600 Digital Camera (Tokyo, Japan).
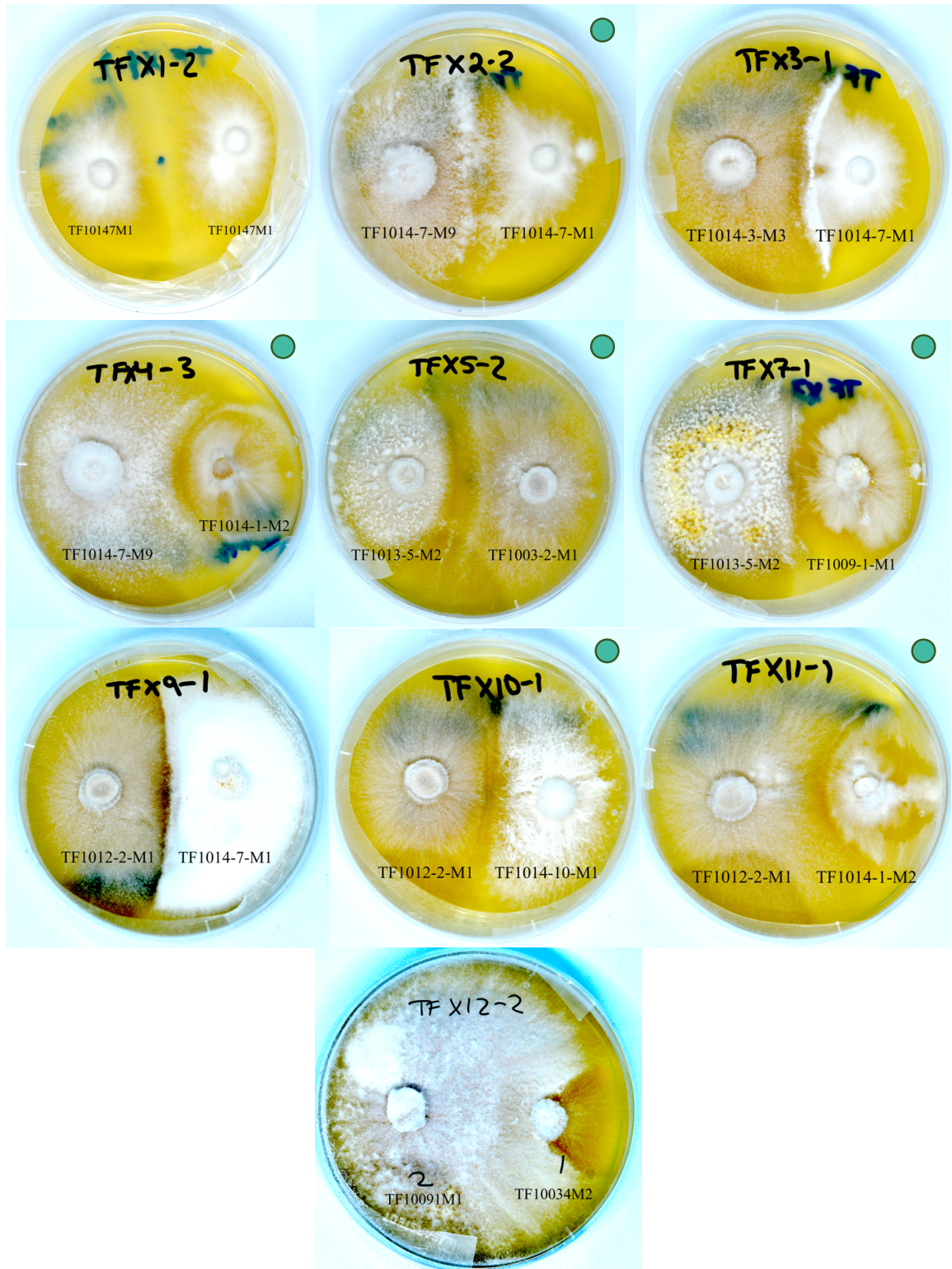
Figure S3. Crossings between individuals of *T. fuscoviolaceum* mated as expected. Photographs of cultures from the crossing experiments (one of the three replicates for each cross). Crossings between individuals were as expected (see Table S2). Successful crossings are marked with a green circle. The cross name is indicated at the top (number after the dashed line indicates replicate number), and the individuals are noted on the bottom. TF = *T. fuscoviolaceum*. Photographs were taken with a Nikon D600 Digital Camera (Tokyo, Japan).

48
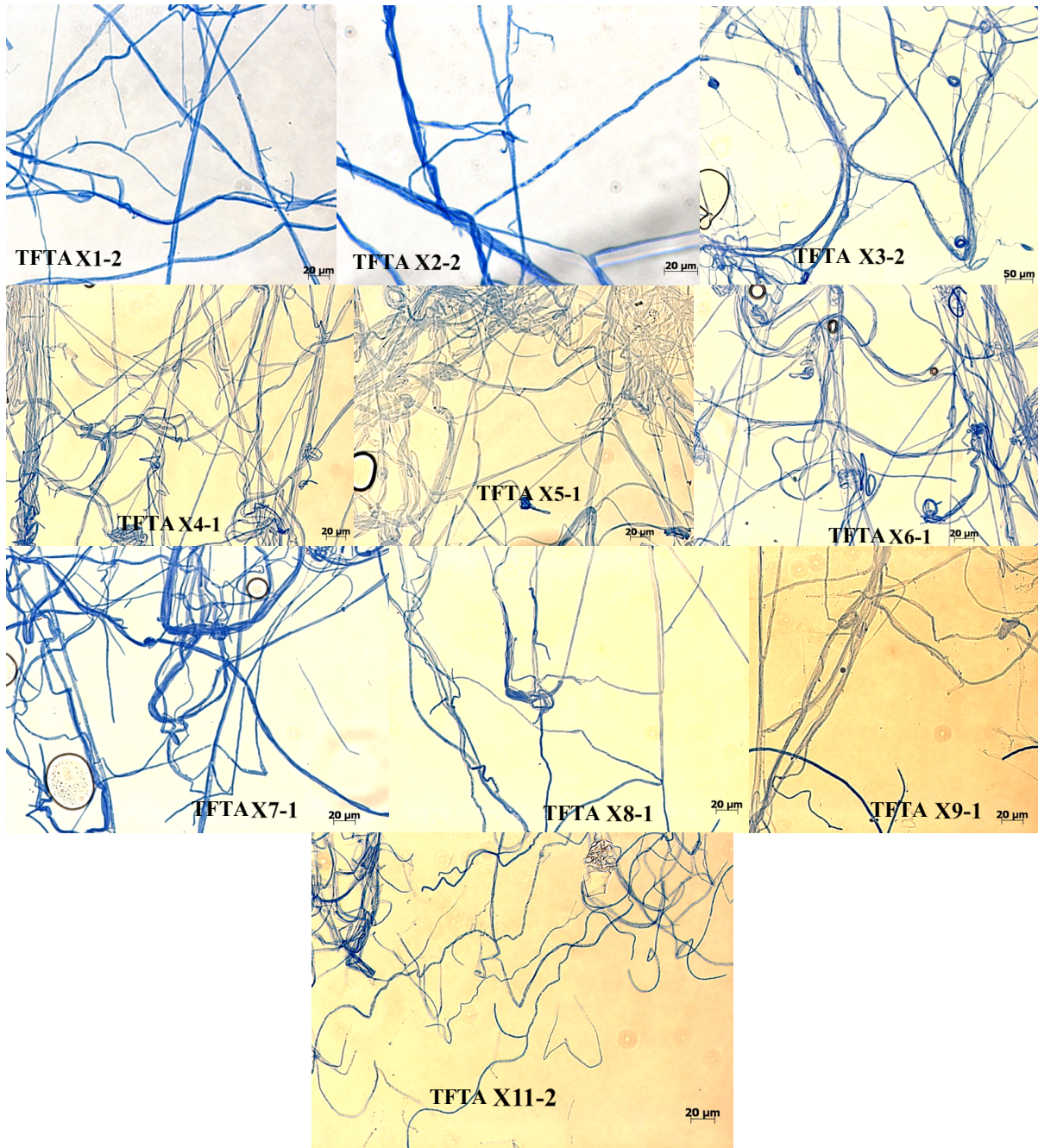
Figure S4. No clamp connections in the crossings between *Trichaptum abietinum* and *T. fuscoviolaceum*. Microscope photographs of the crossing experiments. Cross names are indicated at the bottom of the pictures and the number after the dashed lines indicate replicate number. A scale bar is positioned at the bottom right of every picture. TA = *T. abietinum* and TF = *T. fuscoviolaceum*. Photographs were taken using Zeiss Axioplan 2 imaging light microscope (Güttingen, Germany) with[49] Zeiss AxioCam HRc (Güttingen, Germany).
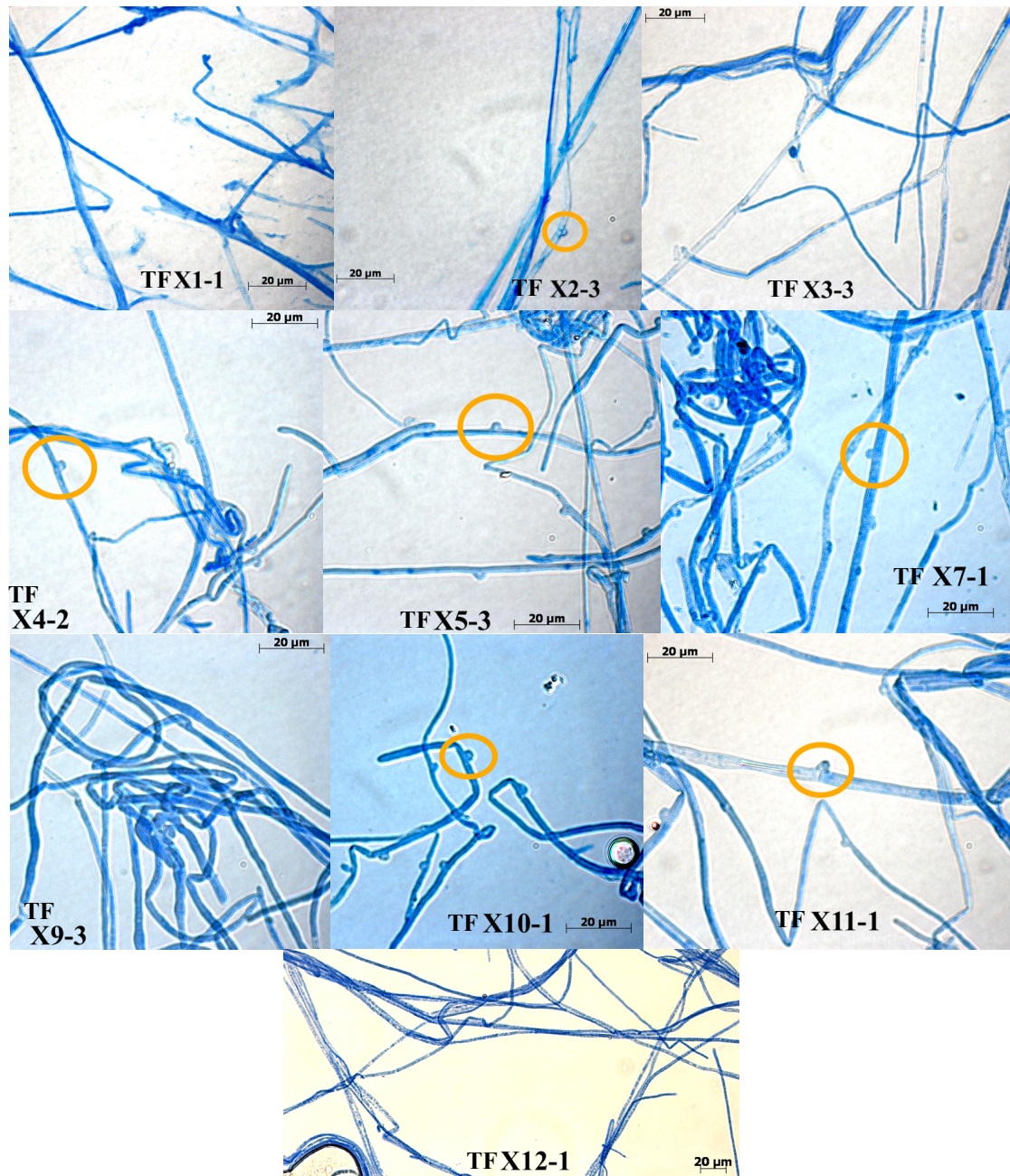
Figure S5. Clamp connections between successful matings as expected in the crossings between individuals of *T. fuscoviolaceum*. Microscope photographs of the crossing experiments. Clamp connections are marked with an orange circle. Cross names are indicated at the bottom of the pictures and the number after the dashed lines indicate replicate number. A scale bar is positioned at the bottom right of every picture. TF = *T. fuscoviolaceum*. Photographs were taken using Zeiss Axioplan 2 imaging light microscope (Güttingen, Germany) with Zeiss AxioCam HRc (Güttingen, Germany).
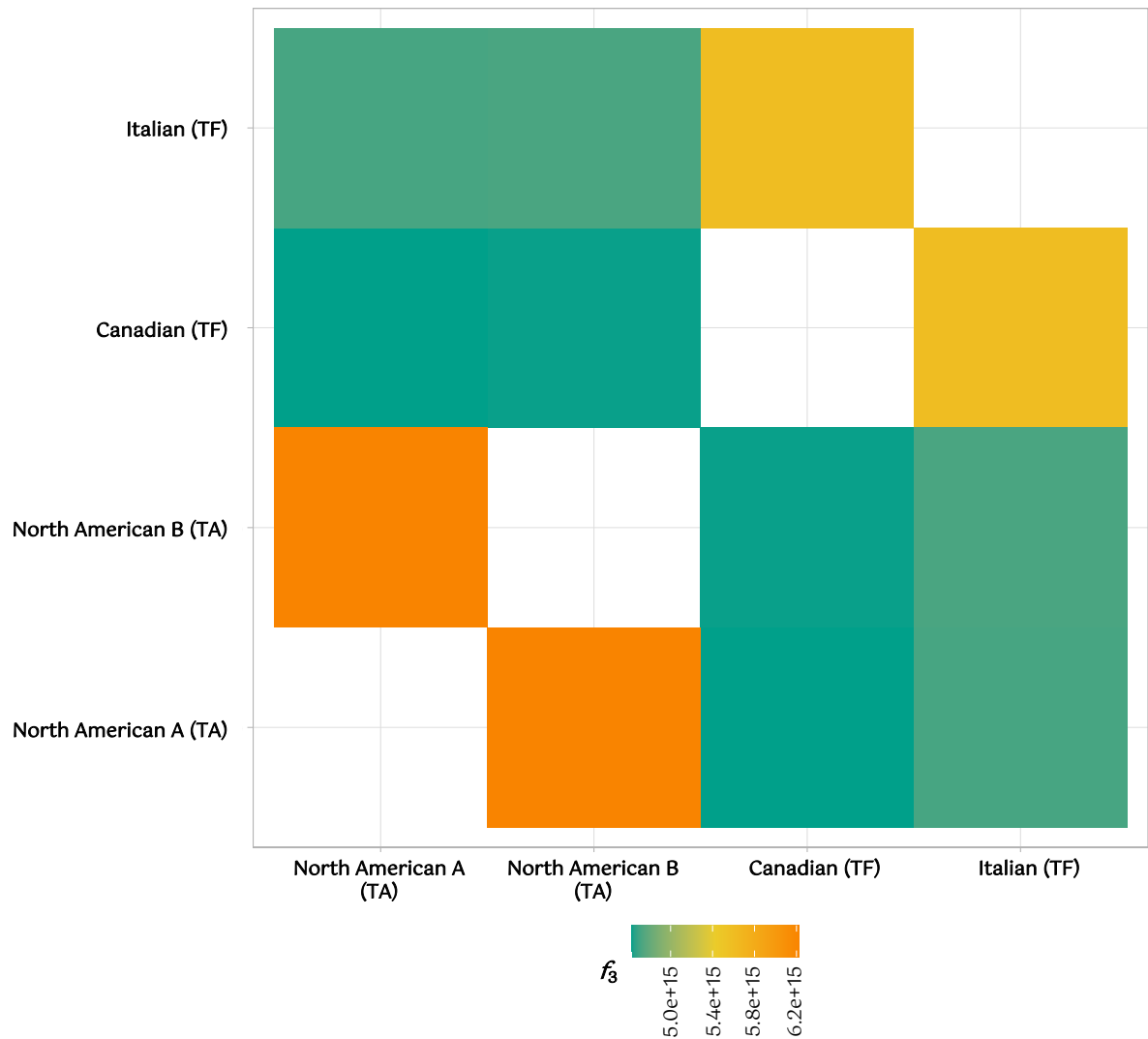
Figure S6. *Trichaptum abietinum* **populations split more recently than the** *T. fuscoviolaceum* **populations in the three-population *f* statistic (*f₃*) with outgroup.** The analysis is based on a single nucleotide polymorphism (SNP) dataset of 3 118 957 SNPs. The figure shows a pairwise comparison of populations colored by amount of shared evolutionary history. The analysis is based upon the phylogenetic hypothesis ((A, B), C)), where the branch lengths of A and B are estimated relative to C. All populations have been tested at position A and B, while C is kept constant as the outgroup *Trichaptum biforme*. The color legend at the bottom depicts relative split (amount of shared genetic drift) between the two populations compared. Higher values indicate a later split than lower values. In Figure 2 these results are illustrated by the time arrow and dotted lines showing the relative split of *T. fuscoviolaceum* populations compared to *T. abietinum* populations from their common ancestor. TF = *T. fuscoviolaceum* and TA = *T. abietinum*. The figure is made in *R v4.0.2* using the packages *admixr* (Petr et al. 2019), *tidyverse* (Wickham et al. 2019) and *wesanderson* (Ram and Wickham 2018).
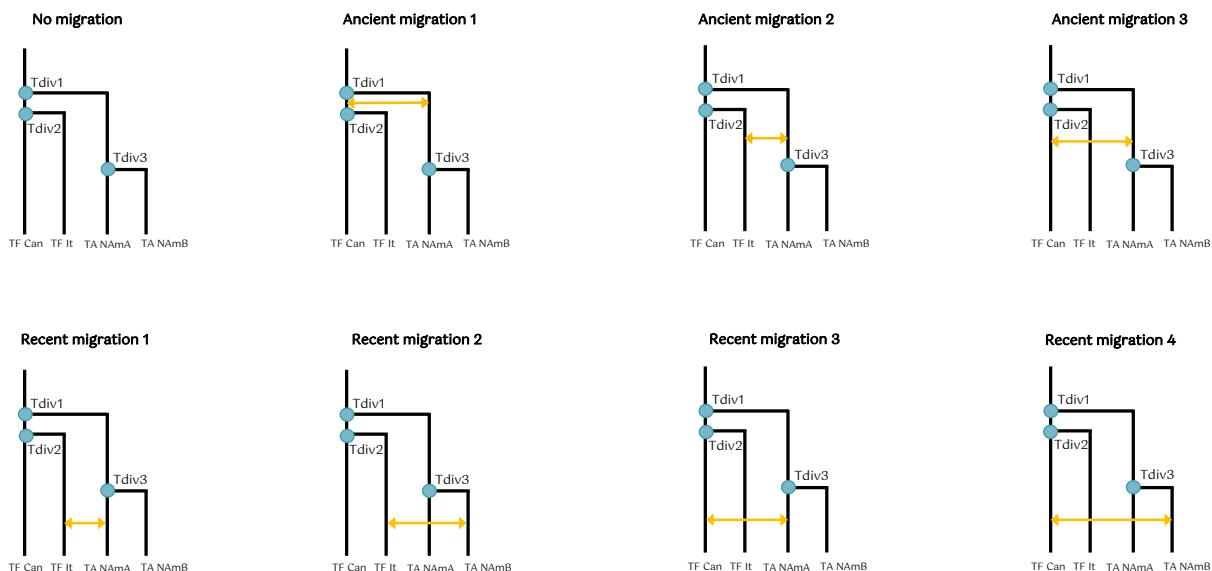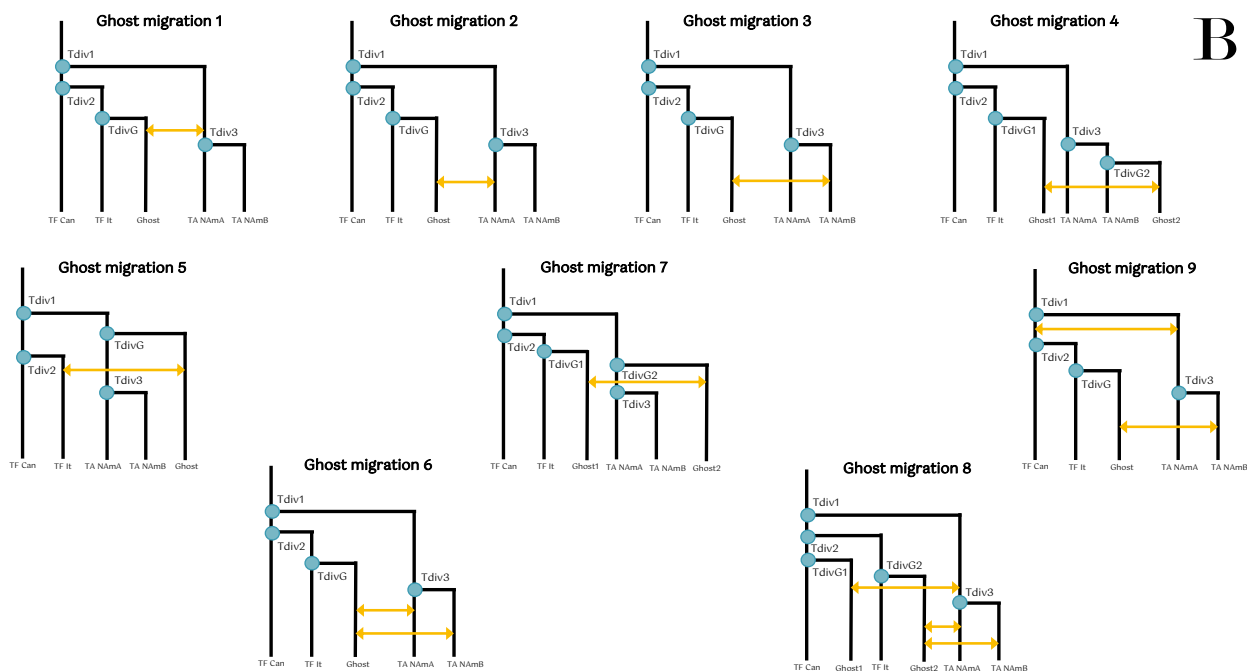
Figure S7. Models used to test divergence and introgression (migration) times. Illustrations of how the different models tested were set up in *fastsimcoal2* (Excoffier et al. 2021). Blue dots are the divergence times being estimated and the yellow arrows are the migrations times being estimated. (A) Models without ghost populations. (B) Models with ghost populations. Tdiv = time of divergence, TF = *Trichaptum fuscoviolaceum*, TA = *T. abietinum*, Can = Canadian, It = Italian, NAmA = North American A and NAmB = North American B.
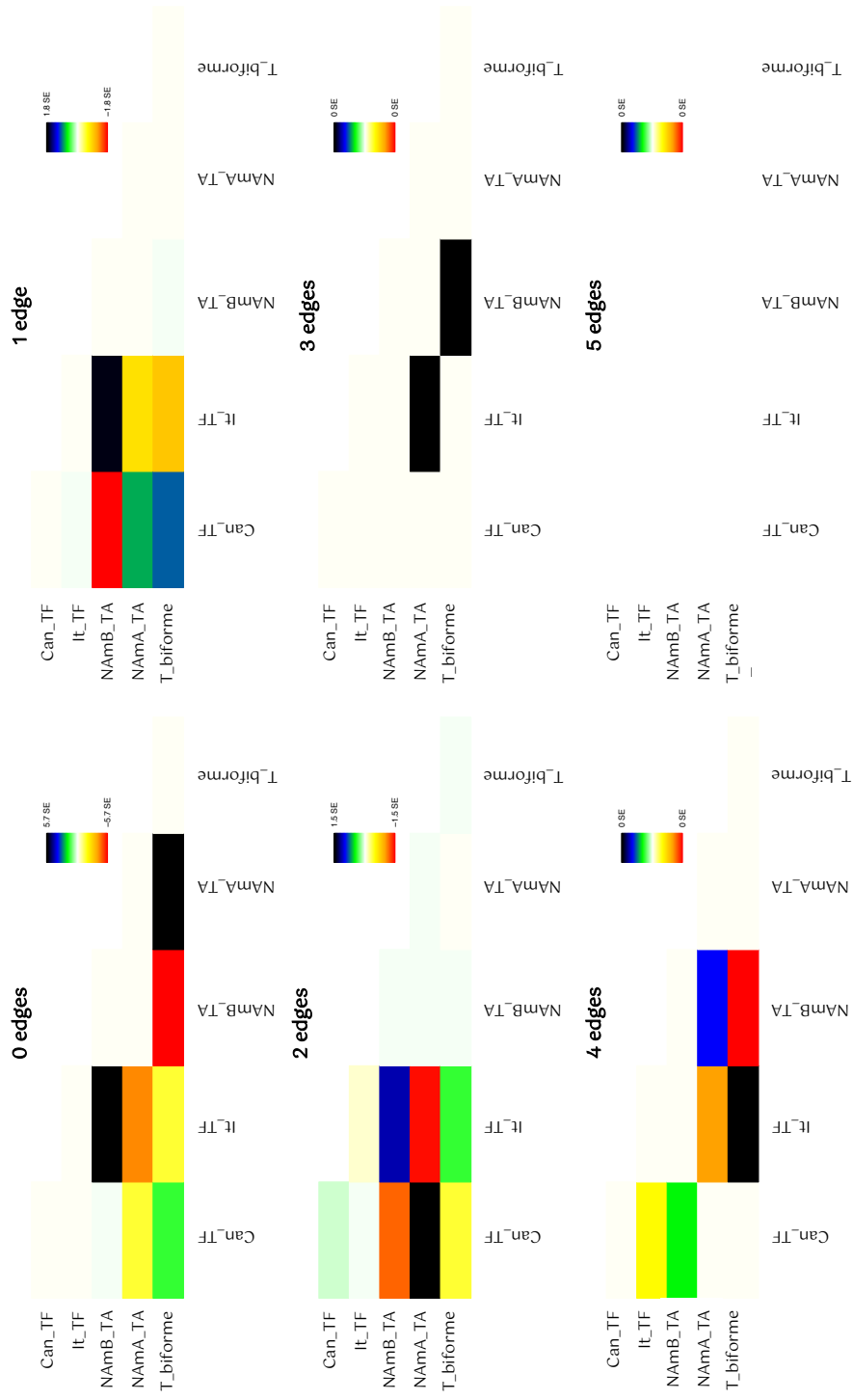
Figure S8. Residual plots from the *TreeMix* analysis. Residuals plotted for different edges based on the *TreeMix* (Fitak 2021) analysis using a block size of 700. The colour bar on the right shows the standard error (SE). A higher SE (e.g., black square) indicates that a large portion of the residuals are not accounted for in the model. TF = *Trichaptum fuscoviolaceum*, TA = *T. abietinum*, Can = Canadian, It = Italian, NAmA = North American A and NAmB = North American B. *T. biforme* is the outgroup.
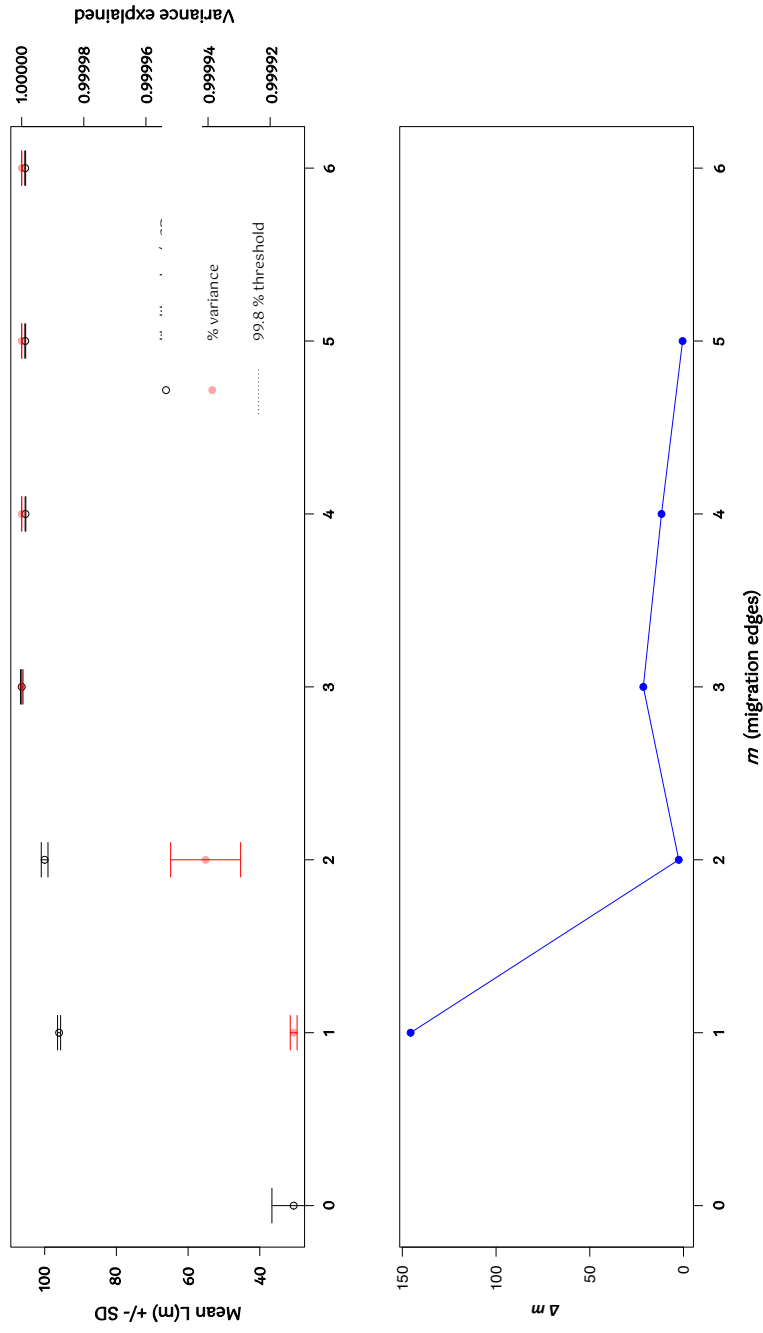
Figure S9. The *optM* (Fitak 2021) analysis indicates that the *TreeMix* model with 1 edge is the most optimal. The upper panel shows the mean and standard deviation for the composite likelihood on the left y-axis and the proportion of variance explained on the right y-axis. The bottom panel depicts the second-order rate of change (Δm) across values of m on the y-axis. The x-axis in both panels indicates the number of migration edges. The peak at edge 1 in the bottom panel is considered to represent the most optimal edge number.
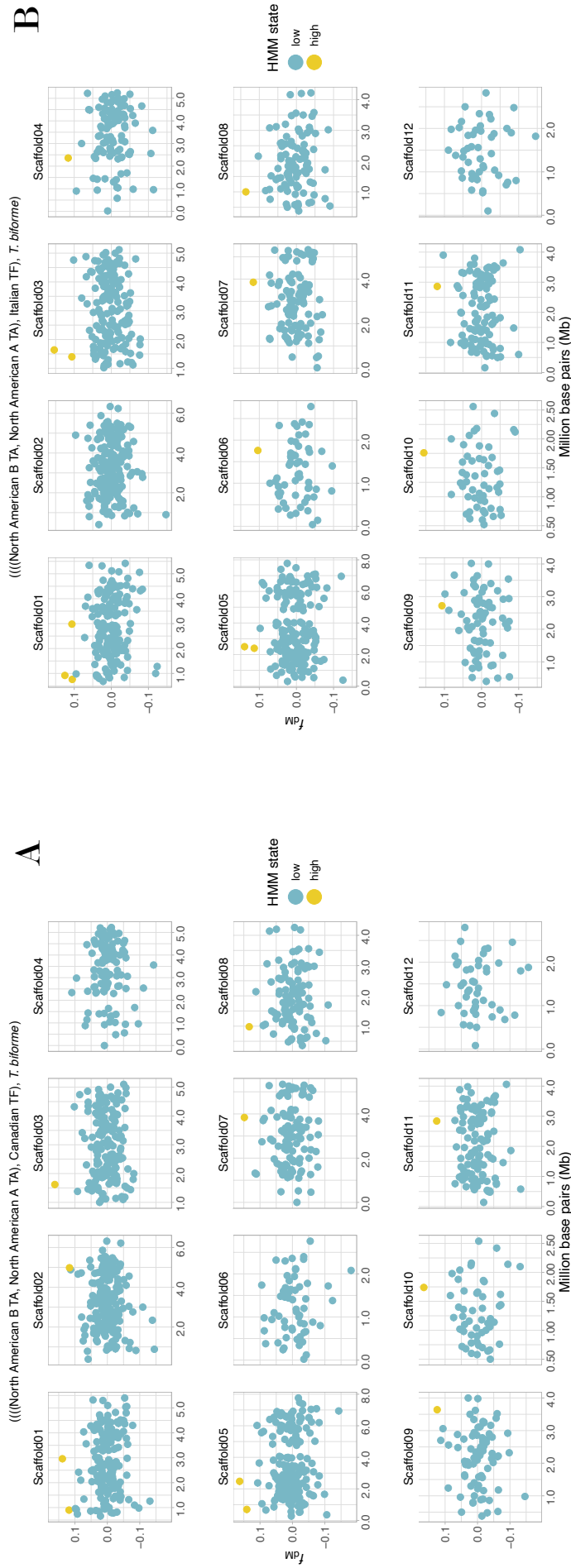
Figure S10. Signs of scattered introgression throughout the genome. A proportion of introgression (f_{dM}) sliding window analysis based on a single nucleotide polymorphism (SNP) dataset of 3 118 957 SNPs, where windows with at least 100 SNPs are included. The main headers depict the phylogenetic hypothesis, ((((P1, P2), P3), O), where P1, P2 and P3 are populations investigated for introgression and O is the outgroup. A positive value indicates more shared derived polymorphisms than expected between P2 and P3, while a negative value indicates the same for P1 and P3. Each point is the f_{dM} value of a window (window size = 20 000 base pairs). Y-axes show the f_{dM} value and x-axes represent million base pair (Mb) position of the windows on the scaffolds. The legend shows the Hidden Markov-model (HMM) state of the windows. Blue colored points (low) indicate insignificant amount of introgression, while yellow-colored points (high) are outlier windows with significant introgression from the HMM analysis. Annotated genes in the outlier windows can be found in Table S3. TA = *Trichaptum abietinum* and TF = *T. fuscoviolaceum*. The figure is made in *R v4.0.2* using the packages *tidyverse* (Wickham et al. 2019) and *wesanderson* (Ram and Wickham, 2018).

10
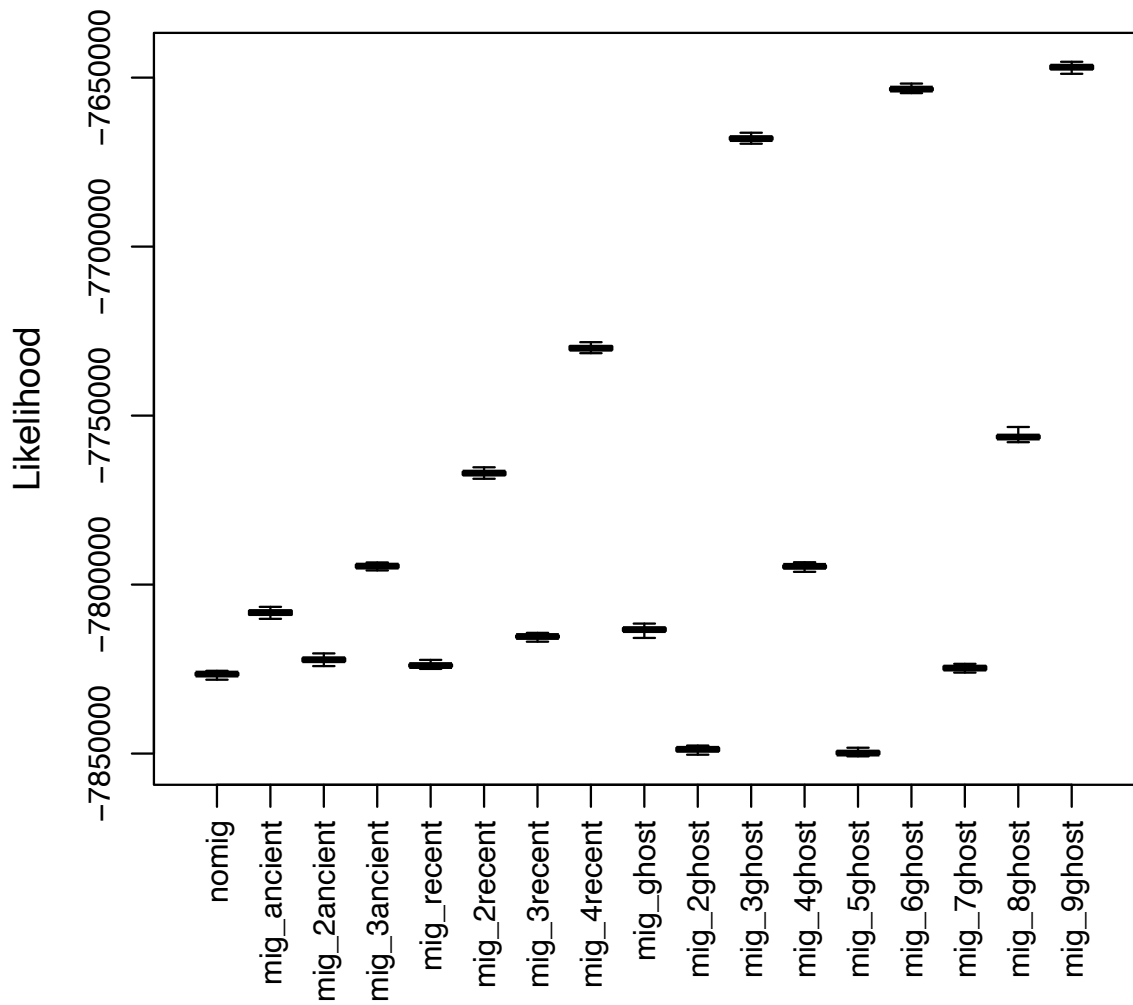
Figure S11. The likelihood distributions support the best model based on AIC. Likelihood distributions plotted for the different models tested in *fastsimcoal2* (Excoffier et al. 2021). The best model (mig_9ghost) does not have an overlapping distribution with the second best model (mig_6ghost). The distributions are plotted as boxplots with the likelihood on the y-axis and the name of the models on the x-axis. The model with the best likelihood (mig_9ghost) is also supported by AIC. See Figure S6 and Table S5 for further details on the models.

Table S1. Overview of individuals collected and used for bioinformatic analyses. The table includes information on name (collection ID), species designation (based on morphology, ITS sequence and Illumina sequence), collection site (area, longitude, latitude, and elevation), host substrate (substrate), name of collector and date of collection. TF = *Trichaptum fuscoviolaceum*, TA = *T. abietinum* and the outgroup TB = *T. biforme*.

Collection ID	Species morphology	Species ITS	Species Illumina	Area	Latitude	Longitude	Elevation	Substrate	Collector	Collection date
TF-1000-1-M1	*T. fuscoviolaceum*	*T. fuscoviolaceum*	*T. fuscoviolaceum*	CA, N.B., Charlotte County	45.13741 N	66.46785 W	120 m	*Abies balsamea*	David Malloch	09.10.2018
TF-1002-2-M1	*T. fuscoviolaceum*	*T. abietinum*	*T. abietinum*	CA, N.B., Charlotte County	45.129722 N	66.523889 W	50 m	*Abies balsamea*	Inger Skrede & Dabao Lu	09.10.2018
TF-1002-3-M1	*T. fuscoviolaceum*	*T. fuscoviolaceum*	*T. fuscoviolaceum*	CA, N.B., Charlotte County	45.126944 N	66.526944 W	50 m	*Abies balsamea*	Inger Skrede & Dabao Lu	09.10.2018
TF-1002-4-M3	*T. fuscoviolaceum*	*T. fuscoviolaceum*	*T. fuscoviolaceum*	CA, N.B., Charlotte County	45.128611 N	66.525000 W	50 m	*Abies balsamea*	Inger Skrede & Dabao Lu	09.10.2018
TF-1002-5-M3	*T. fuscoviolaceum*	*T. fuscoviolaceum*	*T. fuscoviolaceum*	CA, N.B., Charlotte County	45.126944 N	66.527222 W	50 m	*Abies balsamea*	Inger Skrede & Dabao Lu	09.10.2018
TF-1002-7-M2	*T. fuscoviolaceum*	*T. fuscoviolaceum*	*T. fuscoviolaceum*	CA, N.B., Charlotte County	45.130833 N	66.525000 W	50 m	*Abies balsamea*	Inger Skrede & Dabao Lu	09.10.2018
TF-1002-10-M3	*T. fuscoviolaceum*	*T. fuscoviolaceum*	*T. fuscoviolaceum*	CA, N.B., Charlotte County	45.12927 N	66.52469 W	50 m	*Abies balsamea*	Amanda Bremner	09.10.2018
TF-1003-2-M1	*T. fuscoviolaceum*	*T. fuscoviolaceum*	*T. fuscoviolaceum*	CA, N.B., Charlotte County	45.169444 N	66.459167 W	10 m	*Picea rubens*	Inger Skrede & Dabao Lu	09.10.2018
TF-1003-3-M2	*T. fuscoviolaceum*	*T. fuscoviolaceum*	*T. fuscoviolaceum*	CA, N.B., Charlotte County	45.169167 N	66.458889 W	10 m	*Picea rubens*	Inger Skrede & Dabao Lu	09.10.2018

Sample ID	Species 1	Species 2	Species 3	Location	Latitude	Longitude	Elevation	Host	Collector	Date
TF-1003-4-M2	T. fuscoviolaceum	T. fuscoviolaceum	T. fuscoviolaceum	CA, N.B., Charlotte County	45.173611 N	66.465556 W	10 m	Abies balsamea	Inger Skrede & Dabao Lu	09.10.2018
TF-1003-5-M2	T. fuscoviolaceum	T. fuscoviolaceum	T. fuscoviolaceum	CA, N.B., Charlotte County	45.168611 N	66.461389 W	10 m	Picea rubens	Inger Skrede & Dabao Lu	09.10.2018
TF-1004-1-M1	T. fuscoviolaceum	T. fuscoviolaceum	T. fuscoviolaceum	CA, N.B., Sunbury County	45.993333 N	66.307500 W	80 m	Abies balsamea	Inger Skrede & Dabao Lu	10.10.2018
TF-1004-3-M1	T. fuscoviolaceum	T. fuscoviolaceum	T. fuscoviolaceum	CA, N.B., Sunbury County	45.992222 N	66.307222 W	80 m	Abies balsamea	Inger Skrede & Dabao Lu	10.10.2018
TF-1004-4-M2	T. fuscoviolaceum	T. fuscoviolaceum	T. fuscoviolaceum	CA, N.B., Sunbury County	45.992222 N	66.307222 W	80 m	Abies balsamea	Inger Skrede & Dabao Lu	10.10.2018
TF-1004-6-M2	T. fuscoviolaceum	T. fuscoviolaceum	T. fuscoviolaceum	CA, N.B., Sunbury County	46.00562 N	66.40087 W	130 m	Abies balsamea	Stephen R. Clayden	10.10.2018
TF-1004-6-dup-M2	T. fuscoviolaceum	T. fuscoviolaceum	T. fuscoviolaceum	CA, N.B., Sunbury County	45.991389 N	66.307500 W	80 m	Abies balsamea	Inger Skrede & Dabao Lu	10.10.2018
TF-1005-1-M1	T. fuscoviolaceum	T. abietinum	T. abietinum	CA, N.B., Sunbury County	46.035000 N	66.325000 W	130 m	Picea mariana	Inger Skrede & Dabao Lu	10.10.2018
TF-1007-1-M3	T. fuscoviolaceum	T. fuscoviolaceum	T. fuscoviolaceum	CA, N.B., Gloucester County	47.636944 N	65.610833 W	20 m	Abies balsamea	Inger Skrede & Dabao Lu	11.10.2018
TF-1007-2-M2	T. fuscoviolaceum	T. fuscoviolaceum	T. fuscoviolaceum	CA, N.B., Gloucester County	47.636944 N	65.610833 W	20 m	Abies balsamea	Inger Skrede & Dabao Lu	11.10.2018
TF-1009-1-M1	T. fuscoviolaceum	T. fuscoviolaceum	T. fuscoviolaceum	CA, N.B., Restigouche County	47.41611 N	66.868056 W	380 m	Abies balsamea	Inger Skrede & Dabao Lu	12.10.2018
TF-1009-2-M2	T. fuscoviolaceum	T. fuscoviolaceum	T. fuscoviolaceum	CA, N.B., Restigouche County	47.416111 N	66.866944 W	360 m	Abies balsamea	Inger Skrede & Dabao Lu	12.10.2018

ID			Location	Latitude	Longitude	Elevation	Host	Collector	Date
TF-1009-3-M3	*T. fuscoviolaceum*	*T. abietinum*	CA, N.B., Restigouche County	47.418333	66.868889 W	240 m	*Abies balsamea*	Inger Skrede & Dabao Lu	12.10.2018
TF-1009-4-M2	*T. fuscoviolaceum*	*T. fuscoviolaceum*	CA, N.B., Restigouche County	47.418056 N	66.867778 W	290 m	*Abies balsamea*	Inger Skrede & Dabao Lu	12.10.2018
TF-1011-1-M1	*T. fuscoviolaceum*	*T. fuscoviolaceum*	CA, N.B., Victoria County	46.894167 N	67.398333 W	170 m	*Abies balsamea*	Inger Skrede & Dabao Lu	13.10.2018
TF-1011-3-M1	*T. fuscoviolaceum*	*T. fuscoviolaceum*	CA, N.B., Victoria County	46.893333 N	67.398333 W	170 m	*Abies balsamea*	Inger Skrede & Dabao Lu	13.10.2018
TF-1011-4-M1	*T. fuscoviolaceum*	*T. fuscoviolaceum*	CA, N.B., Victoria County	46.892778 N	67.398611 W	170 m	*Abies balsamea*	Inger Skrede & Dabao Lu	13.10.2018
TF-1011-7-M2	*T. fuscoviolaceum*	*T. fuscoviolaceum*	CA, N.B., Victoria County	46.892500 N	67.399167 W	170 m	*Picea sp.*	Inger Skrede & Dabao Lu	13.10.2018
TF-1011-8-M1	*T. fuscoviolaceum*	*T. fuscoviolaceum*	CA, N.B., Victoria County	46.902222 N	67.400556 W	170 m	*Abies balsamea*	Inger Skrede & Dabao Lu	13.10.2018
TF-1012-1-M2	*T. fuscoviolaceum*	*T. fuscoviolaceum*	CA, N.B., Northumberland County	46.783333 N	66.516667 W	390 m	*Abies balsamea*	Inger Skrede & Dabao Lu	13.10.2018
TF-1012-2-M1	*T. fuscoviolaceum*	*T. fuscoviolaceum*	CA, N.B., Northumberland County	46.783889 N	66.525556 W	390 m	*Abies balsamea*	Inger Skrede & Dabao Lu	13.10.2018
TF-1012-3-M3	*T. fuscoviolaceum*	*T. fuscoviolaceum*	CA, N.B., Northumberland County	46.784444 N	66.526111 W	400 m	*Abies balsamea*	Inger Skrede & Dabao Lu	13.10.2018
TF-1012-5-M2	*T. fuscoviolaceum*	*T. fuscoviolaceum*	CA, N.B., Northumberland County	46.785556 N	60.525278 W	380 m	*Abies balsamea*	Inger Skrede & Dabao Lu	13.10.2018
TF-1013-1-M2	*T. fuscoviolaceum*	*T. fuscoviolaceum*	CA, N.B., York County	45.9564 N	66.6668 W	50 m	*Abies balsamea*	Stephen R. Clayden	12.10.2018

14

ID				Location	Latitude	Longitude	Elevation	Host	Collector	Date
TF-1013-2-M2	*T. fuscoviolaceum*	*T. fuscoviolaceum*	*T. fuscoviolaceum*	CA, N.B., York County	45.9564 N	66.6668 W	50 m	*Abies balsamea*	Stephen R. Clayden	12.10.2018
TF-1013-4-M2	*T. fuscoviolaceum*	*T. fuscoviolaceum*	*T. abietinum*	CA, N.B., York County	45.9564 N	66.6668 W	50 m	*Abies balsamea*	Stephen R. Clayden	12.10.2018
TF-1013-5-M2	*T. fuscoviolaceum*	*T. fuscoviolaceum*	*T. fuscoviolaceum*	CA, N.B., York County	45.9564 N	66.6668 W	50 m	*Abies balsamea*	Stephen R. Clayden	12.10.2018
TF-1013-8-M2	*T. fuscoviolaceum*	*T. fuscoviolaceum*	*T. fuscoviolaceum*	CA, N.B., York County	45.9564 N	66.6668 W	50 m	*Abies balsamea*	Stephen R. Clayden	12.10.2018
TF-1014-1-M2	*T. fuscoviolaceum*	*T. fuscoviolaceum*	*T. fuscoviolaceum*	ITL, Pavia, Menconico	44.80662 N	9.31246 E	–	*Pinus nigra*	Carolina Girometta	06.10.2018
TF-1014-2-M1	*T. fuscoviolaceum*	*T. fuscoviolaceum*	*T. fuscoviolaceum*	ITL, Pavia, Menconico	44.80744 N	9.31063 E	–	*Pinus nigra*	Carolina Girometta	06.10.2018
TF-1014-3-M1	*T. fuscoviolaceum*	*T. fuscoviolaceum*	*T. fuscoviolaceum*	ITL, Pavia, Menconico	44.80754 N	9.31042 E	–	*Pinus nigra*	Carolina Girometta	06.10.2018
TF-1014-6-M3	*T. fuscoviolaceum*	*T. fuscoviolaceum*	*T. fuscoviolaceum*	ITL, Pavia, Menconico	44.81126 N	9.30588 E	–	*Pinus nigra*	Carolina Girometta	06.10.2018
TF-1014-7-M1	*T. fuscoviolaceum*	*T. fuscoviolaceum*	*T. fuscoviolaceum*	ITL, Pavia, Menconico	44.81134 N	9.30589 E	–	*Pinus nigra*	Carolina Girometta	06.10.2018
TF-1014-7-M9	*T. fuscoviolaceum*	*T. fuscoviolaceum*	*T. fuscoviolaceum*	ITL, Pavia, Menconico	44.81134 N	9.30589 E	–	*Pinus nigra*	Carolina Girometta	06.10.2018
TF-1014-8-M2	*T. fuscoviolaceum*	*T. fuscoviolaceum*	*T. fuscoviolaceum*	ITL, Pavia, Menconico	44.81137 N	9.30661 E	–	*Pinus nigra*	Carolina Girometta	06.10.2018
TF-1014-9-M3	*T. fuscoviolaceum*	*T. fuscoviolaceum*	*T. fuscoviolaceum*	ITL, Pavia, Menconico	44.80795 N	9.30954 E	–	*Pinus nigra*	Carolina Girometta	06.10.2018
TF-1014-10-M1	*T. fuscoviolaceum*	*T. fuscoviolaceum*	*T. fuscoviolaceum*	ITL, Pavia, Menconico	44.80520 N	9.31298 E	–	*Pinus nigra*	Carolina Girometta	06.10.2018
TA-1002-8-M1	*T. abietinum*	*T. abietinum*	*T. abietinum*	CA, N.B., Charlotte County	45.129444 N	66.524167 W	50 m	*Abies balsamea*	Inger Skrede & Dabao Lu	09.10.2018

15

ID			Location	Latitude	Longitude	Elevation	Host	Collector	Date
TA-1002-13-M2	*T. abietinum*	*T. abietinum*	CA, N.B., Charlotte County	45.128611 N	66.524444 W	50 m	*Picea rubens*	Inger Skrede & Dabao Lu	09.10.2018
TA-1002-19-M1	*T. abietinum*	*T. abietinum*	CA, N.B., Charlotte County	45.128611 N	66.524722 W	50 m	*Abies balsamea*	Inger Skrede & Dabao Lu	09.10.2018
TA-1002-27-M1	*T. abietinum*	*T. abietinum*	CA, N.B., Charlotte County	45.128333 N	66.526111 W	50 m	*Picea rubens*	Inger Skrede & Dabao Lu	09.10.2018
TA-1002-35-M2	*T. abietinum*	*T. abietinum*	CA, N.B., Charlotte County	45.128611 N	66.526389 W	50 m	*Picea rubens*	Inger Skrede & Dabao Lu	09.10.2018
TA-1003-1-M1	*T. abietinum*	*T. abietinum*	CA, N.B., Sunbury County	45.991944 N	66.306944 W	60 m	*Abies balsamea*	Inger Skrede & Dabao Lu	10.10.2018
TA-1003-8-M1	*T. abietinum*	*T. abietinum*	CA, N.B., Sunbury County	45.992778 N	66.306944 W	60 m	*Abies balsamea*	Inger Skrede & Dabao Lu	10.10.2018
TA-1003-17-M1	*T. abietinum*	*T. abietinum*	CA, N.B., Sunbury County	45.990278 N	66.306667 W	60 m	*Abies balsamea*	Inger Skrede & Dabao Lu	10.10.2018
TA-1003-20-M2	*T. abietinum*	*T. abietinum*	CA, N.B., Sunbury County	45.990278 N	66.307222 W	60 m	*Picea rubens*	Inger Skrede & Dabao Lu	10.10.2018
TA-1003-22-M2	*T. abietinum*	*T. abietinum*	CA, N.B., Sunbury County	45.99056 N	66.30690 W	60 m	*Picea rubens*	Stepen R. Clayden	10.10.2018
TA-1007-1	*T. abietinum*	*T. abietinum*	CA, N.B., Gloucester County	47.637500 N	65.610000 W	20 m	*Pinus cf. strobus*	Inger Skrede & Dabao Lu	11.10.2018
TA-1007-3	*T. abietinum*	*T. abietinum*	CA, N.B., Gloucester County	47.631389 N	65.616111 W	20 m	*Picea cf. glauca*	Inger Skrede & Dabao Lu	11.10.2018
TA-1007-5	*T. abietinum*	*T. abietinum*	CA, N.B., Gloucester County	47.637500 N	65.610556 W	30 m	*Picea cf. glauca*	Inger Skrede & Dabao Lu	11.10.2018

Sample ID			Location	Latitude	Longitude	Elevation	Host	Collector	Date
TA-1007-6	*T. abietinum*	*T. abietinum*	CA, N.B., Gloucester County	47.649722 N	65.610833 W	30 m	*Picea cf. glauca*	Inger Skrede & Dabao Lu	11.10.2018
TA-1007-17	*T. abietinum*	*T. abietinum*	CA, N.B., Gloucester County	47.603333 N	65.610833 W	30 m	*Abies balsamea*	Inger Skrede & Dabao Lu	11.10.2018
TA-1009-1-M3	*T. abietinum*	*T. abietinum*	CA, N.B., Restigouche County	47.418056 N	66.866667 W	240 m	*Picea cf. glauca*	Inger Skrede & Dabao Lu	12.10.2018
TA-1009-4-M1	*T. abietinum*	*T. abietinum*	CA, N.B., Restigouche County	47.417778 N	66.866944 W	330 m	*Picea cf. glauca*	Inger Skrede & Dabao Lu	12.10.2018
TA-1009-8-M3	*T. abietinum*	*T. abietinum*	CA, N.B., Restigouche County	47.416944 N	66.867222 W	330 m	*Picea cf. glauca*	Inger Skrede & Dabao Lu	12.10.2018
TA-1009-12-M1	*T. abietinum*	*T. abietinum*	CA, N.B., Restigouche County	47.417778 N	66.866111 W	340 m	*Picea sp.*	Inger Skrede & Dabao Lu	12.10.2018
TA-1009-19-M3	*T. abietinum*	*T. abietinum*	CA, N.B., Restigouche County	47.418611 N	66.881389 W	220 m	*Picea sp*	Inger Skrede & Dabao Lu	12.10.2018
TA-1011-12-M2	*T. abietinum*	*T. abietinum*	CA, N.B., Victoria County	46.894167 N	67.398333 W	170 m	*Picea cf. rubens*	Inger Skrede & Dabao Lu	13.10.2018
TA-1011-19-M1	*T. abietinum*	*T. abietinum*	CA, N.B., Victoria County	46.892500 N	67.399444 W	170 m	*Picea cf. rubens*	Inger Skrede & Dabao Lu	13.10.2018
TA-1011-23-M1	*T. abietinum*	*T. abietinum*	CA, N.B., Victoria County	46.900833 N	67.400278 W	170 m	*Picea sp.*	Inger Skrede & Dabao Lu	13.10.2018
TA-1011-26-M1	*T. abietinum*	*T. abietinum*	CA, N.B., Victoria County	46.893611 N	67.400000 W	170 m	*Abies balsamea*	Inger Skrede & Dabao Lu	13.10.2018
TA-1011-31-M1	*T. abietinum*	*T. abietinum*	CA, N.B., Victoria County	46.893333 N	67.401111 W	170 m	*Abies balsamea*	Inger Skrede & Dabao Lu	13.10.2018

ID			Location	Latitude	Longitude	Elevation	Host	Collector	Date
TA-1012-3-M1		*T. abietinum*	*T. abietinum*	CA, N.B., Northumberland County	46.783889 N	66.525556 W	390 m	*Picea rubens*	Inger Skrede & Dabao Lu	13.10.2018
TA-1012-5-M1		*T. abietinum*	*T. abietinum*	CA, N.B., Northumberland County	46.784167 N	66.059444 W	390 m	*Abies balsamea*	Inger Skrede & Dabao Lu	13.10.2018
TA-1012-7-M1		*T. abietinum*	*T. abietinum*	CA, N.B., Northumberland County	46.784167 N	66.059444 W	390 m	*Picea rubens*	Inger Skrede & Dabao Lu	13.10.2018
TA-1012-11-M1		*T. abietinum*	*T. abietinum*	CA, N.B., Northumberland County	46.785556 N	66.525556 W	390 m	*Picea rubens*	Inger Skrede & Dabao Lu	13.10.2018
TA-1012-17-M1		*T. abietinum*	*T. abietinum*	CA, N.B., Northumberland County	46.784722 N	66.524722 W	380 m	*Picea sp.*	Inger Skrede & Dabao Lu	13.10.2018
TA-1013-3-M2		*T. abietinum*	*T. abietinum*	CA, N.B., York County	45.9564 N	66.6668 W	50 m	*Abies balsamea*	Stephen R. Clayden	12.10.2018
TA-1013-4-M1		*T. abietinum*	*T. abietinum*	CA, N.B., York County	45.9564 N	66.6668 W	50 m	*Tsuga canadensis*	Stephen R. Clayden	12.10.2018
TA-1013-5-M1		*T. abietinum*	*T. abietinum*	CA, N.B., York County	45.9564 N	66.6668 W	50 m	*Tsuga canadensis*	Stephen R. Clayden	12.10.2018
TA-1013-7-M1		*T. abietinum*	*T. abietinum*	CA, N.B., York County	45.9564 N	66.6668 W	50 m	*Picea rubens*	Stephen R. Clayden	12.10.2018
TA-1013-9-M1		*T. abietinum*	*T. abietinum*	CA, N.B., York County	45.9564 N	66.6668 W	50 m	*Picea rubens*	Stephen R. Clayden	12.10.2018
TB-1013-1-M2	*T. fuscoviolaceum*	*T. biforme*	*T. biforme*	CA, N. B., York County	45.9564 N	66.6668 W	50 m	*Abies balsamea*	Stephen R. Clayden	12.10.2018

18

Table S2, *Trichaptum fuscoviolaceum* individuals mated as predicted, while *T. abietinum* crossed with T. fuscoviolaceum individuals did not. The table includes cross name (Cross ID), monokaryotic individuals crossed (Mate pairs), mating loci differences and similarities between the crosses (Mating type (MAT)), populations crossed (Populations), expected outcome (Prediction) and actual outcome by observation (Yes) or no observation (No) of clamp connections (Clamp).

T. abietinum × *T. fuscoviolaceum*					
Cross ID	**Mate pairs**	**Mating type (*MAT*)**	**Populations**	**Prediction**	**Clamp**
TFTAX1	TF10147M9 × TA10264M3	Ident. *MATA*, dist. *MATB*	It × Eu	Incompatible	No
TFTAX2	TF10147M9 × TA10058M1	Ident. *MATA*, dist. *MATB*	It × NAmA	Incompatible	No
TFTAX3	TF101410M1 × TA10355M3	Dist. *MATA*, ident. *MATB*	It × Eu	Incompatible	No
TFTAX4	TF10141M2 × TA10139M1	Dist. *MATA*, ident. *MATB*	It × NAmB	Incompatible	No
TFTAX5	TF101410M1 × TA10264M3	Dist. *MAT*s	It × Eu	Compatible	No
TFTAX6	TF101410M1 × TA10139M1	Dist. *MAT*s	It × NAmB	Compatible	No
TFTAX7	TF101410M1 × TA10058M1	Dist. *MAT*s	It × NAmA	Compatible	No
TFTAX8	TF10034M2 × TA10139M1	Dist. *MAT*s	Can × NAmB	Compatible	No
TFTAX9	TF10034M2 × TA10264M3	Dist. *MAT*s	Can × Eu	Compatible	No
TFTAX11	TF10034M2 × TA10058M1	Dist. *MAT*s	Can × NAmA	Compatible	No

T. fuscoviolaceum × *T. fuscoviolaceum*					
Cross ID	**Mate pairs**	**Mating type (*MAT*)**	**Populations**	**Prediction**	**Clamp**
TFX1	TF10147M1 × TF10147M1	Ident. *MAT*	It × It	Incompatible	No
TFX2	TF10147M1 × TF10147M9	Dist. α *MATA* and *MATB*	It × It	Compatible	Yes
TFX3	TF10147M1 × TF10143M3	Dist. α *MATA,* ident. *MATB*	It × It	Incompatible	No
TFX4	TF10147M9 × TF10141M2	Dist. β *MATA* and *MATB*	It × It	Compatible	Yes
TFX5	TF10032M1 × TF10135M2	Dist. β *MATA* and *MATB*	Can × Can	Compatible	Yes
TFX7	TF10091M1 × TF10135M2	Dist. α *MAT* and *MATB*	Can × Can	Compatible	Yes
TFX9	TF10034M2 × TF10091M1	Dist. *MATA*, ident. *MATB*	Can × Can	Incompatible	No
TFX10	TF10122M1 × TF10147M1	Dist. β *MATA* and *MATB*	It × Can	Compatible	Yes
TFX11	TF101410M1 × TF10122M1	Dist. α *MATA* and *MATB*	It × Can	Compatible	Yes
TFX12	TF10141M2 × TF10122M1	Dist. *MATA*, ident. *MATB*	It × Can	Incompatible	No

Dist. = distinct, ident. = identical, TA = *Trichaptum abietinum*, TF = *T. fuscoviolaceum*, It = Italian, Can = Canadian, Eu = European, NAmA = North American A, NAmB = North American B

Table S3. An overview of the Hidden Markov-model outliers from the f_{dM} genome scan. The headers indicate the phylogenetic typology of the test; (((P1, P2), P3), O), where P1, P2 and P3 are populations investigated for introgression and O is the outgroup. A positive f_{dM} value indicates more shared derived polymorphisms than expected between P2 and P3, while a negative value indicates more shared derived polymorphisms than expected between P1 and P3. The table also denotes which scaffold the genes are in, including the start and end of the gene on that scaffold, how many sites that were used to estimate the f_{dM} value in the specific window (Sites used), the name of the genes from the annotated genome (Gene), and a note on the function of the genes. TA = *Trichaptum abietinum* and TF = *T. fuscoviolaceum*.

					(((North American B TA, North American A TA), Canadian TF), *T. biforme*)	
Scaffold	**Start**	**End**	**Sites used**	**f_{dM}**	**Gene**	**Note**
Scaffold01	745385	745484	122	0.1053	trnascan-Scaffold01-noncoding-Thr_TGT-gene-7.19	Protein of unknown function
Scaffold01	745617	746235	122	0.1053	maker-Scaffold01-snap-gene-7.9	Protein of unknown function
Scaffold01	746860	752545	122	0.1053	snap_masked-Scaffold01-processed-gene-7.3	Similar to ATP1A1: Sodium/potassium-transporting ATPase subunit alpha-1 (Equus caballus OX=9796)
Scaffold01	752671	755023	122	0.1053	maker-Scaffold01-snap-gene-7.12	Protein of unknown function
Scaffold01	900331	900881	133	0.1251	maker-Scaffold01-snap-gene-9.2	Protein of unknown function
Scaffold01	901468	902851	133	0.1251	snap_masked-Scaffold01-processed-gene-9.31	Protein of unknown function
Scaffold01	902938	905385	133	0.1251	snap_masked-Scaffold01-processed-gene-9.39	Protein of unknown function
Scaffold01	906260	907088	133	0.1251	maker-Scaffold01-exonerate_protein2genome-gene-9.29	Protein of unknown function
Scaffold01	909092	910780	133	0.1251	maker-Scaffold01-snap-gene-9.4	Protein of unknown function
Scaffold01	912091	912422	133	0.1251	snap_masked-Scaffold01-processed-gene-9.40	Similar to EMC4: ER membrane protein complex subunit 4 (Saccharomyces cerevisiae (strain ATCC 204508 / S288c) OX=559292)

20

Scaffold	Start	End	Length	Value	Gene ID	Description
Scaffold01	913710	915599	133	0.1251	maker-Scaffold01-snap-gene-9.5	Similar to COX17: Cytochrome c oxidase copper chaperone (Homo sapiens OX=9606)
Scaffold01	915247	917294	133	0.1251	maker-Scaffold01-snap-gene-9.16	Similar to GST: Glutathione S-transferase (Plasmodium vivax OX=5855)
Scaffold01	918913	919817	133	0.1251	snap_masked-Scaffold01-abinit-gene-9.25	Protein of unknown function
Scaffold01	2975369	2975485	137	0.1062	maker-Scaffold01-exonerate_protein2genome-gene-29.152	Protein of unknown function
Scaffold03	1384980	1386656	110	0.1063	genemark-Scaffold03-processed-gene-13.17	Similar to NEP1: Ribosomal RNA small subunit methyltransferase NEP1 (Candida albicans OX=5476)
Scaffold03	1387398	1388969	110	0.1063	maker-Scaffold03-snap-gene-13.15	Similar to fmdA: Formamidase (Methylophilus methylotrophus OX=17)
Scaffold03	1391654	1393548	110	0.1063	maker-Scaffold03-snap-gene-14.40	Protein of unknown function
Scaffold03	1620126	1622452	106	0.1538	genemark-Scaffold03-processed-gene-16.4	Similar to GRC3: Polynucleotide 5'-hydroxyl-kinase GRC3 (Cryptococcus neoformans var. neoformans serotype D (strain JEC21 / ATCC MYA-565) OX=214684)
Scaffold03	1624233	1626184	106	0.1538	maker-Scaffold03-snap-gene-16.42	Similar to DAL1: Allantoinase (Saccharomyces cerevisiae (strain ATCC 204508 / S288c) OX=559292)
Scaffold03	1626282	1628466	106	0.1538	maker-Scaffold03-snap-gene-16.43	Protein of unknown function
Scaffold03	1628643	1630494	106	0.1538	maker-Scaffold03-snap-gene-16.53	Protein of unknown function
Scaffold03	1631588	1632628	106	0.1538	maker-Scaffold03-snap-gene-16.54	Protein of unknown function
Scaffold03	1632377	1632573	106	0.1538	maker-Scaffold03-exonerate_est2genome-gene-16.2	Protein of unknown function
Scaffold03	1632645	1632758	106	0.1538	maker-Scaffold03-exonerate_protein2genome-gene-16.47	Protein of unknown function

Scaffold	Start	End	Number	Value	Gene ID	Description
Scaffold03	1633221	1636431	106	0.1538	maker-Scaffold03-exonerate_protein2genome-gene-16.48	Similar to RDR1: Probable RNA-dependent RNA polymerase 1 (Oryza sativa subsp. japonica OX=39947)
Scaffold03	1637269	1638202	106	0.1538	maker-Scaffold03-exonerate_protein2genome-gene-16.56	Similar to SEC14: SEC14 cytosolic factor (Saccharomyces cerevisiae (strain ATCC 204508 / S288c) OX=559292)
Scaffold05	2389277	2393966	111	0.1137	maker-Scaffold05-snap-gene-24.8	Protein of unknown function
Scaffold05	2480001	2500000	125	0.1398	No annotated genes	No annotated genes
Scaffold06	1741847	1744120	353	0.104	maker-Scaffold06-snap-gene-17.52	Similar to obg: GTPase Obg (Rippkaea orientalis (strain PCC 8801) OX=41431)
Scaffold06	1745220	1746591	353	0.104	snap_masked-Scaffold06-processed-gene-17.7	Similar to cell: Cellulose-growth-specific protein (Agaricus bisporus OX=5341)
Scaffold06	1747150	1751861	353	0.104	maker-Scaffold06-snap-gene-17.53	Similar to fgenesh1_kg.2_#_1379_#_Locus12621v1rpkm4.09: 4-O-methyltransferase 1 (Phanerochaete chrysosporium (strain RP-78 / ATCC MYA-4764 / FGSC 9002) OX=273507)
Scaffold06	1752774	1753281	353	0.104	maker-Scaffold06-exonerate_protein2genome-gene-17.176	Protein of unknown function
Scaffold06	1753385	1755997	353	0.104	maker-Scaffold06-exonerate_protein2genome-gene-17.72	Protein of unknown function
Scaffold06	1756135	1759294	353	0.104	snap_masked-Scaffold06-processed-gene-17.23	Similar to SPBC530.05: Uncharacterized transcriptional regulatory protein C530.05 (Schizosaccharomyces pombe (strain 972 / ATCC 24843) OX=284812)
Scaffold06	1759332	1759612	353	0.104	snap_masked-Scaffold06-processed-gene-17.24	Protein of unknown function

Scaffold	Start	End	Count	Value	Gene ID	Description
Scaffold07	3840869	3842531	163	0.1162	maker-Scaffold07-snap-gene-38.79	Similar to ZFAND2A: AN1-type zinc finger protein 2A (Pongo abelii OX=9601)
Scaffold07	3843021	3844141	163	0.1162	snap_masked-Scaffold07-abinit-gene-38.26	Protein of unknown function
Scaffold07	3845137	3848851	163	0.1162	maker-Scaffold07-snap-gene-38.49	Similar to PHB2: Prohibitin-2 (Saccharomyces cerevisiae (strain ATCC 204508 / S288c) OX=559292)
Scaffold07	3848870	3850190	163	0.1162	maker-Scaffold07-exonerate_protein2genome-gene-38.89	Protein of unknown function
Scaffold07	3850825	3853362	163	0.1162	maker-Scaffold07-snap-gene-38.40	Similar to exosc3: Putative exosome complex component rrp40 (Dictyostelium discoideum OX=44689)
Scaffold07	3853955	3856502	163	0.1162	maker-Scaffold07-snap-gene-38.51	Protein of unknown function
Scaffold07	3857642	3858990	163	0.1162	maker-Scaffold07-snap-gene-38.52	Protein of unknown function
Scaffold08	987415	988459	107	0.1363	maker-Scaffold08-snap-gene-9.25	Protein of unknown function
Scaffold08	990031	990293	107	0.1363	maker-Scaffold08-exonerate_protein2genome-gene-10.96	Protein of unknown function
Scaffold08	990859	996481	107	0.1363	maker-Scaffold08-snap-gene-10.13	Protein of unknown function
Scaffold08	996522	997428	107	0.1363	maker-Scaffold08-snap-gene-10.18	Protein of unknown function
Scaffold08	997701	997814	107	0.1363	trnascan-Scaffold08-noncoding-Gly_GCC-gene-10.66	Protein of unknown function
Scaffold08	998636	999746	107	0.1363	maker-Scaffold08-exonerate_protein2genome-gene-10.16	Similar to ECI3: Enoyl-CoA delta isomerase 3 (Arabidopsis thaliana OX=3702)
Scaffold09	2701803	2709281	146	0.1067	maker-Scaffold09-augustus-gene-27.6	Similar to KES1: Protein KES1 (Ustilago maydis (strain 521 / FGSC 9021) OX=237631)
Scaffold09	2709464	2710425	146	0.1067	maker-Scaffold09-augustus-gene-27.1	Protein of unknown function
Scaffold09	2711618	2713971	146	0.1067	maker-Scaffold09-snap-gene-27.54	Similar to can: Carbonic anhydrase 2 (Shigella flexneri OX=623)

Scaffold	Start	End	Sites used	f_{dM}	Gene	Note
Scaffold10	1740426	1741739	222	0.1555	snap_masked-Scaffold10-abinit-gene-17.14	Protein of unknown function
Scaffold10	1742972	1744147	222	0.1555	snap_masked-Scaffold10-processed-gene-17.14	Protein of unknown function
Scaffold10	1744941	1746814	222	0.1555	snap_masked-Scaffold10-processed-gene-17.15	Similar to MEL: Alpha-galactosidase (Saccharomyces mikatae OX=114525)
Scaffold11	2845048	2846438	110	0.1194	maker-Scaffold11-snap-gene-28.58	Similar to SLC25A17: Peroxisomal membrane protein PMP34 (Homo sapiens OX=9606)
Scaffold11	2846866	2850623	110	0.1194	genemark-Scaffold11-processed-gene-28.8	Protein of unknown function

(((North American B TA, North American A TA), Italian TF), *T. biforme*)

Scaffold	Start	End	Sites used	f_{dM}	Gene	Note
Scaffold01	900331	900881	153	0.1181	maker-Scaffold01-snap-gene-9.2	Protein of unknown function
Scaffold01	901468	902851	153	0.1181	snap_masked-Scaffold01-processed-gene-9.31	Protein of unknown function
Scaffold01	902938	905385	153	0.1181	snap_masked-Scaffold01-processed-gene-9.39	Protein of unknown function
Scaffold01	906260	907088	153	0.1181	maker-Scaffold01-exonerate_protein2genome-gene-9.29	Protein of unknown function
Scaffold01	909092	910780	153	0.1181	maker-Scaffold01-snap-gene-9.4	Protein of unknown function
Scaffold01	912091	912422	153	0.1181	snap_masked-Scaffold01-processed-gene-9.40	Similar to EMC4: ER membrane protein complex subunit 4 (Saccharomyces cerevisiae (strain ATCC 204508 / S288c) OX=559292)
Scaffold01	913710	915599	153	0.1181	maker-Scaffold01-snap-gene-9.5	Similar to COX17: Cytochrome c oxidase copper chaperone (Homo sapiens OX=9606)
Scaffold01	915247	917294	153	0.1181	maker-Scaffold01-snap-gene-9.16	Similar to GST: Glutathione S-transferase (Plasmodium vivax OX=5855)
Scaffold01	918913	919817	153	0.1181	snap_masked-Scaffold01-abinit-gene-9.25	Protein of unknown function

Scaffold	Start	End	Length	Value	Gene ID	Description
Scaffold01	2975369	2975485	145	0.1387	maker-Scaffold01-exonerate_protein2genome-gene-29.152	Protein of unknown function
Scaffold02	4983917	4984611	104	0.1167	maker-Scaffold02-snap-gene-50.56	Protein of unknown function
Scaffold02	4986952	4987539	104	0.1167	maker-Scaffold02-exonerate_protein2genome-gene-49.140	Protein of unknown function
Scaffold02	4989157	4990010	104	0.1167	maker-Scaffold02-exonerate_protein2genome-gene-50.0	Similar to imp1: Mitochondrial inner membrane protease subunit 1 (Schizosaccharomyces pombe (strain 972 / ATCC 24843) OX=284812)
Scaffold02	4991580	4993084	104	0.1167	snap_masked-Scaffold02-processed-gene-50.18	Similar to TTC1: Tetratricopeptide repeat protein 1 (Bos taurus OX=9913)
Scaffold02	4993236	4994391	104	0.1167	snap_masked-Scaffold02-processed-gene-50.3	Similar to dltE: Uncharacterized oxidoreductase DltE (Bacillus subtilis (strain 168) OX=224308)
Scaffold02	4994792	4998647	104	0.1167	maker-Scaffold02-snap-gene-50.59	Similar to bglX: Periplasmic beta-glucosidase (Escherichia coli (strain K12) OX=83333)
Scaffold03	1620126	1622452	117	0.1619	genemark-Scaffold03-processed-gene-16.4	Similar to GRC3: Polynucleotide 5'-hydroxyl-kinase GRC3 (Cryptococcus neoformans var. neoformans serotype D (strain JEC21 / ATCC MYA-565) OX=214684)
Scaffold03	1624233	1626184	117	0.1619	maker-Scaffold03-snap-gene-16.42	Similar to DAL1: Allantoinase (Saccharomyces cerevisiae (strain ATCC 204508 / S288c) OX=559292)
Scaffold03	1626282	1628466	117	0.1619	maker-Scaffold03-snap-gene-16.43	Protein of unknown function
Scaffold03	1628643	1630494	117	0.1619	maker-Scaffold03-snap-gene-16.53	Protein of unknown function
Scaffold03	1631588	1632628	117	0.1619	maker-Scaffold03-snap-gene-16.54	Protein of unknown function
Scaffold03	1632377	1632573	117	0.1619	maker-Scaffold03-exonerate_est2genome-gene-16.2	Protein of unknown function

Scaffold	Start	End	Length	Value	Gene	Function
Scaffold03	1632645	1632758	117	0.1619	maker-Scaffold03-exonerate_protein2genome-gene-16.47	Protein of unknown function
Scaffold03	1633221	1636431	117	0.1619	maker-Scaffold03-exonerate_protein2genome-gene-16.48	Similar to RDR1: Probable RNA-dependent RNA polymerase 1 (Oryza sativa subsp. japonica OX=39947)
Scaffold03	1637269	1638202	117	0.1619	maker-Scaffold03-exonerate_protein2genome-gene-16.56	Similar to SEC14: SEC14 cytosolic factor (Saccharomyces cerevisiae (strain ATCC 204508 / S288c) OX=559292)
Scaffold05	701713	703798	153	0.1408	maker-Scaffold05-exonerate_protein2genome-gene-7.20	Similar to pik-1: Pelle-like serine/threonine-protein kinase pik-1 (Caenorhabditis elegans OX=6239)
Scaffold05	704321	707167	153	0.1408	maker-Scaffold05-snap-gene-7.2	Similar to STY8: Serine/threonine-protein kinase STY8 (Arabidopsis thaliana OX=3702)
Scaffold05	707896	709998	153	0.1408	maker-Scaffold05-snap-gene-7.3	Similar to splB: Dual specificity protein kinase splB (Dictyostelium discoideum OX=44689)
Scaffold05	710120	711718	153	0.1408	maker-Scaffold05-snap-gene-7.16	Similar to SPAC4A8.06c: AB hydrolase superfamily protein C4A8.06c (Schizosaccharomyces pombe (strain 972 / ATCC 24843) OX=284812)
Scaffold05	712158	713655	153	0.1408	maker-Scaffold05-exonerate_protein2genome-gene-7.31	Similar to SPAC5D6.12: Uncharacterized protein C5D6.12 (Schizosaccharomyces pombe (strain 972 / ATCC 24843) OX=284812)
Scaffold05	713908	718057	153	0.1408	maker-Scaffold05-snap-gene-7.17	Similar to ARHGAP39: Rho GTPase-activating protein 39 (Homo sapiens OX=9606)
Scaffold05	718805	719236	153	0.1408	maker-Scaffold05-exonerate_protein2genome-gene-7.39	Protein of unknown function
Scaffold05	2480001	2500000	122	0.1628	No annotated genes	No annotated genes

Scaffold						
Scaffold07	3840869	3842531	167	0.1483	maker-Scaffold07-snap-gene-38.79	Similar to ZFAND2A: AN1-type zinc finger protein 2A (Pongo abelii OX=9601)
Scaffold07	3843021	3844141	167	0.1483	snap_masked-Scaffold07-abinit-gene-38.26	Protein of unknown function
Scaffold07	3845137	3848851	167	0.1483	maker-Scaffold07-snap-gene-38.49	Similar to PHB2: Prohibitin-2 (Saccharomyces cerevisiae (strain ATCC 204508 / S288c) OX=559292)
Scaffold07	3848870	3850190	167	0.1483	maker-Scaffold07-exonerate_protein2genome-gene-38.89	Protein of unknown function
Scaffold07	3850825	3853362	167	0.1483	maker-Scaffold07-snap-gene-38.40	Similar to exosc3: Putative exosome complex component rrp40 (Dictyostelium discoideum OX=44689)
Scaffold07	3853955	3856502	167	0.1483	maker-Scaffold07-snap-gene-38.51	Protein of unknown function
Scaffold07	3857642	3858990	167	0.1483	maker-Scaffold07-snap-gene-38.52	Protein of unknown function
Scaffold08	987415	988459	111	0.1333	maker-Scaffold08-snap-gene-9.25	Protein of unknown function
Scaffold08	990031	990293	111	0.1333	maker-Scaffold08-exonerate_protein2genome-gene-10.96	Protein of unknown function
Scaffold08	990859	996481	111	0.1333	maker-Scaffold08-snap-gene-10.13	Protein of unknown function
Scaffold08	996522	997428	111	0.1333	maker-Scaffold08-snap-gene-10.18	Protein of unknown function
Scaffold08	997701	997814	111	0.1333	trnascan-Scaffold08-noncoding-Gly_GCC-gene-10.66	Protein of unknown function
Scaffold08	998636	999746	111	0.1333	maker-Scaffold08-exonerate_protein2genome-gene-10.16	Similar to ECI3: Enoyl-CoA delta isomerase 3 (Arabidopsis thaliana OX=3702)
Scaffold09	3641160	3641384	185	0.1245	maker-Scaffold09-exonerate_protein2genome-gene-36.137	Protein of unknown function
Scaffold09	3642465	3643793	185	0.1245	maker-Scaffold09-snap-gene-36.38	Similar to truC: tRNA pseudouridine synthase C (Yersinia pestis OX=632)

Scaffold09	3643841	3652490	185	0.1245	maker-Scaffold09-snap-gene-36.30	Similar to PPO1: Polyphenol oxidase 1 (Agaricus bisporus OX=5341)
Scaffold09	3652622	3655673	185	0.1245	maker-Scaffold09-exonerate_protein2genome-gene-36.142	Similar to pr1: Aspartic protease (Phaffia rhodozyma OX=5421)
Scaffold09	3657013	3657605	185	0.1245	maker-Scaffold09-exonerate_protein2genome-gene-36.152	Protein of unknown function
Scaffold09	3658494	3659856	185	0.1245	maker-Scaffold09-snap-gene-36.41	Similar to SPBC216.03: UPF0659 protein C216.03 (Schizosaccharomyces pombe (strain 972 / ATCC 24843) OX=284812)
Scaffold10	1740426	1741739	241	0.1656	snap_masked-Scaffold10-abinit-gene-17.14	Protein of unknown function
Scaffold10	1742972	1744147	241	0.1656	snap_masked-Scaffold10-processed-gene-17.14	Protein of unknown function
Scaffold11	2845048	2846438	113	0.1259	maker-Scaffold11-snap-gene-28.58	Similar to SLC25A17: Peroxisomal membrane protein PMP34 (Homo sapiens OX=9606)
Scaffold11	2846866	2850623	113	0.1259	genemark-Scaffold11-processed-gene-28.8	Protein of unknown function

Table S4. Overview of which cellular processes the outlier genes from the f_{dM} genome scan are involved in, according to the GO enrichment analysis. The headers indicate the phylogenetic typology of the f_{dM} analysis; (((P1, P2), P3), O), where P1, P2 and P3 are populations investigated for introgression and O is the outgroup. The table also denotes which processes the genes are involved in (Process), number of annotated genes assigned to the process (Annotated), number of observed significant genes within the process (Significant), the expected number of significant genes (Expected), and a raw p-value. Only processes with a raw p-value of less than 1 are included in the table.

(((North American B TA, North American A TA), Canadian TF), *T. biforme*)				
Process	**Annotated**	**Significant**	**Expected**	**Raw p-value**
Metabolic process	2388	7	6.12	0.0012
Pantothenate biosynthetic process	2	1	0.01	0.0016
Copper ion transport	6	1	0.02	0.0049
Organic substance metabolic process	1597	4	4.09	0.0912

(((North American B TA, North American A TA), Italian TF), *T. biforme*)				
Process	**Annotated**	**Significant**	**Expected**	**Raw p-value**
Metabolic process	2388	11	7.83	0.00058
Protein phosphorylation	303	3	0.99	0.00672
Copper ion transport	6	1	0.02	0.00797
Pseudouridine synthesis	7	1	0.02	0.00929
Carbohydrate metabolic process	213	2	0.7	0.03156
Biological_process	3020	13	9.9	0.08499
Primary metabolic process	1512	7	4.96	0.09572

TA = *Trichaptum abietinum* and TF = *T. fuscoviolaceum*

29

Table S5. Estimated model parameters. The table shows estimated parameters for the different models tested with *fastsimcoal2* (Excoffier et al. 2021). See Figure S7 for illustrations of the models. The headers (except for the model names) are parameters estimated in the analysis. Parameters for the best model is highlighted in light grey. See the GitHub page for example of parameter estimation and template files (*GitHub address will be added*).

Model name	Npop0	Npop1	Npop2	Npop3	Npop4	Npop5	Tbw1	Tbw2	Tbw3	Tbw4
nomig	68673	166871	68367	106745	NA	NA	205058	91189	NA	NA
mig_ancient	153841	130988	60260	83895	NA	NA	365698	322921	NA	NA
mig_2ancient	175064	184860	69203	109285	NA	NA	241285	81014	NA	NA
mig_3ancient	243968	321735	202473	161661	NA	NA	2076408	2149569	NA	NA
mig_recent	151618	185172	57426	103008	NA	NA	233029	49995	NA	NA
mig_2recent	249507	94388	68676	75408	NA	NA	212338	1765	2453	NA
mig_3recent	151271	176916	69615	101860	NA	NA	191882	68860	NA	NA
mig_4recent	137582	92899	61313	116092	NA	NA	1335026	1242192	75564	NA
mig_ghost	147810	117770	62165	70301	66357	NA	27347	3802	3980	4082
mig_2ghost	203693	139057	85727	91339	31008	NA	192683	4625	5820	12291
mig_3ghost	99441	199627	44307	95489	8697	NA	162766	4127	1416	81597
mig_4ghost	174489	253190	72062	67152	48214	68623	106757	2104	3613	33568
mig_5ghost	221459	152231	100595	96448	60168	NA	212413	6967	6943	4007
mig_6ghost	98582	239473	36571	114641	7912	NA	128055	2285	2473	2885
mig_7ghost	193661	144624	85059	87238	62942	67945	20960	4950	3950	4421
mig_8ghost	162493	202415	28410	10958	1809	42420	1328373	22786	882	25308
mig_9ghost	112995	215613	44149	112356	8660	NA	423851	1521	4565	2276

Model name	Tbw5	Tbw6	Tbw7	Resize1	Resize2	Resize3	Resize4	Resize5	mig1	mig2
nomig	NA	NA	NA	3.7842780	1.1040148	3.6354008	NA	NA	NA	NA
mig_ancient	NA	NA	NA	16.44110324	0.5678387	1.4124021	NA	NA	4.99E-06	NA
mig_2ancient	NA	NA	NA	3.6690522	1.1947390	7.8242162	NA	NA	8.61E-08	NA

		mig3									
mig_3ancient	NA	NA	NA	NA	0.1919287	1.1519895	0.9203845	NA	NA	6.32E-06	NA
mig_recent	NA	NA	NA	NA	3.3260382	1.3913197	20.1069646	NA	NA	1.18E-07	NA
mig_2recent	NA	NA	NA	NA	8.4229324	0.4118750	51.4254188	NA	NA	6.53E-04	NA
mig_3recent	NA	NA	NA	NA	4.1196068	1.1041883	2.8093411	NA	NA	8.65E-08	NA
mig_4recent	NA	NA	NA	NA	1.0985346	0.2360949	4.3962013	NA	NA	0.0078392	NA
mig_ghost	NA	NA	NA	NA	51.8148929	0.1368533	32.3027957	0.1561456	NA	6.52E-04	NA
mig_2ghost	NA	NA	NA	NA	4.4920900	0.8896129	7.1859076	25.4961475	NA	5.25E-08	NA
mig_3ghost	NA	NA	NA	NA	6.0686313	1.0769703	0.9589982	30.4487050	NA	1.63E-04	NA
mig_4ghost	34844	NA	NA	NA	7.7386680	0.6000751	0.2393338	16.1633557	0.1060555	7.51E-06	NA
mig_5ghost	NA	NA	NA	NA	4.2221569	0.9345088	0.8658364	45.1669535	NA	8.66E-08	NA
mig_6ghost	71762	NA	NA	NA	6.1241827	0.8451170	39.8374047	0.1301317	NA	1.77E-05	2.12E-04
mig_7ghost	3388	NA	NA	NA	38.3739585	0.1179344	2.8160925	0.1217412	6.2096511	4.42E-06	NA
mig_8ghost	1676	6450	0	NA	2.7466486	20.3338345	0.1026876	15.1573729	3.7116024	2.29E-04	6.57E-05
mig_9ghost	86633	NA	NA	NA	15.9185133	0.6420154	63.8036232	0.4883565	NA	1.48E-04	1.42E-06

Model name	mig3	prop1	prop2	prop3	prop4	prop5	prop6	Tdiv1	Tdiv2	TdivG1	Tdiv3	TdivG2
nomig	NA	NA	NA	NA	NA	NA	NA	354864	149806	NA	263675	NA
mig_ancient	NA	1.3078478	NA	NA	NA	NA	NA	508998	143300	NA	186077	NA
mig_2ancient	NA	0.0292000	NA	NA	NA	NA	NA	397096	155811	NA	316082	NA
mig_3ancient	NA	1.1130978	NA	NA	NA	NA	NA	2430173	353765	NA	280604	NA
mig_recent	NA	0.0226856	NA	NA	NA	NA	NA	369159	136130	NA	319164	NA
mig_2recent	NA	0.0104990	0.0147415	NA	NA	NA	NA	380536	168198	NA	166433	NA
mig_3recent	NA	0.0150532	NA	NA	NA	NA	NA	339611	147729	NA	270751	NA
mig_4recent	NA	0.3266287	NA	NA	NA	NA	NA	1473540	138514	NA	231348	NA
mig_ghost	NA	0.0251477	0.0270060	0.0284662	NA	NA	NA	178536	151189	147387	139325	NA
mig_2ghost	NA	0.0215626	0.0277299	0.0602280	NA	NA	NA	407206	214523	209898	204078	NA
mig_3ghost	NA	0.0440787	0.0158260	0.9262883	NA	NA	NA	256400	93634	89507	88091	NA

mig_4ghost	NA	NA	0.0135652	0.0236086	0.3007132	NA	261914	155157	153053	149440	115872
mig_5ghost	NA	NA	0.0294961	0.0302867	0.0180291	NA	448629	236216	229249	218299	NA
mig_6ghost	NA	NA	0.0278678	0.0310197	0.9648624	NA	210074	82019	79734	77261	NA
mig_7ghost	NA	NA	0.0247137	0.0202190	0.0181224	NA	221273	200313	195363	183604	191413
mig_8ghost	3.22E-04	0.2733081	0.3068627	0.0171475	0.0663087	0.0175051	1402628	74255	51469	50587	23603
mig_9ghost	NA	NA	0.0151238	0.0460690	0.9390135	NA	524473	99101	94536	92260	NA

Model name	Tmig1	Tmig2	Tmig3	MaxEstLhood	MaxObsLhoos	AIC
nomig	NA	NA	NA	-7823474.701	-6809027.340	1014447.361
mig_ancient	187414	NA	NA	-7805505.703	-6809510.928	995994.775
mig_2ancient	4549	NA	NA	-7819058.784	-6809510.928	1009547.856
mig_3ancient	312339	NA	NA	-7792802.746	-6809510.928	983291.818
mig_recent	3088	NA	NA	-7821830.565	-6809510.928	1012319.637
mig_2recent	163980	NA	NA	-7764526.353	-6809510.928	955015.425
mig_3recent	4075	NA	NA	-7813832.196	-6809510.928	1004321.268
mig_4recent	155784	NA	NA	-7728427.346	-6809510.928	918916.418
mig_ghost	143407	NA	NA	-7808874.997	-6809510.928	999364.069
mig_2ghost	191787	NA	NA	-7845453.172	-6809510.928	1035942.244
mig_3ghost	6494	NA	NA	-7664400.865	-6809510.928	854889.937
mig_4ghost	81028	NA	NA	-7791242.284	-6809510.928	981731.356
mig_5ghost	222306	NA	NA	-7845664.969	-6809510.928	1036154.041
mig_6ghost	74376	2614	NA	-7650743.782	-6809510.928	841232.854
mig_7ghost	186992	NA	NA	-7821700.972	-6809510.928	1012190.044
mig_8ghost	25279	17153	17153	-7754011.360	-6809510.928	944500.432
mig_9ghost	100622	5627	NA	-7644407.431	-6809510.928	834896.503

Literature cited

Benoit, K. and A. Obeng. 2020. Readtext: Import and Handling for Plain and Formatted Text Files. R package version 0.80. Available at https://CRAN.R-project.org/package=readtext. Accessed March 2, 2022.

Dowle, M. and A. Srinivasan. 2020. Data.table: Extension of `data.frame`. R package version 1.13.0. Available at https://CRAN.R-project.org/package=data.table. Accessed March 2, 2022.

Fitak, R. R. 2021. OptM: estimating the optimal number of migration edges on population trees using Treemix. Biol. Methods Protoc. 6:bpab017.

Garnier, S. 2018. Viridis: Default Color Maps from 'matplotlib'. R package version 0.5.1. Available at https://CRAN.R-project.org/package=viridis. Accessed March 2, 2022.

Petr, M., B. Vernot, and J. Kelso. 2019. admixr—R package for reproducible analyses using ADMIXTOOLS. Bioinformatics. 35:3194–3195.

R Core Team. 2020. R: A language and environment for statistical computing, R Foundation for Statistical Computing. Vienna, Austria. Available at https://www.R-project.org/. Accessed March 2, 2022.

Rudis, B. 2020. hrbrthemes: Additional Themes, Theme Components and Utilities for 'ggplot2'. R package version 0.8.0. Available at https://CRAN.R-project.org/package=hrbrthemes. Accessed March 2, 2022.

Wickham, H. 2016. ggplot2: Elegant Graphics for Data Analysis. 2nd ed. Springer, New York, NY.

Wickham, H., M. Averick, J. Bryan, W. Chang, L. D. McGowan, R. François, G. Grolemund, A. Hayes, L. Henry, J. Hester, M. Kuhn, T. L. Pedersen, E. Miller, S. M. Bache, K. Müller, J. Ooms, D. Robinson, D. P. Seidel, V. Spinu, K. Takahashi, D. Vaughan, C. Wilke, K. Woo, and H. Yutani. 2019. Welcome to the tidyverse. J. Open Source Softw. 4:1686.

Paper III

"(...) you sprout all your worth and you
woof your wings, so if you want to be
Phoenixed, come and be parked."
(James Joyce, Finnegans Wake)

Paper IV

Front cover: illustration of Njord from a 1680 icelandic manuscript of the Younger Edda. Árni Magnússon Institute for Icelandic Studies (Public Domain)

"The light dove cleaving the air in her free flight and feeling its resistance, might imagine that its flight would be easier still in empty space."
(Immanuel Kant, Critique of Pure Reason)