UNIVERSITÉ DE LORRAINE | BIBLIOTHÈQUES UNIVERSITAIRES

AVERTISSEMENT
DROIT D'AUTEUR

Ce document est le fruit d'un long travail approuvé par le jury de soutenance et mis à disposition de l'ensemble de la communauté universitaire élargie.

Il est soumis à la propriété intellectuelle de l'auteur. Ceci implique une obligation de citation et de référencement lors de l'utilisation de ce document. Toute contrefaçon, plagiat ou reproduction illicite encourt une poursuite pénale.

Contact bibliothèque : ddoc-theses-contact@univ-lorraine.fr
(Cette adresse ne permet pas de contacter les auteurs)

Code de la Propriété Intellectuelle. articles L 122. 4
Code de la Propriété Intellectuelle. articles L 335.2- L 335.10

UNIVERSITÉ DE LORRAINE
34 COURS LÉOPOLD — CS 25233
54052 NANCY CEDEX
TÉL : 03 54 50 54 00
CONTACT@UNIV-LORRAINE.FR
WWW.UNIV-LORRAINE.FR

Université de Lorraine, Collégium Sciences et Techniques
Ecole Doctorale Sciences et Ingénierie des Ressources Naturelles
UMR 1136 INRAE - Université de Lorraine, Interactions Arbres – Microorganismes

Thèse

Présentée et soutenue publiquement pour l'obtention du titre de

Docteur de l'Université de Lorraine

Mention : Biologie et écologie des forêts et des agrosystèmes

par **Ammar ABDALRAHEM MOHAMMAD ABDALRAHEM**

Evolutionary and functional investigations of reproductive mode polymorphism in the poplar rust fungus

Soutenance le 22 octobre 2025 devant le Jury composé de :

Tatiana GIRAUD	Directrice de recherche, CNRS, Gif-sur-Yvette	Rapporteure, Présidente du jury
Sophie ARNAUD-HAOND	Chercheure HDR, Ifremer, Sète	Rapporteure
Frédéric SUFFERT	Ingénieur de recherche, INRAE, Palaiseau	Examinateur
Pascal FREY	Directeur de recherche, INRAE, Nancy	Directeur de thèse
Sébastien DUPLESSIS	Directeur de recherche, INRAE, Nancy	Co-directeur de thèse
Fabien HALKETT	Chargé de recherche, INRAE, Nancy	Co-encadrant

Declaration

This thesis is structured around three main chapters, each of which has been prepared in the form or a foundation of research articles. I am the first author of all chapters, except for one section in Chapter III, where I share co-first authorship. The status of each chapter and section is as follows:

- Chapter I consists of two sections.
    - Section I has been submitted to ***Molecular Ecology*** and is currently under peer reviewing
    - Section II is in the final stages of preparation and will soon be submitted to ***Heredity***.

- Chapter II forms the foundation of a manuscript that will be submitted to ***Phytopathology***.

- Chapter III is divided into two sections.
    - Section I will form the basis of a co-authored manuscript
    - Section II will serve as the foundation of a manuscript.

In preparing this thesis, I used AI-based tools for supportive purposes. Elicit and Perplexity were primarily used for bibliography searches, with Perplexity also occasionally assisting in writing and coding. Grammarly and ChatGPT were employed for grammar checking, proofreading, and language improvement, while Perplexity, ChatGPT, and GitHub Copilot were additionally used for coding improvement, debugging, and code documentation in R and Python. All writings and analyses were carefully reviewed by me and my supervisors.

Acknowledgements

I am sincerely grateful to my supervisors, **Pascal Frey**, **Fabien Halkett**, and **Sébastien Duplessis**, for the invaluable experience I have gained from you over the past three years. Our discussions, especially during coffee breaks, have had a significant impact on improving this thesis through your scientific guidance.

I would like to sincerely thank the members of my thesis advisory committee, **Annegret Kohler**, **Solenn Stoecke**l, **Ronan Becheler**, and **Frédéric Suffert**, for their valuable guidance throughout my PhD journey.

I would also like to thank **Aurélie Deveau** from the Ecogenomics of Interactions team for her guidance with the microscopy experiments. I am sincerely grateful to my colleagues on the team, **Jérémy Pétrowski**, **Anaïs Gillet**, **Bénédicte Fabre**, **Christophe Bailly,** and **Benoit Marçais**. Besides lab work, I felt truly supported on a personal level. I want to express my deepest gratitude to **Axelle Andrieux**, **Mireia Gomez-Gallego**, and **Irène Hummel**, whose support went far beyond the scientific aspects, both personal and professional, and was essential for me throughout this experience.
I am also grateful to my colleagues, students, and interns in the team, **Elodie Muller**, **Méline Poch**, **Anatole Courouble**, **Kadiatou Schiffer-Forsyth, Scott Heslop,** and **Thomas Tempestini**. Our discussions were both scientifically fruitful and personally enriching.

Alhamdulillah, I am truly grateful to my family, **Asem**, **Alyaa**, and my parents, and I appreciate their prayers and support. I owe you all for enabling me to be where I am today. Your love and unwavering support have been the foundation of this journey.

I am deeply grateful to my best friends, **Mohamed Sarhan**, **Mohamed Abdelfadil**, **Elhussein Mourad**, **Mahmoud Mabrouk**, and **Mohamed Saleh**. Your support and feedback have meant a lot to me, and I'm grateful for your friendship.

To

Asmaa, my loving wife

Owais Ammar, my little wolf

Table of Contents

Résumé des travaux de thèse..... 1
Introduction générale..... 1
Objectifs de recherche..... 2
Méthodes..... 3
Résultats principaux..... 4
Contributions méthodologiques et portée générale..... 5
Abstract..... 7
Résumé..... 9
General introduction..... 11
1. The evolution of reproductive systems: insights into sexual and asexual strategies..... 13
1.1 Evolutionary history of polymorphism in reproductive modes among eukaryotes..... 13
1.2 Cost and benefits of sexual and asexual reproduction..... 13
1.3 Reproductive mode diversity and mating systems in Fungi..... 14
1.4 Mating-Type architecture in fungi..... 15
2. Rust fungi as models for reproductive complexity..... 18
Life cycle and host alternation..... 18
3. The heteroecious rust: Melampsora larici-populina..... 21
3.1 Biological and economic importance of poplar rust..... 21
3.2 Life cycle of M. larici-populina..... 21
3.3 Genomic Landmarks of M. larici-populina..... 23
3.4 Previous findings: evidence of persistent clonal lineages..... 23
4. Challenges and open questions..... 26
4.1 Population structure and genetic isolation of reproductive modes..... 26
4.2 Sexual competence and functional capacity for reproduction..... 26
4.3 Genomic and transcriptional basis of reproductive mode polymorphism..... 26
5. Thesis objectives and structure..... 27
5.1 Evolutionary ecology approach to map sexual vs. asexual populations..... 27
5.2 Phenotypic assays of sexual capacity in asexual lineages..... 28
5.3 Genomic and transcriptomic analysis of reproductive mode variation..... 28
5.4 Overview of thesis chapters and their interconnections..... 29
Chapter one..... 31
Population Structure and Theoretical Modeling of Sexual vs. Asexual Reproduction..... 31
Section one..... 32

Long-lasting coexistence of multiple asexual lineages alongside their sexual counterparts in a fungal plant pathogen....32
Abstract....34
1 Introduction....35
2 Material and Methods....37
2.1 Isolate collections....38
2.2 Genotyping....38
2.3 Population genetic and statistical analyses....39
3 Results....43
3.1 Delineation of multilocus lineages....43
3.2 Reproductive mode assessment....43
3.3 Spatial distribution of reproductive modes....47
Figure 1....47
3.4 Genetic relationships between sexual and asexual lineages....50
3.5 Genetic structure and persistence of asexual lineages....51
4 Discussion....53
Data....59
Acknowledgments....59
References....61
Supplementary figures....75
Supplementary Tables....77
Section two....78
Behind the scenes of a shadow play: illuminating the population genetics of partially clonal organisms through the prism of clone age distribution....78
Abstract....80
Introduction....81
Materials and Methods....85
Simulation design....85
Analysis of population genetic indices across rates of clonality....88
Temporal analysis of genetic indices across clonality rates....88
The effect of rates of clonality on clone age distribution....89
Results....90
Introducing the effect of rates of clonality on the temporal change of population genetic indices....90
How rates of clonality drive the variability in evolutionary trajectories....92
The rate of clonality shapes clone age structure....94

Abundance of clones responds to changes in rates of clonality..... 96
Summaries of clone age distribution correlates with the variation of some population genetic indices..... 98
Detailed examination of the relationships between metrics of clone age distribution provides further explanation on the variability of population genetic indices among replicates.....100
Discussion..... 104
References..... 109
Supplementary figures..... 116
Chapter two.....122
Can Asexual Melampsora larici-populina Isolates Undergo Sexual Reproduction? Experimental Assessment of Life Cycle Completion and Nuclear Fusion.....122
1 Introduction..... 124
1.1 Sexual and asexual life cycle of poplar rust.....124
1.2 Discovery of asexual lineages.....127
1.3 Cryptic sexuality: the need for experimental validation..... 127
2. Materials and Methods..... 128
2.1 Microscopic analysis of teliospore fusion dynamics..... 128
2.1.1 Inoculation protocol.....128
2.1.2 Sample fixation procedure.....128
2.1.3 Tissue embedding and sectioning.....128
2.1.4 Staining and microscopic analysis..... 129
2.2 Heteroecious cycle completion by asexual lineages..... 129
2.2.1 Experimental overview..... 129
2.2.2 Poplar host preparation..... 130
2.2.3 Poplar inoculation (asexual cycle).....130
2.2.4 Telia collection and dormancy breaking.....131
2.2.5 Basidiospore production (sexual cycle).....131
2.2.6 Larch inoculation.....132
2.2.7 Assessing life cycle completion..... 133
3 Results..... 134
3.1 Nuclear fusion dynamics.....134
3.2 Life cycle completion assessment..... 135
3.2.1 Isolate selection and poplar inoculation..... 135
3.2.2 Telia formation and basidiospore production..... 136
3.2.3 Sexual stage development on larch and secondary inoculation..... 137

3.2.4 Confirmation of sexual capacity in asexual lineages..... 138
4 Discussion..... 140
5 Challenges and future directions..... 143
5.1 Nuclear fusion observation in teliospores..... 143
5.1 Assessment of life cycle completion..... 144
Chapter three..... 145
Genomic and Transcriptomic Signatures of Reproductive Mode Variation in the Poplar Rust Fungus..... 145
Section one..... 149
Mating-Type Locus Diversity and Genomic Variation in Sexual and Asexual Lineages..... 149
1 Introduction..... 150
1.1 Basidiomycota mating systems..... 150
1.2 The tetrapolar homeodomain and pheromone/pheromone receptors mating systems 152
1.3 Poplar rust as a model system..... 153
2 Material and Methods..... 154
2.1 Population genomics datasets and reference isolates..... 154
2.2 Identification and annotation of MAT genes..... 156
2.3 Comparative analysis of MAT loci across the new reference phased genomes..... 156
2.4 Population-level comparative analysis of MAT genes..... 157
3 Results..... 157
3.1 Organization of mating-type (MAT) genes in sexual and asexual phased genomes..157
3.2 Diversity of the HD locus..... 159
3.3 Pairings of homeodomain alleles (HD1–HD2) across chromosomes, within isolates..... 160
3.5 Allelic diversity of STE3 pheromone receptor genes..... 162
4 Discussion..... 163
Data availability..... 167
Supplementary figures..... 168
Section two..... 170
Comparative Developmental Transcriptomics of Sexual and Asexual Lineages in the Poplar Rust Fungus Melampsora larici-populina..... 170
1 Introduction..... 171
2 Material and Methods..... 173
2.1 Experimental Design..... 173
2.2 Plant and fungal material and infection protocol..... 173

2.3 RNA isolation and quality control..... 174
2.4 Quality filtering and gene quantification..... 174
2.5 Exploratory analyses of isolate- and stage-specific expression.....175
2.6 Differential expression analysis plans..... 175
3 Results..... 177
3.1 Sequencing quality assessment..... 177
3.2 Read mapping quality metrics and potential reference bias..... 178
3.3 Samples variation aligns with developmental stage and reproductive strategy.....180
3.4 Isolate- and timepoint-specific gene expression.....182
4 Discussion..... 184
4.1 The impact of developmental stage on transcriptional variation..... 184
4.2 Reproductive mode imprints on gene expression..... 185
5 Limitations and future directions.....186
Supplementary figures..... 188
General Discussion & Perspectives.....189
Discussion.....190
Revisiting the classical view of poplar rust life cycle.....191
Imperfect fungi or imperfect observation?.....193
Beyond indices: The importance of experimental validation.....196
MAT expression equivalence in sexual and asexual isolates..... 198
Diversity in Mating-Type genes.....201
Expression programs reflect reproductive history..... 202
Opportunities and challenges..... 204
Bibliography..... 206
Appendices.....235
List of Appendices.....236
Appendix I:.....237
Appendix II:..... 250

Résumé des travaux de thèse

Introduction générale

L'étude de l'évolution et du maintien de la reproduction sexuée demeure un élément central en biologie évolutive. Bien que la reproduction sexuée entraîne des coûts — comme la nécessité de trouver un partenaire, le coût énergétique et la complexité développementale — elle est conservée chez la plupart des eucaryotes. Ce paradoxe est particulièrement évident chez les organismes qui alternent entre reproduction sexuée et asexuée. Dans ces cas, les deux modes diffèrent non seulement dans leurs conséquences épidémiologiques immédiates, mais aussi dans leurs conséquences évolutives à long terme. Comprendre pourquoi les deux modes de reproduction coexistent au sein d'une même espèce, ou même dans une même population, reste une énigme en biologie évolutive.

Les champignons responsables de la rouille des plantes (ordre des Pucciniales) constituent un modèle adapté pour étudier ce paradoxe. Leurs cycles de vie peuvent inclure jusqu'à cinq stades de spores. Chez les espèces hétéroïques, le cycle de vie alterne entre deux plantes hôtes non apparentées, chacune ayant un rôle reproductif différent. Certaines lignées fongiques ont été autrefois décrites comme des « champignons imparfaits » parce que seule la propagation clonale avait été observée dans la nature. Dans de nombreux cas, cependant, cette désignation reflétait les limites de l'observation plutôt qu'une absence réelle de reproduction sexuée. En particulier pour les Pucciniales, supposer que la capacité sexuée est perdue peut induire en erreur. Une approche plus prudente consiste à considérer que le potentiel sexué demeure sauf en cas de preuve claire de son absence.

Cette thèse se concentre sur l'espèce *Melampsora larici-populina* (*Mlp*), un des champignons responsables de la rouille du peuplier. Il est généralement décrit comme ayant un cycle sexué annuel : la reproduction sexuée a lieu sur le mélèze (*Larix* spp.) au printemps, et une multiplication clonale rapide prend place sur le peuplier (*Populus* spp.) durant l'été. Mettre en

évidence la coexistence de lignées sexuées et asexuées et étudier leurs conséquences écologiques et génétiques est l'objectif principal de cette thèse, afin de comprendre comment différentes stratégies reproductives peuvent persister dans une même espèce et un même environnement.

Objectifs de recherche

Le travail présenté dans cette thèse est structuré autour de quatre objectifs principaux :

1. Persistance des lignées asexuées

Nous avons cherché à démontrer l'existence de lignées asexuées persistantes dans les populations naturelles de *M. larici-populina* et à mesurer leur récurrence dans différents sites et années en utilisant le génotypage de lignées multilocus (MLL). Nous avons également étudié dans quelles conditions environnementales les lignées peuvent contourner la phase sexuée et persister de façon clonale, notamment la survie hivernale sur les feuilles de peuplier.

2. Limites d'inférence dans les populations partiellement clonales

Nous avons évalué la pertinence des indices de génétique des populations standards, tels que le F_{IS}, pour étudier la clonalité dans des populations partiellement clonales (où individus sexués et asexués coexistent). Ces indices peuvent donner des résultats trompeurs selon le taux de clonalité. Pour clarifier ce point, nous avons examiné la dynamique génétique des populations sous différents taux de clonalité au fil des générations, afin de mieux comprendre la fiabilité de ces indices pour détecter les modes de reproduction dans des populations partiellement clonales.

3. Test expérimental de la capacité sexuée des lignées asexuées

Nous avons voulu vérifier expérimentalement si les lignées identifiées comme asexuées dans les populations naturelles de *M. larici-populina* avaient conservé la capacité de réinvestir et de réaliser le cycle sexué hétéroïque complet dans des conditions contrôlées en serre et laboratoire. Des observations microscopiques ont également été utilisées pour vérifier la fusion des noyaux lors de la formation des téliospores.

4. Différences moléculaires et fonctionnelles entre lignées sexuées et asexuées

Nous avons comparé la structure de la région du type sexuel (MAT) entre les lignées sexuées et asexuées afin de déterminer s'il existe des différences génétiques pouvant expliquer la perte de sexualité. De plus, nous avons analysé les transcriptomes, à la fois entre lignées sexuées et asexuées et entre différents stades de développement. Cela nous a permis de détecter des programmes d'expression spécifiques aux lignées ainsi que des gènes spécifiques aux stades de développement, apportant un aperçu de la manière dont la régulation façonne les différentes stratégies reproductives.

Méthodes

Enquêtes spatiales et temporelles sur les lignées asexuées

Des collectes sur le terrain ont été effectuées à travers la France au cours des dernières décennies pour capturer la diversité naturelle de *M. larici-populina*. Des marqueurs microsatellites ont été utilisés pour définir les lignées multilocus (MLLs), et leur répartition a été suivie à la fois à des échelles spatiales et temporelles. Cette approche a permis la détection indirecte de lignées asexuées grâce à l'analyse de la structure des populations à l'aide d'indices de génétique des populations, et la détection directe de récurrences des mêmes clones, fournissant ainsi des preuves de leur persistance à long terme.

Test expérimental de la capacité sexuée

Les lignées identifiées comme asexuées dans la nature ont été testées en conditions de serre et de laboratoire. Nous avons réalisé des infections contrôlées sur peuplier et mélèze pour examiner si ces lignées pouvaient réaliser un cycle de vie hétéroïque complet. La complétude du cycle a été évaluée en suivant la production et le développement des différents stades de spores sur chaque hôte. De plus, des observations microscopiques de la fusion des noyaux ont été effectuées pour vérifier la maturation des téliospores.

Analyses comparatives génomiques et transcriptomiques

Le séquençage du génome complet d'isolats représentatifs des deux modes de reproduction a été réalisé pour comparer la structure et la variation de la région MAT. En parallèle, des données de RNA-seq couvrant différents stades de développement ont permis de définir des profils d'expression génique associés à des lignées présentant différentes stratégies reproductives. Cela a fourni une vision intégrée de l'arsenal génomique et de sa régulation.

Résultats principaux

1. Un « pont vert » écologique permet la persistance clonale

La détection répétée des mêmes lignées multilocus sur plusieurs années a montré que des populations sexuées et asexuées coexistent à long terme. Les enquêtes de terrain ont confirmé qu'un scénario dans lequel des feuilles de peuplier infectées survivent pendant des hivers doux fournit un « pont vert » (*green bridge*) qui pourrait permettre la survie hivernale de l'inoculum asexué. Cette voie serait la plus à même d'expliquer comment certaines lignées asexuées peuvent contourner la phase sexuée sur mélèze et persister par voie clonale sur plusieurs années. Les expériences en serre ont également révélé que certaines lignées identifiées comme asexuées étaient encore capables de réaliser leur cycle sexué, tandis que d'autres s'arrêtaient à certains stades de développement. Ces résultats montrent que le potentiel sexué peut persister même chez des lignées qui se comportent de manière asexuée dans la nature, soulignant l'importance de la validation expérimentale.

2. L’architecture MAT est conservée entre lignées

Les analyses génomiques comparatives ont montré que les loci MAT sont intacts chez les isolats sexués comme asexués, incluant les gènes codant des facteurs de transcription homéodomaines HD1/HD2 et des récepteurs de phéromones. L’analyse de plus de 200 isolats a identifié sept variants des gènes HD, indiquant une diversité génétique dans les régions liées aux loci MAT. L’analyse de profils transcriptomiques a en outre révélé des différences dans les programmes d’expression entre lignées sexuées et asexuées, avec de nombreux gènes exprimés uniquement à des stades spécifiques lors de la production d’urédiniospores ou de teliospores ; lors de la différentiation des télies et enfin lors de l’entrée en phase de dormance.

Contributions méthodologiques et portée générale

Cette thèse intègre quatre approches complémentaires : (i) le suivi temporel de lignées sexuées et asexuées, (ii) des analyses statistiques interprétées par des simulations sous différents taux de clonalité, (iii) la validation expérimentale directe de la capacité sexuée, et (iv) des analyses génomiques et transcriptomiques comparatives. En combinant ces approches, nous réduisons la dépendance à une seule source de preuve et proposons un cadre applicable à d’autres Pucciniales et à des pathosystèmes similaires.

Les résultats de cette thèse montrent que la persistance de l’asexualité chez les Pucciniales n’implique pas nécessairement une perte irréversible de sexualité. Le mode de reproduction peut être façonné par des différences dans les programmes génétiques exprimés ou par le contexte environnemental. Cela apporte de nouvelles perspectives pour comprendre comment les lignées sexuées et asexuées coexistent dans une même espèce et un même environnement.

Dans une perspective plus large, cette étude montre que la persistance asexuée influence directement l’épidémiologie des champignons phytopathogènes. Chez *M. larici-populina*, le succès de l’infection précoce dépend fortement de la phase asexuée, qui permet à l’inoculum de survivre entre les saisons et accélère la propagation initiale de la maladie. En revanche, la reproduction sexuée réorganise la variation génétique et constitue une menace plus grande

pour la durabilité de la résistance des hôtes. Enfin, en clarifiant le comportement des indices de génétique des populations dans des populations partiellement clonales, ce travail contribue à un cadre général qui dépasse le cas des Pucciniales. Il fournit des indications pratiques pour interpréter la clonalité dans des populations naturelles présentant des stratégies reproductives mixtes et leurs conséquences évolutives associées. Dans l'ensemble, cette étude améliore notre compréhension de *M. larici-populina* et ouvre de nouvelles perspectives pour explorer le maintien évolutif de la reproduction sexuée.

Abstract

The evolution and maintenance of sex in eukaryotes remain a central question in evolutionary biology. Sexual reproduction has costs, such as finding a partner and the energy it requires, whereas asexual reproduction provides a direct, presumed less costly, alternative. This paradox can be further investigated in organisms that reproduce both sexually and asexually, such as rust fungi. In this thesis, we focus on *Melampsora larici-populina* (causing poplar rust), a plant pathogenic fungus, which alternates between a clonal stage on poplar trees and a sexual stage on larch trees. Notably, some lineages appear to bypass the sexual phase , raising evolutionary and functional questions about the balance between sexual and asexual reproduction within the same species.

We address this phenomenon on three complementary axes. First, at the population level, we investigated the presence of distinct asexual lineages in natural populations of *M. larici-populina*. Our surveys revealed that asexual lineages are more prevalent in Southern France, potentially due to mild winters that create a "green bridge," offering the opportunity for quasi-continuous clonal multiplication on green poplar leaves. To better interpret the population genetic indices used in our empirical analyses, we conducted a complementary theoretical study using simulations. This allowed us to explore how these indices do vary in the light of the pace of clone turnover, which is particularly relevant for apprehending the evolutionary dynamics of partially clonal populations.

Second, we tested whether and to what extent asexual lineages are able to reinvest in sexual reproduction under favorable conditions. We explored this through greenhouse experiments assessing the completeness of the heteroecious lifecycle and, in parallel, by microscopic examinations of nuclear fusion during teliospore formation. Some asexual lineages are still able to complete the sexual phase, while others were stopped at different developmental stages.

Third, we conducted comparative genomic and transcriptomic analyses to investigate the molecular basis of the differences between sexual and asexual lineages. We found that mating-type (MAT) genes, essential for sexual reproduction, are intact in both sexual and asexual lineages. This supports the hypothesis that asexuality in these lineages is not due to MAT gene loss. Transcriptome comparisons across developmental stages further revealed specific expression patterns linked to reproduction. Additionally, we found unique genes associated with key stages of fungal development.

In summary, this thesis reveals the existence of several long-lasting asexual lineages in *M. larici-populina* with some of them being nonetheless able to invest into sexual reproduction. In addition we report diversity at MAT-Type loci which is shared among sexual and asexual lineages. Our analyses further identify stage and lineage specific expression programs that characterize evolutionary developmental imprints along the life cycle. These findings provide new insights into the biology of rust species which is of interest for the study of some major

crop pathogens, for which disease outbreaks fully rely on the asexual phase. Finally, by investigating the dynamics of population genetic indices across generations, it contributes a broader perspective on the interpretation of partial clonality in natural populations. Altogether, this thesis advances the understanding of the life cycle of the poplar rust fungus while opening new directions for a better understanding ofthe evolutionary maintenance of sex.

Résumé

L'évolution et le maintien de la reproduction sexuée chez les eucaryotes demeurent une question centrale en biologie évolutive. La reproduction sexuée présente des coûts, tels que la recherche d'un partenaire et l'énergie qu'elle requiert, tandis que la reproduction asexuée constitue une alternative directe, présumée moins coûteuse. Ce paradoxe peut être exploré plus en détail chez des organismes qui se reproduisent à la fois de façon sésuée et asesuée, comme les champignons responsables de rouilles. Dans cette thèse, nous nous intéressons à *Melampsora larici-populina* (responsable de la rouille du peuplier), un champignon phytopathogène qui alterne entre un stade clonal sur les peupliers et un stade sexué sur les mélèzes. Il est à noter que certaines lignées semblent contourner le phasesexuée, soulevant des questions évolutives et fonctionnelles quant à l'équilibre entre reproduction sexuée et asexuée au sein d'une même espèce.

Nous abordons ce phénomène selon trois axes complémentaires. Premièrement, à l'échelle des populations, nous avons relevé la présence de lignées asexuées distinctes dans des populations naturelles de *M. larici-populina*. Nos analyses ont révélé que les lignées asexuées sont plus fréquentes dans le sud de la France, probablement en raison d'hivers plus doux créant un « pont vert », permettant une multiplication clonale quasi continue sur les feuilles de peuplier restant vertes plus longtemps. Afin d'interpréter au mieux les indices génétiques de population utilisés dans nos analyses empiriques, nous avons mené une étude théorique complémentaire à l'aide de simulations. Cela nous a permis d'examiner la variation de ces indices en fonction du rythme de renouvellement des clones, ce qui s'est révélé pertinent pour appréhender les dynamiques évolutives des populations partiellement clonales.

Deuxièmement, nous avons testé si, et dans quelle mesure, les lignées asexuées sont capables de réaliser la reproduction sexuée. Nous avons exploré cette question par des expériences en serre évaluant l'accomplissement de la phase sexuée du cycle hétéroïque avec, en parallèle, par des observations microscopiques de la fusion nucléaire lors de la formation des téliospores.

Certaines lignées asexuées sont encore capables d'accomplir la partie sexuée du cycle, tandis que d'autres sont stoppées à différents stades de développement.
Troisièmement, nous avons réalisé des analyses génomiques et transcriptomiques comparatives afin d'examiner les bases moléculaires des différences entre lignées sexuées et asexuées. Nous avons constaté que les gènes de type sexuel (gènes MAT), essentiels à la réalisation de la reproduction sexuée, sont intacts chez les lignées aussi bien sexuées qu'asexuées. Cela soutient l'hypothèse selon laquelle l'asexualité observée chez ces lignées n'est pas due à la perte des gènes MAT. Les comparaisons de transcriptomes à différents stades de développement ont également révélé des profils d'expression spécifiquement liés au mode de reproduction des lignes, hors gènes MAT. De plus, nous avons identifié des gènes uniques associés à des étapes clés du développement.

En résumé, cette thèse met en évidence l'existence de plusieurs lignées asexuées persistantes chez *M. larici-populina*, dont certaines sont néanmoins capables d'investir dans la reproduction sexuée. Nous rapportons également une diversité des loci de type MAT, diversité partagée entre les lignées sexuées et asexuées. Nos analyses identifient en outre des programmes d'expression spécifiques à certains stades et lignées, qui traduisent des empreintes développementales spécifiques tout au long du cycle de vie. Ces résultats apportent un éclairage nouveau sur la biologie des espèces de rouille, ce qui est pertinent pour l'étude d'autres agents pathogènes majeurs des cultures, dont les épidémies reposent entièrement sur la phase asexuée. Enfin, en étudiant la dynamique des indices génétiques de population à travers les générations, cette thèse apporte un cadre élargi pour l'interprétation de l'effet de la clonalité partielle dans les populations naturelles. Dans l'ensemble, cette thèse fait progresser la compréhension du cycle de vie de la rouille du peuplier tout en ouvrant de nouvelles perspectives sur une meilleure compréhension due maintien évolutif de la reproduction sexuée.

General introduction

The diversity of reproductive strategies, ranging from strict clonality to obligate sexuality, remains a central enigma in evolutionary biology (Otto, 2009; Hartfield & Keightley, 2012). Between these two poles, many taxa across the tree of life, including plants, fungi, invertebrates, and some vertebrates, exhibit the ability to undergo both sexual and asexual reproduction (Halkett et al., 2005; Kobayashi & Hoshi, 2011; Combosch & Vollmer, 2013; Dyer & Kück, 2017). Reproductive mode not only shapes the genetic structure of populations but also improves adaptability and drives the ecological success of species (Agrawal, 2006; Neiman et al., 2009).

The two key mechanisms of the sexual reproductive mode to produce viable offspring are: i) halving the parental genome to produce haploid gametes (n) through meiotic reduction, and ii) the fusion (syngamy) of two gametes (2n) (Williams, 1975; Ju et al., 2023). The main features of sexual reproduction are chromosome recombination and the production of distinct gamete types, which are regulated by mating-type or sex-determination systems (Goodenough & Heitman, 2014). This genetic shuffling generates new genotypic combinations, facilitating adaptation to environmental changes, but this comes with costs such as the “twofold cost of sex” and the energy and time needed to find a mate (Otto & Lenormand, 2002; Lehtonen et al., 2012).

Asexual reproduction, conversely, results in offspring that are identical and arise from a single parent, regardless of somatic mutation (de Meeûs et al., 2007). This not only provides an ecological advantage that enables lineages to quickly colonize favourable niches and preserve well-adapted genotypes, but it also risks the accumulation of deleterious mutations, known as Muller's ratchet (Muller, 1964; Lynch et al., 1993). Understanding the mechanisms by which organisms switch or maintain particular reproductive modes over evolutionary timescales is therefore crucial for predicting population dynamics, disease emergence, and the evolution of life-history traits (Lively & Morran, 2014).

1. The evolution of reproductive systems: insights into sexual and asexual strategies

1.1 Evolutionary history of polymorphism in reproductive modes among eukaryotes

The evolutionary history of sexual and asexual reproduction in eukaryotes represents one of the most enduring questions in evolutionary biology. While sexual reproduction is predominant across the eukaryotic tree of life, asexual lineages frequently emerge independently in many eukaryotic taxa (Bell, 1983; Barton & Charlesworth, 1998). Sexuality is hypothesized to have emerged as an ancient trait in early eukaryotic evolution, indicating that the Last Eukaryotic Common Ancestor (LECA) had a biological prerequisite for sexual reproduction (Heitman, 2015; C. Fu et al., 2019). Asexual reproduction has often and independently emerged across many of the eukaryotic taxa, showing how adaptable and important it is in different ecological contexts (Ross et al., 2013; van der Kooi & Schwander, 2014). Throughout reproductive evolutionary history, transitions from sexuality to asexuality have often occurred in association with multiple factors. These trigger factors could be changes in environmental conditions or hybridization events, which are frequently linked to ecological factors such as niche specialization and habitat stability (Rushworth et al., 2018; Lamelza et al., 2019; Yan et al., 2020).

1.2 Cost and benefits of sexual and asexual reproduction

Sexual reproduction provides several evolutionary advantages despite its costs. However, the main benefit lies in genetic recombination, which increases genetic variation and enables populations to adapt to fluctuating environmental conditions (Maynard Smith, 1978; Hörandl, 2009). Sexual reproduction also facilitates the purging of deleterious mutations through outcrossing, helping maintain genome integrity over time (Heitman, 2010; Hartfield & Keightley, 2012). The genetic diversity generated through sexual reproduction provides populations with the flexibility to respond to environmental challenges, pathogen pressure, and other selective forces (Calo et al., 2013). However, sexual reproduction carries substantial costs. These include the cost of meiosis, which breaks up co-gene combinations, the cost of males and genome dilution where sexual parents transmit only half of their genes to offspring, risks associated with

fertilization, and costs related to sexual selection (Hörandl, 2009; Maynard Smith, 1978). Additionally, sexual reproduction requires the energy and time investment necessary for finding compatible mates and producing gametes (Lehtonen et al., 2012).

Asexual reproduction offers many other advantages, particularly for organisms in stable environments or during rapid expansion phases. Asexual reproduction enables the propagation of successful genotypes without disrupting favourable gene combinations (Hörandl, 2009). It avoids the need to find a partner and reduces the energetic costs associated with gamete production (Drenth et al., 2019). The efficacy of transferring genetic material gives asexual reproduction an upper hand over sexual reproduction, especially during epidemics and invasion events, enabling rapid colonization and population growth (Smith, 1964; Drenth et al., 2019; Warren & Mokadam, 2025). Despite these advantages, asexual reproduction suffers from significant long-term disadvantages. Asexual lineages experience limited adaptive flexibility in the short-term and suffer long-term genome damage from the accumulation of deleterious mutations through generations (Lynch et al., 1993; Hartmann et al., 2021; Jiang et al., 2011; Doellman et al., 2024a). Without the benefits of genetic recombination, asexual populations become vulnerable to environmental changes and their biological enemies' evolution, potentially leading to evolutionary dead ends (Calo et al., 2013; Hojsgaard & Hörandl, 2015). Many extant species reproduce asexually, and their evolutionary trajectories are not strictly dead-end pathways; instead, they could experience multiple alternative possibilities (Tripp, 2016; Bast et al., 2018; Hoshino et al., 2024).

1.3 Reproductive mode diversity and mating systems in Fungi

One of the most fascinating models for studying the diversity of reproduction is the fungal kingdom, a broad diversity of reproductive strategies and mating systems, with many species capable of both sexual and asexual reproduction, each with distinct mechanisms, genetic control systems, and evolutionary implications (Billiard et al., 2012; Nieuwenhuis & James, 2016; Hartmann et al., 2021). This reproductive flexibility allows fungi to adapt their reproductive strategies in response to changes in the environment. (Lee et al., 2010). Fungal

sexual reproduction is controlled by Mating-Type (MAT) loci found on chromosomes labelled as sex chromosomes (Morrow & Fraser, 2009; Heitman, 2015). This system determines compatibility between individuals but does not establish male or female categories like sex chromosomes XX/XY in humans (Hartmann et al., 2021). Fungal strains are classified by mating types (e.g., "A" and "a" or "+" and "–") rather than sexual forms. These labels indicate the genetic compatibility required for sexual mating and fusion of haploid stages. Fungi display three primary mating strategies (Dodge, 1936; Langton & Elliott, 1980; Giraud et al., 2008; Billiard et al., 2011, 2012): i) Heterothallic fungi (self-sterile) require compatible partners of different mating types (e.g., MAT A x MAT a) to complete sexual reproduction, leading to obligate outcrossing, for example, in the ascomycete *Neurospora crassa* and in basidiomycota *Microbotryum violaceum* (opposite mating types (A1 × A2) (Beadle & Coonradt, 1944; Glass et al., 1988; Giraud et al., 2008); 2) Homothallic species can complete sexual reproduction from a single individual without a partner (self-fertile), such as in the ascomycete *Sordaria fimicola* (Nauta & Hoekstra, 1992; Krobanan & Shen, 2018*);* 3) Pseudo-homothallic fungi: a single individual already carries two nuclei of opposite mating types (e.g., MATa + MATα together in one spore), so when it germinates, its mycelium is compatible, similar to *Neurospora tetrasperma* (Grognet & Silar, 2015; Wilson et al., 2015; Meunier et al., 2018).

Many phylogenetic analyses demonstrate that heterothallism represents the ancestral state in most fungal lineages, with transitions to homothallism occurring repeatedly and independently throughout fungal evolution (Yun et al., 1999; Gioti et al., 2012). These transitions involve recombination events that bring both mating-type genes into a single nucleus (Gioti et al., 2012).

1.4 Mating-Type architecture in fungi

Basidiomycete fungi exhibit particularly complex mating systems that extend beyond simple binary compatibility: i) Bipolar mating system (Fig. 1 A; Passer et al., 2022) that is determined by just one genetic region (single mating-type locus), mating can only occur between individuals with different types, for example, type "A" and type "a" (Bakkeren & Kronstad, 1994; Heitman, 2015); ii) Tetrapolar mating system (Fig. 1B; Passer et al., 2022) that is determined by two

separate genetic regions (two genetically unlinked MAT loci). There are many possible mating types. To mate successfully, individuals must differ at both genetic regions, making compatibility more complex but leading to more genetic variety (Coelho et al., 2010). Nearly all Ascomycetes are bipolar, and the majority of Basidiomycota are tetrapolar (Bakkeren & Kronstad, 1994; Yun et al., 1999; Bian et al., 2025).

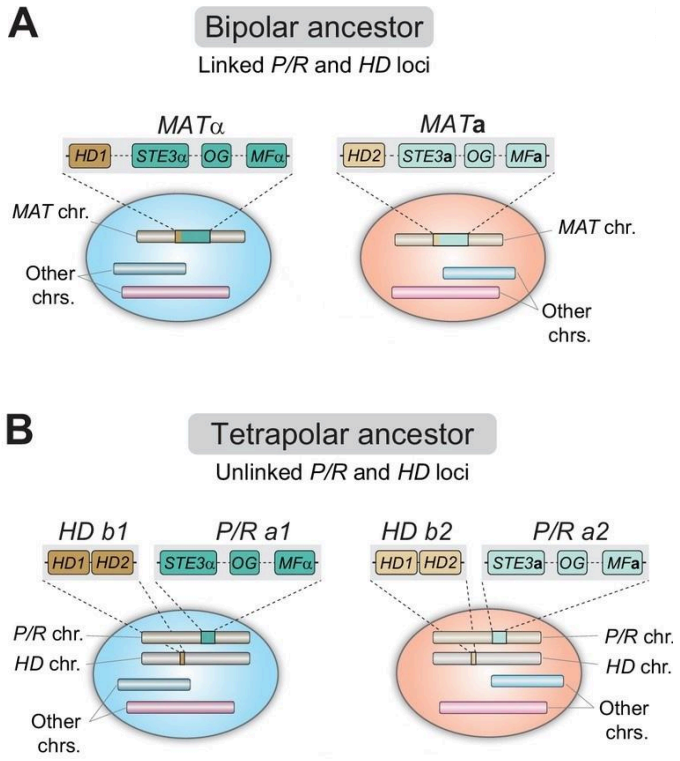


Fig. 1 The diagram illustrates the bipolar and tetrapolar Mating-Type locus organization in Basidiomycota. (**A**) Bipolar system, where the P/R (pheromone/receptor) locus and the HD (homeodomain transcription factor) locus are physically linked on the same chromosome. Two mating types are shown: MATα (carrying HD1, STE3α, OG, MFα) and MATa (carrying HD2, STE3a, OG, MFa). (**B**) Tetrapolar system, where the P/R and HD loci are unlinked and located on separate chromosomes, resulting in four possible mating-type combinations. HD1/HD2: genes encoding homeodomain transcription factors that must heterodimerize to regulate post-mating development. STE3a/STE3α: genes encoding pheromone receptors that detect mating pheromones from the opposite mating type. MFa/MFα: genes encoding mating pheromone precursors, which are recognized by STE3 receptors. OG: other genes.
Adapted from Passer et al., 2022

In tetrapolar basidiomycetes, the two unlinked loci are pheromone/pheromone-receptor (P/R) locus and homeodomain (HD) locus (Fig. 1B). During the pre-mating phase, haploid cells release

lipopeptide pheromones encoded at their P/R allele and display membrane receptors that specifically recognize pheromones from non-self-partners; reciprocal pheromone–receptor binding activates cAMP-PKA and MAP-kinase cascades that direct growth, arrest the cell cycle, and drive plasmogamy (Kües, 2000; Kaffarnik et al., 2003; Hsueh et al., 2007). Once cytoplasms fuse, the two parental nuclei coexist in a dikaryon, but further development only occurs if the nuclei carry different HD1 and HD2 alleles (non-self recognition): heteroallelic bW (HD1) and bE (HD2) proteins heterodimerize to form a key transcription factor that activates dikaryon-specific genes, governing clamp-connection formation and polarized hyphal extension (Kämper et al., 1995). Because each locus is multiallelic and segregates independently, their combinations can produce numerous compatible types, providing a powerful genetic filter that maximizes genetic diversity.

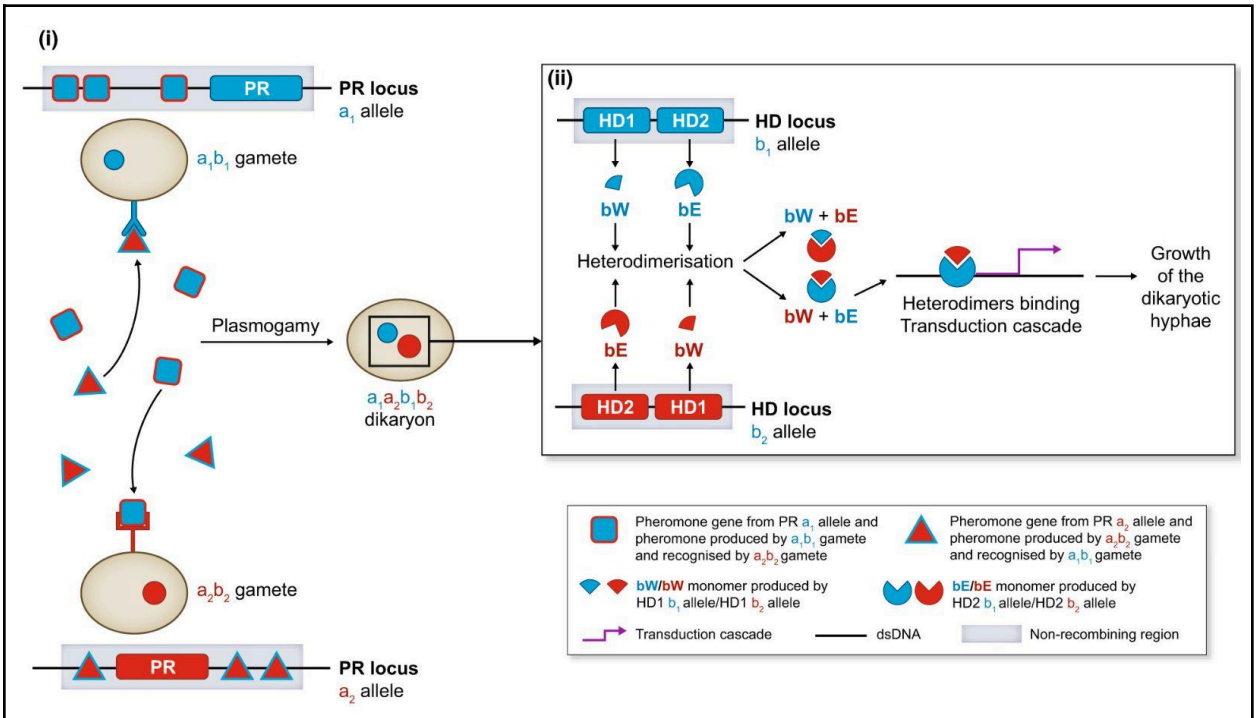


Fig. 2 Illustration showing a successful mating between two haploid cells with a tetrapolar mating system in basidiomycota. (i) PR (pheromone/receptor) locus: controls pre-fusion partner recognition step, pheromones produced by one gamete (a_1) are detected by the receptor of the opposite gamete (a_2), which leads to plasmogamy (cell fusion). (ii) HD (homeodomain) locus: regulates post-fusion development, proteins encoded by HD1 (bW)

and HD2 (bE) from different alleles (b_1 and b_2) heterodimerize. These bW/bE heterodimers activate a transcriptional cascade required for dikaryotic hyphal growth. Adapted from Figure 2 in Hartmann et al., 2021.

Many plant pathogenic fungi have evolved mixed strategies that optimize both epidemic spread through asexual reproduction and long-term adaptation through sexual reproduction (Drenth et al., 2019). The study of fungal reproductive diversity as a model could reveal novel mechanisms and evolutionary strategies. Understanding these strategies provides crucial insights for managing fungal diseases and conserving biodiversity.

2. Rust fungi as models for reproductive complexity

The order Pucciniales represents one of the most compelling examples of reproductive complexity in the fungal kingdom, encompassing approximately > 7,000 species of obligate biotrophic plant pathogens that have evolved extraordinarily intricate life cycles (Duplessis et al., 2021a; Aime & McTaggart, 2021; Kaishian et al., 2024). These fungi serve as exceptional models for understanding the evolution and mechanisms of complex reproductive strategies, particularly due to their unique combination of multiple spore stages (mainly five) and host alternation patterns (Kolmer et al., 2018; Duplessis et al., 2021a).

Life cycle and host alternation

The life cycle architecture of rust fungi represents the most complex reproductive system known among fungi. The complete macrocyclic life cycle of rust fungi involves five distinct spore stages that alternate between haploid and dikaryotic nuclear conditions, representing a level of developmental complexity unmatched in other fungal groups (Duplessis et al., 2021a). The sexual life cycle of rust fungi begins with spermatia (stage 0, n; also called pycniospores), which are produced on the gametothallus (alternate host). Fertilization at this stage gives rise to aeciospores (stage I, dikaryotic; n+n), which initiate infection of the sporothallus (primary host). Then continues by production of urediniospores (stage II, dikaryotic; n+n), which are formed asexually on the sporothallus and undergo rapid clonal multiplication during the growing season. This is followed by teliospores (stage III, initially dikaryotic but becoming diploid (2n)

after karyogamy), which function as the survival structures. Finally, after meiosis, basidiospores (stage IV, n) are produced and complete the cycle by infecting the gametothallus once again (Littlefield & Heath, 1979; Aime & McTaggart, 2021; Duplessis et al., 2021a). The rust life-cycle length is measured by the number of obligatory spore stages produced (Fig. 3; Lorrain et al., 2019): i) macrocyclic rusts generate all five spores (spermatia, aeciospores, urediniospores, teliospores, basidiospores); ii) demicyclic forms omit the uredinial stage; iii) hemicyclic forms lack the aecial phase and often the spermatial phase as well; iv) microcyclic taxa retain only teliospores/basidiospores (sometimes with residual spermatial activity) (Lorrain et al., 2019; Duplessis et al., 2021a). With two categories of host range (Fig. 3; Lorrain et al., 2019), i) heteroecious rust fungi alternate between unrelated hosts, e.g., *Puccinia graminis* alternates between barberry and cereals; ii) autoecious rust fungi complete all spore stages on a single host, e.g., *Melampsora lini* on flax species (Lawrence et al., 2007; Aime et al., 2018; Lorrain et al., 2019; Villegas et al., 2022).

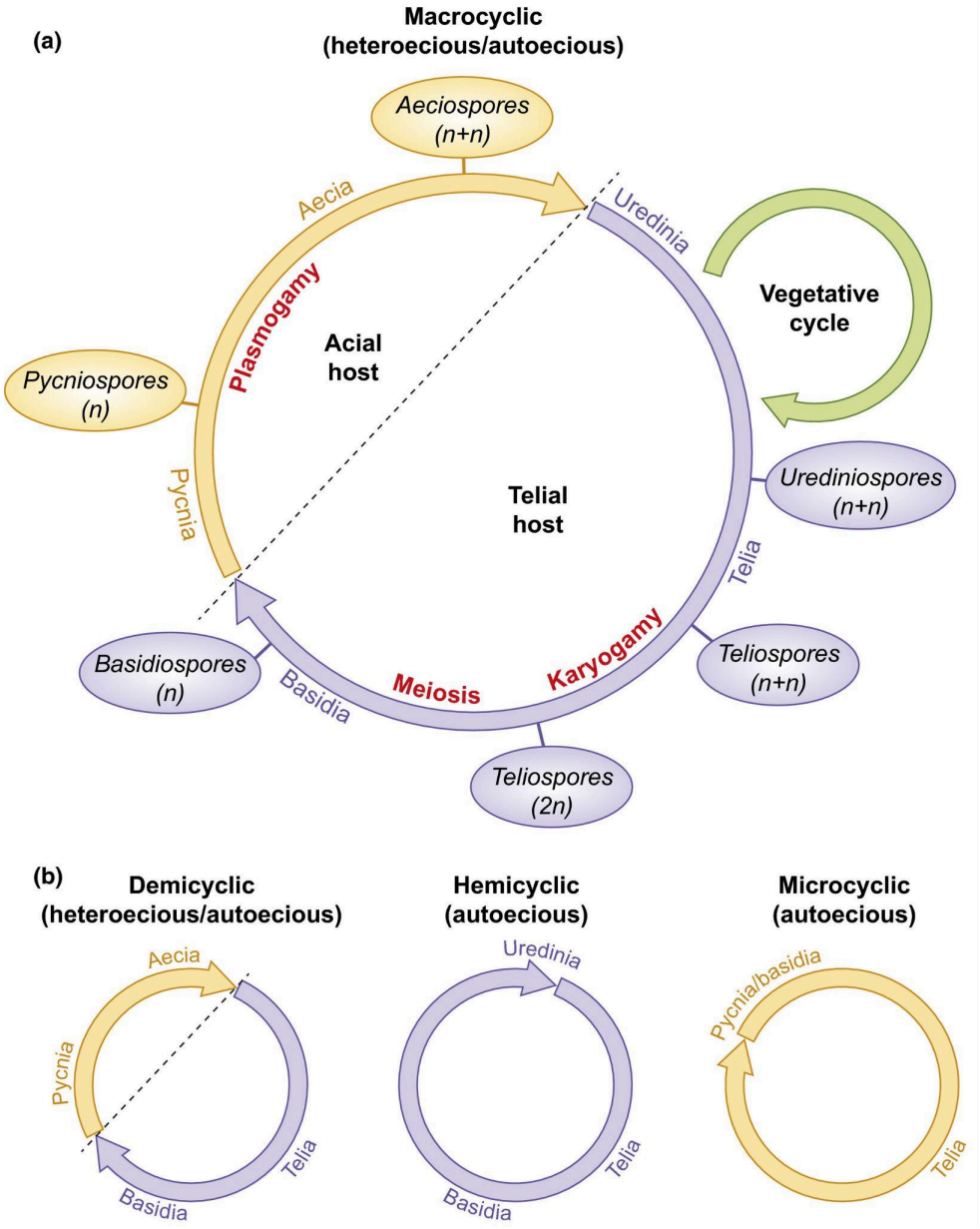


Fig. 3 The various life cycles of rust fungi. The circular schematic in the figure illustrates the rust life cycle, divided into developmental stages marked by nuclear events (plasmogamy, karyogamy, meiosis) and host shifts. (**a**) Macrocyclic cycle, the left semicircle (yellow) shows pycnial and aecial stages on the aecial (or gametothallus) host, with plasmogamy occurring when spermatia issued from a pycnial droplet fertilize receptive hyphae in another compatible pycnial droplet. The right semicircle (purple) displays dikaryotic urediniospores and teliospores on the telial (or sporothallus) host; teliospores undergo karyogamy and meiosis, producing basidiospores that infect the alternate host. (**b**) Demicyclic rusts lack

uredinia, hemicyclic rusts lack sexual phases, and microcyclic rusts only exhibit telia and basidia. Adapted from Lorrain et al., 2019.

Understanding these life-cycle architectures not only clarifies pathogen evolution but also guides integrated disease management and resistance-breeding strategies in agriculture and forestry, because it helps to understand pathogen spread, anticipate the emergence of new variants, and direct breeding for durable resistance. Continuous genomic surveillance coupled with ecological monitoring is essential to anticipate how rust fungi may further adapt their cycles in response to climate change and shifting host landscapes.

3. The heteroecious rust: *Melampsora larici-populina*

3.1 Biological and economic importance of poplar rust

Melampsora larici-populina (Mlp) (Basidiomycota, Pucciniales) is a plant pathogen that causes poplar rust disease. This disease has a widespread impact on poplar trees (Aime & McTaggart, 2021; Fabre et al., 2021). To better understand its biology and impact, it is crucial to outline the life cycle and host alternation of this rust fungus. This rust fungus is heteroecious and macrocyclic, alternating between poplar (*Populus* spp.) and larch (*Larix* spp.) trees. Asexual reproduction occurs on poplar trees (telial/sporothallus host), giving rise to both urediniospores and teliospores, while sexual reproduction takes place on larch trees (aecial/gametothallus host), resulting in the production of pycniospores and aeciospores. In contrast, other *Melampsora* species, such as *M. aecidioides*, are classified as hemicyclic rusts, featuring a truncated, two-stage asexual reproductive cycle (Lorrain et al., 2019; Duplessis et al., 2021a).

3.2 Life cycle of *M. larici-populina*

The life cycle of *M. larici-populina* can be described as follows (Fig. 4; Barrès et al., 2008; Hacquard et al., 2011). In spring and summer, *M. larici-populina* forms uredinia on poplar (the telial host), the hallmark symptom of the disease, which manifests as yellow pustules on the abaxial surface of mature leaves. Large quantities of urediniospores (the dikaryotic phase, n+n) are released in the air from the uredinia and dispersed over great distances. Throughout spring

and summer, several clonal multiplication cycles can occur on poplar leaves, and in the autumn, black telia pustules containing dikaryotic teliospores (n+n) differentiate on senescent poplar leaves. Then, teliospores undergo karyogamy (2n) and begin meiosis in decaying leaves. After overwintering, meiosis resumes, and teliospores produce basidia, which in turn produce haploid basidiospores (n). Upon spreading by wind or water, these spores infect larch needles (the aecial host). The infection on the aecial host leads to the production of the sexual structures, spermogonia, and spermatia, also called pycniospores (n) are formed. Spermatia, which are non-motile gametes, are produced in liquid nectar droplets issued from spermogonia. Spermatia from one spermogonium can be transmitted by small insects, wind, or water to another spermogonium of the opposing mating type containing haploid receptive hyphae. The plasmogamy of haploid cells from contrasting mating types results in the formation of dikaryotic aecia and dikaryotic aeciospores (n+n). These wind-borne spores then infect poplar leaves, leading to the formation of uredinia.

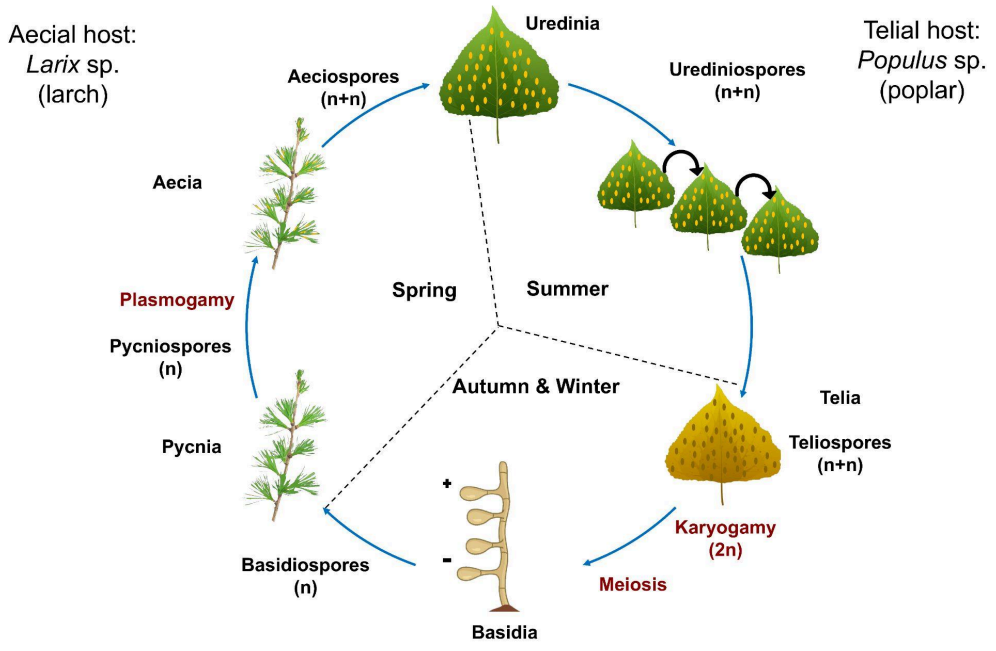


Fig. 4 The heteroecious life cycle of *Melampsora larici-populina*.

3.3 Genomic Landmarks of *M. larici-populina*

Melampsora larici-populina was one of the first two rust fungi whose genome was sequenced (Duplessis, Cuomo, et al., 2011). Two curated reference versions of the genome were successively available through the Joint Genome Institute (JGI) MycoCosm portal (Duplessis, Cuomo, et al., 2011; Persoons et al., 2022). The genome releases featured ~101 then ~109 Mbp assemblies derived from the sexual reference isolate 98AG31. The genomic landscape analysis showed that transposable elements constitute approximately 45% of the total genome size.

3.4 Previous findings: evidence of persistent clonal lineages

The team *Ecology of Forest Pathogenic Fungi* in the Department *Interactions Arbres-Microorganismes* (UMR 1136 IAM) conducted several population genetic analyses to investigate the consequences of poplar resistance breakdown on the *M. larici-populina* populations (Xhaard et al., 2011; Persoons et al., 2017; Louet et al., 2023). By doing so, Xhaard et al. (2011) made a peculiar discovery. This study focused on populations sampled on wild poplars, where it was suspected that admixture occurred between two genetic groups: the native “wild” genotypes that co-evolved with wild poplars and the new “cultivated” genotypes that derived from the foundation of a population after the *RMlp7* resistance breakdown. The research involved the phenotyping of 1,050 individuals for their pathogenicity profiles, and a subsample of 476 individuals was genotyped using 25 microsatellite markers. The population genetic analysis on the genotyped individuals found a partition between two main genetic groups: group 1, consisting of virulent individuals with signs of selection and severe demographic changes, and group 2 consisting of individuals from an ancestral population of *M. larici-populina* in the native range. But this study further highlighted the existence of a third genetic group. Unlike the two other groups, group 3 showed no direct evidence for gene flow, with clear-cut assignment profiles (Fig. 5; Xhaard et al., 2011).

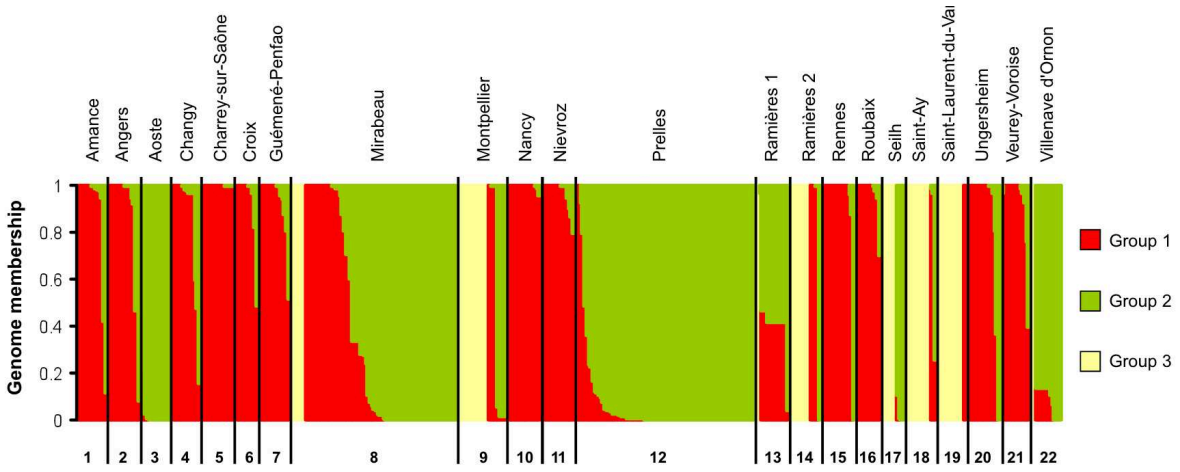


Fig. 5 Population genetic structure of *Melampsora larici-populina* in France and northern Italy. A total of 514 individuals sampled from 22 locations (collected August–November 2009) were genotyped at microsatellite markers. Each vertical bar represents one individual, with colors indicating proportional genome assignment to three genetic groups (K = 3). Most individuals showed strong membership in a single group, defining three main genetic clusters. Adapted from Fig. 2 in Xhaard et al., 2011.

In addition, a detailed analysis of population genetic indices indicated that group 3 displays the hallmark of strong asexual reproduction: reduced genotypic diversity, significant linkage disequilibrium, and above all a negative value of inbreeding coefficient F_{IS} value which is indicative of an excess of heterozygotes, a distinctive feature of long-lasting asexual reproduction (Balloux et al., 2003; Halkett et al., 2005; Stoeckel & Masson, 2014). This typically contrasted with group 1, which showed no significant departure from genetic equilibrium. Group 3 displayed genetic signatures of a potentially long-lasting asexual reproductive mode, which was previously unreported in this species.

Further analyses using unpublished data from the lab revealed that the asexual group consists of several lineages (Table 1). At least four of these lineages have been resampled through the years. This provides direct evidence for their survival without the need to reproduce sexually once a year, as is the case for some hemicyclic *Melampsora* species (Vialle et al., 2013). Some lineages have been resampled at 15–20 year intervals, which provides evidence of certain ecological success.

Table 1: Details of the genotypes of different asexual lineages

MLL	Lineage Size	N(MLG)	R	c(Pareto)	Year
1	43	5 (1;2;36;2;2)	0.10	1.05068	1996/2008/2009/2011
2	129	12 (1;3;1;4;25;56;4;1;1;4;2)	0.09	1.529933	1992/2008/2009/2011/2014
3	23	8 (3;1;2;1;1;2;1;2)	0.32	1.276653	1996/2008/2009/2011
4	37	13 (1;3;1;2;1;6;1;1;1;1;1;1;1)	0.33	1.27695	1996/2008/2009/2011
5	3	1	-	-	2011
6	1	1	-	-	2009
7	1	1	-	-	1994
8	1	1	-	-	1993
9	3	1	-	-	1995
10	6	5 (2;1;1;3;1)	0.57	-	1993/1995

MLL: Multilocus Lineage. Lineage Size: Number of isolates belonging to the lineage. N(MLG): Number of distinct Multilocus Genotypes within the lineage. *R*: Clonal Richness Index. c(Pareto): Pareto distribution parameter. Year: Years in which isolates of the lineage were sampled

These findings provide important insights into the population genetics of asexual lineages and shed light on the mechanisms driving the existence of asexual lineages. Further research is needed to understand the implications of clonal interference in the evolution of pathogenicity and disease transmission.

4. Challenges and open questions

The mechanisms behind transitions between sexual and asexual reproduction in *M. larici-populina* remain poorly understood, highlighting a knowledge gap that is important for our understanding of rust biology, but also more broadly in the evolutionary biology of fungi. This thesis outlines three interconnected research priorities that emerged from current knowledge gaps in the field.

4.1 Population structure and genetic isolation of reproductive modes

An essential question concerns whether asexual and sexual lineages of *M. larici-populina* represent genetically distinct populations or if rare sexual events introduce genetic variation. What is the turnover rate of these asexual lineages? A comprehensive sampling collection and a population genetic study are needed to determine this.

4.2 Sexual competence and functional capacity for reproduction

Even if mating-type machinery is genomically intact, a critical question is whether asexual lineages retain the functional capacity to complete the sexual cycle under appropriate conditions. This involves testing whether such isolates can initiate mating responses, progress through meiosis, and produce sexual spores when subjected to favorable conditions. Many fungi require specific environmental conditions, nutrient starvation stress, or availability of host to trigger sexual reproduction (Meng et al., 2015; Wallen & Perlin, 2018a), yet the ecological contexts that favor sexual versus asexual reproduction in *M. larici-populina* have not been systematically characterized. The microscopic examination to evaluate the potential for nuclear fusion, and the laboratory experiment to assess the ability of asexual isolates to undergo sexual reproduction are needed.

4.3 Genomic and transcriptional basis of reproductive mode polymorphism

At the molecular level, key questions address whether reproductive mode differences result from genomic degradation essential for sexual reproduction. Do asexual lineages exhibit loss of

function of mating-type (MAT) loci or other essential genes, or do they retain intact genetic architecture with a different transcriptional program? A comparative analysis of MAT locus integrity and expression profiles across various developmental stages of fungi is essential to understand the genomic basis of reproductive mode variation in poplar rust.

These interconnected questions form the foundation for understanding how reproductive mode polymorphism and the asexual lines arise and persist in natural populations of poplar rust.

5. Thesis objectives and structure

5.1 Evolutionary ecology approach to map sexual vs. asexual populations

The evolutionary approach will employ population genetics and phylogeographic methods to characterize the geographic distribution and evolutionary relationships of sexual and asexual *M. larici-populina* lineages. This objective addresses fundamental questions about the evolutionary origins and maintenance of reproductive diversity within a pathogenic fungal species.

1. Population sampling strategy: Comprehensive sampling across the natural range of *M. larici-populina* will target diverse ecological contexts across France, including different climatic zones and regions with varying intensities of poplar cultivation.
2. Molecular marker: High-resolution genetic analysis requires markers capable of distinguishing between sexual and asexual reproduction and detecting fine-scale population structure. For that, we used microsatellite markers specifically developed for *M. larici-populina*.
3. Comprehensive population genetic study: The analyses will reconstruct the evolutionary history of reproductive mode transitions, testing whether asexual lineages represent single or multiple independent origins from sexual ancestors.
4. Spatial analysis: A Map of France will integrate genetic data to identify the potential geographic ranges of sexual *versus* asexual lineages under current climates.

5.2 Phenotypic assays of sexual capacity in asexual lineages

The phenotypic approach represents an important experimental component designed to directly assess the reproductive capacity of putatively asexual lineages identified through population genetic analysis. This objective addresses the fundamental question of whether asexual *M. larici-populina* lineages retain functional sexual machinery or have lost the ability for sexual reproduction entirely. By using microscopic examination techniques and controlled infection experiments, this method will assess the sexual capacity of representative asexual lineages.

1. Microscopic analysis of nuclear dynamics: The primary experimental axis employs cytological techniques to examine nuclear fusion dynamics and karyogamy in asexual versus sexual lineages.
2. Life cycle completion assays: The second experimental focus is on testing the ability of asexual lineages to complete the full heteroecious life cycle under controlled conditions. Representative isolates from major asexual lineages identified in the population analysis will be subjected to systematic infection studies designed to assess each critical stage of sexual development.

5.3 Genomic and transcriptomic analysis of reproductive mode variation

The functional approach will use genomic and transcriptomic methods to identify the molecular basis of reproductive mode differences and characterize the genetic architecture underlying sexual versus asexual reproduction in *M. larici-populina*. This objective represents the most technically challenging but potentially most informative component of the research program.

1. Whole-genome sequencing of representative sexual and asexual isolates will identify structural variants. Particular focus will be placed on Mating-Type loci controlling sexual development.
2. Transcriptomic profiling: RNA sequencing across different developmental stages that will characterize gene expression programs associated with sexual *versus* asexual reproduction modes. Time-course experiments will identify the regulatory cascades and

determine whether asexual lineages show altered expression in comparison to a sexual reference isolate.

5.4 Overview of thesis chapters and their interconnections

This thesis integrates evolutionary, phenotypic, and functional approaches to provide comprehensive insights into reproductive mode polymorphism in *M. larici-populina.* The three research axes are designed to reinforce each other and started independently and in parallel at the beginning of the thesis project.

- **Chapter integration strategy:** The evolutionary approach (5.1) will identify candidate populations and lineages for detailed phenotypic and genomic analysis. Phenotypic assays (5.2) will test reproductive mode classifications derived from population genetic analysis and provide qualitative/quantitative traits for genomic and transcriptomic studies. Functional genomics (5.3) will test hypotheses generated by evolutionary and phenotypic analyses and provide a genomic overview and a transcriptomic program as background to create a comprehensive picture.
- **Temporal dimension:** All approaches will include temporal sampling to capture the dynamic nature of reproductive mode evolution.
- **Methodological innovation**: This research will improve approaches for studying reproductive mode polymorphism in rust fungi by developing new population genetic analysis workflows, phenotyping protocols, and bioinformatic pipelines that can be used in other pathosystems.
- **Broader Implications:** Beyond understanding the reproductive biology of poplar rust fungi, this research will inform applied questions about pathogen evolution, disease management, and the evolutionary consequences of agricultural practices.

The interconnected nature of these approaches ensures that the thesis will provide novel insights into the reproductive mode polymorphism evolution while advancing both theoretical understanding and practical applications in plant pathology and evolutionary biology.

Chapter one

Population Structure and Theoretical Modeling of Sexual vs. Asexual Reproduction

Section one

Long-lasting coexistence of multiple asexual lineages alongside their sexual counterparts in a fungal plant pathogen

Declaration

The work presented in this section one has been submitted as a research article to the journal ***Molecular Ecology***, where it is currently under peer review.

Long-lasting coexistence of multiple asexual lineages alongside their sexual counterparts in a fungal plant pathogen

Ammar Abdalrahem[1], Axelle Andrieux[1], Ronan Becheler[1,2], Sébastien Duplessis[1], Pascal Frey[1], Benoit Marçais[1], Kadiatou Schiffer-Forsyth[1], Solenn Stoeckel[3], Fabien Halkett[1]

[1] Université de Lorraine, INRAE, IAM, F-54000 Nancy, France
[2] INRAE, Agrocampus Ouest, Université de Rennes, IGEPP, Le Rheu, France
[3] DECOD (Ecosystem Dynamics and Sustainability), INRAE, Institut Agro, IFREMER, Rennes, France

Abstract

Cases of sexual to asexual transitions within species are key to understanding the evolution of reproductive modes, yet the coexistence of both modes in the same environment is poorly documented. Here, we reported this transition in a plant pathogen species coexisting in the same geographical environment. Our biological model is the poplar rust fungus *Melampsora larici-populina,* which displays a complex life cycle typical of rust fungi (Pucciniales). It alternates between two unrelated hosts to complete an obligate sexual life cycle once a year. We conducted a comprehensive population genetic analysis, utilizing 21 microsatellite markers and data from 2,122 individuals gathered over 30 years from various locations in France. Our results demonstrate the existence of many distinct lineages that reproduce asexually through the years, skipping the sexual phase. A clustering analysis identified a group of multilocus lineages that displayed all hallmarks of the genetic consequences of asexual reproduction, including highly negative and large variance among loci of the inbreeding coefficient (F_{IS}). This indirect evidence for asexual reproduction was confirmed by the direct observation of these asexual lineages being repeatedly sampled across multiple years. This result demonstrates the coexistence of these lineages with their sexual counterparts in the same ecological niche, challenging conventional assumptions about geographical sorting of reproductive modes. These considerations are of paramount importance for understanding the contemporary evolution of major pathogen species. This switch from sexual to asexual reproduction has contributed to devastating epidemics worldwide.

Keywords: sexual-asexual transition | evolution of reproductive modes | population genetics | rust fungi | time samples.

1 Introduction

The evolution of sex remains a fundamental question and an intriguing topic in evolutionary biology (Culotta & Hines, 1998; Otto, 2009; Hartfield & Keightley, 2012). Sexual reproduction, which has benefits and costs, is widespread in eukaryotic organisms (Schurko et al., 2009). The key benefit is increased genetic diversity, purging deleterious mutations, and enhancing adaptation to changing environments (Pierre et al., 2022). However, it incurs costs such as time and energy to find mates, combined with the biological cost of meiosis (Roughgarden, 1991; Aanen & Hoekstra, 2007; Hörandl, 2013). Asexual reproduction, on the other hand, produces offspring without a mate, allowing for efficient reproduction, maintenance of lineages adapted to local stable conditions, and quick multiplication (Shick et al., 1979; Lodé, 2013). The offspring is genetically identical to the parent, barring somatic mutations, which renders asexual lineages more vulnerable to the accumulation of deleterious mutations due to a lack of recombination (Lynch et al., 1993; Peck, 1994; Keightley & Eyre-Walker, 2000), and more prone to loose allele diversity by genetic drift (Rouger et al., 2016). The study of species displaying both sexual and asexual reproduction enables pinpointing the pros and cons of each mode of reproduction (Butlin, 2002; Simon et al., 2010; Billiard et al., 2012). In particular, the study of cases where there is a dynamic transition from sexual to asexual reproductive modes is likely to highlight scenarios in which each reproductive mode could be favored.

The success of asexual reproduction frequently relies on specific environmental conditions. Remarkably, invasive species often shift towards asexual reproduction in their exotic range, while reproducing sexually in their native environment (Song et al., 2011; Platt & Jeschke, 2014; Tilquin & Kokko, 2016; Gladieux et al., 2015). Ant species can exhibit such diverse reproductive strategies, with asexual reproduction frequently being dominant in invasive populations, as observed in *Wasmannia auropunctata* and *Cerapachys biroi* (Foucaud et al., 2006; Kronauer et al., 2012). Plant pathogenic fungi also adopt diverse reproductive strategies, with asexual reproduction being responsible for the most damaging epidemics (Ashu & Xu, 2015; Gladieux et al., 2015; Drenth et al., 2019). Potato late blight is a devastating disease caused by the oomycete pathogen *Phytophthora infestans*. Both sexual and asexual reproduction take place in the native range, but it is asexual reproduction only that led to the devastating Irish Potato Famine in the 1840s (Gavino et al., 2000; Ristaino, 2002; Fry, 2020). The wheat yellow rust fungus *Puccinia*

striiformis f. sp. *tritici* displays a similar pattern, with different reproductive modes depending on the region. The fungus reproduces sexually in China and Pakistan, which is highlighted by its high genotypic diversity, whereas it reproduces asexually in the rest of the world (Bahri et al., 2009; Mboup et al., 2009; Ali et al., 2014). The rice blast fungus *Magnaporthe oryzae* (Syn: *Pyricularia oryzae*) reproduces sexually in a limited area of Southeast Asia, its native range, whereas epidemics globally are driven by asexual lineages (Saleh et al., 2012).

The coexistence of both reproductive modes in the same geographical environment is still poorly documented, despite it is offering a unique opportunity to decipher the evolutionary and ecological dynamics of sexually or asexually derived lineages. While such studies are rare on fungal biological models, they have been documented in other taxa, such as in aphids (Halkett et al., 2006; Simon et al., 2010), daphnia (Lehto & Haag, 2010), and a cyprinid fish *Carassius gibelio* (Barbuti et al., 2012). The phytopathogenic fungus *Melampsora larici-populina*, causing poplar rust, represents an ideal candidate for elucidating the transition from one reproductive mode to another, as both modes can be found in the same environment. *M. larici-populina* belongs to the taxonomic order of Pucciniales (Steenackers et al., 1996). This species follows a complex life cycle with an obligate sexual reproduction that alternates with rounds of clonal multiplication (Figure S1). Two unrelated host plants are required to complete the life cycle (Hacquard et al., 2011). Clonal multiplication occurs on poplar trees (*Populus* spp.), while sexual reproduction takes place on larch trees (*Larix* spp.) (Lorrain et al., 2019). In the following, lineages that complete the entire life cycle, including the obligate sexual reproduction phase, are referred to as sexual lineages. The sexual reproduction on larch ensures genetic recombination and obliterates all traces of the previous rounds of clonal multiplication (Xhaard et al., 2011; Rouger et al., 2016). The impact of this fungal disease on larch is insignificant, whereas the many rounds of clonal multiplication that occur on poplar leaves from spring to autumn can be responsible for strong epidemics with large nutrient uptake by the fungus and perturbation of photosynthesis (Jiang et al., 2016). Disease symptoms and damage are greatest on cultivated poplars and resemble those of cereal rusts caused by several *Puccinia* species. A previous population genetic study (Xhaard et al., 2011) has provided evidence for some variability in reproductive mode, with an obligate sexual genetic group coexisting with the genetically distant group that displays all hallmarks of long-lasting asexual reproduction. This suggests that some

lineages of the poplar rust fungus could survive alone through clonal multiplication on poplar. Yet, this preliminary study did not provide detailed investigations on the ecological and evolutionary dynamics of this asexual group, largely because it was based on a single year of sampling.

In this study, we provide an in-depth characterization of the reproductive strategies of *M. larici-populina*, with an emphasis on the (spatio-temporal) persistence of asexual lineages and their genetic characteristics. Our study also examines the ecological and spatial factors influencing the distribution of sexual and asexual populations, thereby offering insights into the evolution of reproductive modes in rust fungi. This study encompasses the analysis of 2122 individuals sampled across France over 30 years of population surveys that have been genotyped using 21 microsatellite markers.

We used a two-fold approach to assess the reproductive mode variation among poplar rust individuals. First, we employed an indirect population genetic approach based on the clustering among lineages and subsequent evaluation of the variation in population genetic indices known to be signatures of asexual reproductive mode, namely: repeated genotypes, linkage disequilibrium, and overall negative F_{IS} value and large variance over loci (Balloux et al., 2003; Halkett et al., 2005; Stoeckel & Masson, 2014). Second, we benefited from time-samples to investigate the resampling through years of the same multilocus lineage (MLL), which is direct evidence for their asexual survival. Furthermore, the genetic relatedness of all individuals was analyzed to investigate evolutionary relationships among asexual lineages and with their sexual counterparts, which points to multiple events of the emergence of asexual lineages. Finally, examining the coexistence of sexual and asexual lineages, their long-term persistence, and their spatial distribution suggested that environmental and ecological factors potentially play a role in shaping the survival and maintenance of asexual lineages in this species.

2 Material and Methods

2.1 Isolate collections

In this study, we examined 2122 individuals of *M. larici-populina* sampled from 61 “départements” across France over 32 years (1992-2024). The dataset (Table S1) includes previously published genotypes from the study by Xhaard et al. (2011), sampled from a countrywide survey conducted in France in 2009 (23 locations | 428 samples), and more recent samples with genotype data original to this study. These recent additions comprise i) samples from a second countrywide survey conducted in 2011 (60 locations | 1374 samples), ii) individuals from the laboratory collection (1992, 1993, 1994, 1996, 1998, 2008, and 2014 | 34 samples), and genotypes from a project monitoring poplar rust epidemics in the Durance River valley (Xhaard et al., 2012; Becheler et al., 2016) that were suspected to belong to the asexual group and thus not included in the previously cited publications (2004, 2008 and 2020 | 246 samples). All these samples were collected in late summer, at the end of the clonal multiplication phase, therefore before leaf fall and lineage engagement into sexual reproduction (Figure S1). In addition, we benefit from the observation of a distinctive poplar stand in Cavalaire-sur-Mer (south of France) that maintains green leaves until mid-winter, where samples were collected in February 2024 (40 samples).

2.2 Genotyping

The DNA extraction process was guided by a protocol that employs the BioSprint 96 DNA Plant Kit and the BioSprint 96 automated workstation (Qiagen). Samples were genotyped using 21 microsatellite loci [Mlp_91, Mlp_83, Mlp_50, Mlp_71, Mlp_96, Mlp_95, Mlp_55, Mlp_66, Mlp_12, Mlp_68b Mlp_54, Mlp_57, Mlp_58, Mlp_49, Mlp_93, Mlp_82b Mlp_56, Mlp_87, Mlp_97, Mlp_77, Mlp_94 (Xhaard et al., 2009)]. These loci have been extensively utilized in several population genetic studies of this species (Xhaard et al., 2011; Becheler et al., 2017; Persoons et al., 2017; Louet et al., 2023; Saubin et al., 2024). The amplification through PCR with the Multiplex PCR Kit (Qiagen) was performed as described by Xhaard et al. (2011). It was conducted across three multiplex reactions containing eight, five, and eight loci, ranked by amplification size, respectively. The PCR products were then pooled and analyzed using an ABI 3730 Genetic Analyzers (Applied Biosystems) by the company GenoScreen. Fragment sizing

utilized a LIZ-1200 size standard, and allele scoring was performed with GENEMAPPER V4.0. Samples that failed to amplify at more than six loci were excluded from further analysis.

2.3 Population genetic and statistical analyses

First, to handle genetic data of microsatellite markers in the R programming language v.4.4.1, we converted the data format from a data frame to a genuine object using the `df2genind` function from the **adegenet** package v.2.1.10. This function optimizes multilocus genotype data and accommodates missing data if it is found in some individuals or loci. The automated workflow code provides full details on subsequent analyses (https://github.com/ammarabdalrahem/poplar_rust.git). Moreover, specific analyses were executed using the GenAPoPop 1.0 software due to specific methodological requirements.

Identification of repeated genotypes: The prerequisite to an accurate assessment of reproductive mode is the proper delineation of repeated genotypes derived from clonal multiplication (Halkett et al., 2005; Arnaud-Haond et al., 2007). Individuals with the same combination of alleles (the so-called multilocus genotypes, MLG) were consistently identified through both an ad hoc procedure in Excel during the manual curation of chromatographs and using `mlg.filter` from the **poppr** package v.2.9.3 (Kamvar et al., 2014). Genotypes differing by a few alleles were double checked by returning to the original chromatographs. All remaining differences between genotypes are therefore confirmed and reliable. We computed (*Pidu*) and (*N_Pidu)* to test for the discrimination power of our marker set, and to rule out the probability of finding repeated MLGs by chance (that is deriving from sexual reproduction) using GenAPoPop 1.0 (Evett & Weir, 1998; Waits et al., 2001; Stoeckel et al., 2024).

Delineation of multilocus lineages: As the accumulation of somatic mutations over time can lead to slight differences among multilocus genotypes within a given asexual lineage (Arnaud-Haond et al., 2007), we subsequently grouped MLGs into multilocus lineages (MLL). So as not to apply an arbitrary threshold for MLL delineation, we first computed the distribution of pairwise shared allele distance as advised by Arnaud-Haond et al. (2007), and Bailleul et al. (2016). As intra lineage differentiation can vary across MLLs, we chose to use the UPGMA clustering method, which aggregates MLGs one by one by examining the nearest neighbors until the MLL

boundaries are found (Kamvar et al., 2014, 2015). The threshold on genetic distance between neighbor genotypes was set using the function `cutoff_predictor` from **poppr**.

Clustering procedure to indirectly identify asexual lineages: As the asexual reproductive mode freezes genotypes, it retains ancestral allelic states. Asexual lineages can thus rapidly diverge from their sexual counterparts, the latter being still influenced by genetic drift (Rouger et al., 2016). In line with this statement, we previously found large genetic differences between sexual and asexual groups of *M. larici-populina* (Xhaard et al., 2011). Assigning poplar rust individuals would thus be an effective approach to identify asexual lineages without any *a priori*. The main pitfall of this method being that the clustering algorithm should not rely on gathering individuals according to Hardy-Weinberg expectations, which is the common assumption of any Structure-like approach (Pritchard et al., 2000). We thus carry out our clustering analysis using a PCA and discriminant analysis approach based on allelic distribution (Jombart et al., 2010). We used the `find.clusters` function from the **adegenet** package to reduce the dimensionality of the dataset. Discriminant analysis of principal components (DAPC) was then used to refine the identification of clusters by maximizing the variance between two clusters, while minimizing within cluster variance (Jombart et al., 2010). Nonetheless, 97.86% of genetic variation among individuals was captured by retaining 90 principal components (n.pca = 90). Individuals were grouped into two clusters using the `dpca` function with K=2. To ensure the reliability of the assignment, only individuals with a high membership coefficient (≥0.8) were retained for further analyses of within-cluster genetic characteristics.

Resampling multilocus lineages (MLLs) as direct evidence for asexual reproduction. Observing the same genotype over several years is an indication that this lineage survived asexually - unless sexual reproduction could have resulted in the same combination of alleles by chance, which can be ruled out by a *Pidu* analysis (Evett & Weir, 1998; Stoeckel et al., 2024). To track the persistence of MLLs over time, we recorded the years each MLL was detected and calculated its frequency across different years. A lineage that persisted for at least two years was considered asexual. We tested the overlap between the direct and indirect approaches for assessing asexual reproductive mode by performing Fisher's exact test (Clarkson et al., 1993) on the contingency table of MLL sorting according to the direct or indirect approach by using the `fisher.test` function from the **stats** package v.4.4.1. In the following, a given MLL is considered asexual if it

meets at least one of the two criteria: the assignment to an asexual cluster or the inter-annual resampling over a minimum of two years.

Computation of classical population genetic indices: As asexual reproductive mode distorts allele repartition within and among individuals, we computed several classical population genetic indices used to characterize within cluster and within MLL genetic signatures. Observed heterozygosity (H_O), unbiased expected heterozygosity (H_E) (Nei, 1978), genotype richness (R), and multilocus linkage disequilibrium corrected for the number of loci $\bar{r}_d$ were calculated using the **poppr** package (basic function). Allelic richness (A_R), and F-statistics, including the inbreeding coefficient (F_{IS}) (Weir & Cockerham, 1984) were computed using the **hierfstat** package v.0.5-11. The clonal evenness was measured by calculating the Pareto β index, which summarizes the distribution in size of MLLs using GenAPoPop 1.0 (Arnaud-Haond et al., 2007; Stoeckel et al., 2024).

Spatial analysis of the repartition of sexual and asexual lineages: To illustrate the geographical distribution of sexual and asexual individuals across France, the **maps** package v3.4 was utilized. This enabled us to obtain information not only on spatial distribution, but also on the potential coexistence of the two reproduction modes within the same geographical area. Using the `glmer` function from the **lme4** v.1.1-35.5 package, a generalized linear mixed model (GLMM) with a binomial distribution and a logit as a link function was employed to assess the regression coefficients (Bolker et al., 2009). To further test the relationship between the reproductive mode variation and the geographic locations, we performed a regression analysis on the count of sexual and asexual lineages per site. The independent variables (fixed effect) were the geographic location represented by latitude and longitude. The effect of the year of sampling was added as a random effect. This analysis was performed on the 2009 and 2011 data originating from the two countrywide surveys, so as not to overrepresent certain locations. To further assess the distribution and relative abundance of asexual MLLs across time and space, we performed Fisher's exact test for fluctuation in MLL abundance across years and locations. Given the large size of the data sets, we employed a Monte Carlo simulation approach with 100,000 replicates to estimate p-values.

Genetic relationships between and among sexual and asexual lineages: To further elucidate the evolutionary trajectories of sexual and asexual individuals, and to address the question of the dynamics of transition from sexual to asexual reproduction and vice versa, a discrete matrix based on the relative dissimilarity between individuals was calculated by `diss.dist` with the **poppr** package to create a neighbor-joining tree (Saitou & Nei, 1987) using `nj` from the **ape** package v5.8.1.

Inference of the rate of clonality: Finally, to estimate the rate of clonality in the *M. larici-populina* populations, we employed the ClonEstiMate Bayesian method. This method analyzes the change of genotype frequencies from the same population over generations (Becheler et al., 2017). We calculated based on two scenarios: i) merging all individuals within a mixed population (that is any individuals can survive asexually by chance), or ii) considering two distinct sexual and asexual populations, based on the lineage sorting detailed above. Since the ClonEstiMate method should be applied on temporal samplings of the same kind of populations, we used the data collected during the countrywide surveys carried out in the years 2009 and 2011.

3 Results

3.1 Delineation of multilocus lineages

We identified 1475 multilocus genotypes (MLGs) over the 2121 genotyped individuals. The 21 markers we used provided enough statistical power to ensure unambiguous discrimination between genotypes. The overall probability of obtaining the same combination of alleles by chance under panmictic sexual reproduction (*Pidu*) is approximately 10^{-16}. This implies that 118 million individuals would have to be genotyped before observing two sexually produced descendants with identical genotypes (*N_Pidu*). We thus confidently considered that all the repeated genotypes we found were clones. We then grouped closely related MLGs into MLLs to account for the few somatic mutations that could occur during clonal multiplication. The delineation of MLLs was undertaken using a genetic distance threshold of 4.6, as determined by the UPGMA clustering algorithm. This grouping led to the delineation of 1384 MLLs, so slightly fewer than the number of MLGs, showing that the vast majority of MLGs have no closely related genotype.

3.2 Reproductive mode assessment

The clustering approach indicated that almost all individuals exhibited a high probability of membership in one of two genotypic clusters: 445 individuals were unambiguously assigned to cluster one, while 1676 individuals were assigned to cluster two (Figure S2). Only one isolate had membership coefficients below the threshold and was excluded from further analyses.

A comprehensive population genetic analysis of each cluster (Table 1) was conducted to assess the value of key population genetic indices, thereby revealing a contrast between the two clusters. Cluster 1 exhibits a highly negative F_{IS} value, low genotypic diversity, and high linkage disequilibrium (LD), all of which are indicative of a strong rate of clonal reproduction (Balloux et al., 2003; Halkett et al., 2005; Arnaud-Haond et al., 2007). The high variance of F_{IS} across loci further points to rare events of recombination within a population reproducing essentially asexually over the long term. In contrast, cluster 2 exhibited slightly positive F_{IS} values and higher genetic diversity, suggesting a higher degree of allelic recombination, a feature of sexual reproduction. The reduced linkage disequilibrium in this cluster further supported this

classification. Consequently, cluster 1 was labeled as asexual, and cluster 2 as sexual. It is noteworthy that similar levels of observed heterozygosity were found in both clusters, but a much reduced allelic richness and gene diversity in the asexual cluster 1. Further, the *Pareto β* index was low (~ 0.2), indicating a small number of dominant MLGs, which was consistent with high clonality in cluster 1. In the sexual cluster 2, the *Pareto β* was greater than 1, reflecting a more even distribution of genotypes due to genetic recombination.

Table 1. Characteristics of the two genetic clusters identified in *M. larici-populina* individuals.

	N	MLG	MLL	*R*	*Pareto β*	A_R	H_O	H_E	F_{IS}	*Var* (F_{IS})	$\bar{r}_d$
Clustering approach											
Cluster 1	445	64	11	0.142	0.226	4.372	0.525	0.360	-0.461	0.085	0.356
Cluster 2	1676	1411	1373	0.842	1.078	9.529	0.506	0.543	0.070	0.004	0.016
Resampling approach											
Multi-years	488	90	19	0.183	0.248	5.364	0.516	0.390	-0.322	0.129	0.337
Single-year	1633	1385	1365	0.848	1.057	9.700	0.508	0.545	0.068	0.004	0.012
Combination of approaches											
Asex	499	96	25	0.191	0.252	5.414	0.516	0.393	-0.315	0.123	0.327
Sex	1622	1379	1359	0.850	1.051	9.749	0.508	0.544	0.068	0.004	0.012

MLG: number of multilocus genotypes, MLL: number of multilocus lineages, *N*: number of individuals, *R*: genotypic diversity, *Pareto β*: evenness among repeated genotypes; A_R: allelic richness, H_O: observed heterozygosity, H_E: expected heterozygosity, F_{IS}: inbreeding coefficient, Var (F_{IS}): variance of inbreeding coefficient, and $\bar{r}_d$: multilocus linkage disequilibrium.

The resampling approach revealed two groups: 488 individuals with lineages spanning more than one year due to asexual reproduction and 1,633 sexual individuals without resampled lineages exceeding two years. The contingency table (Table 2), which compares the clustering and resembling approaches, shows a discrepancy. Among the 54 individuals identified as sexual through cluster analysis, subsequent resampling evidence validated their clonality. In contrast, 11 individuals were clustered as asexual; however not resampled.

Table 2. Contingency table showing the relationship between clustering and resampling approaches for reproductive mode definition in the dataset. Values represent the number of individuals assigned to each category.

	Asex (Resampling approach)	**Sex (Resampling approach)**
Asex (Clustering approach)	434	11
Sex (Clustering approach)	54	1622

A Fisher's exact test was performed to evaluate the correlation between the assignment of reproductive modes from clustering and resampling analyses. The results were highly significant (p-value $< 2.2 \times 10^{-16}$), thereby rejecting the null hypothesis of no association, with a 95% confidence interval. This indicated a robust relationship between the two classification methods. Therefore, a combination of clustering and resampling approaches was used to determine the reproductive mode of each individual. The subsequent analysis involved the calculation of the population genetic indices for both groups. The inference of reproductive mode using two approaches as well as a combination of both, yielded consistent results over all population genetic indices, confirming the robustness of the classification (Table 1). However, there are slight differences in the values of population genetic indices among classification methods, with the asexual group determined by the clustering approach revealing a stronger signature of asexual reproduction.

3.3 Spatial distribution of reproductive modes

The spatial distribution of both sexual and asexual lineages of *M. larici-populina* individuals in France revealed distinct patterns (Figure 1). Sexual lineages were widespread across most regions, predominating in the northern and central parts of the country. In contrast, asexual lineages were more concentrated in the southern regions.

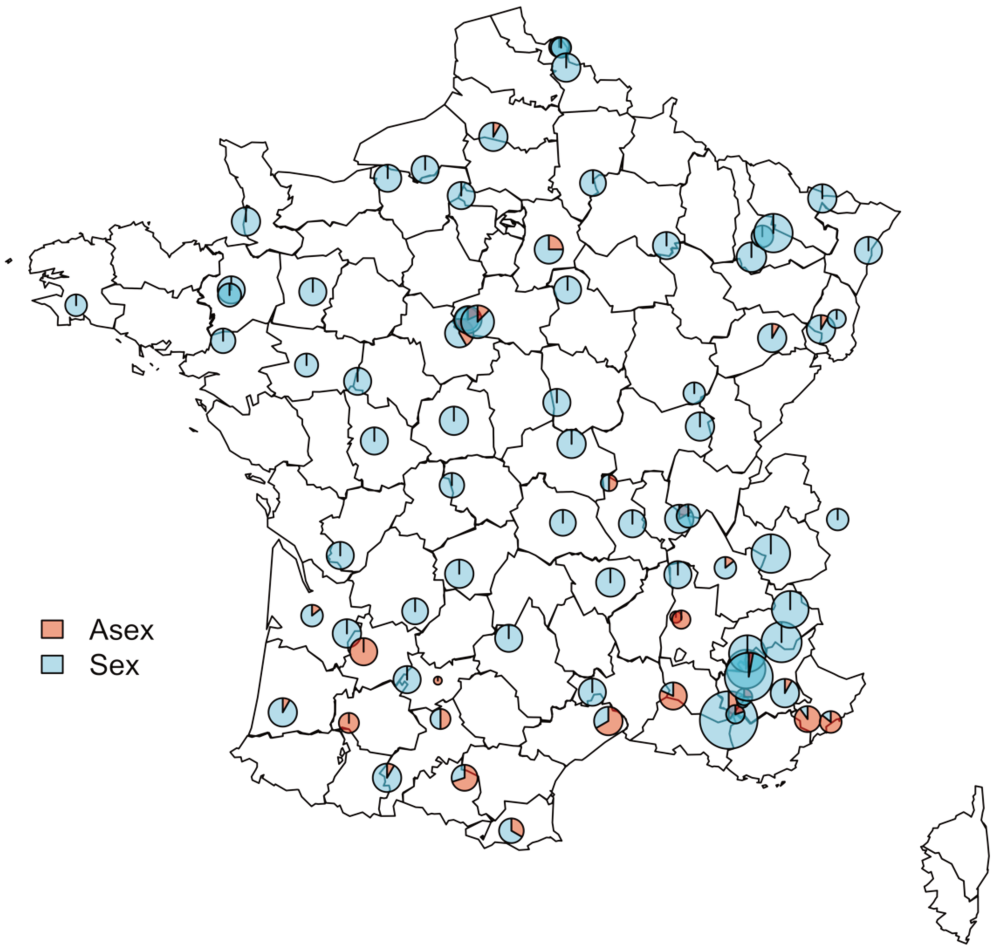


Figure 1

Geographic distribution of sexual and asexual lineages of *Melampsora larici-populina* individuals collected in 2009 and 2011 across France. The pie charts show the proportions of asexual (red) and sexual (blue) lineages at each sampling location. The size of each pie chart corresponds to the total number of individuals collected from that site.

To further investigate the factors that influence the prevalence of asexual lineages, a GLMM binomial regression analysis was conducted. The analysis (Table 3) revealed a significant effect of both latitude and longitude (Figure S3) on the proportion of sexual and asexual lineages, with positive coefficients indicating that the relative abundance of asexual individuals increases with lower latitude (the closer we get to the south of the country). Additionally, the random effect of years (variance = 18.78, sd = 4.333) indicated that the temporal distribution of individuals was not associated with a specific reproductive mode.

Table 3. Results of the binomial regression model predicting the effect of spatial distribution on the reproductive mode of *M. larici-populina*

Predictor	***Estimate***	***Std. Error***	***z-value***	***p-value***	***Significance***
(Intercept)	-22.99815	2.66959	-8.615	< 2e-16	***
Lat	0.46035	0.04730	9.732	< 2e-16	***
Long	0.09671	0.02912	3.321	0.000896	***

Estimate (regression coefficient), Std. Error (standard error of the estimation), z-value (z-statistic), p-value (statistical significance), and Significance power indicated by asterisks (0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1).

The ClonEstiMate Bayesian analysis yielded insights into the rate of asexual reproduction in *M. larici-populina*. The 99% credible interval from the posterior probability obtained for each group's clonal rates. In the context of a mixed group, where sexual and asexual lineages are merged into one population, the estimated clonal rate was between 0.22 to 0.31, indicating that about 27% of the population reproduces asexually. When the analysis was performed under the assumption of two distinct populations, the clonal rate was between 0.93 to 0.99 for the asexual group and between 0 to 0.21 for the sexual group (Figure 2).

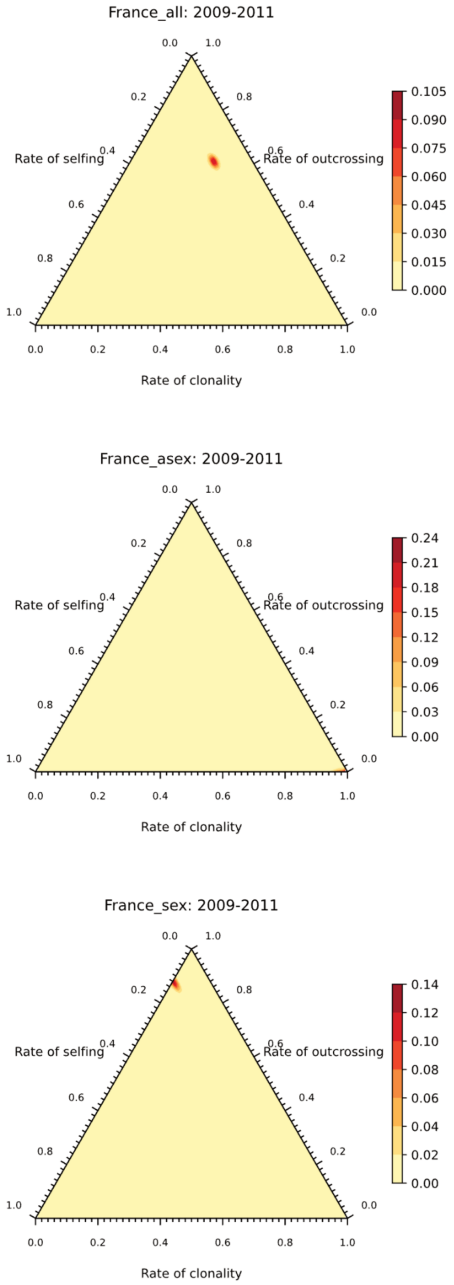


Figure 2

Ternary plot representing the relative proportions of selfing, outcrossing, and clonality rates in three different groups of mixed reproductive, asexual reproductive, and sexual reproductive in France (2009–2011). The color gradient reflects the posterior probability estimates.

3.4 Genetic relationships between sexual and asexual lineages

The dendrogram tree shows different clades for sexual and asexual lineages (Figure 3), highlighting the seven main asexual lineages. However, some asexual lineages also appeared as closely related sister branches from the sexual clade. Barring MLL9, one of the most abundant asexual lineages, all of these asexual MLLs, closely related to the sexual clade are of small size. These small asexual lineages were thus sporadic, unlike the seven most abundant asexual MLLs that show persistence through time (Figure 3, Table 4). These main asexual MLLs also displayed some reticulation, indicating they consist of different MLGs. In contrast, MLL5 (a most prevalent MLL but classified as sexual) consists of individuals sampled only in 2024 that have exactly the same genotype. These observations underscored the evolutionary relationships and temporal dynamics between reproductive modes within and among *M. larici-populina* lineages.

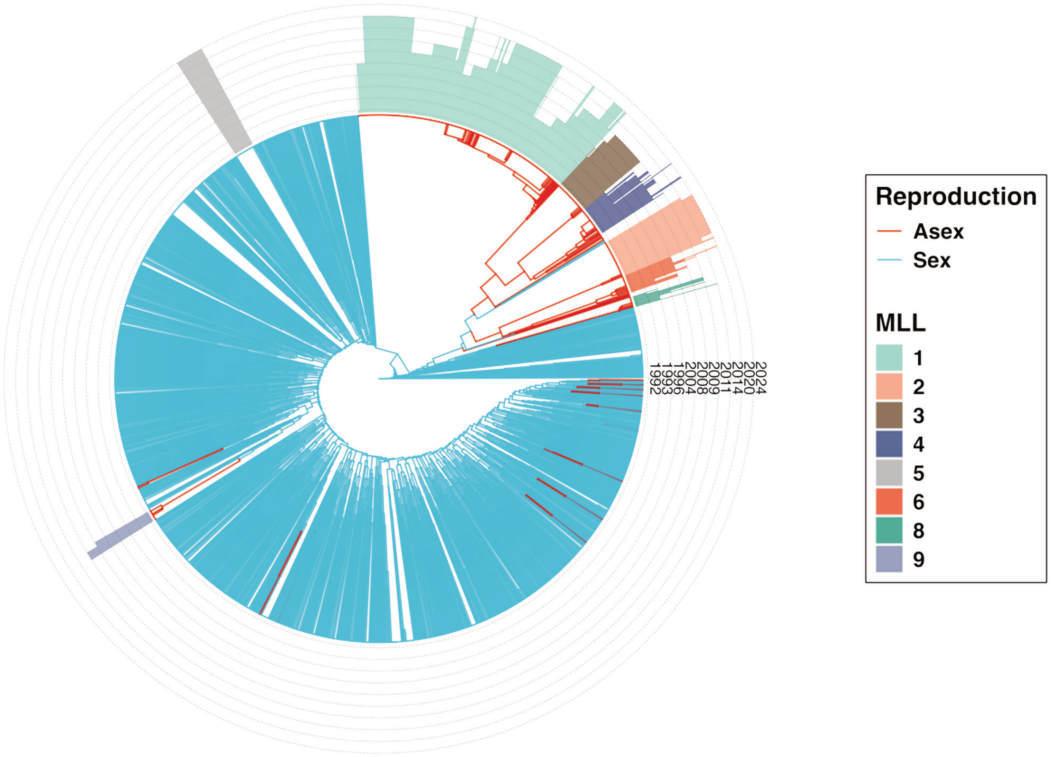


Figure 3

Circular Neighbor-joining (NJ) tree based on microsatellite data (21 markers) illustrating the genetic relationships among *M. larici-populina* individuals. The color of the branches indicates the reproductive mode, with blue representing sexual and red representing asexual modes. The outer rings represent the year of sample collection, with each height corresponding to a year. The nine most abundant multilocus lineages (MLLs) are color-coded.

3.5 Genetic structure and persistence of asexual lineages

A more detailed analysis of the asexual group identified several distinct genetic lineages. Among these, seven lineages were found to be the most abundant. The dominant lineages exhibited varying levels of persistence over the years, as indicated by their range of collection years (Table 4). In particular, MLL1 was consistently detected in almost all collection years and across all regions, highlighting its long-term persistence and widespread occurrence.

Table 4. Characteristics of the seven most abundant asexual multilocus lineages (MLLs) of *M. larici-populina*.

MLL	*MLG*	N	Year range	*R*	A_R	H_O	H_E
1	36	280	1992:2024	0.125	1.274	0.519	0.274
2	3	49	2014:2020	0.042	1.290	0.571	0.287
3	4	44	1996:2009	0.070	1.316	0.619	0.312
4	9	38	1996:2020	0.216	1.266	0.499	0.263
6	6	23	1996:2020	0.227	1.196	0.354	0.192
8	5	14	1993:2014	0.308	1.306	0.575	0.296
9	3	14	2009:2011	0.154	1.189	0.350	0.183

MLL: number of multilocus lineages, *MLG:* number of multilocus genotypes, N: number of individuals, *R*: Genotypic diversity, A_R: Allelic richness, H_O: Observed heterozygosity, H_E: Expected heterozygosity.

The distribution of these lineages in France differed with some lineages being widespread and others more restricted to some areas (Figure 4).

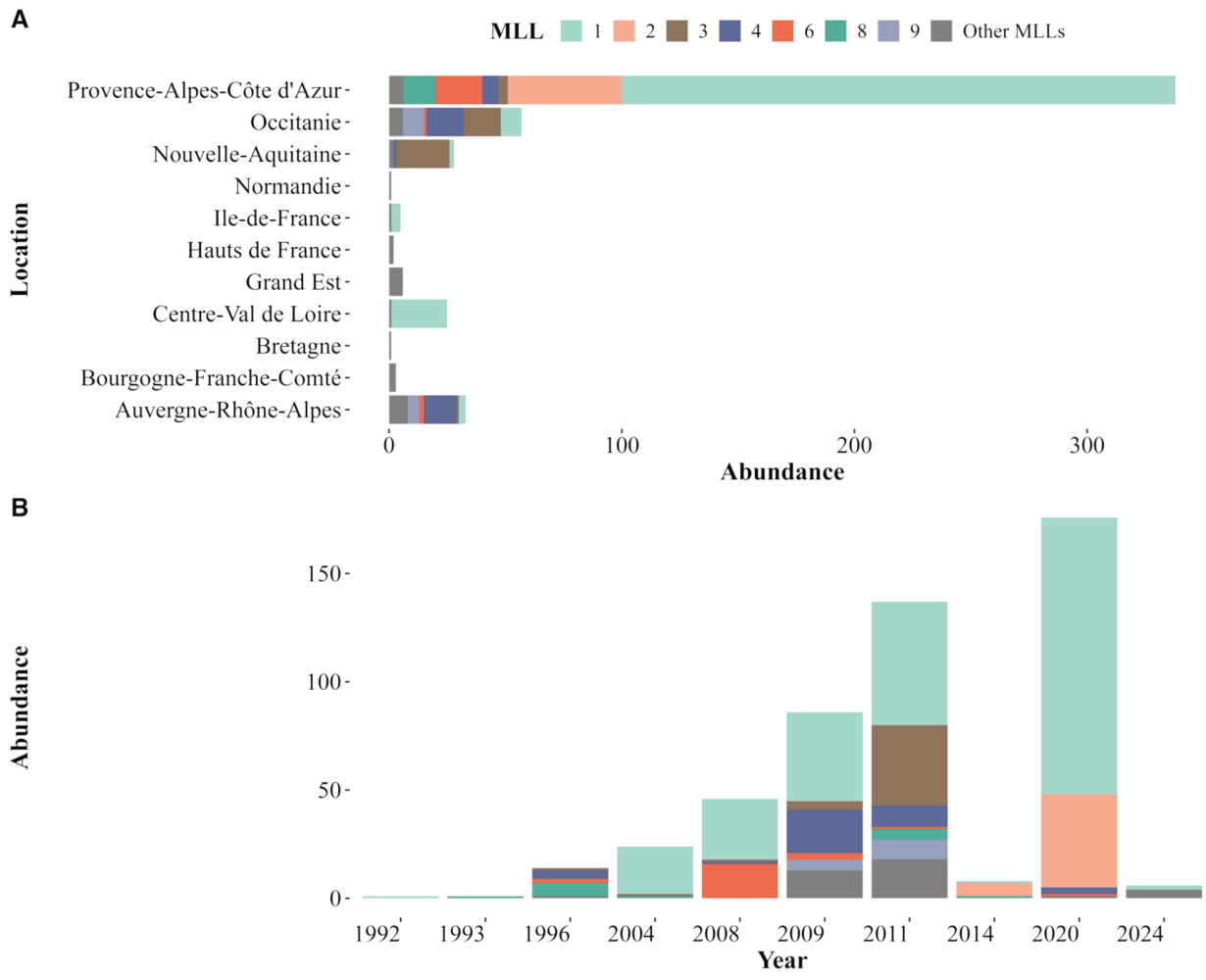


Figure 4

Distribution of asexual lineages over time and across locations. (A) The abundance of asexual MLLs in different French administrative regions. (B) The abundance of asexual MLLs in different years.

Fisher's exact test revealed a highly significant association between MLL distribution and both temporal and spatial factors (P = 1e-05). The analysis revealed that certain MLLs were repeatedly observed in the same geographic regions and over multiple sampling years, suggesting a pattern of persistence. Conversely, other MLLs appeared sporadic.

4 Discussion

The poplar rust fungus *M. larici-populina* is commonly regarded as an obligate sexual species that needs the presence of both poplars and larches to perform its annual life cycle (Pinon & Frey, 2005; Hacquard et al., 2011). However, a previous population genetic study suggested the existence of a distinct group of individuals that would survive asexually only on poplar hosts (Xhaard et al., 2011). Here i) we characterized the existence of an asexual group maintaining on the poplar hosts over the years, ii) we found that this asexual group is made up of several genetic lineages. We documented that some of these asexual lineages have persisted for nearly three decades despite their coexistence with sexual lineages in some places. These findings challenge conventional assumptions about the evolutionary dominance of sexual recombination in this species and offer a unique opportunity to investigate the transitions from sexual to asexual reproduction in a plant pathogenic fungus.

In the present study, two complementary approaches were employed to reveal and characterize asexual individuals. A clustering approach based on differences in allelic frequency and probability of membership was combined with a resampling approach aiming to identify the persistence of lineages over the years. We observed a strong overlap between these indirect and direct approaches, which enabled a sound delineation of asexual lineages among *M. larici-populina* individuals. This result further supports the reliability of the clustering approach performed in the initial study of (Xhaard et al., 2011) to identify asexual lineages indirectly. Applying such an individual based clustering approach is all the more relevant that sexual and asexual lineages can be admixed in the same location (Halkett et al., 2006). Consequently, the clustering approach can be regarded as a robust and independent method for detecting cryptic population genetic structures. In addition to the genetic differentiation between sexual and asexual clusters caused by the lack of gene flow, several genetic signatures clearly distinguished asexual lineages thereby reflecting the fundamental differences in evolutionary trajectories between these reproductive modes (Balloux et al., 2003; Halkett et al., 2005; Rouger et al., 2016). Sexual reproduction ensures the convergence toward Hardy-Weinberg equilibrium (HWE) with recombination among loci that reduces linkage disequilibrium (LD). Asexual reproduction, on the other hand, distorts both allelic and genotypic frequencies mostly because it freezes genotypes (Stoeckel & Masson, 2014; Rouger et al., 2016). As a result, the best way to identify

asexual lineages is to apply a multi-criteria method (Kronauer et al., 2012) in which the (diploid) population displays the following hallmarks (Halkett et al., 2005; Arnaud-Haond et al., 2007): a reduced genotypic diversity, combined with a significant excess of heterozygotes (negative F_{IS} values). It should be noted that we also observed a very high variance of F_{IS} across loci, as expected when some recombination events seldom occur (Balloux et al., 2003; Stoeckel & Masson, 2014; Stoeckel et al., 2021). Here, the values found for these indices are consistent with a very high asexuality rate as inferred by ClonEstiMate analysis. The asexual group also displayed significant linkage disequilibrium among loci, which is often reported in asexual and partially asexual species, despite not being consistently expected in neutrally evolving populations (Stoeckel et al., 2021). Linkage disequilibrium was also observed in the sexual group along with a positive F_{IS} value as the result of a Wahlund effect. Indeed, the sexual group encompasses different genetic subgroups that have been identified by previous population genetics studies (Xhaard et al., 2011; Persoons et al., 2017) and which are not considered here. We observed clone mates in the sexual group too, as these can form lineages as abundant as those in the asexual group, with the notable difference that they occur always at a unique sampling site and that, above all, they are never resampled over the years in spite of their local abundance. This local abundance of clone-mate lineages is in accordance with the sampling mostly done in the autumn, at the end of the clonal epidemic phase of *M. larici-populina* life cycle. With regards to this, it should be recalled that the mandatory sexual reproduction that takes place on larch resets the population genetic indices in the sexual group each year (Xhaard et al., 2011; Rouger et al., 2016).The contrast in population genetic indices provide a sound framework for distinguishing reproductive modes, supporting the reliable assignment of individuals into either sexual or asexual groups and providing the first insights into their respective evolutionary history (Xhaard et al., 2011).

In accordance with the previous study of Xhaard et al. (2011), we found that the asexual group was unevenly distributed across the country, and mostly occurred in southern France. Asexual lineages have often been described as distributed at the margins of a species' natural range (Simon et al., 1999; Halkett et al., 2005; Drenth et al., 2019; Pereyra et al., 2023), a phenomenon called the geographic parthenogenesis (Peck et al., 1998). Poplar rust is intimately dependent upon the presence of its two hosts. Although the native range of European Larch (*Larix decidua*)

is restricted to the Alps and some other Central-European Mountain ranges, it has been widely planted in many parts of France in mountains and plains since the 18th century. Nevertheless, it is less frequent in the Mediterranean area, which can therefore be considered as a margin for the poplar rust fungus. If we compare our results with those of other rust fungi, particularly those infecting wheat, we can see a larger spatial segregation between the sexual lineages found only in the disease's center of origin and the asexual lineages that have invaded the crops worldwide. Beyond rusts, this observation is common to many pathogenic species as for the ascomycete fungus responsible for rice blast (Saleh et al., 2014), or for the oomycete responsible for potato late blight (Ristaino, 2002). Interestingly, asexual lineages in these cases are specialized on different cultivars, and therefore on different host genotypes (Goyeau et al., 2007), and the relative abundance of asexual lineages therefore fluctuates over time with the use or disuse of these cultivars (Fontyn et al., 2022). This shows to which extent the asexual evolution of these plant pathogen species is associated with human practices. By contrast, in *M. larici-populina*, asexual lineages are sampled mostly on wild poplar trees and are thus found on different host genotypes. We thus can argue that this rust species enables the incipient stage of sexual to asexual transition to be documented.

A puzzling observation is that asexual lineages of *M. larici-populina* coexist with their sexual counterparts in most sampling sites. The absence of larch trees in some areas can be compensated for by the very high dispersal abilities of this species (Barrès et al., 2008), which can be wind dispersed over hundreds of kilometers in a year (Saubin et al., 2024).This can readily explain why sexual and asexual lineages are admixed even if no larch tree exists in the close vicinity of the sampling site. Another non-exclusive hypothesis to account for the long-lasting coexistence of sexual and asexual lineages is that of a temporal fluctuation in the environmental conditions that favor the ones over the others. This temporal fluctuation could be responsible for the fluctuations in abundance we observed, with almost half of the asexual lineages sampled in the year 2020 alone. These fluctuations in relative sexual vs. asexual abundances could be linked to peculiar meteorological events that would desynchronize the alternation between poplar and larch during the sexual life cycle and disadvantage sexual lineages that have to complete this complex life cycle. An outstanding example of the effect of a hazardous event on the fate of sexual and asexual lineages was documented in the red alga

Agarophyton chilense along the Chilean coasts (Becheler et al., 2020). This study monitored wild (mostly sexual) and cultivated (asexually propagated) populations of this red alga before and after a devastating earthquake. Although sexual populations paid a large tribute immediately after the catastrophe, the artificial maintenance of an asexual population caused the algae culture to collapse in the long term, suggesting that each mode of reproduction plays a role in maintaining a dynamic equilibrium.

The distribution of asexual lineages of *M. larici-populina* prevailing in southern France points to the first mechanism that can explain their maintenance and diversification. In milder climates the asexual survival could be facilitated by a longer growing period of poplars, creating a so-called green bridge that would enable the rust lineages to jump over the winter from poplar leaves to poplar leaves. According to this hypothesis, the asexual lineages would passively bypass the sexual phase on larch, without the need for any genetic determinism. In support of this hypothesis, it should be noted that *M. larici-populina* urediniospores can survive for 10 months and still be able to infect poplar leaves (Taris, 1968). But in these experiments, rust-infected poplar leaves were stored either in a greenhouse, or in a wire cage 2 meters above ground level, which does not mimic real natural overwintering conditions. According to this neutral hypothesis, both the onset and the survival of asexual lineages would be a matter of chance only. In support of this hypothesis, we observed that some asexual lineages branched among the sexual cluster and that these lineages have only been revealed through the resampling over years, indicating they are in very close genetic proximity with the lineages that undergo obligate sexual reproduction each year.

The persistence of several asexual lineages over decades, attesting to their evolutionary maintenance and relative success, could also be indicative of some genetic determinism favoring their survival over winter. In support of this hypothesis, the trade-off between overwintering ability and multiplication during the epidemic season has been documented in other fungal species (Hamelin et al., 2016; Suffert et al., 2018), and proven to account for the coexistence and evolutionary maintenance of sexual and asexual lineages (Castel et al., 2014). It is also likely that those long-lasting asexual lineages could have lost, at least to some degree, their ability to invest in sexual reproduction, maybe to the benefit of greater asexual survival. Our population genetic analysis, and especially the high variance of F_{IS}, would nonetheless indicate that recombination

does not occur among asexual lineages and thus that they do not evolve under strict asexuality but rather as a succession of clones maintaining on poplar leaves over winter, renewed over the years. Furthermore, the rare engagement of asexual lineages into the sexual reproduction can bring additional explanations for the diversity of asexual lineages we observed. In line with the concept of contagious asexuality observed in other organisms with mixed reproductive modes, such as aphids and daphnia (Simon et al., 2002, 2003; Halkett et al., 2006, 2008), asexual lineages that invest in sexual reproduction could be able to produce offspring that inherit their propensity to survive asexually (Halkett et al., 2008). Deciphering further the history and evolutionary success of these asexual lineages calls for a dedicated experiment to test their performance and investment throughout the life cycle. While the complex life cycle of *M. larici-populina* can be implemented in the laboratory (Pernaci et al., 2014), it remains a challenging experiment that is still ongoing.

The main original result of our study was to document the diversity among asexual lineages in *M. larici-populina*, with some lineages persisting over decades and others appearing more sporadically. The fact that two sampling campaigns covered the entire French territory, and that other sites were sampled regularly over more than 30 years is evidence for this pattern not being due to sampling bias. Notably, differences in abundance correlate with differences in the genotypic structure between lineages. We documented that asexual MLLs can consist of different genotypes (MLGs), unlike lineages within the sexual group that primarily consist of a singular repeated MLG. While some persistent asexual lineages, such as MLL1, showed relatively high genotypic diversity over time. Here it is important to state that only slight differences distinguish MLGs within asexual MLLs, in the vast majority of cases up to three loci differ only. Moreover, the changes consist in either heterozygous to homozygous conversion or to increase or decrease in allele length by a one step mutation. This regular pattern of allele changes, found across all MLGs within MLLs, supports the hypothesis that the asexual lineages evolved through single mutation steps. Although mutation rates for the microsatellite markers used in *M. larici-populina* have not yet been estimated, this regular pattern could present an opportunity to apply a tip calibration approach of molecular clock analysis to estimate the timing of the emergence and persistence of these asexual lineages (Rieux & Balloux, 2016). Yet, the differential abundance and genotypic diversity among asexual lineages reflects heterogeneity

within asexual population dynamics and emphasizes the necessity to consider each lineage independently, rather than presuming that the whole asexual population follows the same evolutionary trajectory. The most parsimonious hypothesis is that this diversity has been generated by the repeated emergence of asexual lineages through the evolutionary mechanisms previously discussed (chance of survival or contagious asexuality). Among all asexual lineages, MLL1 stood out as the most abundant lineage, being repetitively sampled from 1992 and showing an outbreak in 2020. It was detected in all regions where asexual lineages were sampled, and forms the only contribution to the unexpected occurrence of asexual lineages northward in the center of France, nearby Paris. The genome of a reference asexual isolate, representative of this MLL, has recently been sequenced, and analysis of its genome structure, polymorphisms, and gene content will further document the tale of this dominant asexual lineage.

Data

To ensure reproducibility, all analyses were executed using R version 4.2.2. The relevant scripts and data can be accessed at https://github.com/ammarabdalrahem/poplar_rust.git.

Author Contributions

Sébastien Duplessis, Pascal Frey, and Fabien Halkett conceived and designed the study. Axelle Andrieux designed and performed most of the molecular experiments. Ronan Becheler, Ammar Abdalrahem, Axelle Andrieux, and Fabien Halkett performed the manual curation of genotyping results. Axelle Andrieux, Ronan Becheler, and Fabien Halkett did the preliminary analyses of the data. Ammar Abdalrahem performed the final data analyses with help and advice from Benoit Marçais, Ronan Becheler, Solenn Stoeckel, and Fabien Halkett. Ammar Abdalrahem, Kadiatou Schiffer-Forsyth, and Fabien Halkett prepared the manuscript. All authors revised and approved the manuscript.

Acknowledgments

We thank Jean Pinon for having created the historical poplar rust collection at INRAE Nancy in the 1980s and for having sampled the Cavalaire-sur-Mer population in 2014 and 2024. We warmly thank Asmaa A. Eltahlawy and Scott Heslop for proofreading an earlier version of the manuscript. This long-term research program was supported by several grants from the French National Research Agency (ANR-11-BSV7-0007, Clonix project; ANR-12-ADAP-0009, GANDALF project; ANR-13-BSV7-0011, FUNFIT project; ANR-18-CE32-0001, Clonix2D project; ANR-23-CE20-0032, ENDURANCE project). Ronan Becheler was supported by a postdoctoral fellowship from both the Region Lorraine and the French National Research Agency (CLONIX project, ANR-11-BSV7-007). Ammar Abdalrahem was supported by a PhD fellowship from the French Ministry of Education and Research (MESR).

We have used DeepL, Grammarly, and ChatGPT to proofread, suggest synonyms, check grammar and spelling, and debug R script. All suggestions made by these tools have been scrutinized and cross-checked by the authors.

Conflicts of Interest

The authors declare no conflicts of interest

Benefit-Sharing Statement

The advantage of this study is that its analysis code is an automated workflow and data are publicly accessible.

References

Aanen, D. K., & Hoekstra, R. F. (2007). Why Sex Is Good: On Fungi and Beyond. In *Sex in Fungi* (pp. 527–534). John Wiley & Sons, Ltd. https://doi.org/10.1128/9781555815837.ch32

Ali, S., Gladieux, P., Rahman, H., Saqib, M. S., Fiaz, M., Ahmad, H., Leconte, M., Gautier, A., Justesen, A. F., Hovmøller, M. S., Enjalbert, J., & de Vallavieille-Pope, C. (2014). Inferring the contribution of sexual reproduction, migration and off-season survival to the temporal maintenance of microbial populations: A case study on the wheat fungal pathogen uccinia striiformis f.sp. tritici. *Molecular Ecology*, *23*(3), 603–617. https://doi.org/10.1111/mec.12629

Arnaud-Haond, S., Duarte, C. M., Alberto, F., & Serrão, E. A. (2007). Standardizing methods to address clonality in population studies. *Molecular Ecology*, *16*(24), 5115–5139. https://doi.org/10.1111/j.1365-294X.2007.03535.x

Ashu, E. E., & Xu, J. (2015). The roles of sexual and asexual reproduction in the origin and dissemination of strains causing fungal infectious disease outbreaks. *Infection, Genetics and Evolution*, *36*, 199–209. https://doi.org/10.1016/j.meegid.2015.09.019

Bahri, B., Leconte, M., Ouffroukh, A., De Vallavieille-Pope, C., & Enjalbert, J. (2009). Geographic limits of a clonal population of wheat yellow rust in the Mediterranean region. *Molecular Ecology*, *18*(20), 4165–4179. https://doi.org/10.1111/j.1365-294X.2009.04267.x

Bailleul, D., Stoeckel, S., & Arnaud-Haond, S. (2016). RClone: A package to identify MultiLocus Clonal Lineages and handle clonal data sets in r. *Methods in Ecology and Evolution*, *7*(8), 966–970. https://doi.org/10.1111/2041-210X.12550

Balloux, F., Lehmann, L., & de Meeûs, T. (2003). The Population Genetics of Clonal and Partially Clonal Diploids. *Genetics*, *164*(4), 1635–1644. https://doi.org/10.1093/genetics/164.4.1635

Barbuti, R., Mautner, S., Carnevale, G., Milazzo, P., Rama, A., & Sturmbauer, C. (2012). Population dynamics with a mixed type of sexual and asexual reproduction in a fluctuating environment. *BMC Evolutionary Biology*, *12*(1), 49. https://doi.org/10.1186/1471-2148-12-49

Barrès, B., Halkett, F., Dutech, C., Andrieux, A., Pinon, J., & Frey, P. (2008). Genetic structure of the poplar rust fungus Melampsora larici-populina: Evidence for isolation by distance in Europe and recent founder effects overseas. *Infection, Genetics and Evolution*, *8*(5), 577–587. https://doi.org/10.1016/j.meegid.2008.04.005

Becheler, R., Guillemin, M.-L., Stoeckel, S., Mauger, S., Saunier, A., Brante, A., Destombe, C., & Valero, M. (2020). After a catastrophe, a little bit of sex is better than nothing: Genetic consequences of a major earthquake on asexual and sexual populations. *Evolutionary Applications*, *13*(8), 2086–2100. https://doi.org/10.1111/eva.12967

Becheler, R., Masson, J.-P., Arnaud-Haond, S., Halkett, F., Mariette, S., Guillemin, M.-L., Valero, M., Destombe, C., & Stoeckel, S. (2017). ClonEstiMate, a Bayesian method for

quantifying rates of clonality of populations genotyped at two-time steps. *Molecular Ecology Resources*, *17*(6), e251–e267. https://doi.org/10.1111/1755-0998.12698

Becheler, R., Xhaard, C., Klein, E. K., Hayden, K. J., Frey, P., De Mita, S., & Halkett, F. (2016). Genetic signatures of a range expansion in natura: When clones play leapfrog. *Ecology and Evolution*, *6*(18), 6625–6632. https://doi.org/10.1002/ece3.2392

Billiard, S., López-Villavicencio, M., Hood, M. E., & Giraud, T. (2012). Sex, outcrossing and mating types: Unsolved questions in fungi and beyond. *Journal of Evolutionary Biology*, *25*(6), 1020–1038. https://doi.org/10.1111/j.1420-9101.2012.02495.x

Bolker, B. M., Brooks, M. E., Clark, C. J., Geange, S. W., Poulsen, J. R., Stevens, M. H. H., & White, J.-S. S. (2009). Generalized linear mixed models: A practical guide for ecology and evolution. *Trends in Ecology & Evolution*, *24*(3), 127–135. https://doi.org/10.1016/j.tree.2008.10.008

Butlin, R. (2002). The costs and benefits of sex: New insights from old asexual lineages. *Nature Reviews Genetics*, *3*(4), 311–317. https://doi.org/10.1038/nrg749

Castel, M., Mailleret, L., Andrivon, D., Ravigné, V., & Hamelin, F. M. (2014). Allee Effects and the Evolution of Polymorphism in Cyclic Parthenogens. *The American Naturalist*, *183*(3), E75–E88. https://doi.org/10.1086/674828

Clarkson, D. B., Fan, Y., & Joe, H. (1993). A remark on algorithm 643: FEXACT: an algorithm for performing Fisher's exact test in r x c contingency tables. *ACM Trans. Math. Softw.*, *19*(4), 484–488. https://doi.org/10.1145/168173.168412

Culotta, E., & Hines, P. (1998). The Evolution of Sex. *Science*, *281*(5385), 1979–1979. https://doi.org/10.1126/science.281.5385.1979

Drenth, A., McTaggart, A. R., & Wingfield, B. D. (2019). Fungal clones win the battle, but recombination wins the war. *IMA Fungus*, *10*(1), 18. https://doi.org/10.1186/s43008-019-0020-8

Evett, I., & Weir, B. S. (Bruce S.). (1998). *Interpreting DNA evidence: Statistical genetics for forensic scientists*. Sinauer Associates. https://cir.nii.ac.jp/crid/1130282269131400192

Fontyn, C., Zippert, A.-C., Delestre, G., Marcel, T. C., Suffert, F., & Goyeau, H. (2022). Is virulence phenotype evolution driven exclusively by Lr gene deployment in French Puccinia triticina populations? *Plant Pathology*, *71*(7), 1511–1524. https://doi.org/10.1111/ppa.13599

Foucaud, J., Jourdan, H., Breton, J. L., Loiseau, A., Konghouleux, D., & Estoup, A. (2006). RARE SEXUAL REPRODUCTION EVENTS IN THE CLONAL REPRODUCTION SYSTEM OF INTRODUCED POPULATIONS OF THE LITTLE FIRE ANT. *Evolution*, *60*(8), 1646–1657. https://doi.org/10.1111/j.0014-3820.2006.tb00509.x

Fry, W. E. (2020). Phytophthora infestans: The itinerant invader; “late blight”: the persistent disease. *Phytoparasitica*, *48*(1), 87–94. https://doi.org/10.1007/s12600-019-00778-3

Gavino, P. D., Smart, C. D., Sandrock, R. W., Miller, J. S., Hamm, P. B., Lee, T. Y., Davis, R. M., & Fry, W. E. (2000). Implications of Sexual Reproduction for Phytophthora infestans in

the United States: Generation of an Aggressive Lineage. *Plant Disease*, *84*(7), 731–735. https://doi.org/10.1094/PDIS.2000.84.7.731

Gladieux, P., Feurtey, A., Hood, M. E., Snirc, A., Clavel, J., Dutech, C., Roy, M., & Giraud, T. (2015). The population biology of fungal invasions. *Molecular Ecology*, *24*(9), 1969–1986. https://doi.org/10.1111/mec.13028

Goyeau, H., Halkett, F., Zapater, M.-F., Carlier, J., & Lannou, C. (2007). Clonality and host selection in the wheat pathogenic fungus *Puccinia triticina*. *Fungal Genetics and Biology*, *44*(6), 474–483. https://doi.org/10.1016/j.fgb.2007.02.006

Hacquard, S., Petre, B., Frey, P., Hecker, A., Rouhier, N., & Duplessis, S. (2011). The Poplar-Poplar Rust Interaction: Insights from Genomics and Transcriptomics. *Journal of Pathogens*, *2011*, 716041. https://doi.org/10.4061/2011/716041

Halkett, F., Kindlmann, P., Plantegenest, M., Sunnucks, P., & Simon, J. C. (2006). Temporal differentiation and spatial coexistence of sexual and facultative asexual lineages of an aphid species at mating sites. *Journal of Evolutionary Biology*, *19*(3), 809–815. https://doi.org/10.1111/j.1420-9101.2005.01055.x

Halkett, F., Plantegenest, M., Bonhomme, J., & Simon, J.-C. (2008). Gene flow between sexual and facultatively asexual lineages of an aphid species and the maintenance of reproductive mode variation. *Molecular Ecology*, *17*(12), 2998–3007. https://doi.org/10.1111/j.1365-294X.2008.03798.x

Halkett, F., Simon, J.-C., & Balloux, F. (2005). Tackling the population genetics of clonal and partially clonal organisms. *Trends in Ecology & Evolution*, *20*(4), 194–201. https://doi.org/10.1016/j.tree.2005.01.001

Hamelin, F. M., Bisson, A., Desprez‐Loustau, M., Fabre, F., & Mailleret, L. (2016). Temporal niche differentiation of parasites sharing the same plant host: Oak powdery mildew as a case study. *Ecosphere*, *7*(11), e01517. https://doi.org/10.1002/ecs2.1517

Hartfield, M., & Keightley, P. D. (2012). Current hypotheses for the evolution of sex and recombination. *Integrative Zoology*, *7*(2), 192–209. https://doi.org/10.1111/j.1749-4877.2012.00284.x

Hörandl, E. (2013). Meiosis and the Paradox of Sex in Nature. In *Meiosis*. IntechOpen. https://doi.org/10.5772/56542

Jiang, Y., Ye, J., Veromann, L.-L., & Niinemets, Ü. (2016). Scaling of photosynthesis and constitutive and induced volatile emissions with severity of leaf infection by rust fungus (Melampsora larici-populina) in Populus balsamifera var. Suaveolens. *Tree Physiology*, *36*(7), 856–872. https://doi.org/10.1093/treephys/tpw035

Jombart, T., Devillard, S., & Balloux, F. (2010). Discriminant analysis of principal components: A new method for the analysis of genetically structured populations. *BMC Genetics*, *11*(1), 94. https://doi.org/10.1186/1471-2156-11-94

Kamvar, Z. N., Brooks, J. C., & Grünwald, N. J. (2015). Novel R tools for analysis of genome-wide population genetic data with emphasis on clonality. *Frontiers in Genetics*, *6*. https://www.frontiersin.org/articles/10.3389/fgene.2015.00208

Kamvar, Z. N., Tabima, J. F., & Grünwald, N. J. (2014). Poppr: An R package for genetic analysis of populations with clonal, partially clonal, and/or sexual reproduction. *PeerJ*, *2*, e281. https://doi.org/10.7717/peerj.281

Keightley, P. D., & Eyre-Walker, A. (2000). Deleterious Mutations and the Evolution of Sex. *Science*, *290*(5490), 331–333. https://doi.org/10.1126/science.290.5490.331

Kronauer, D. J. C., Pierce, N. E., & Keller, L. (2012). Asexual reproduction in introduced and native populations of the ant Cerapachys biroi. *Molecular Ecology*, *21*(21), 5221–5235. https://doi.org/10.1111/mec.12041

Lehto, M. P., & Haag, C. R. (2010). Ecological differentiation between coexisting sexual and asexual strains of Daphnia pulex. *Journal of Animal Ecology*, *79*(6), 1241–1250. https://doi.org/10.1111/j.1365-2656.2010.01726.x

Lodé, T. (2013). Adaptive Significance and Long-Term Survival of Asexual Lineages. *Evolutionary Biology*, *40*(3), 450–460. https://doi.org/10.1007/s11692-012-9219-y

Lorrain, C., Gonçalves dos Santos, K. C., Germain, H., Hecker, A., & Duplessis, S. (2019). Advances in understanding obligate biotrophy in rust fungi. *New Phytologist*, *222*(3), 1190–1206. https://doi.org/10.1111/nph.15641

Louet, C., Saubin, M., Andrieux, A., Persoons, A., Gorse, M., Pétrowski, J., Fabre, B., De Mita, S., Duplessis, S., Frey, P., & Halkett, F. (2023). A point mutation and large deletion at the candidate avirulence locus AvrMlp7 in the poplar rust fungus correlate with poplar RMlp7 resistance breakdown. *Molecular Ecology*, *23*(10), 2472–2483. https://doi.org/10.1111/mec.16294

Lynch, M., Bürger, R., Butcher, D., & Gabriel, W. (1993). The Mutational Meltdown in Asexual Populations. *Journal of Heredity*, *84*(5), 339–344. https://doi.org/10.1093/oxfordjournals.jhered.a111354

Mboup, M., Leconte, M., Gautier, A., Wan, A. M., Chen, W., De Vallavieille-Pope, C., & Enjalbert, J. (2009). Evidence of genetic recombination in wheat yellow rust populations of a Chinese oversummering area. *Fungal Genetics and Biology*, *46*(4), 299–307. https://doi.org/10.1016/j.fgb.2008.12.007

Nei, M. (1978). Estimation of Average Heterozygosity and Genetic Distance from a Small Number of Individuals. *Genetics*, *89*(3), 583–590.

Otto, S. P. (2009). The Evolutionary Enigma of Sex. *The American Naturalist*, *174*(S1), S1–S14. https://doi.org/10.1086/599084

Peck, J. R. (1994). A Ruby in the Rubbish: Beneficial Mutations, Deleterious Mutations and the Evolution of Sex. *Genetics*, *137*(2), 597–606.

Peck, J. R., Yearsley, J. M., & Waxman, D. (1998). Explaining the geographic distributions of sexual and asexual populations. *Nature*, *391*(6670), 889–892. https://doi.org/10.1038/36099

Pereyra, R. T., Rafajlović, M., De Wit, P., Pinder, M., Kinnby, A., Töpel, M., & Johannesson, K. (2023). Clones on the run: The genomics of a recently expanded partially clonal species. *Molecular Ecology*, *32*(15), 4209–4223. https://doi.org/10.1111/mec.16996

Pernaci, M., De Mita, S., Andrieux, A., Pétrowski, J., Halkett, F., Duplessis, S., & Frey, P. (2014). Genome-wide patterns of segregation and linkage disequilibrium: The construction of a linkage genetic map of the poplar rust fungus Melampsora larici-populina. *Frontiers in Plant Science*, *5*. https://www.frontiersin.org/articles/10.3389/fpls.2014.00454

Persoons, A., Hayden, K. J., Fabre, B., Frey, P., De Mita, S., Tellier, A., & Halkett, F. (2017). The escalatory Red Queen: Population extinction and replacement following arms race dynamics in poplar rust. *Molecular Ecology*, *26*(7), 1902–1918. https://doi.org/10.1111/mec.13980

Pierre, J.-S., Stoeckel, S., & Wajnberg, E. (2022). The advantage of sex: Reinserting fluctuating selection in the pluralist approach. *PLOS ONE*, *17*(8), e0272134. https://doi.org/10.1371/journal.pone.0272134

Pinon, J., & Frey, P. (2005). Interactions between poplar clone and Melampsora populations and their implications for breeding for durable resistance. In *Rust Diseases of Willow and Poplar* (pp. 139–154). https://doi.org/10.1079/9780851999999.0139

Platt, V., & Jeschke, J. M. (2014). Are exotic species red queens? *Ethology Ecology & Evolution*, *26*(2–3), 101–111. https://doi.org/10.1080/03949370.2014.898699

Pritchard, J. K., Stephens, M., & Donnelly, P. (2000). Inference of Population Structure Using Multilocus Genotype Data. *Genetics*, *155*(2), 945–959. https://doi.org/10.1093/genetics/155.2.945

Rieux, A., & Balloux, F. (2016). Inferences from tip‐calibrated phylogenies: A review and a practical guide. *Molecular Ecology*, *25*(9), 1911–1924. https://doi.org/10.1111/mec.13586

Ristaino, J. B. (2002). Tracking historic migrations of the Irish potato famine pathogen, Phytophthora infestans. *Microbes and Infection*, *4*(13), 1369–1377. https://doi.org/10.1016/S1286-4579(02)00010-2

Rouger, R., Reichel, K., Malrieu, F., Masson, J. P., & Stoeckel, S. (2016). Effects of complex life cycles on genetic diversity: Cyclical parthenogenesis. *Heredity*, *117*(5), Article 5. https://doi.org/10.1038/hdy.2016.52

Roughgarden, J. (1991). The Evolution of Sex. *The American Naturalist*, *138*(4), 934–953. https://doi.org/10.1086/285261

Saitou, N., & Nei, M. (1987). The neighbor-joining method: A new method for reconstructing phylogenetic trees. *Molecular Biology and Evolution*, *4*(4), 406–425. https://doi.org/10.1093/oxfordjournals.molbev.a040454

Saleh, D., Milazzo, J., Adreit, H., Fournier, E., & Tharreau, D. (2014). South-East Asia is the center of origin, diversity and dispersion of the rice blast fungus, Magnaporthe oryzae. *New Phytologist*, *201*(4), 1440–1456. https://doi.org/10.1111/nph.12627

Saleh, D., Xu, P., Shen, Y., Li, C., Adreit, H., Milazzo, J., Ravigné, V., Bazin, E., Nottéghem, J.-L., Fournier, E., & Tharreau, D. (2012). Sex at the origin: An Asian population of the rice blast fungus Magnaporthe oryzae reproduces sexually. *Molecular Ecology*, *21*(6), 1330–1344. https://doi.org/10.1111/j.1365-294X.2012.05469.x

Saubin, M., Coville, J., Xhaard, C., Frey, P., Soubeyrand, S., Halkett, F., & Fabre, F. (2024). A mechanistic-statistical approach to infer dispersal and demography from invasion dynamics, applied to a plant pathogen. *Peer Community Journal*, *4*. https://doi.org/10.24072/pcjournal.356

Schurko, A., Neiman, M., & Logsdon, J. (2009). Signs of sex: What we know and how we know it. *Trends in Ecology & Evolution*, *24*, 208–217. https://doi.org/10.1016/j.tree.2008.11.010

Shick, J. M., HOFFMANN, R. J., & LAMB, A. N. (1979). Asexual Reproduction, Population Structure, and Genotype-Environment Interactions in Sea Anemones. *American Zoologist*, *19*(3), 699–713. https://doi.org/10.1093/icb/19.3.699

Simon, J.-C., Baumann, S., Sunnucks, P., Hebert, P. D. N., Pierre, J.-S., Le Gallic, J.-F., & Dedryver, C.-A. (1999). Reproductive mode and population genetic structure of the cereal aphid Sitobion avenae studied using phenotypic and microsatellite markers. *Molecular Ecology*, *8*(4), 531–545. https://doi.org/10.1046/j.1365-294x.1999.00583.x

Simon, J.-C., Delmotte, F., Rispe, C., & Crease, T. (2003). Phylogenetic relationships between parthenogens and their sexual relatives: The possible routes to parthenogenesis in animals: ROUTES TO PARTHENOGENESIS IN ANIMALS. *Biological Journal of the Linnean Society*, *79*(1), 151–163. https://doi.org/10.1046/j.1095-8312.2003.00175.x

Simon, J.-C., Rispe, C., & Sunnucks, P. (2002). Ecology and evolution of sex in aphids. *Trends in Ecology & Evolution*, *17*(1), 34–39. https://doi.org/10.1016/S0169-5347(01)02331-X

Simon, J.-C., Stoeckel, S., & Tagu, D. (2010). Evolutionary and functional insights into reproductive strategies of aphids. *Comptes Rendus. Biologies*, *333*(6–7), 488–496. https://doi.org/10.1016/j.crvi.2010.03.003

Song, Y., Scheu, S., & Drossel, B. (2011). Geographic parthenogenesis in a consumer-resource model for sexual reproduction. *Journal of Theoretical Biology*, *273*(1), 55–62. https://doi.org/10.1016/j.jtbi.2010.12.020

Steenackers, J., Steenackers, M., Steenackers, V., & Stevens, M. (1996). Poplar diseases, consequences on growth and wood quality. *Biomass and Bioenergy*, *10*(5), 267–274. https://doi.org/10.1016/0961-9534(95)00121-2

Stoeckel, S., Becheler, R., Bocharova, E., & Barloy, D. (2024). GenAPoPop 1.0: A user-friendly software to analyse genetic diversity and structure from partially clonal and selfed autopolyploid organisms. *Molecular Ecology Resources*, *24*(1), e13886. https://doi.org/10.1111/1755-0998.13886

Stoeckel, S., & Masson, J.-P. (2014). The Exact Distributions of FIS under Partial Asexuality in Small Finite Populations with Mutation. *PLOS ONE*, *9*(1), e85228. https://doi.org/10.1371/journal.pone.0085228

Stoeckel, S., Porro, B., & Arnaud-Haond, S. (2021). The discernible and hidden effects of clonality on the genotypic and genetic states of populations: Improving our estimation of clonal rates. *Molecular Ecology Resources*, *21*(4), 1068–1084. https://doi.org/10.1111/1755-0998.13316

Suffert, F., Goyeau, H., Sache, I., Carpentier, F., Gélisse, S., Morais, D., & Delestre, G. (2018). Epidemiological trade‐off between intra‐ and interannual scales in the evolution of aggressiveness in a local plant pathogen population. *Evolutionary Applications*, *11*(5), 768–780. https://doi.org/10.1111/eva.12588

Taris, B. (1968). Contribution à l'étude des rouilles des Populus observées en France. *Annales Des Épiphyties*, *19*, 5–54.

Tilquin, A., & Kokko, H. (2016). What does the geography of parthenogenesis teach us about sex? *Philosophical Transactions of the Royal Society B: Biological Sciences*, *371*(1706), 20150538. https://doi.org/10.1098/rstb.2015.0538

Waits, L. P., Luikart, G., & Taberlet, P. (2001). Estimating the probability of identity among genotypes in natural populations: Cautions and guidelines. *Molecular Ecology*, *10*(1), 249–256. https://doi.org/10.1046/j.1365-294X.2001.01185.x

Weir, B. S., & Cockerham, C. C. (1984). Estimating F-Statistics for the Analysis of Population Structure. *Evolution*, *38*(6), 1358–1370. https://doi.org/10.2307/2408641

Xhaard, C., Andrieux, A., Halkett, F., & Frey, P. (2009). Characterization of 41 microsatellite loci developed from the genome sequence of the poplar rust fungus, Melampsora larici-populina. *Conservation Genetics Resources*, *1*, 21–25. https://doi.org/10.1007/s12686-009-9005-z

Xhaard, C., Barrès, B., Andrieux, A., Bousset, L., Halkett, F., & Frey, P. (2012). Disentangling the genetic origins of a plant pathogen during disease spread using an original molecular epidemiology approach. *Molecular Ecology*, *21*(10), 2383–2398. https://doi.org/10.1111/j.1365-294X.2012.05556.x

Xhaard, C., Fabre, B., Andrieux, A., Gladieux, P., Barrès, B., Frey, P., & Halkett, F. (2011). The genetic structure of the plant pathogenic fungus Melampsora larici-populina on its wild host is extensively impacted by host domestication. *Molecular Ecology*, *20*(13), 2739–2755. https://doi.org/10.1111/j.1365-294X.2011.05138.x

Supplementary figures

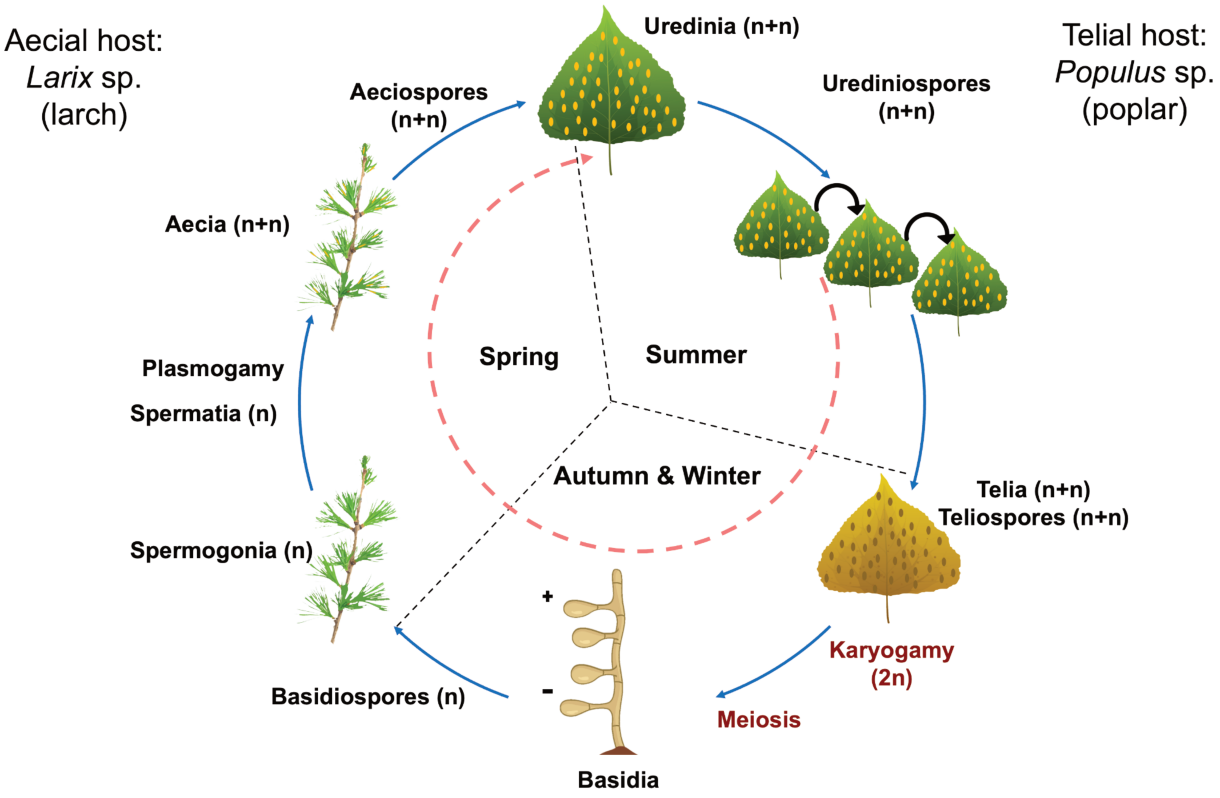


Figure S1

The *Melampsora larici-populina* life cycle. The blue arrows indicate the heteroecious life cycle, the clonal multiplication phase on poplar tree leaves during summer, and the sexual phase on the larch tree. The dashed red arrow indicates pure asexual multiplication on poplar, skipping the sexual phase.

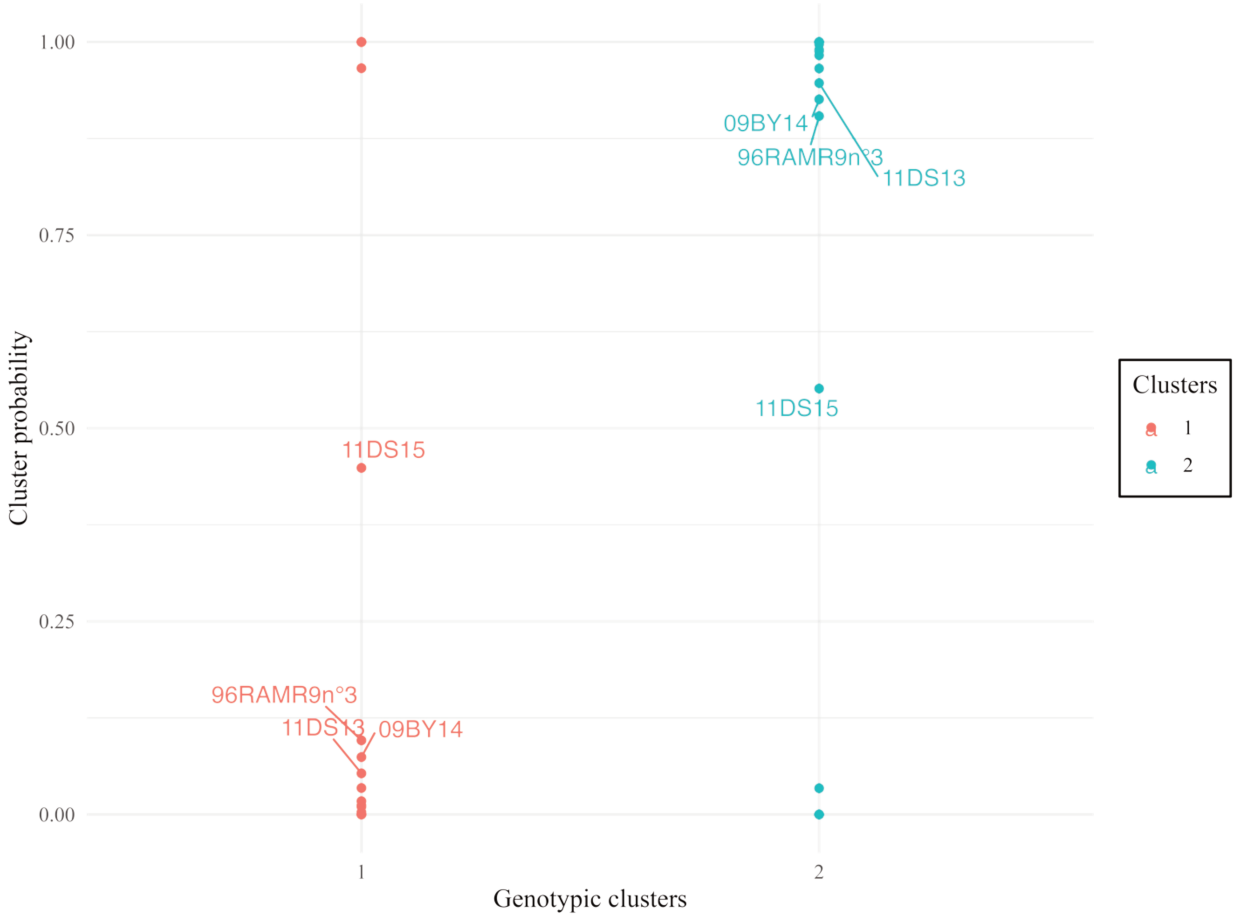


Figure S2

Scatter plot illustrating the membership probability for individuals within each genotypic cluster, with the names of individuals of lower probability.

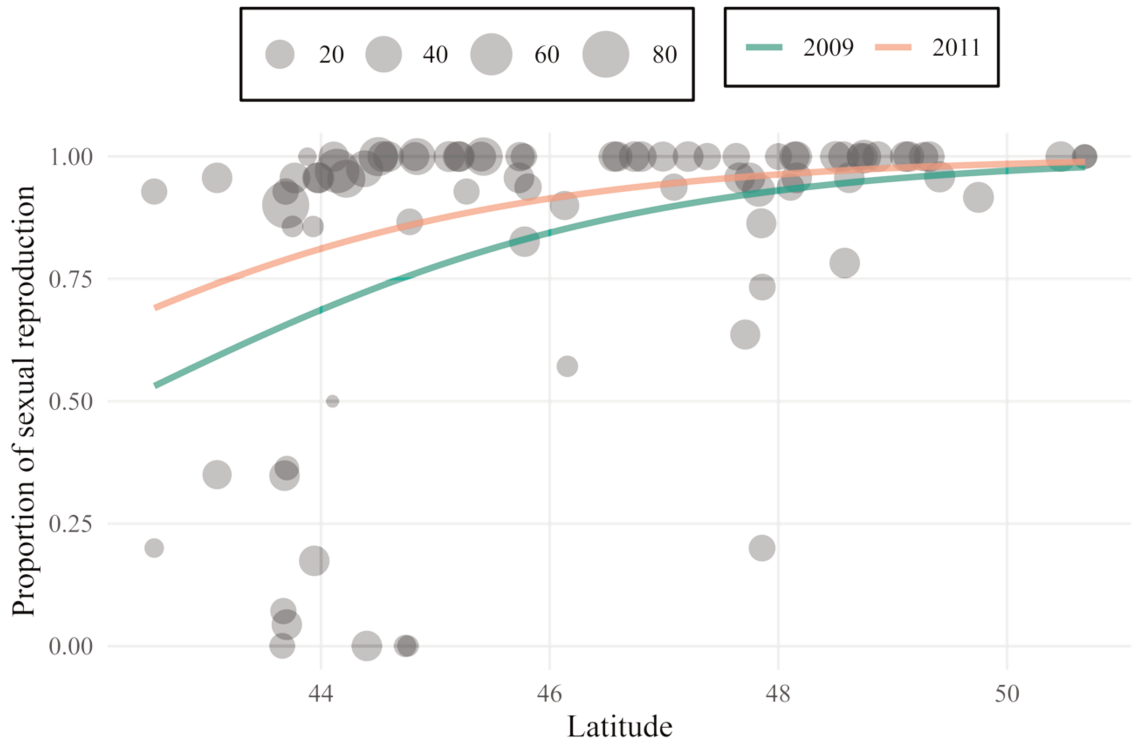


Figure S3

The regression plot illustrates the relationship between latitude and the proportion of sexual reproduction. This relationship was analyzed using a binomial generalized linear mixed model (GLMM). It was based on data from 2009 and 2011. The size of the points represents sample sizes.

Supplementary Tables

Table S1. The data set for 2,122 *Melampsora larici-populina* individuals used in population genetic analyses includes collection year, geographic location, and genotypes across 21 microsatellite markers.

Section two

Behind the scenes of a shadow play: illuminating the population genetics of partially clonal organisms through the prism of clone age distribution

Declaration

The work presented in this Section two is being finalized in the form of a research article and will be submitted soon to ***Heredity***.

Behind the scenes of a shadow play: illuminating the population genetics of partially clonal organisms through the prism of clone age distribution

Ammar Abdalrahem[1], Sébastien Duplessis[1], Pascal Frey[1], Solenn Stoeckel[2*], Fabien Halkett[1*]

1- Université de Lorraine, INRAE, IAM, F-54000 Nancy, France

2- DECOD (Ecosystem Dynamics and Sustainability), INRAE, Institut Agro, IFREMER, Rennes, France

* these authors contributed equally to this work.

Abstract

Partial clonality is widespread in natural populations; however, the effects of clonality rate and clone age dynamics on genetic patterns over time remain unexplored. This gap undermines our ability to interpret genotypic and genetic indices in both theoretical and empirical studies of such populations. To investigate how rates of clonality impact population structure, we simulated partially clonal populations over 10,000 generations across a range of clonality and mutation rates. Using a forward-time and individual-based model, we evaluated genetic and genotypic indices alongside measures of clone age distribution, to examine how clonality rates influence changes in population structure and genetic diversity over time. Our results reveal two distinct trajectories: (i) at low to moderate rates of clonality evolution is driven by sharp and predictable patterns of rapid clone turnovers, resulting in predominantly young clones which relative abundances well align with genotypic indices ; (ii) at extreme rates of clonality demographic stochasticity generates unpredictable clone age distributions which mirror the high variance of mean and variance of F_{IS} that summarize gene reshuffling between individuals. Interestingly, a portion of the variability of these indices can be explained by differences in clone age distribution that occur by chance. Our results, therefore, complement previous theoretical studies on the population genetics of partial clonality by providing biological insights into how clonal turnover dynamics shape the temporal evolution and variability of highly clonal

populations. These results have practical implications for inferring evolutionary trajectories and managing partially clonal species.

Introduction

Clonal reproduction is common throughout the Tree of Life. It has profound consequences for the evolution of organisms and shapes genetic diversity within and between populations. In particular, it distorts genetic equilibrium through the burst of repeated genotypes. However, how genotype turnover (that is, the distribution of clone ages) structures genetic diversity remains largely unknown.

A diverse array of reproductive strategies exists, ranging from obligate sexuality to strict clonality, with many species employing mixed strategies, which is often referred to as facultative sexual reproduction or partial clonality. In the following, we use the latter expression as it has broader acceptance both in experimental and theoretical literature. Partial clonality is a common reproductive strategy whereby individuals are capable of both sexual reproduction and clonal propagation. In these populations, genetic individuals (genets) persist across generations through asexual ramet production, while sexual recruitment introduces novel genets via recombination (Hartfield, 2016; Stoeckel, Porro, et al., 2021; Orive & Krueger-Hadfield, 2021). Clonality refers to the reproduction of identical individuals (with the exception of somatic mutations) from one generation to the next. For a detailed lexicon of species specific means to achieve clonal reproduction please see the review of Orive & Krueger-Hadfield (2021). Partial clonality has been documented in many plant species, such as seagrasses and duckweed *Spirodela polyrhiza* (Ho et al., 2019; Arnaud-Haond et al., 2020), invertebrates like aphids and coral reefs (Halkett et al., 2005; Carpenter et al., 2008), fungi like cereal rusts (Pucciniales) (Ali et al., 2014; Duplessis et al., 2021) and pathogenic yeasts (e.g. *Candida albicans*) (Nébavi et al., 2006), and protozoan parasites such as *Plasmodium falciparum* (Early et al., 2022). These organisms exhibit a diverse range of ecological niches, spanning from primary producers to the highest trophic level, and a large array of life-style, including free-living organisms but also many plant or animal (including human) pathogens.

Partial clonality impacts the ecological and evolutionary trajectories of species populations. On one hand, clonal reproduction leads to rapid demographic expansion, exploiting stable environments, without paying the twofold cost of (biparental) sex (Lehtonen et al., 2012; Lodé, 2013). On the other hand, sexual reproduction results in allelic shuffling through recombination, which is crucial for adaptation (Roze, 2012; Pierre et al., 2022). Partial clonality could take advantage of the benefits of each reproduction mode, as evidenced for many parasites and pests (McDonald & Linde, 2002; Bazin et al., 2014). For instance, *Phytophthora infestans*, the oomycete responsible for the Irish Potato Famine by causing potato late blight disease, relies on clonal reproduction to spread rapidly, while sexually reproducing to generate new lineages, posing an ongoing threat to global agriculture (Gavino et al., 2000; Fry et al., 2015). Malaria is a human life-threatening disease caused by the *P. falciparum*. The infection is dominated by rapid clonal multiplication within human red blood cells, while sexual reproduction occurs in the mosquito vector (Chang et al., 2013; Early et al., 2022).

Genotypic and genetic measures are crucial for understanding the evolutionary trajectory and ecological potential of organisms that reproduce both clonally and sexually (Stoeckel, Arnaud-Haond, et al., 2021; Stoeckel, Porro, et al., 2021). Clonal lineages can serve as an "evolutionary memory" of the population's history, maintaining genetic signatures of past events, such as bottlenecks and founder effects, through the freezing of ancestral genotypes (Brakefield, 1989; Reichel et al., 2016). Theoretical and empirical studies indicate the existence of clonality within the population through genotypic metrics using traditional and genome-wide molecular markers (e.g., SSR and SNPs, respectively) (Stoeckel, Porro, et al., 2021; Abdalrahem et al., 2025), such as clonal richness (*R*), which estimates the proportion of unique multilocus genotypes (MLGs), and the Pareto β index, which estimates the distribution of clonal sizes, highlighting the dominance of certain MLGs (Arnaud-Haond et al., 2007). Furthermore, genetic measures such as the inbreeding coefficient (F_{IS}), where negative values may indicate an excess of observed heterozygotes (H_O). Clonality often increases linkage disequilibrium (LD), indicating the presence of allelic associations due to limited recombination among lineages (Balloux et al., 2003; Halkett et al., 2005; Stoeckel & Masson, 2014).

Clonality within partial clonal populations slows down the convergence to the Hardy-Weinberg equilibrium (HWE) state, creating temporary deviations in genetic patterns that influence both genotypic and genetic indices (Reichel et al., 2016). This “transient” genetic deviation highlights the importance of estimating the rate of clonality (*c*) to measure the proportion of clonal and sexual reproduction within populations, which influences our interpretation of genetic diversity and equilibrium dynamics (Stoeckel, Porro, et al., 2021). For example, at low *c*, populations often follow HWE expectations, and the reduction of genotypic diversity indicates historical bottlenecks or founder effects (Hanko et al., 2023), but at high *c*, it may indicate dominance by a few clonal lineages (Reynes et al., 2021; Abdalrahem et al., 2025). Similarly, a negative F_{IS} value at low c implies outcrossing (CABRERA-TOLEDO et al., 2008), whereas under a high clonality rate, negative F_{IS} likely reflects an excess of heterozygosity from multiple clonal propagation, effectively “freezing” heterozygous genotypes (Abdalrahem et al., 2025). Although the effect of clonality rate has been largely established theoretically (Becheler et al., 2017; Halkett et al., 2005; Stoeckel, Porro, et al., 2021; Stoeckel & Masson, 2014), it is seldom used in empirical studies dealing with partial clonal organisms, leading to potential misinterpretations of key population genetic indices. Here, we propose to better highlight the link between the rate of clonality (*c*) and the variation of population genetic indices by examining the clonal turnover, which denotes the pace at which alleles are reshuffled within and between individuals.

Despite the recognition of the influence of partial clonality on population genetic structure, the evolutionary consequences of clonal age distributions (i.e., the variation and structure of clone longevity) remain unexamined across rates of clonality, a critical gap given its implications for mutation accumulation and adaptive potential. Early work proposed that clone age, due to mutation accumulation, would profoundly impact genetic diversity (the 'Meselson effect'; Birky, 1996; Mark Welch & Meselson, 2000; Balloux et al., 2003), yet no study addresses how increasing *c* values reshape clone age distributions or their downstream diversity effects. We hypothesize that high rates of clonality shift age structure toward older lineages, amplifying mutation accumulation and decoupling genotypic from allelic diversity. Specifically, increased *c* values may reduce genotypic richness (*R*) through dominance of persistent clonal lineages while

paradoxically maintaining allelic diversity (H_E) through genotype freezing that prevents genetic drift.

To resolve this gap and test our hypothesis, we employ individual-based forward simulations to quantify how clonality rate (*c*), mutation rate (μ) and population size (*N*) jointly shape genetic diversity indices through clone age turnover dynamics. This represents the first exploration of the interplay among four factors c, μ, *N* and clone age distributions. Our aim here is to disentangle the deterministic effects of clone age distribution on population genetic indices from its random component due to the stochasticity of their evolutionary trajectories using to 100 simulation replicates. We simulate populations under an annual life cycle where each discrete generation corresponds to one reproductive season. Our model tracks clone age distributions across the full range of rates of clonality, from strictly sexual (*c*=0) to strictly clonal (*c*=1) reproduction, over 10,000 generations, while dynamically quantifying genetic diversity indices (e.g., expected heterozygosity H_E) and genotypic metrics (e.g., genotypic richness *R*) at regular time points. In the following, we first revisit the effect of clonality rate (c) on the temporal variation of population genetic indices, focusing on two iconic mutation rates (classically set at the scale of the genetic drift i.e. μ=1/N, corresponding here $\mu=10^{-3}$ for a population size N=1000 or at a reduced but biologically relevant mutation rate of 10^{-6} thus when genetic drift is dominant). Then, we characterize the changes in clone age distributions according to clonality rate (*c*) and the distortion of predictability in clone age structure. Last, we examine the correlations between descriptors of clone age distributions and the values of population genetic indices obtained both within and among *c* values. The strengths and shapes of these correlations are discussed from the perspective of inferring clone turnover in empirical studies.

Materials and Methods

Simulation design

We performed individual-based, forward-time simulation to investigate how clonality rate (*c*: proportion of asexual reproduction) and mutation rate (μ: mutations per locus per generation) shape population genetic indices under clone age structure dynamics (Fig. 1). Each simulation modeled a population of diploid individuals with a constant size (N = 1,000 individuals) maintained across generations, thereby isolating the effects of *c* and μ on genetic diversity, while avoiding confounding demographic fluctuations (Gazave et al., 2013; Raynes et al., 2018). Reproduction occurred via clonal propagation (producing genetically identical offspring) and/or sexual outcrossing (generating recombinant offspring), with the relative contribution of each mode controlled by *c*.

We systematically varied clonality rate across 19 values, spanning from fully sexual (*c* = 0) to fully clonal (*c* = 1), with increased resolution at high *c* values (0.95, 0.96, ..., 0.9999) to capture thresholds where clonal dominance alters genetic dynamics. We implemented a k-allele mutation model (KAM) (Bhargava & Fuentes, 2010; Putman & Carbone, 2014), where each locus can harbor up to a number of distinct alleles per locus (Nballele), and mutations transition stochastically between these alleles at rate μ (Stoeckel, Porro, et al., 2021). For that, we varied the number of alleles (Nballele) to 2 (biallelic), 4, 8, or 99 (approximating an infinite alleles model) to capture differing levels of allelic richness. Mutation rates (μ) were scaled relative to population size *N*, to examine systems where mutation dominates ($\mu = 10/N = 10^{-2}$), balances ($\mu = 1/N = 10^{-3}$), or is dominated by genetic drift ($\mu = 1/(10N) = 10^{-4}$). To ground simulations in realistic systems, we focussed our analyses on two mutation rates: $\mu = 10^{-6}$ representing the mutation rate of a genome-wide marker data, such as single-nucleotide polymorphisms (SNPs), and $\mu = 10^{-3}$ as the rate of a traditional marker data like microsatellites (SSR) (Payseur & Cutter, 2006; McCulloch & Kunkel, 2008; Stoeckel, Porro, et al., 2021). All loci evolve independently, exhibiting no linkage, and are selectively neutral to sort out the effects of c and μ on population genetic indices.

Simulation ran for 10,000 generations, with genetic data collected at 10-generation intervals (0, 10, 20, ..., 10,000) to resolve transient and equilibrium states and reduce data redundancy. For each parameter combination (c, μ, Nballele), 100 replicate runs were performed to quantify stochastic variance, which is sufficient to capture 95% confidence intervals for diversity metrics (Rijnsoever, 2017). Simulation was implemented in a custom Python 3 script, which tracked clone ages (the number of generations a clone has maintained clonally in the population, corresponding to the number of generations since this genotype was produced by a sexual event) and evolutionary parameters including μ, c, and generational transitions (Fig. 1).

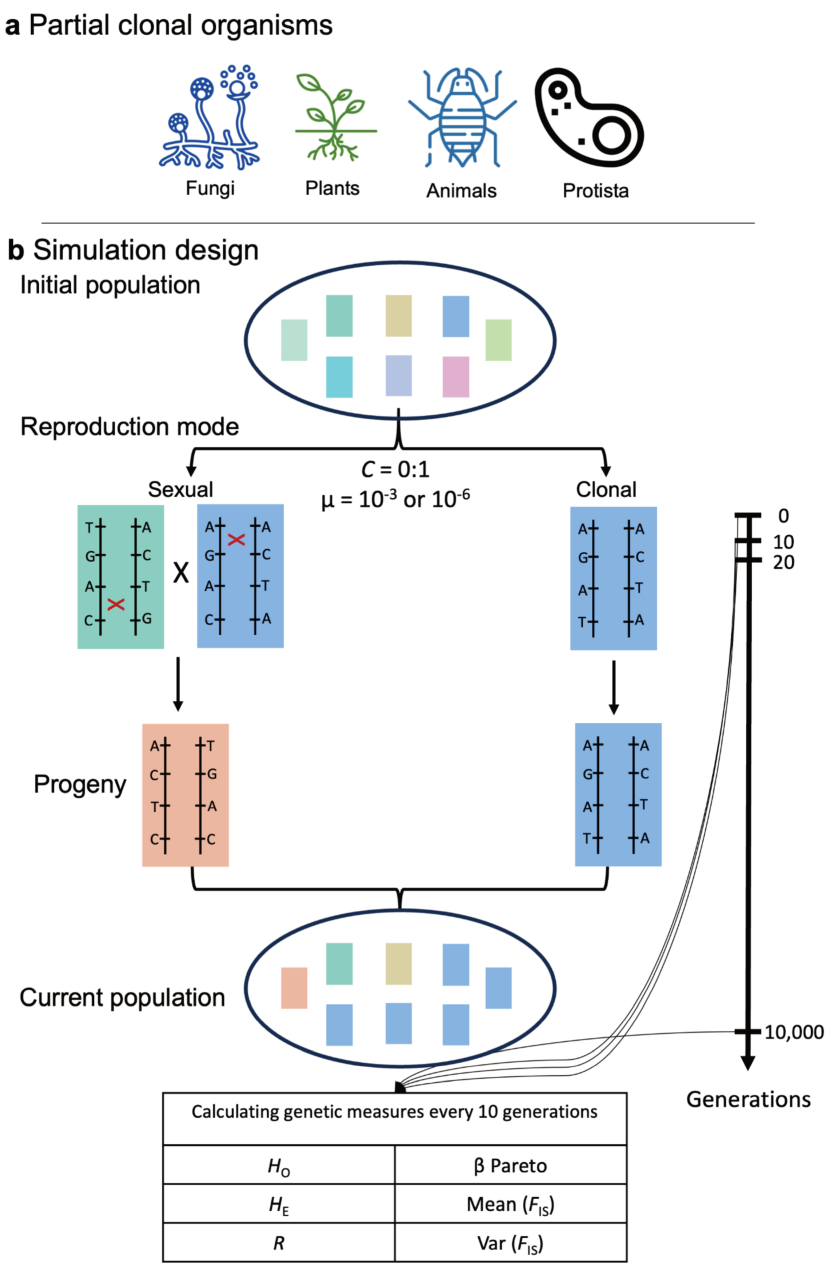


Fig. 1 Simulation workflow for investigating clone age structure and genetic diversity in a partial clonal system. The simulation begins with a population of 1,000 diploid individuals at Hardy-Weinberg equilibrium (HWE). Each generation, individuals reproduce either sexually or clonally, depending on the clonality rate (c). Mutation is introduced per locus at rates of $\mu = 10^{-3}$ or $\mu = 10^{-6}$, with each locus initially harboring up to four alleles. The simulation proceeds for 10,000 generations with 100 replicate runs per parameter combination. Population genetic indices (e.g., inbreeding coefficient (F_{IS}), expected heterozygosity (H_E), mean observed heterozygosity (H_O), clonal richness (R), Pareto β index), calculated every 10 generations.

Analysis of population genetic indices across rates of clonality

As outputs, we calculate several population genetic indices for each replicate. The mean expected heterozygosity (H_E) and mean observed heterozygosity (H_O) were calculated as measures of genetic variation at the allele level (Nei, 1978). Additionally, the mean and variance of the inbreeding coefficient (F_{IS}) were computed, indicating the deviations from the Hardy–Weinberg equilibrium (Wright, 1931, 1949; Stoeckel & Masson, 2014). The total number of alleles and the number of fixed loci were also captured as summaries of allelic diversity and fixation dynamics. To assess clonal structure, the total number of distinct genotypes, clonal richness (R), was derived as an index of clonal diversity, and Pareto β was estimated to capture the distribution of genotype frequencies (Arnaud-Haond et al., 2007). We used the Python library from GenAPoPop software (Stoeckel et al., 2024), to compute population genetic indices from the simulated data . Clone age structure was captured by recording the number of individuals of each age (clone age abundance) for each simulation replicate. By definition in our system, the age of all individuals produced by a sexual event in the last considered generation was set tozero.

Temporal analysis of genetic indices across clonality rates

To analyze the temporal dynamics of population genetic indices under different rates of clonality, we filtered the simulation data to focus on a representative set of c values: 0.1, 0.3, 0.5, 0.7, 0.9, 0.95, 0.98, 0.99, 999, and 0.9999. This selection aims to improve clarity in visual comparisons of the key transitions of dynamics of reproductive modes within populations and their impact on genetic structure. We summarized the trend in the evolution of population genetic indices (calculated using the methods described above) by calculating the median value across all simulation replicates for each combination of clonality rate, mutation rates (10^{-4} or 10^{-6}), at each time point (generation).

After quantifying the temporal dynamics of population genetic indices across generations, we selected one generation (time point) to examine the relationship between population genetic indices and the distribution of clone ages. We set this generation in such a way as to tip the

balance towards describing a stable distribution for most population genetic indices, while limiting the effect of allele fixation over time.

The effect of rates of clonality on clone age distribution

For each replicate, we recorded the number of individuals associated with each clone age at the end of the simulation (I.e., after 10,000 generations). We analysed clone age and abundance of clone ages with varying rates of clonality and mutation rates to assess how the persistence and prevalence of clones of different ages varied. Overall distributions of clone ages for each rate were visualised using joy plots (ridge plots) to compare age structure patterns.

The distribution of clone ages for each replicate was then summarized using basic descriptors :

- the median, maximum and variance of clone age distribution.
- the slope and the coefficient of determination (R^2) values of the linear regression between the logarithm of the abundance (number of individuals) and clone age. These last two descriptors encapsulate both the shape (slope) and the predictability (R^2) of the relationship between clone age and abundance across different rates of clonality.

To examine how clone age-abundance relationships relate to population genetic structure, we computed a pairwise Spearman correlation coefficient matrix using pairwise complete observations. At the fixed generation, we merged the regression outcomes (slope and R^2) values with corresponding genotypic and genetic indices for each simulation replicate, including the mean and variance of inbreeding coefficients (F_{IS}), observed and expected heterozygosity (H_O, H_E), clonal richness (R), and the *Pareto* β parameter. This indicates the strength and direction of monotonic associations between the regression outcomes and the various population genetic indices. Additionally, we generated scatter plots to visually explore the relationships between each clone age-abundance metric (slope and R^2) and the population indices. These plots enabled us to detect potential nonlinear patterns or outliers that may not be apparent from correlation coefficients alone.

Results

Introducing the effect of rates of clonality on the temporal change of population genetic indices.

First, we analyzed the dynamics of variation of population genetic indices over generations (time). By tracking these indices across 10,000 generations, we shed light on how their values dynamically change in response to both clonality rate (c) and mutation rate (μ). This approach captures the overall patterns of temporal changes in both genotype composition and allelic states within the population, providing a comprehensive view of how partial clonality shapes population genetics across diverse evolutionary scenarios (Fig. 2).

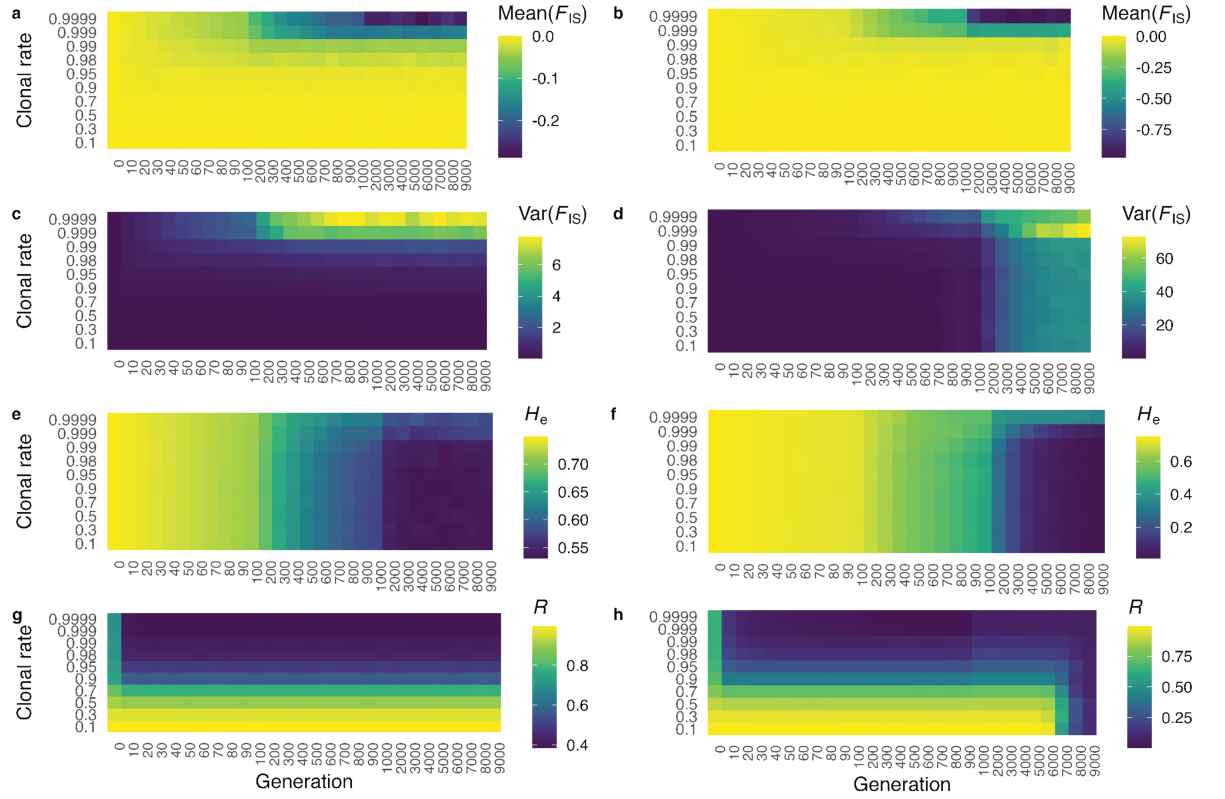


Fig. 2 Effect of clonality and mutation rates on genetic and genotypic diversity. Heatmaps showing the effects of clonality rate (c) and mutation rate (μ) on genetic and genotypic diversity indices across simulated generations. The left column (a, c, e, g) presents results for high mutation rate ($\mu = 10^{-3}$), while the right column (panels b, d, f, h) represents low mutation rate ($\mu = 10^{-6}$). From top to bottom, the panels represent: mean value of inbreeding coefficient (Mean (F_{IS})), variance in inbreeding coefficient (Var (F_{IS})), expected heterozygosity (H_E), and clonal richness (R). Within each graph, the y-axis represents the rate of clonality (c),

ranging from 0.1 to 0.9999, and the x-axis shows generation number (0 to 9,000). Color intensity indicates the mean value over 100 simulation replicates, with yellow representing higher and purple lower values. Note the difference in scale values between mutation rates.

At low to moderately high clonal rates ($c \leq 0.95$), the mean inbreeding coefficient Mean (F_{IS}) at a mutation rate of $\mu = 10^{-3}$ (Fig. 3a) remains close to zero or slightly positive, reflecting minimal deviation from HW expectations over generations even when populations are observed for up to 9,000 generations. However, at higher clonality ($c \geq 0.98$), a pronounced temporal shift emerges, with Mean (F_{IS}) becoming strongly negative after generation 1,000. When the mutation rate is lower $\mu = 10^{-6}$, so when genetic drift is the dominant force, a similar qualitative trend is observed, but Mean (F_{IS}) becomes even more negative at very high clonality ($c \geq 0.999$). Then, the scale of mean (F_{IS}) differs between mutation rates, reaching values close to -0.8 at low mutation rates and around -0.3 at high mutation rates. This illustrates that mutation accumulation is not the main driver of heterozygote excess in partially clonal populations. Values of mean (F_{IS}) have to be examined in the light of the marked genetic drift that apply at low mutation rate ($\mu = 10^{-6}$). Except for very high rates of clonality, expected genetic diversity (H_E) collapses. After generation 6,000, most loci tend to become fixed, and genetic diversity is only preserved within the genotypes frozen by clonal reproduction when heterozygotes by chance, which generate these extreme values of mean (F_{IS}), almost artefactually.

Notably, the variance of the inbreeding coefficient, Var (F_{IS}), exhibits a slightly different pattern, with difference between mutation rates. Under a high mutation rate ($\mu = 10^{-3}$, Fig. 3c) variance remains relatively low across generations, rising moderately to values between approximately 2 and 6 at extreme clonality ($c \geq 0.999$). In contrast, under a low mutation rate ($\mu = 10^{-6}$, Fig. 3d), Var (F_{IS}) stays consistently low over most clonality rates and generations, but exhibits notable increase after about 1,000 generations. At these later stages and higher clonality, Var (F_{IS}) escalates substantially, reaching much larger values ranging roughly from 20 up to 60. In accordance with previous theoretical studies, Var (F_{IS}) does not increase monotonously with clonality rate, the highest value being observed for very high ($c = 0.999$) but not the most extreme c values ($c \geq 0.9999$). Clonal richness (R) measured over the entire population exhibited distinct temporal patterns that varied with both clonality rate and mutation rate (Fig. 3g,h).

Initially, all populations began with high *R* values across the clonality gradient, but these *R* values computed on all individuals almost immediately align with rates of clonality rate. For moderate to high clonality rates (c ≥ 0.5), *R* values declined sharply within the first 10 generations under both mutation rates, decreasing progressively with increasing c, reflecting that the population is composed of fewer clonal lineages. At high mutation rate ($\mu = 10^{-3}$, Fig. 3g), this pattern is invariant throughout the 9,000 generations. At low mutation rate ($\mu = 10^{-6}$, Fig. 3h), low to moderate clonality rates (*c* = 0.1: 0.7) maintained relatively high clonal richness (R ≈ 0.75) up to generation 5,000, but declined after generation 6,000, reaching R ≈ 0.25 by generation 9,000. Here, again, this is an effect of genetic drift that causes a loss of discrimination among individuals because of a reduced set of polymorphic loci. This difference in scale of population indices values reflects how the mutation rate interacts with clonal reproduction to shape the genetic structure and variability over time.

In the following, we focus our analyses on the highest mutation rate ($\mu = 10^{-3}$) since artifactual variations of population genetic indices are observed for low mutation rate ($\mu = 10^{-6}$). We chose to consider generation 6,000 as a good compromise to allow observing very old clones while not enabling genetic drift to fix only one clone at high rates of clonality.

How rates of clonality drive the variability in evolutionary trajectories.

To document the effect of increased rates of clonality on the stochastic variation in population genetic indices, we analyzed the genetic states of populations observed at a single time point (generation 6,000), considering a mutation rate of $\mu = 10^{-3}$ and four alleles per locus (Fig.3). Notably, the variability among replicates of both the mean and variance of inbreeding coefficient (F_{IS}) across loci increases at high rates of clonality (*c* > 0.99), when simulation departs from Hardy-Weinberg expectations. Variation of expected heterozygosity (H_E) and observed heterozygosity (H_O) remain relatively even, with only a slight increase at higher clonality rate (*c* ≤ 0.999), more pronounced for observed heterozygosity (H_O). Notably the variability of H_O is maximal at *c* = 0.999, with simulation replicates that either display values consistent with moderate or high clonality rates.

The clonal richness (*R*) and the clonal evenness (*Pareto* β parameter) are both inversely related to the rate of clonality, but with marked difference at low to moderate *c* values. Whereas *R* displays a bell shape decreasing with *c* values, with no difference in variability across *c* values, *Pareto* β displays an exponential decrease with marked variability among replicates at low clonality rate ($c \leq 0.3$). The variation in these indices indicates that the clonal structure varies greatly even at low values of *c*.

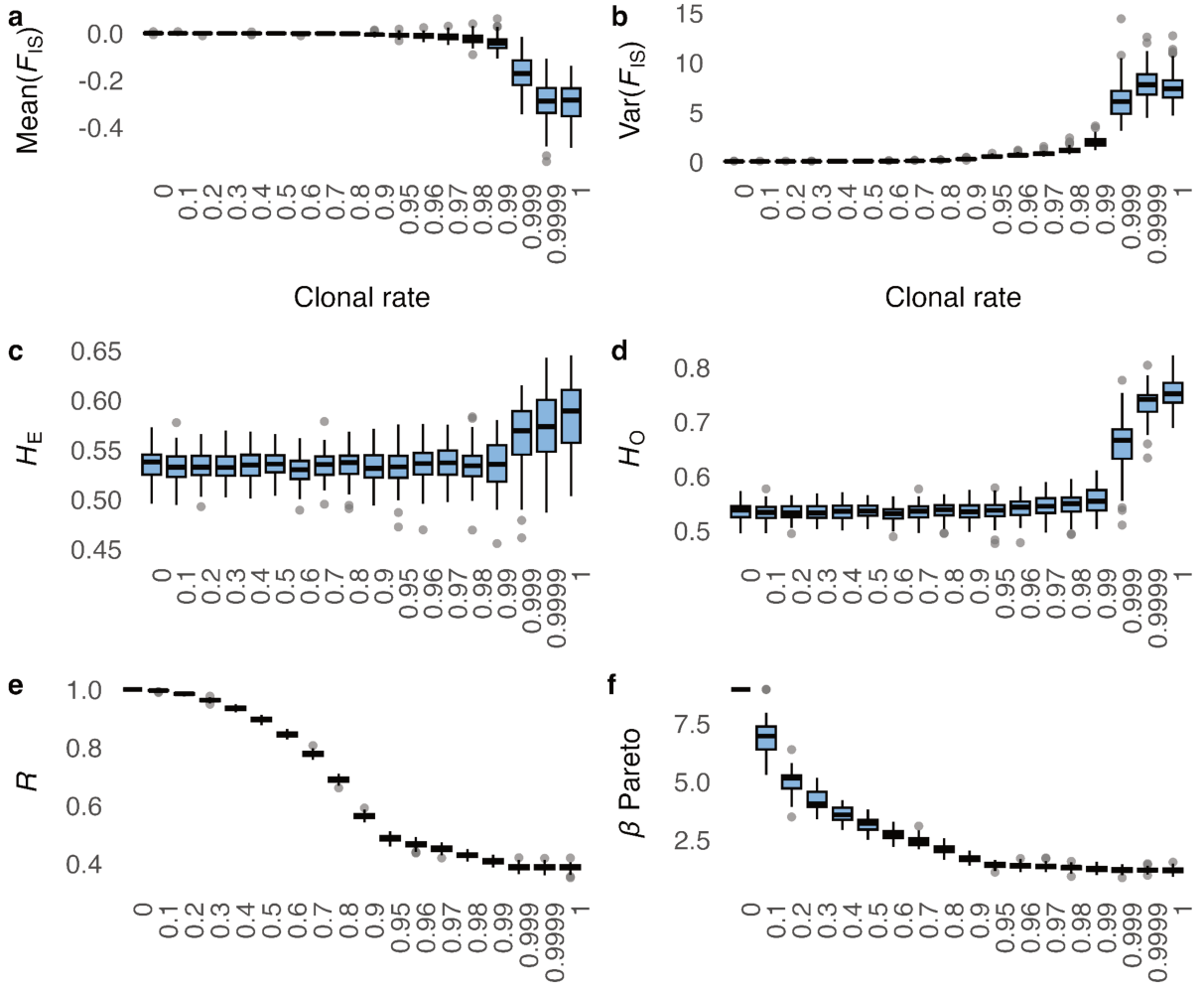


Fig. 3 Effects of clonal rate on population genetic indices. Boxplots show the distribution of population genetic indices based on 100 simulation data replicates per different clonal rates *c*, at generation 6,000. The population genetic characteristics: (a) mean, and (b) variance of the inbreeding coefficient F_{IS}, (c) mean expected heterozygosity H_E, (d) mean observed heterozygosity H_O, (e) clonal richness *R*, (f) Pareto β index The simulation was conducted with a population size of N=1,000, 100 loci, mutation rate $\mu = 10^{-3}$, and four alleles.

The rate of clonality shapes clone age structure

To investigate the impact of the rate of clonality on population age structure, we examined the distribution of clone ages across a broad gradient of clonality rates throughout 10,000 generations (Fig. 4). At low to moderately high clonality ($c \leq 0.8$), clonal age was limited, with clones rarely surviving more than 100 generations. Even at very high clonality ($c = 0.99$), maximum clone age remained below 1,000 generations. In contrast, at $c = 0.999$, a different pattern emerged, with a more flattened distribution of clone ages. Here the population is composed of a mixture of rather young to much older clones, with some of them that could persist by chance for over 9,000 generations. When $c > 0.999$ the population is mostly composed of very old clones, with the extreme case of $c = 1$ in which the population consists solely of clones of the maximum age of 10,000 generations. This sharp transition across the gradient of clonality rates demonstrates that while increasing clonality gradually extends clone ages, the turnover of clones remains very high. Extremely high clonality is required to enable the long-term persistence of clones.

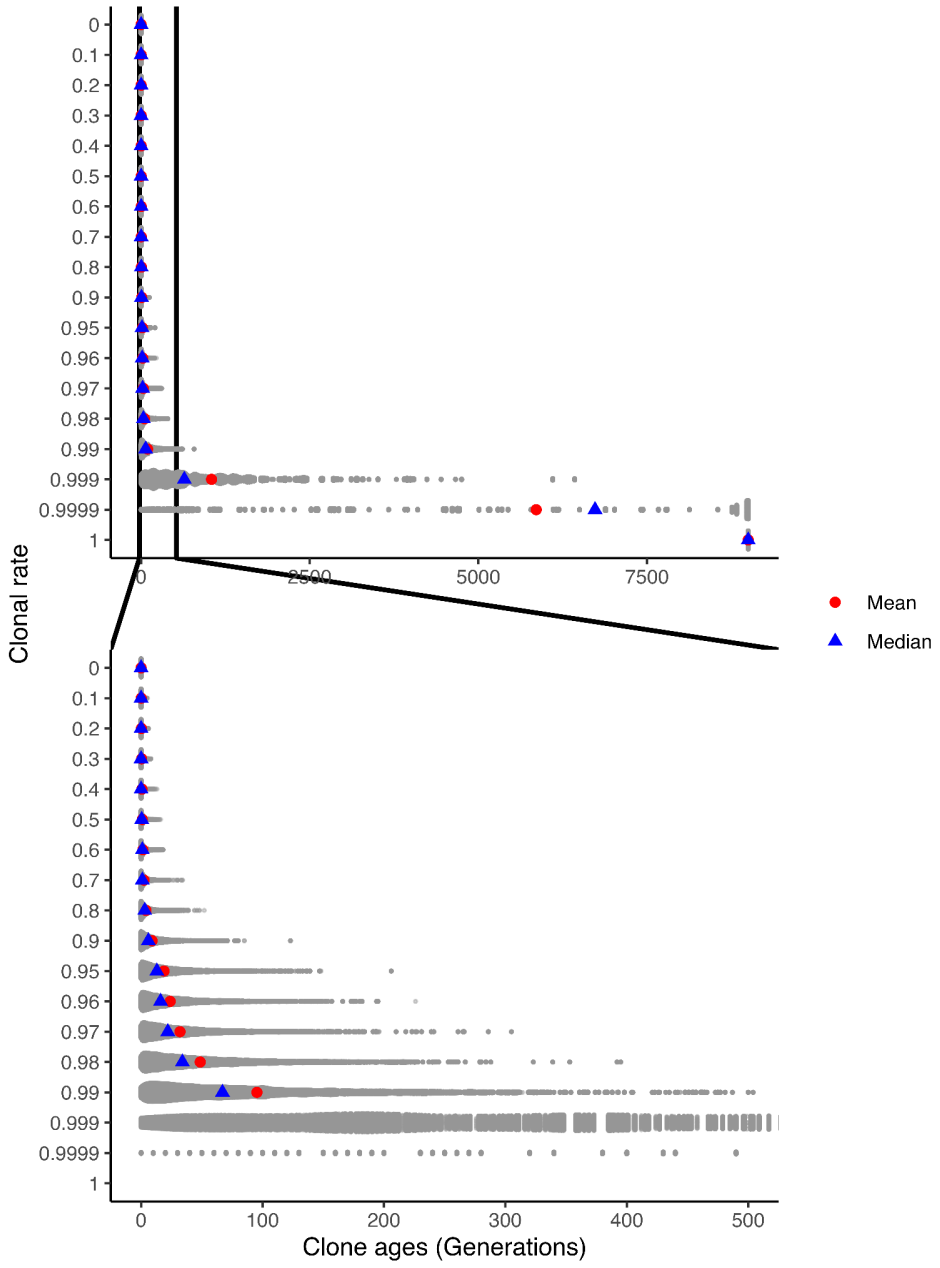


Fig. 4 Clone age distributions across clonal rate gradients. Joyplot showing the distribution of clone ages (generations) across clonality rates. Each density curve represents the age distribution of clones at a specific clonality rate (*c*), with curve height indicating the distribution of individuals at each age. Additional zoom-in on the generation range (0 to 2000) for clearer visualization.

Abundance of clones responds to changes in rates of clonality

To examine the relationship between clone age and clone abundance of individuals under a gradient of *c* values, we analyzed their joint distribution at generation 9,000 (Fig. 5). At low to moderate clonality rates ($c \leq 0.7$), nearly all clones were young, with few individuals persisting beyond 30 generations. Importantly, these clones showed a visible pattern of the distribution of abundance according to their ages, with the logarithm of abundance that fits a linear regression with age (Fig. S1 & S2).

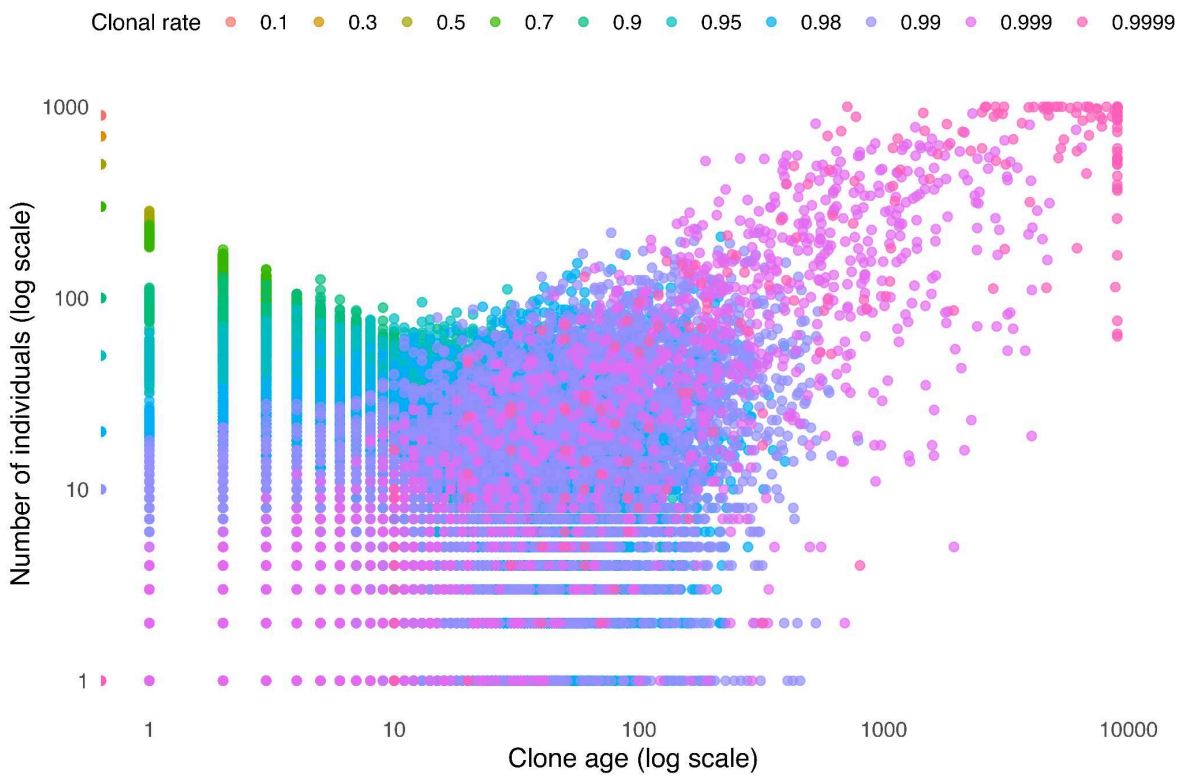


Fig. 5 Clone age distribution across a gradient of clonal rate. Scatterplot illustrating the relationship between clone age (log scale; in generations) and clone size (log scale; number of individuals per clone) at generation 9,000. Each point represents a distinct clone observed in the population at this age. Colors indicate clonality rates (*c*).

However, at higher clonal rates ($c \geq 0.98$), two key shifts emerged: 1) populations contained numerous large clones persisting for hundreds of generations, and 2) the distribution became diffuse, with clone ages and sizes exhibiting a stochastic rather than a structured correlation pattern as observed at lower clonality rates (Fig.S3). This diffusion indicates a change in

age-abundance distribution patterns at extreme clonality, where the structured relationships observed at lower clonality rates become unpredictable (Fig. 6c).

To delve deeper and to quantify the observed transition from structured to stochastic age/abundance patterns, we summarized the distribution of clone ages using four indices : (i) the slope and (ii) the coefficient of determination of the linear regression between the logarithm of clone abundance and clone age at generation 6,000 and (iii) the variance and (iv) the maximum age of clones. The two latter indices highlight how clonality rates shape the longevity of clones, while R^2 and slope values capture the predictability and steepness of age/abundance relationship (Fig. 6).

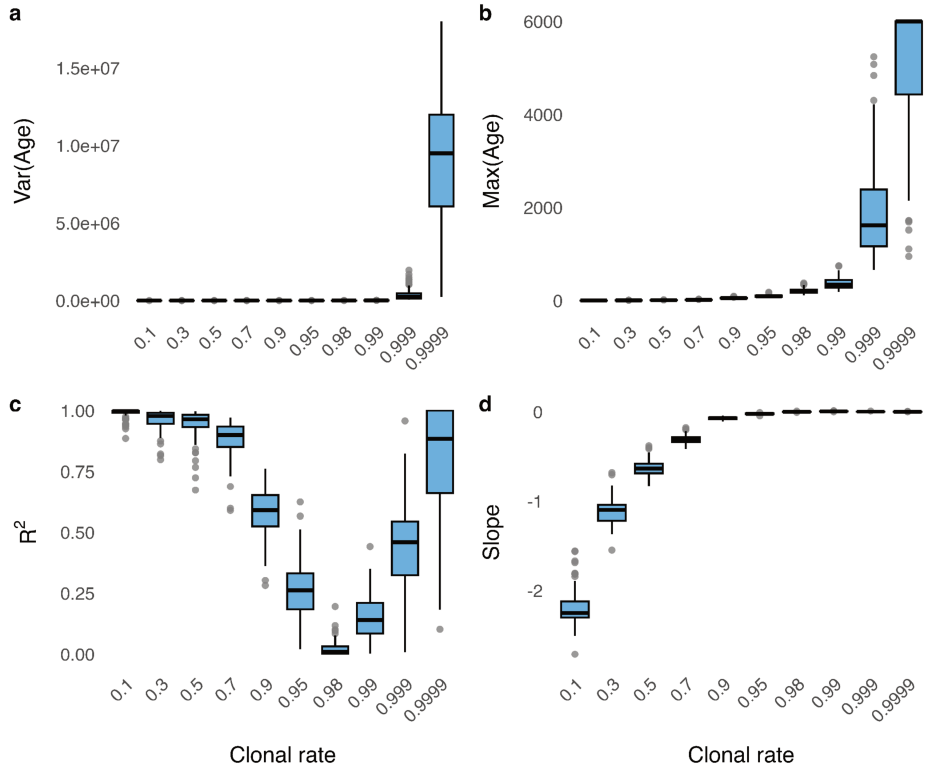


Fig. 6 Clone age distribution metrics across the clonality rate gradient at generation 6,000. Boxplots show (a) variance of clone ages, (b) maximum clone age, (c) coefficient of determination (R^2), and (d) slope; from regression analysis of the distribution of number of individuals and clonal ages across the clonality rate gradient. Each box represents results from 100 replicated simulations per clonality rate.

The variance in clone ages dramatically increased at high clonality rate ($c \geq 0.999$), indicating a wider range of clone age, in line with the increase in maximum clone age that often reached the maximal value of 6,000 generations for c=0.9999 . The coefficient of determination (R^2) was

consistently high at low clonality rates (c = 0.1 to 0.7), indicating well predictable relationships. As the clonality rate increased beyond 0.7, R^2 gradually declined, becoming null at c = 0.98, indicating that abundance of clone age distributed randomly all over the possible clone age in different simulations. Interestingly, at the extreme clonality rates ($c \geq 0.999$), the R^2 values increased again, implying a recovery of the age/abundance relationship. However, visual inspection of the scatter plots (Fig. 5) revealed considerable noise in age/abundance relationship, the correlations being mostly driven by few extremely-old and abundant clones that were present in all simulation replicates (Fig. S4).

Similarly, the slope of the regression line was negative and steepest under low clonality rates, reflecting an exponential decrease in clone abundance with increasing clone age. This slope became flatter (less negative) as clonality rate increased around c = 0.98, when age/abundance became uncorrelated. At the higher clonality rates ($c \geq 0.999$), the slopes are null. Together, these results highlight that the strength and sign of age/abundance relationship vary considerably on either side of a tipping point of maximal unpredictability of clone age distribution.

Summaries of clone age distribution correlates with the variation of some population genetic indices

We then examined how distribution of abundance of clone age impacted population genetic indices, using Spearman correlations among metrics of clone age distributions (Fig. 6) and the value of population genetic indices (Fig. 3), both calculated at generation 6,000 and for simulation data obtained with $\mu = 10^{-3}$ (Fig. 7).

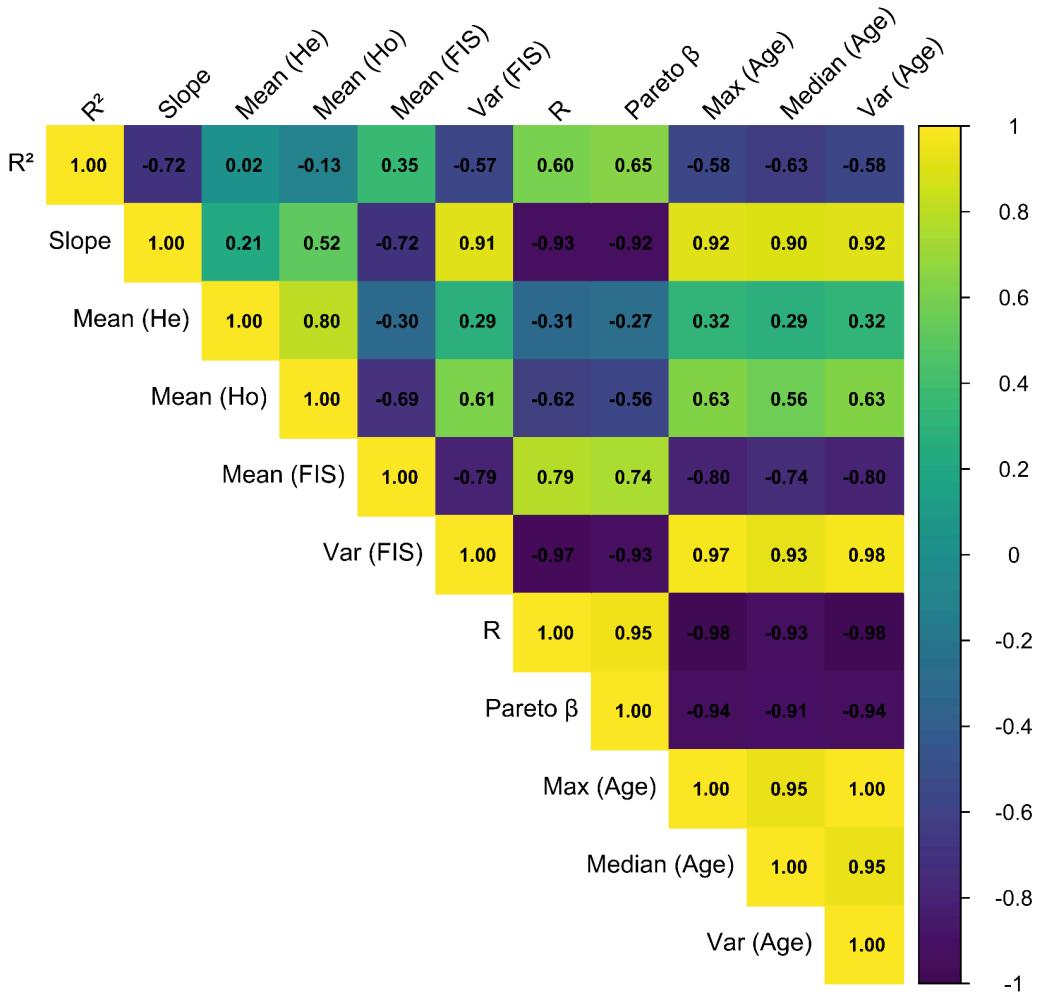


Fig. 7 Relationship between clone age distribution metrics and population genetics indices at generation 6,000. Spearman correlation matrix among population indices and clone age distribution metrics at generation 6,000. Variables include: R^2 and slope (from clone age-size regressions), observed heterozygosity (H_O), expected heterozygosity (H_E), mean and variance of inbreeding coefficient (F_{IS}), clonal richness (R), *Pareto β* (a measure of clone size inequality), and the maximum, median, and variance of clone ages. Each cell shows the correlation coefficient between two variables, color-coded from purple (–1) to yellow (+1).

Both R^2 and slope are correlated with clonal richness (R): R^2 correlates positively ($\rho = 0.60$) while slope correlates highly negatively ($\rho = -0.93$), and similarly with Pareto β ($\rho = 0.65$ and –0.92, respectively). As expected, populations with younger clones maintain higher genotypic diversity (high R) or have more even clone size distributions (high *Pareto β*). Spearman analysis also revealed a positive association ($\rho \geq 0.90$) linking the three clone age longevity measures

(maximum age, median age, and variance of age) to clonal richness (*R*) and to the *Pareto β*, indicating that populations containing very old clones also harbour fewer genotypes, but more evenly abundant. The slope of age/abundance relationship exhibited a strong positive correlation with all longevity measures ($\rho \approx 0.92$), consistent with rapid clone turnover limiting the build-up of ancient clones.

Classic population genetic indices were more weakly connected to the age structure: expected and observed heterozygosity showed modest correlation to clone age measures. Mean and variance of F_{IS} displayed a moderately high correlation with longevity metrics.

Mean F_{IS} shows weak positive correlation with R^2 ($\rho = 0.35$) and moderate negative correlation with slope ($\rho = -0.72$), whereas variance of F_{IS} is associated with both R^2 ($\rho = -0.57$) and slope ($\rho = 0.91$). Both heterozygosity indices (H_O and H_E) display weak correlations with R^2, as well as moderate correlations with the regression slope.

Overall, the correlations show that longevity measures are most strongly linked with clonal structure (*R* and *Pareto β*) and the inbreeding index (F_{IS}), indicating that the persistence of a few long-lasting clones in the population reduces the number of distinct clones and increases genetic similarity among individuals. In contrast, their correlations with genetic diversity among or within individuals (H_E and H_O) are comparatively weak, highlighting that these metrics are less sensitive to clonal age structure.

Detailed examination of the relationships between metrics of clone age distribution provides further explanation on the variability of population genetic indices among replicates

Then we investigate how clone age structure influences overall population genetic diversity, with an emphasis on the variability across simulation replicates. The relationships are illustrated for the slope and R^2 of the clone age distribution (Fig. 8,9) and maximum clone age (Fig. 10) with key genetic indices across clonality rates, at generation 6,000.

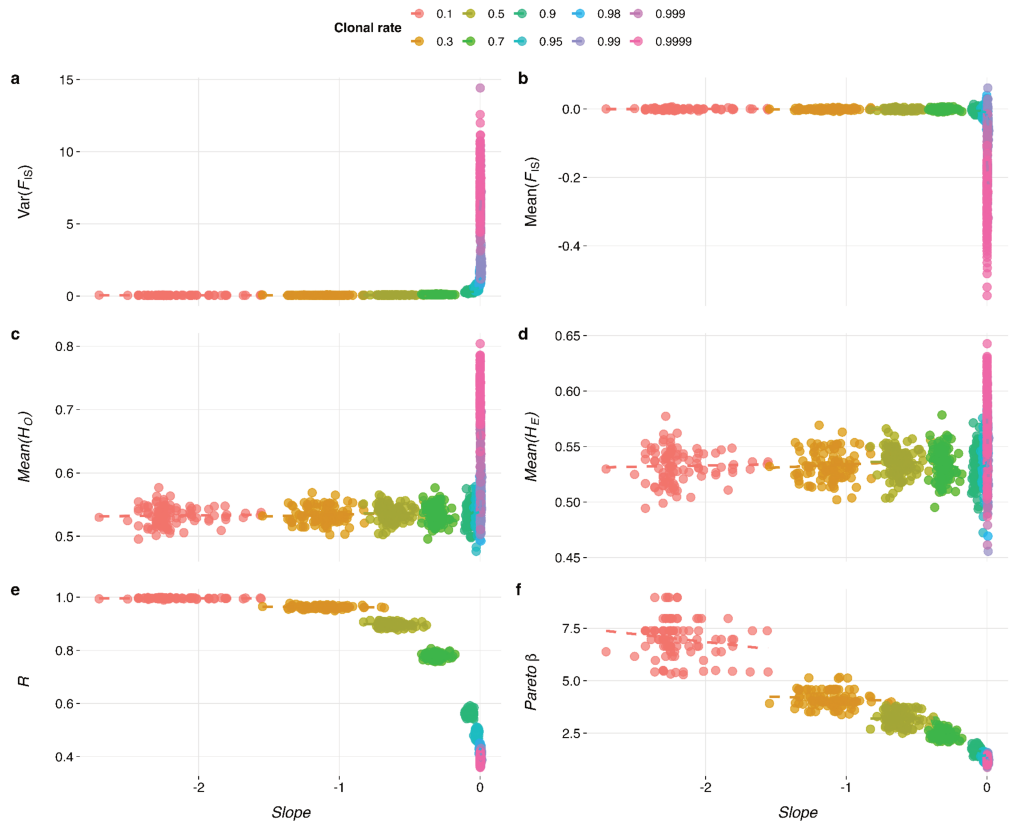


Fig. 8 Relationship between clone age distribution slope and population genetics indices across gradient of clonality rates at generation 6,000. Scatterplots of the slope of the semi-log regression between clone age and log-transformed clone size (x-axis) against six indices (y-axes) computed for each replicate and clonality rate:
(a) variance in F_{IS}, (b) mean F_{IS}, (c) mean observed heterozygosity (H_O), (d) mean expected heterozygosity (H_E), (e) clonal richness (R), and (f) *Pareto* β. Each point represents one of 100 simulation replicates at a given clonality rate. Dashed lines show the best-fit linear regression for each clonality rate distribution.

As illustrated above, the variation in slope values correlates very well with genotypic diversity indices along the clonality gradient (Fig.8). Nevertheless, for low values of c ($c \leq 0.5$), we can observe a lack of correlation among simulation replicates of the same modality. This is very clear for c = 0.1 or 0.2, for *Pareto* β (Fig. 8f). The variance between replicates for Pareto B does not correlate with differences in slope values, which indicates that evenness of genotype abundance is not linked to differences in clone ages (Fig. S1).

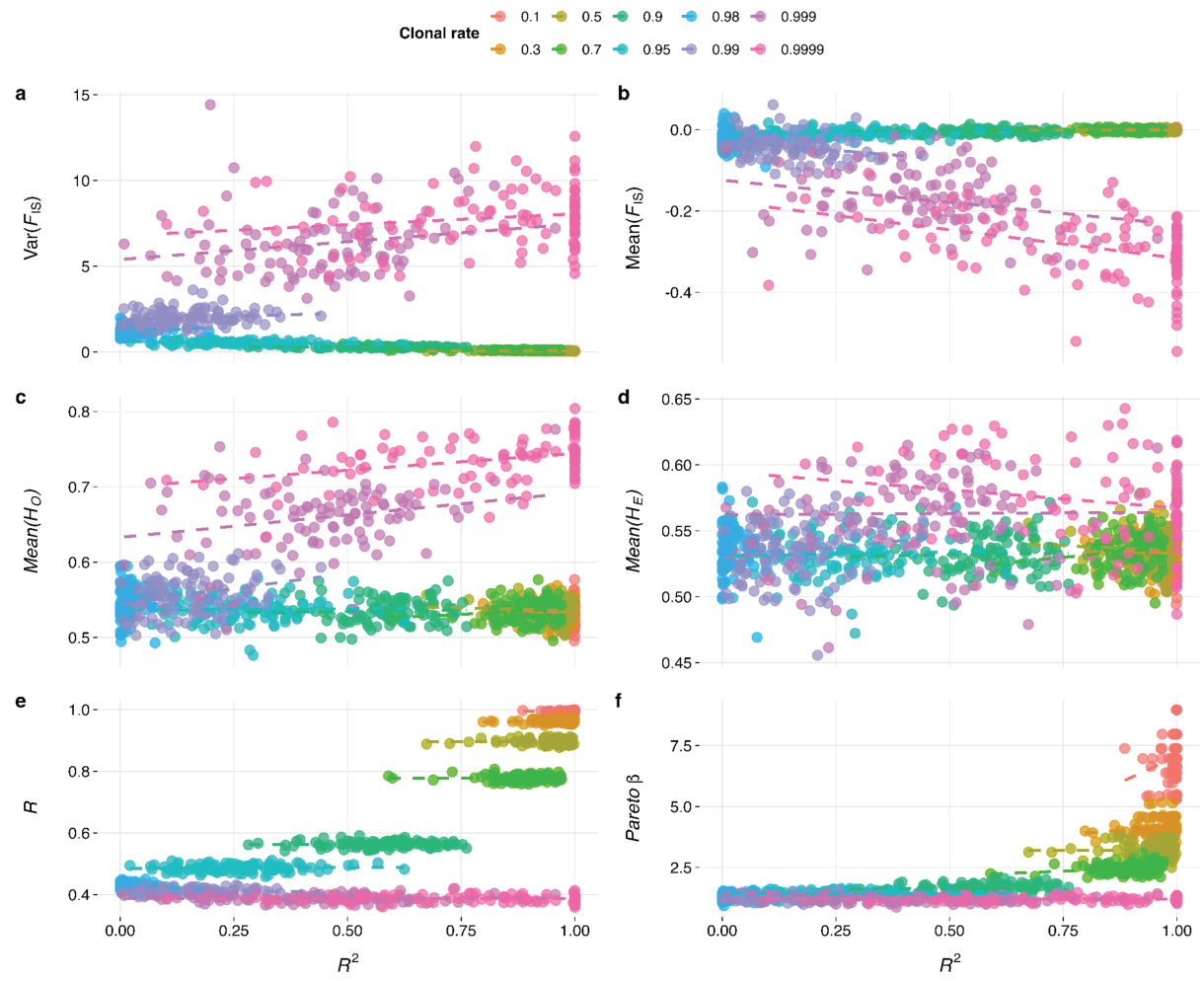


Fig. 9 Relationship between clone age distribution R2 and population genetics indices across a gradient of clonality rates at generation 6,000. Scatterplots of the R^2 of the semi-log regression between clone age and log-transformed abundance (x-axis) against six indices (y-axes) computed for each replicate and clonality rate: (a) variance in F_{IS}, (b) mean F_{IS}, (c) mean observed heterozygosity (H_O), (d) mean expected heterozygosity (H_E), (e) clonal richness (R), and (f) *Pareto* β. Each point represents one of 100 simulation replicates at a given clonality rate. Dashed lines show the best-fit linear regression for each clonality rate distribution.

The other metrics that summarize clone age distribution (R^2 of age/abundance relationship and maximum age) seemed to better explain the variation among simulation replicates, especially for genetic diversity and inbreeding coefficient for high rates of clonality. In particular, there are positive correlations of H_O with R^2 and maximum age for $c \geq 0.99$, which indicates higher heterozygosity with increased clone age among replicates. In parallel, we observed negative

correlations of mean F_{IS} with these metrics of clone age distribution. One final correlation worth noting is that between var (F_{IS}) and R^2, but var (F_{IS}) is not correlated with maximum age. For this relationship, the scatter plots for $c \geq 0.99$ are overlapping. Here, we clearly denote the effect of heterogeneity in clone age distribution among simulation replicates on this key parameter, which is associated with more or less rare recombination events, by chance.

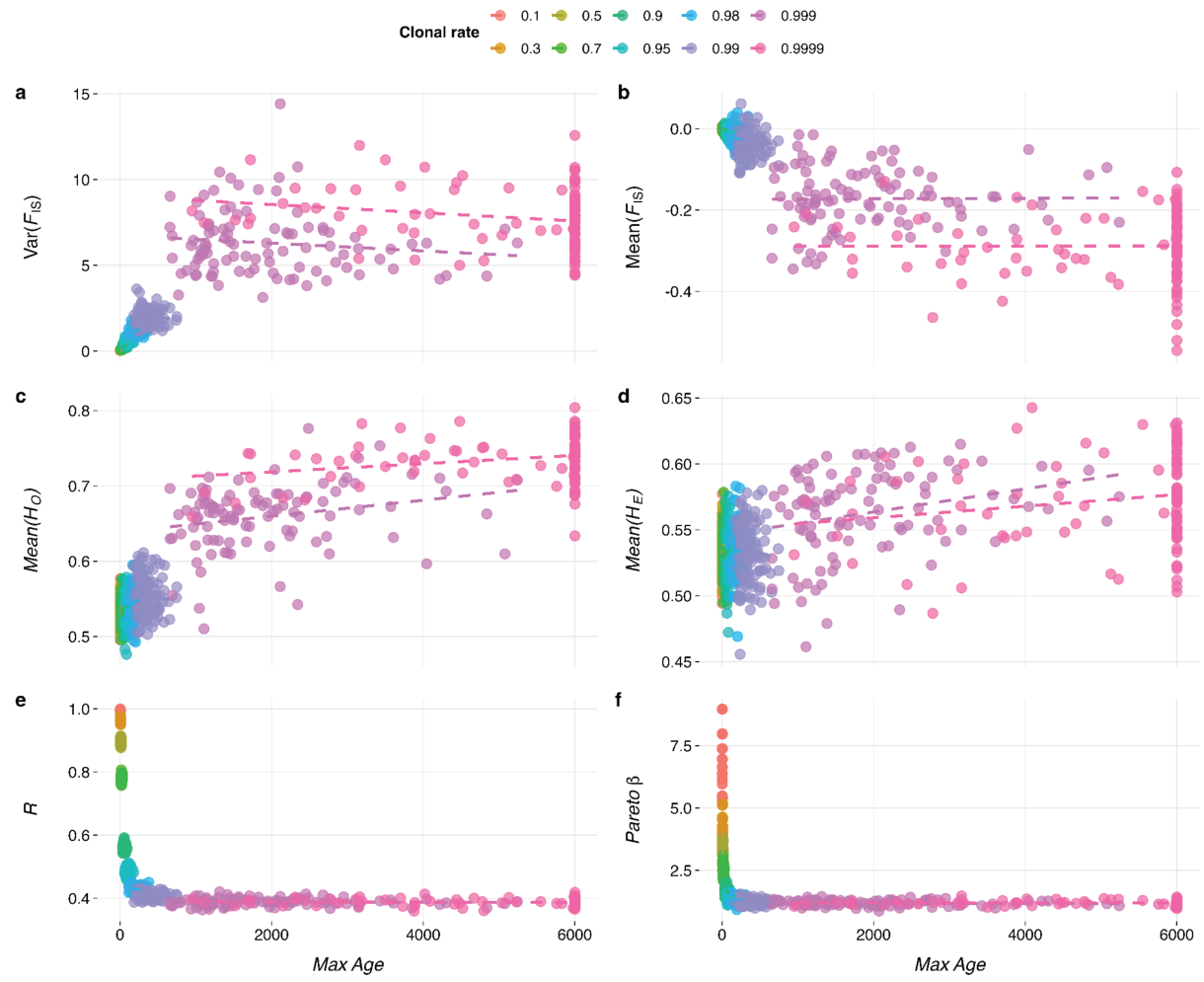


Fig. 10 Relationship between the maximum of clone age distribution and population genetics indices across a gradient of clonality rates at generation 6,000. Scatterplots of the maximum clone age distribution against six indices (y-axes) computed for each replicate and clonality rate: (a) variance in F_{IS}, (b) mean F_{IS}, (c) mean observed heterozygosity (H_O), (d) mean expected heterozygosity (H_E), (e) clonal richness (R), and (f) *Pareto* β. Each point represents one of 100 simulation replicates at a given clonal rate. Dashed lines show the best-fit linear regression for each clonality rate distribution.

Discussion

Our theoretical analysis indicates that patterns of clone age structure vary with the rate of clonality and that different clone age distributions impact in turn population genetic indices. Our results help understand that variations of clone age distributions may explain part of the large ranges of population genetic values found in highly clonal populations. Altogether, these results shed new light on how to better infer clonal structure in natural populations.

The most original aspect of our analysis is the identification of a numerical tipping point in the change of clone age distribution. As the clonality rate increases, we observe a shift from a perfectly compact and determined age structure to completely variable and spread age distributions across replicates. From this tipping point onwards, we observe the emergence of old, even very old, clones, which mix with younger ones. The age structure becomes less variable and spread as the proportion of old clones increases at high clonality rates, but clone age distributions remain highly variable from one simulation to another. This result illustrates, if need be, the high degree of stochasticity in highly clonal populations and the fallacy of adopting deterministically-based interpretations of their evolutionary trajectories (Reichel et al. 2016). This tipping point, at which we observe a completely variable and spread age distribution, actually corresponds to two processes balancing each other out. At low clonality rates, the age distribution is driven by the rapid turnover of clones, whose lifespan is determined to be exponentially decreasing. Indeed, each generation, the proportion of new genotypes created by sexual reproduction is high while the number of genotypes that can pretend to maintain by clonality in the next generation is low. At high clonality rates, the number of surviving clones also decreases, but due to the random fluctuation in clone abundance over time (an analogous of genetic drift, but at the scale of genotypes rather than alleles). This chance of survival applies both for old and young clones, whose establishment and persistence are impeded by the dominance of few (more or less old) clones. Mathematically speaking this tipping point amounts to a diffusion between the two laws that determines the birth and death of clones. In the latter, we discuss how these two processes shape the variation of population genetic indices, hence

offering the opportunity for a better disentangling of the evolutionary trajectories of partially clonal populations.

At low clonality rates, that is below the tipping point, we observe a gradual establishment of clonal structure, with more individuals becoming clonemates. In accordance, we observe a gradual decrease in both clonal richness (*R*) and evenness (*Pareto β*). The best metric that then summarises clone age structure is the slope of the regression between clone abundance and age, which well correlates in turn with the variation of genotypic indices. Notably, the variation in population genetic indices only concerns indices describing genotypic diversity. At low clonality rates, allelic diversity and its apportionment within and between individuals is almost unchanged compared to pure sexual reproduction. The mean inbreeding coefficient (F_{IS}) remains close to zero as the population follows Hardy-Weinberg expectations under near-random mating. Similarly, the variance in F_{IS} is low, reflecting the homogenizing influence of sexual reproduction on the genetic structure of the population towards genotype frequencies expected under Hardy-Weinberg's assumptions. One sexual reproduction event every 100 generations is enough to erase traces of clonal multiplication on allelic diversity and its apportionment within and between individuals (Rouger et al. 2015). At low clonality rates, the pace of clonal turnover imposes transient clonal structure, as a dynamical process dominated by sexual reproduction that continuously reshuffles alleles into new genotypes that may survive as clones for a few generations only, before being replaced by new ones.

[evolutionary trajectory for high c values]

Conversely, as clonality rate exceeds the tipping point ($c \geq 0.98$), populations enter a different evolutionary dynamic characterized by increased unpredictability of clone age distributions that mirrors large variation in population genetic structure (Reichel et al. 2016, Stoeckel et al. 2021). The variation in genotypic indices (clonal diversity, *R,* and evenness, *Pareto β*) is no longer informative since they live very close to their minimum values, while the variation in allele-based indices (H_O, H_E, F_{IS}) live in a range of values typical of high clonality with a very high overall excess of heterozygotes and a high variance in these indices across loci (Balloux et al. 2003, Halkett et al. 2005, Stoeckel & Masson 2014). Differences in clone age distribution, due to a slow turnover of clones making the dynamics of increasing, maintaining and decreasing clones

along generation very stochastic, become the main driver of the variability in genetic indices, which explains why populations with the same rate of clonality can have different genetic signatures. These stochastic trajectories pose major interpretation challenges for applied studies, as the range of possible values of population genetic indices increases and measured values in studied populations lose their bijective reliability at high rates of clonality (Stoeckel et al. 2021). Here, we show that part of this stochasticity, observed for high clonality, is explained by differences in age structure between simulation replicates. Notably, the mean F_{IS} correlates negatively with the maximum age of clones, and the variance of F_{IS} correlates negatively too with the coefficient of determination of the clone age/abundance relationship. These correlations result from different proportions of old and young clones in different populations, with older clones tending to increase individual heterozygosity (H_O) and heterogeneity in genetic composition (variance of F_{IS}) within the population. Beyond the tipping point, we clearly document the combined effect of mutation accumulation and clonal drift (which clones, with the specificities of their allele apportionments when they were produced by sexual gamete mixing, have randomly increased in size along generations) on the evolution of highly clonal populations.

As expected (Reichel et al. 2016, Stoeckel et al. 2021), the temporal change of classical population genetic indices show that it takes much more time at high rates of clonality to reach steady-state values. This is especially true for the dynamics of F_{IS} and leads to the very large disparity between the evolutionary trajectories of low or highly clonal populations. As a result, it is difficult to compare the variation of population genetic indices all else being equal (in time). Typically, the mutation-drift equilibrium depends on the value of c (Reichel et al. 2016), which results in a distortion in time along c values. This strong interplay between time (left so that clonal drift has acted), rate of clonality and values of population genetic indices is difficult to grasp, but can be easily explained by the differences in turnover between clones, depending on the rate of clonality. Up to values of c = 0.99, the age of the oldest clones is negligible compared to the duration of a simulation run. However, this is no longer true at higher c value, with an increasing proportion of old clones, which can by chance reach the maximal age set to the

duration of the simulation run. The disadvantage of such a long simulation duration is that it leaves room for drift to fix alleles if mutation does not compensate for the loss of polymorphism. Typically, this leads to artifact values of population genetics indices for very low mutation rates ($\mu = 10^{-6}$). Because of these extremely slow dynamics and these artifacts, it is difficult to compare theoretical expectations with empirically observed values for low mutation rates. Therefore, we advocate using highly variable molecular markers, such as microsatellite loci, when monitoring partially clonal populations sampled *in natura*.

Our results clearly underline the complexity of inferring clonality rates. The inference of *c* is hindered by a double paradox regarding the variation of population genetic indices along clonality rates.

At low clonality rates, the variation in clonal richness R is fully determined by *c* value (Stoeckel et al. 2020), which would naively suggest that the values of *c* can be easily identified. Unfortunately, the R index is highly biased and subjected to strong variability when considering subsamples of the total population. Our study shows that, unfortunately, the age distribution is quite noisy too, at low *c* values, as evidenced by the variability of the slope of the clone age/abundance relationship. No index derived from the age distribution can be confidently used to infer low *c* values. Worse still, the age structure does not correspond to the relative abundance of genotypes captured by the Pareto B index, which adds to the complexity of analysing such transient clonal structure. Therefore, contrary to what might be expected in a deterministic case (Burt et al., 1996), we cannot use clone resampling to accurately estimate small values of *c* (Ali et al., 2016).

At high clonality rates, the issue of inferring *c* value is reversed. The index that correlates best with the variation in *c* is F_{IS}. With little bias, this index is extremely robust to variation in sample size, but unfortunately it shows high variability between simulations for the same values of *c*, as expected from the study of theoretical distributions obtained analytically (Stockel & Masson 2014). The inference of *c* is hindered here by the high stochasticity of evolutionary trajectories in highly clonal populations. Our detailed analysis shows, however, that part of this variability can be explained by differences in the age structure of clones, particularly through differences

in relative abundance of young and old clones. This signal, linked to clonal drift, could provide additional information in combination with the signal provided by changes in allele frequencies through time, which is already used in the Bayesian method ClonEstiMate (Becheler et al. 2017).

Ultimately, our simulation study indicates that clone age structure is more insightful to decipher long lasting clonality rather than transient genotype turn-over. This property can be used to better understand the origin of clonal structures in species with mixed reproduction, presenting a mixture of sexual and asexual lineages. In this case, an important biological question is whether all individuals form a single population (as in our simulations) or whether there is additional reproductive isolation segregating sexual and asexual lineages (such as determinism for the loss of sex). There are many cases of coexistence of sexual and asexual lineages (Halkett et al. 2005, Abdalrahem et al. 2025). Our study shows that the observation of persistent clones is indicative of a very high rate of clonality, even if these clones are in the minority in a sample. In these cases, the use of clonality inference methods (Ali et al. 2016, Becheler et al. 2017) would have led to the inference of a low rate of clonality, considering that all individuals form the same population. Conversely, resampling of the same multilocus genotypes in poplar rust over nearly thirty years (Abdalrahem et al. 2025) confirms a split within the population between sexual and asexual lineages.

References

Abdalrahem, A., Andrieux, A., Becheler, R., Duplessis, S., Frey, P., Marçais, B., Schiffer-Forsyth, K., Stoeckel, S., & Halkett, F. (2025). *Long-lasting coexistence of multiple asexual lineages alongside their sexual counterparts in a fungal plant pathogen* (p. 2025.03.28.645883). bioRxiv. https://doi.org/10.1101/2025.03.28.645883

Ali, S., Gladieux, P., Rahman, H., Saqib, M. S., Fiaz, M., Ahmad, H., Leconte, M., Gautier, A., Justesen, A. F., Hovmøller, M. S., Enjalbert, J., & de Vallavieille-Pope, C. (2014). Inferring the contribution of sexual reproduction, migration and off-season survival to the temporal maintenance of microbial populations: A case study on the wheat fungal pathogen uccinia striiformis f.sp. tritici. *Molecular Ecology*, *23*(3), 603–617. https://doi.org/10.1111/mec.12629

Ali, S., Soubeyrand, S., Gladieux, P., Giraud, T., Leconte, M., Gautier, A., Mboup, M., Chen, W., de Vallavieille-Pope, C., & Enjalbert, J. (2016). cloncase: Estimation of sex frequency and effective population size by clonemate resampling in partially clonal organisms. *Molecular Ecology Resources*, *16*(4), 845–861. https://doi.org/10.1111/1755-0998.12511

Arnaud-Haond, S., Duarte, C. M., Alberto, F., & Serrão, E. A. (2007). Standardizing methods to address clonality in population studies. *Molecular Ecology*, *16*(24), 5115–5139. https://doi.org/10.1111/j.1365-294X.2007.03535.x

Arnaud-Haond, S., Stoeckel, S., & Bailleul, D. (2020). New insights into the population genetics of partially clonal organisms: When seagrass data meet theoretical expectations. *Molecular Ecology*, *29*(17), 3248–3260. https://doi.org/10.1111/mec.15532

Balloux, F., Lehmann, L., & de Meeûs, T. (2003). The Population Genetics of Clonal and Partially Clonal Diploids. *Genetics*, *164*(4), 1635–1644. https://doi.org/10.1093/genetics/164.4.1635

Bazin, É., Mathé-Hubert, H., Facon, B., Carlier, J., & Ravigné, V. (2014). The effect of mating system on invasiveness: Some genetic load may be advantageous when invading new environments. *Biological Invasions*, *16*(4), 875–886. https://doi.org/10.1007/s10530-013-0544-6

Becheler, R., Masson, J.-P., Arnaud-Haond, S., Halkett, F., Mariette, S., Guillemin, M.-L., Valero, M., Destombe, C., & Stoeckel, S. (2017). ClonEstiMate, a Bayesian method for quantifying rates of clonality of populations genotyped at two-time steps. *Molecular Ecology Resources*, *17*(6), e251–e267. https://doi.org/10.1111/1755-0998.12698

Bhargava, A., & Fuentes, F. F. (2010). Mutational Dynamics of Microsatellites. *Molecular Biotechnology*, *44*(3), 250–266. https://doi.org/10.1007/s12033-009-9230-4

Birky, C. W., Jr. (1996). Heterozygosity, Heteromorphy, and Phylogenetic Trees in Asexual Eukaryotes. *Genetics*, *144*(1), 427–437. https://doi.org/10.1093/genetics/144.1.427

Brakefield, P. M. (1989). The Variance in Genetic Diversity Among Subpopulations is More Sensitive to Founder Effects and Bottlenecks Than is the Mean: A Case Study. In A. Fontdevila (Ed.), *Evolutionary Biology of Transient Unstable Populations* (pp. 145–161). Springer. https://doi.org/10.1007/978-3-642-74525-6_10

Burt, A., Carter, D. A., Koenig, G. L., White, T. J., & Taylor, J. W. (1996). Molecular markers reveal cryptic sex in the human pathogen Coccidioides immitis. *Proceedings of the National Academy of Sciences of the United States of America*, *93*(2), 770–773. https://doi.org/10.1073/pnas.93.2.770

CABRERA-TOLEDO, D., GONZÁLEZ-ASTORGA, J., & VOVIDES, A. P. (2008). Heterozygote excess in ancient populations of the critically endangered Dioon caputoi (Zamiaceae, Cycadales) from central Mexico. *Botanical Journal of the Linnean Society*, *158*(3), 436–447. https://doi.org/10.1111/j.1095-8339.2008.00868.x

Carpenter, K. E., Abrar, M., Aeby, G., Aronson, R. B., Banks, S., Bruckner, A., Chiriboga, A., Cortés, J., Delbeek, J. C., Devantier, L., Edgar, G. J., Edwards, A. J., Fenner, D., Guzmán, H. M., Hoeksema, B. W., Hodgson, G., Johan, O., Licuanan, W. Y., Livingstone, S. R., … Wood, E. (2008). One-third of reef-building corals face elevated extinction risk from climate change and local impacts. *Science (New York, N.Y.)*, *321*(5888), 560–563. https://doi.org/10.1126/science.1159196

Chang, H.-H., Moss, E. L., Park, D. J., Ndiaye, D., Mboup, S., Volkman, S. K., Sabeti, P. C., Wirth, D. F., Neafsey, D. E., & Hartl, D. L. (2013). Malaria life cycle intensifies both natural selection and random genetic drift. *Proceedings of the National Academy of Sciences*, *110*(50), 20129–20134. https://doi.org/10.1073/pnas.1319857110

Duplessis, S., Lorrain, C., Petre, B., Figueroa, M., Dodds, P. N., & Aime, M. C. (2021). Host Adaptation and Virulence in Heteroecious Rust Fungi. *Annual Review of Phytopathology*, *59*(1), 403–422. https://doi.org/10.1146/annurev-phyto-020620-121149

Early, A. M., Camponovo, F., Pelleau, S., Cerqueira, G. C., Lazrek, Y., Volney, B., Carrasquilla, M., de Thoisy, B., Buckee, C. O., Childs, L. M., Musset, L., & Neafsey, D. E. (2022). Declines in prevalence alter the optimal level of sexual investment for the malaria parasite Plasmodium falciparum. *Proceedings of the National Academy of Sciences*, *119*(30), e2122165119. https://doi.org/10.1073/pnas.2122165119

Fry, W. E., Birch, P. R. J., Judelson, H. S., Grünwald, N. J., Danies, G., Everts, K. L., Gevens, A. J., Gugino, B. K., Johnson, D. A., Johnson, S. B., McGrath, M. T., Myers, K. L., Ristaino, J. B., Roberts, P. D., Secor, G., & Smart, C. D. (2015). Five Reasons to Consider Phytophthora infestans a Reemerging Pathogen. *Phytopathology®*, *105*(7), 966–981. https://doi.org/10.1094/PHYTO-01-15-0005-FI

Gavino, P. D., Smart, C. D., Sandrock, R. W., Miller, J. S., Hamm, P. B., Lee, T. Y., Davis, R. M., & Fry, W. E. (2000). Implications of Sexual Reproduction for Phytophthora infestans in the United States: Generation of an Aggressive Lineage. *Plant Disease*, *84*(7), 731–735. https://doi.org/10.1094/PDIS.2000.84.7.731

Gazave, E., Chang, D., Clark, A. G., & Keinan, A. (2013). Population Growth Inflates the Per-Individual Number of Deleterious Mutations and Reduces Their Mean Effect. *Genetics*, *195*(3), 969–978. https://doi.org/10.1534/genetics.113.153973

Halkett, F., Simon, J.-C., & Balloux, F. (2005). Tackling the population genetics of clonal and partially clonal organisms. *Trends in Ecology & Evolution*, *20*(4), 194–201. https://doi.org/10.1016/j.tree.2005.01.001

Hanko, G. R., Vogel, M. T., Negrón-Ortiz, V., & Moore, R. C. (2023). High Prevalence of Clonal Reproduction and Low Genetic Diversity in Scutellaria floridana, a Federally Threatened Florida-Endemic Mint. *Plants*, *12*(4), Article 4. https://doi.org/10.3390/plants12040919

Hartfield, M. (2016). Evolutionary genetic consequences of facultative sex and outcrossing. *Journal of Evolutionary Biology*, *29*(1), 5–22. https://doi.org/10.1111/jeb.12770

Ho, E. K. H., Bartkowska, M., Wright, S. I., & Agrawal, A. F. (2019). Population genomics of the facultatively asexual duckweed Spirodela polyrhiza. *New Phytologist*, *224*(3), 1361–1371. https://doi.org/10.1111/nph.16056

Lehtonen, J., Jennions, M. D., & Kokko, H. (2012). The many costs of sex. *Trends in Ecology & Evolution*, *27*(3), 172–178. https://doi.org/10.1016/j.tree.2011.09.016

Lodé, T. (2013). Adaptive Significance and Long-Term Survival of Asexual Lineages. *Evolutionary Biology*, *40*(3), 450–460. https://doi.org/10.1007/s11692-012-9219-y

Mark Welch, D. B., & Meselson, M. (2000). Evidence for the Evolution of Bdelloid Rotifers Without Sexual Reproduction or Genetic Exchange. *Science*, *288*(5469), 1211–1215. https://doi.org/10.1126/science.288.5469.1211

McCulloch, S. D., & Kunkel, T. A. (2008). The fidelity of DNA synthesis by eukaryotic replicative and translesion synthesis polymerases. *Cell Research*, *18*(1), 148–161. https://doi.org/10.1038/cr.2008.4

McDonald, B. A., & Linde, C. (2002). Pathogen Population Genetics, Evolutionary Potential, and Durable Resistance. *Annual Review of Phytopathology*, *40*(1), 349–379. https://doi.org/10.1146/annurev.phyto.40.120501.101443

Nébavi, F., Ayala, F. J., Renaud, F., Bertout, S., Eholié, S., Moussa, K., Mallié, M., & de Meeûs, T. (2006). Clonal population structure and genetic diversity of Candida albicans in AIDS patients from Abidjan (Côte d'Ivoire). *Proceedings of the National Academy of Sciences*, *103*(10), 3663–3668. https://doi.org/10.1073/pnas.0511328103

Nei, M. (1978). Estimation of Average Heterozygosity and Genetic Distance from a Small Number of Individuals. *Genetics*, *89*(3), 583–590.

Orive, M. E., & Krueger-Hadfield, S. A. (2021). Sex and Asex: A Clonal Lexicon. *Journal of Heredity*, *112*(1), 1–8. https://doi.org/10.1093/jhered/esaa058

Payseur, B. A., & Cutter, A. D. (2006). Integrating patterns of polymorphism at SNPs and STRs. *Trends in Genetics*, *22*(8), 424–429. https://doi.org/10.1016/j.tig.2006.06.009

Pierre, J.-S., Stoeckel, S., & Wajnberg, E. (2022). The advantage of sex: Reinserting fluctuating selection in the pluralist approach. *PLOS ONE*, *17*(8), e0272134. https://doi.org/10.1371/journal.pone.0272134

Putman, A. I., & Carbone, I. (2014). Challenges in analysis and interpretation of microsatellite data for population genetic studies. *Ecology and Evolution*, *4*(22), 4399–4428. https://doi.org/10.1002/ece3.1305

Raynes, Y., Wylie, C. S., Sniegowski, P. D., & Weinreich, D. M. (2018). Sign of selection on mutation rate modifiers depends on population size. *Proceedings of the National Academy of Sciences*, *115*(13), 3422–3427. https://doi.org/10.1073/pnas.1715996115

Reichel, K., Masson, J.-P., Malrieu, F., Arnaud-Haond, S., & Stoeckel, S. (2016). Rare sex or out of reach equilibrium? The dynamics of FISin partially clonal organisms. *BMC Genetics*, *17*(1), 76. https://doi.org/10.1186/s12863-016-0388-z

Reynes, L., Thibaut, T., Mauger, S., Blanfuné, A., Holon, F., Cruaud, C., Couloux, A., Valero, M., & Aurelle, D. (2021). Genomic signatures of clonality in the deep water kelp Laminaria rodriguezii. *Molecular Ecology*, *30*(8), 1806–1822. https://doi.org/10.1111/mec.15860

Rijnsoever, F. J. van. (2017). (I Can't Get No) Saturation: A simulation and guidelines for sample sizes in qualitative research. *PLOS ONE*, *12*(7), e0181689. https://doi.org/10.1371/journal.pone.0181689

Roze, D. (2012). Disentangling the Benefits of Sex. *PLOS Biology*, *10*(5), e1001321. https://doi.org/10.1371/journal.pbio.1001321

Stoeckel, S., Arnaud-Haond, S., & Krueger-Hadfield, S. A. (2021). The Combined Effect of Haplodiplonty and Partial Clonality on Genotypic and Genetic Diversity in a Finite Mutating Population. *Journal of Heredity*, *112*(1), 78–91. https://doi.org/10.1093/jhered/esaa062

Stoeckel, S., Becheler, R., Bocharova, E., & Barloy, D. (2024). GenAPoPop 1.0: A user-friendly software to analyse genetic diversity and structure from partially clonal and selfed autopolyploid organisms. *Molecular Ecology Resources*, *24*(1), e13886. https://doi.org/10.1111/1755-0998.13886

Stoeckel, S., & Masson, J.-P. (2014). The Exact Distributions of FIS under Partial Asexuality in Small Finite Populations with Mutation. *PLOS ONE*, *9*(1), e85228. https://doi.org/10.1371/journal.pone.0085228

Stoeckel, S., Porro, B., & Arnaud-Haond, S. (2021). The discernible and hidden effects of clonality on the genotypic and genetic states of populations: Improving our estimation of clonal rates. *Molecular Ecology Resources*, *21*(4), 1068–1084. https://doi.org/10.1111/1755-0998.13316

Wright, S. (1931). Evolution in Mendelian Populations. *Genetics*, *16*(2), 97–159.

Wright, S. (1949). Adaptation and selection. *Genetics, Paleontology and Evolution*, 365–389.

Supplementary figures

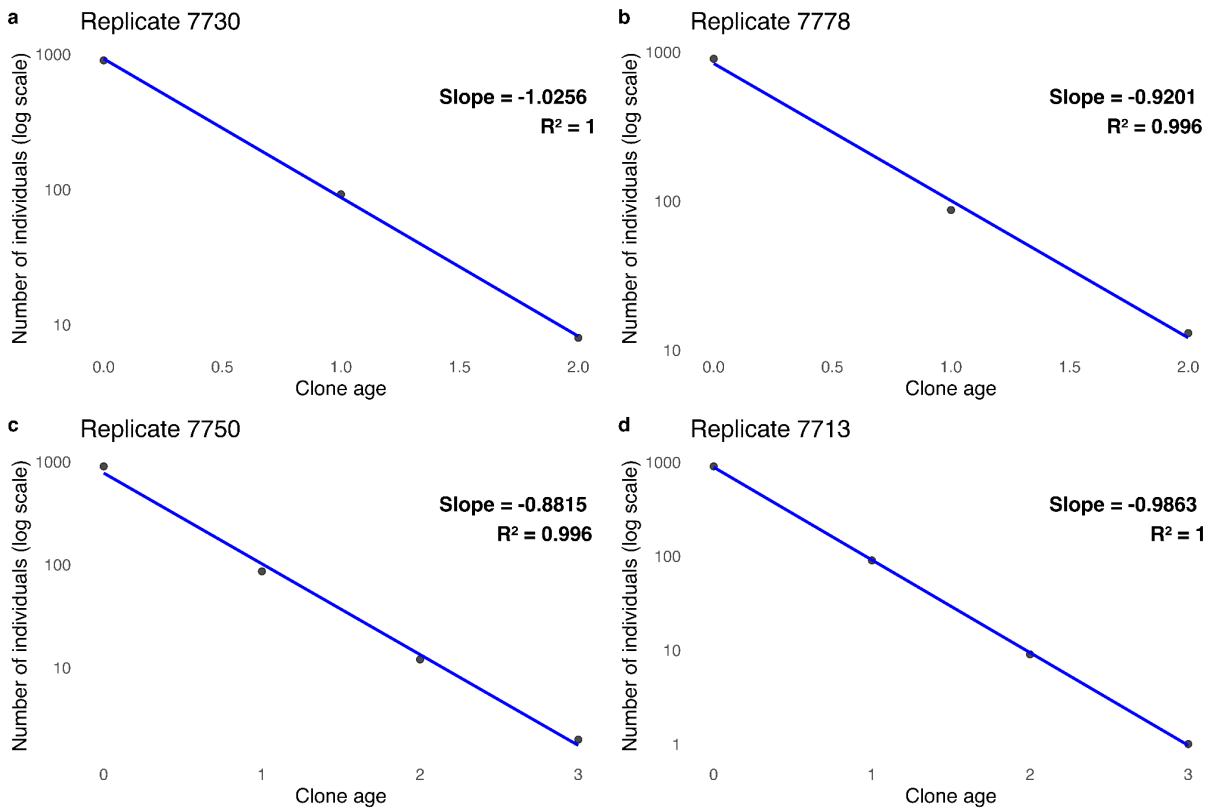


Fig. S1 Distribution of clone ages at generation 6,000 with a clonality rate of 0.1. Scatterplots display the relationship between clone age and the number of individuals per clone (log scale) for four random simulation replicates with a clonality rate of 0.1, at generation 6,000. Each blue line is the best-fit regression, with annotated slope and R^2 values

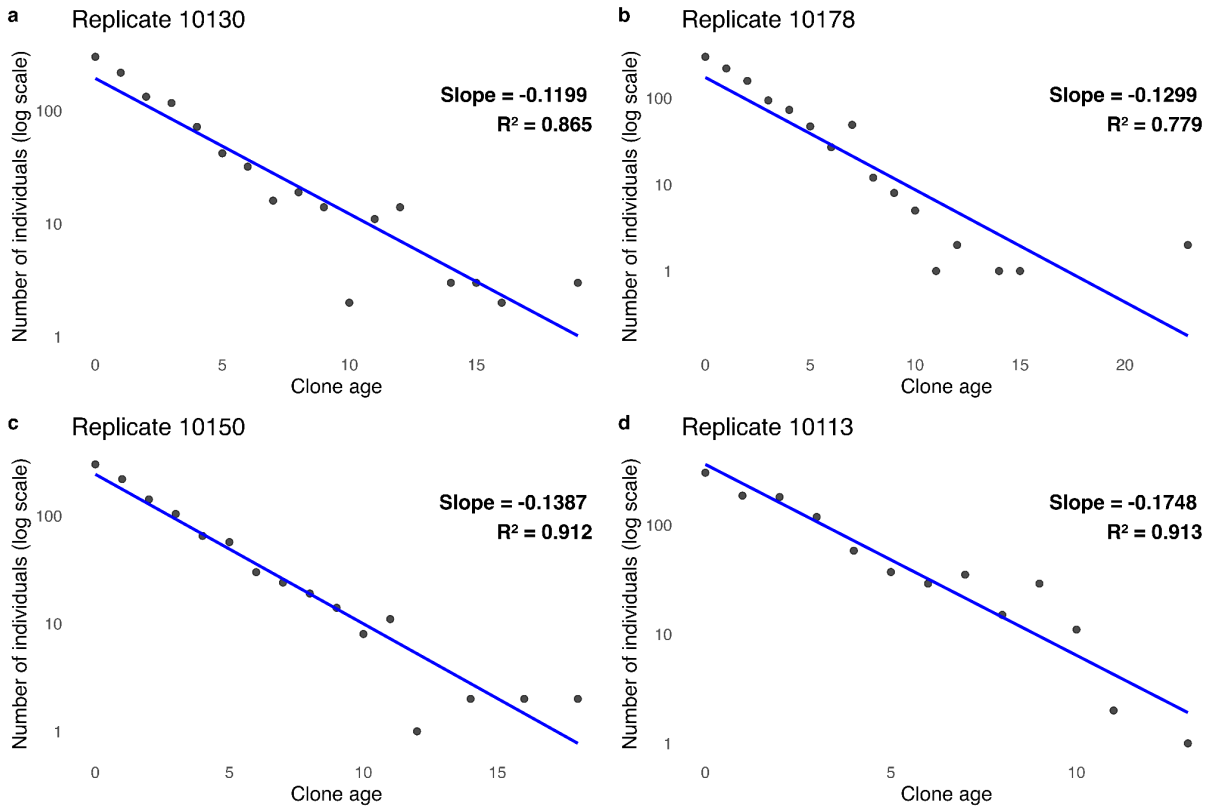


Fig. S2 Distribution of clone ages at generation 6,000 with a clonality rate of 0.7. Scatterplots display the relationship between clone age and the number of individuals per clone (log scale) for four random simulation replicates with a clonal rate of 0.7, at generation 6,000. Each blue line is the best-fit regression, with annotated slope and R^2 values

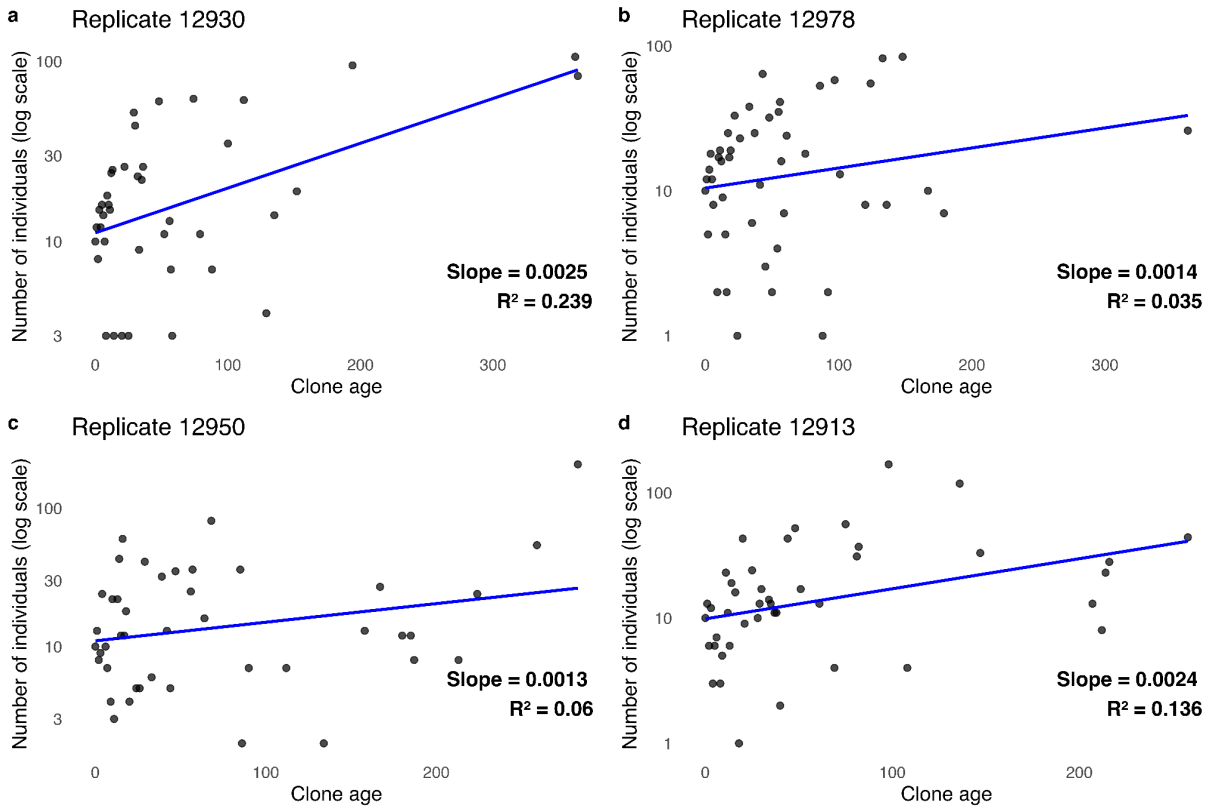


Fig. S3 Distribution of clone ages at generation 6,000 with a clonality rate of 0.99. Scatterplots display the relationship between clone age and the number of individuals per clone (log scale) for four random simulation replicates with a clonal rate of 0.99, at generation 6,000. Each blue line is the best-fit regression, with annotated slope and R^2 values

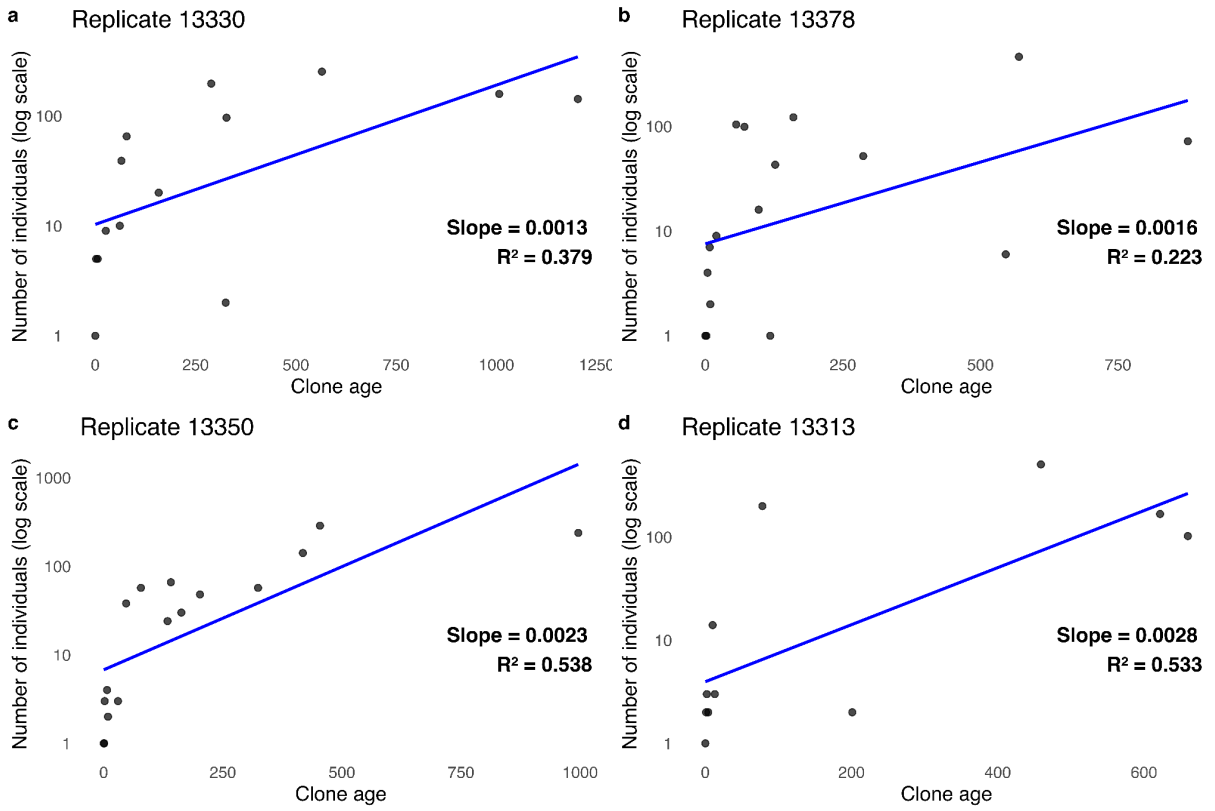


Fig. S4 Distribution of clone ages at generation 6,000 with a clonality rate of 0.999. Scatterplots display the relationship between clone age and the number of individuals per clone (log scale) for four random simulation replicates with a clonal rate of 0.999, at generation 6,000. Each blue line is the best-fit regression, with annotated slope and R^2 values

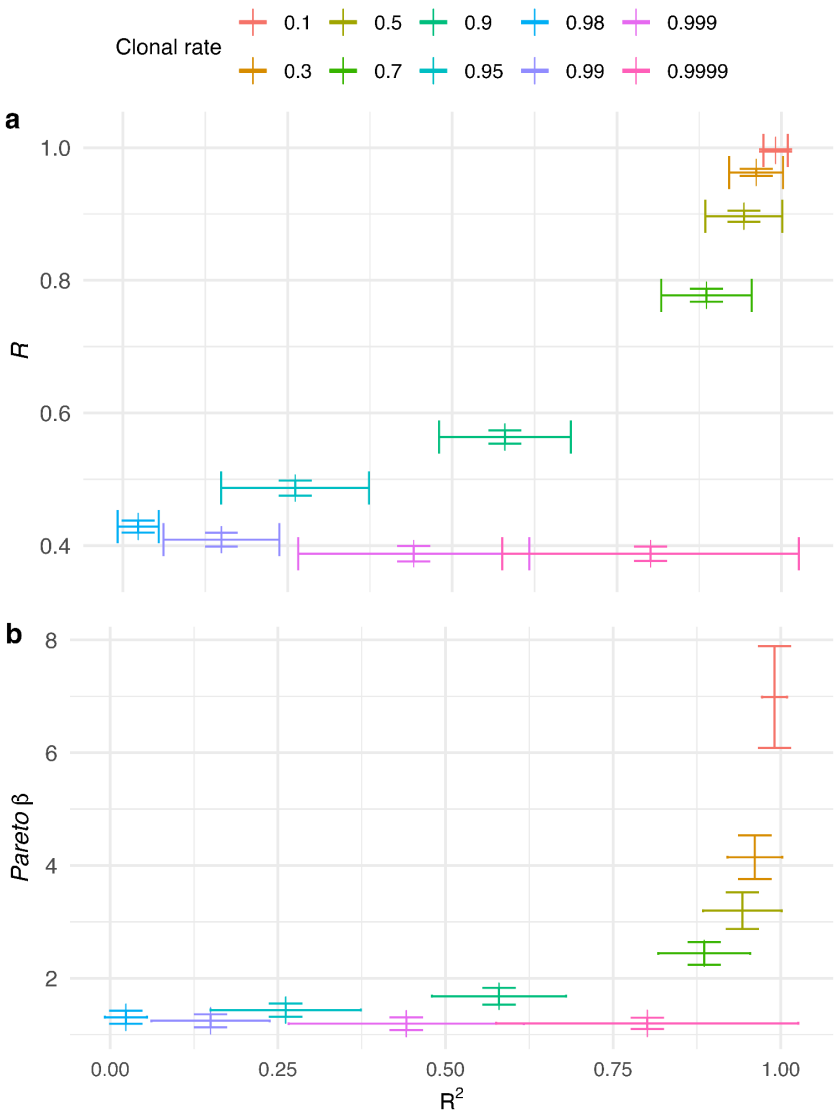


Fig. S5 Relationship between clone age distribution predictability (R^2), clonal richness (*R*), and *Pareto* β across clonality rates at generation 6,000. Scatter plots summarised by cross-shaped markers (+) represent the relationship of clone age distribution predictability (R^2), clonal richness (*R*), and *Pareto* β for each clonality rate at generation 6,000. The horizontal and vertical arms of each cross indicate ±1 standard deviation around the mean for the corresponding metric, across 100 simulated replicates.

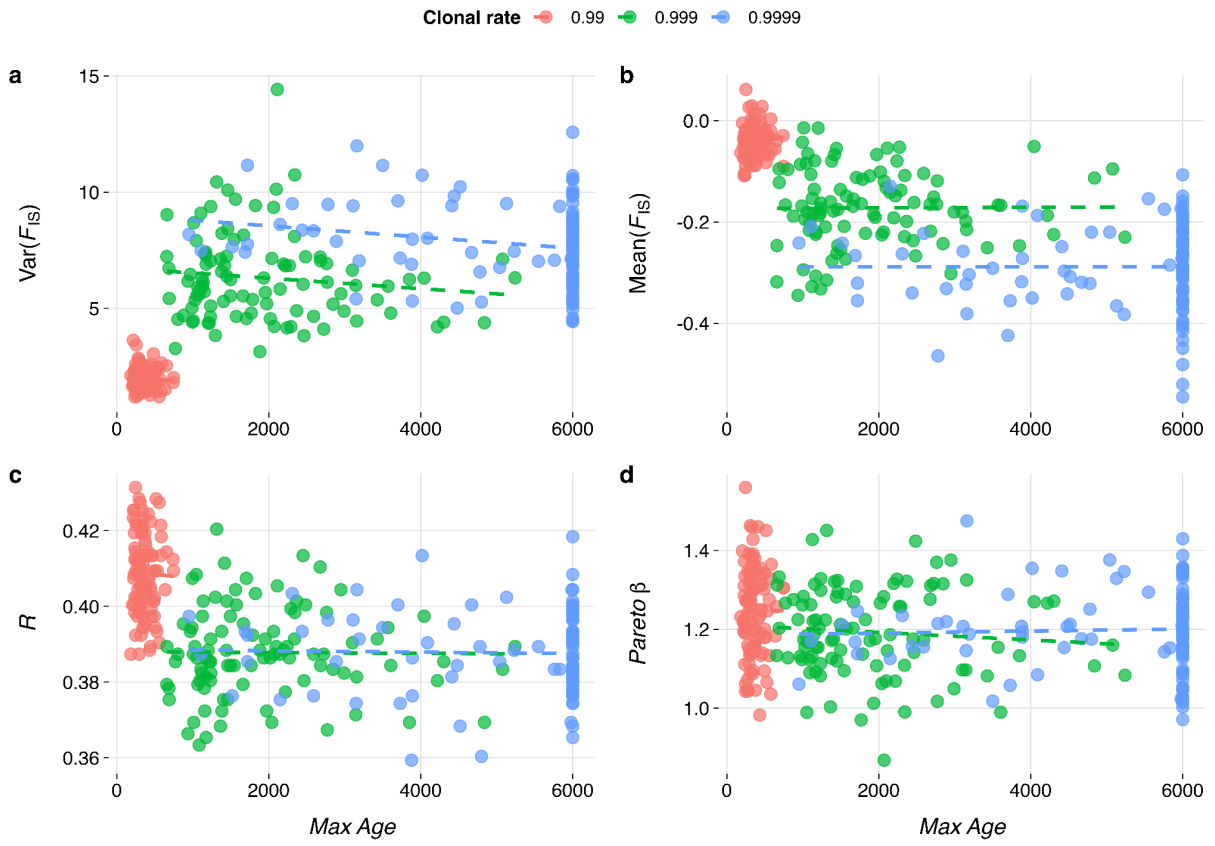


Fig. S6 Relationship between the maximum of clone age distribution and the mean and variance of the inbreeding coefficient (F_{IS}) across specific clonal rates at generation 6,000. Scatterplots of the maximum of clone age distribution against 4 indices (y-axes) computed for each replicate at specific clonal rates: (a) variance in F_{IS}, (b) mean F_{IS}, (c) clonal richness (R), and (d) *Pareto* β. Each point represents one of 100 replicate simulations at a given clonal rate. Dashed lines show the best-fit linear regression for each clonal rate distribution.

Chapter two

Can Asexual *Melampsora larici-populina* Isolates Undergo Sexual Reproduction? Experimental Assessment of Life Cycle Completion and Nuclear Fusion

This chapter addresses the second objective of my PhD work: Phenotypic assays of sexual capacity in asexual lineages **(Objective 5.2)**. Building on the population genetics analysis presented in Chapter One that identified many distinct asexual lineages within poplar rust populations using microsatellite markers, we selected representative asexual isolates from the most prevalent lineages to examine their capacity for sexual reproduction through controlled greenhouse and laboratory experiments.

The experimental framework encompassed two complementary approaches:

1. Microscopic examination of nuclear fusion dynamics to compare asexual and sexual isolates, and
2. Field-based life cycle completion assays to test the ability of asexual isolates to undergo sexual reproduction under semi-natural environmental conditions

Through these phenotypic assays, this study provides experimental evidence for the sexual reproductive capabilities of these asexual lineages, thereby bridging the population genetic inferences from Chapter One with the biological realities addressed in this chapter. This transition from population level to direct phenotypic observation provides a more integrated understanding of the reproductive biology of poplar rust fungi and contributes to the broader knowledge of how asexual lineages either lose or maintain sexual capacity.

The results presented here will serve as the foundation for a manuscript intended for submission to Phytopathology or potentially integrated into the Axis Three publication.

1 Introduction

Rust fungi display diverse reproductive strategies, making them a model for studying the evolution of sexual reproduction. Across the Pucciniales, the diversity of life cycles range from fully sexual, heteroecious species that alternate between two different host species, to strictly clonal species that reproduce only through asexual urediniospores (Duplessis et al., 2021a). Many species exhibit both modes within the same taxonomic group, and in some cases, even within the same population. This coexistence of sexual and asexual lineages (Abdalrahem, Andrieux, et al., 2025), coupled with the ability of some “asexual” isolates to retain hidden sexual capacity, provides a unique chance to explore the ecological, genetic, and evolutionary factors shaping the maintenance, loss, or facultative expression of sex. As plant pathogens with complex life cycles, rust fungi offer an amazing system to study how reproductive mode diversity contributes pathogens to adapt, survive, and influence population structure.

1.1 Sexual and asexual life cycle of poplar rust

The infection cycle of *M. larici-populina* begins when aeciospores produced on larch trees are dispersed by wind and land on susceptible poplar leaves (*Populus* spp.) in late spring to early summer (Fig. 1A) (Lorrain et al., 2018). These dikaryotic aeciospores (n+n) display a bright yellow pigmentation due to accumulated carotenoid compounds (Fig. 1B1) (Zhan et al., 2023). Upon arrival on the poplar leaf surface, aeciospores penetrate through stomatal openings and establish initial infections on the abaxial (lower) surface, marking the onset of the asexual phase (Laurans & Pilate, 1999). Within 7-10 days, uredinia develop and erupt through the lower epidermis as yellow pustules, releasing large quantities of urediniospores (n+n) (Fig. 1B2). These urediniospores also exhibit a yellow-coloured appearance as a result of carotenoid pigmentation (Zhan et al., 2023).

In autumn, the same mycelium that previously produced uredinia begins forming early telia (n+n) under the epidermis of the adaxial surfaces of poplar leaves. These structures appear as light brown to dark brown pustules containing developing teliospores (Fig. 1B3). The subepidermal position of teliospores is a distinctive character of the genus *Melampsora* within

the Pucciniales. Individual teliospores are thick-walled and melanized (Hacquard et al., 2013). During the overwintering period, the sexual phase is initiated: karyogamy occurs within mature teliospores, resulting in transient diploid nuclei (2n). Meiosis begins immediately in autumn and is blocked in prophase I at the diplonema stage when teliospores enter dormancy (Hacquard et al., 2013). In spring, meiosis resumes, generating genetic recombinants while restoring the haploid state. Subsequently, teliospores germinate to produce basidia bearing four basidiospores, light yellow, haploid spores (Fig. 1B4), which are readily discharged into the air and must quickly reach their alternate host, larch (*Larix* spp.) (Lorrain et al., 2018).

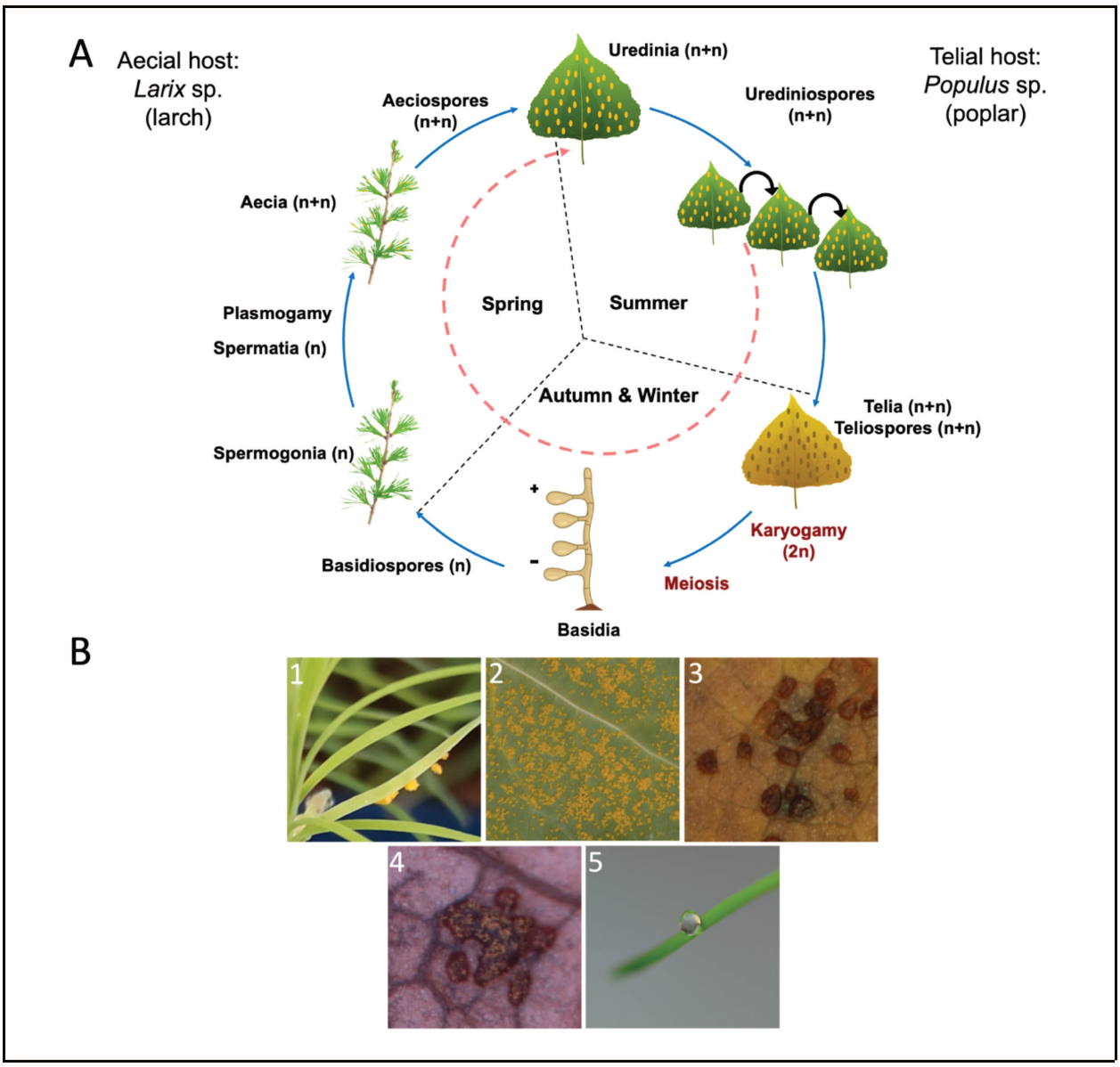


Fig. 1 Life cycle and spore morphology of *Melampsora larici-populina*. (A) Life cycle of *Melampsora larici-populina*. Solid blue arrows depict the heteroecious, macrocyclic sexual cycle, which alternates between an aecial host (larch, *Larix* spp.) and a telial host (poplar, *Populus* spp.), involving all five spore stages. The red dashed arrow represents the asexual cycle, in which the fungus undergoes clonal multiplication on poplar through urediniospores, thereby bypassing larch and sexual recombination. Seasons are shown centrally, with ploidy state indicated in parentheses (n = haploid, n+n = dikaryotic, 2n = diploid). (B) Spore stages: (1) aeciospores: light yellow spores on the abaxial surface of larch needles; (2) urediniospores: yellow spores on the abaxial surface of poplar leaves; (3) teliospores: dark brown, thick-walled spores on senescent poplar leaves; (4) basidiospores: small, yellow spores produced on basidia that germinate from teliospores on senescent leaves; (5) pycniospores: colorless spores released in nectar droplets on the adaxial surface of larch needles.

In April-May, basidiospores infect emerging larch needles, continuing the sexual phase through a mating process. Within 7-10 days, infected needles develop pycnia (also termed spermogonia) on their adaxial surfaces, which produce colourless, sweet nectar droplets containing pycniospores (spermatia) (Pernaci et al., 2014; Lorrain et al., 2018). These sugary droplets attract insects (Fig. 1B5), which facilitate the transfer of pycniospores between pycnia and enable cross-fertilization between opposite mating types (Lorrain et al., 2019). Successful fertilization leads to the development of aecia on the abaxial needle surfaces, which release aeciospores (n+n) that will infect poplar leaves, thereby completing the annual reproductive cycle.

The complete life cycle of *M. larici-populina* takes one year, from the germination of aeciospores to the production of a new generation of aeciospores. The sexual phase, which occurs on larch during spring, promotes genetic recombination and adaptation, whereas the clonal multiplication phase on poplar during summer allows rapid population expansion and disease spread. Overall, this one-year life cycle alternates between poplar and larch hosts; it is considered sexual because genetic recombination occurs during the larch phase.

1.2 Discovery of asexual lineages

Although *M. larici-populina* was traditionally regarded as an obligately sexual species requiring completion of the full heteroecious cycle, recent population genetics studies have revealed the presence of many different asexual lineages within natural populations in France (Xhaard et al., 2011; Abdalrahem, Andrieux, et al., 2025). These lineages appear to skip the sexual phase, persisting instead in an asexual form (probably urediniospores), without the need for the alternate host (Fig. 1A). These asexual lineages raise fundamental questions about their ability to invest in sexual reproduction when provided with suitable environmental conditions and compatible mating partners.

1.3 Cryptic sexuality: the need for experimental validation

Many studies have challenged traditional assumptions about asexuality in fungi, revealing that species thought to reproduce only asexually may retain cryptic sexual mechanisms that can be activated under specific conditions. For example, the wheat stripe rust pathogen *(Puccinia striiformis* f. sp. *tritici*), long regarded as a classic example of epidemic asexuality, was shown experimentally to complete its sexual cycle on barberry (Ali, Gladieux, Rahman, et al., 2014; Du et al., 2023).

A recent study on lichenized fungi revealed that species such as *Lepraria*, long thought to reproduce exclusively asexually, still retain complete mating-type loci and the full set of genes required for meiosis and sexual reproduction (Doellman et al., 2024b). If the genetic machinery for sex is intact, why do these organisms consistently reproduce asexually in nature? This observation is highly relevant to our work on *M. larici-populina*: do the asexual lineages identified through population genetic analysis still possess functional sexual reproduction machinery? These questions highlight the importance of direct experimental validation.

In this chapter, we examine nuclear fusion dynamics during early and late telial stages to better understand the cellular mechanisms underlying sexual competence in these lineages (Hacquard et al., 2013). In parallel, we investigate whether these asexual lineages are capable of completing the full heteroecious life cycle under controlled experimental conditions, specifically

testing their ability to progress from telial development through alternate host infection to aeciospore production.

2. Materials and Methods

2.1 Microscopic analysis of teliospore fusion dynamics

2.1.1 Inoculation protocol

Detached poplar leaves (*Populus deltoides* × *P. nigra* 'Robusta') were inoculated with urediniospores from asexual and sexual isolates of *M. larici-populina* under controlled laboratory conditions. Following inoculation, leaves were incubated at 19°C for 10 days for infection establishment and urediniospore development. Subsequently, infected leaves were inverted (abaxial surface facing upward) and transferred to 10°C to induce telial stage development. Leaf samples with developing telia were collected at specific time points to capture the progression of nuclear fusion events: 18, 25, 32, 39, and 46 days post-inoculation (dpi). These sampling intervals were selected to match key developmental points for teliospore formation in *M. larici-populina*, as described in established protocols (Hacquard et al., 2013).

2.1.2 Sample fixation procedure

Four leaf sections (~2 × 0.5 cm) containing telia were collected at each time point from each isolate, immediately rinsed with sterile 1X PBS buffer, and directly subjected to chemical fixation to arrest biological activity for downstream microscopic examination. Samples were fixed in a 3:1 ratio of 3% paraformaldehyde and 1X PBS buffer. After overnight incubation at 4 °C (~24 h), leaf sections were washed three times with 1X PBS buffer, with gentle vortexing to wash away the remaining. Samples were then transferred to a 1:1 mixture of PBS buffer and 96% ethanol and stored at –20 °C until further processing.

2.1.3 Tissue embedding and sectioning

A 4% agarose solution was prepared and maintained at 60 °C in a heated water bath suitable for microcentrifuge tubes. Fixed leaf pieces were vertically positioned and embedded in molten agarose, which was allowed to solidify at room temperature before storage at 4 °C. Embedded

samples were sectioned using a vibratome (VT1000S, Leica) with automatic feed at a thickness of 10-15 µm in deionized water to prevent tissue drying and to facilitate collection with a fine paintbrush.

2.1.4 Staining and microscopic analysis

To improve the adhesion of sections during DAPI staining and washing steps, slides were gelatinized before use. Slides were immersed for a few seconds in a 0.075% (w/v) gelatine and 0.01% (w/v) chromium(III) potassium sulfate dodecahydrate solution, maintained at 70 °C in a water bath. Slides were then dried under the airflow of the laminar flow for one hour before use (Hacquard et al., 2013).

Approximately 30 sections per leaf sample were transferred onto gelatinized slides and then stained with 4′,6-diamidino-2-phenylindole (DAPI) to visualize fungal nuclei inside teliospores. A volume of 30 µL of DAPI solution (300 nM in PBS) was added to each section, and samples were incubated for 2–3 hours at room temperature in the dark. The standard DAPI protocol was modified due to initial difficulties with nuclear staining penetration in fungal-infected leaf sections to improve dye accessibility through complex tissue barriers. The working solution was prepared by diluting the DAPI stock 1:100 in 1X PBS buffer and supplementing it with 1% (v/v) dimethyl sulfoxide (DMSO) and 0.1% (v/v) Triton X-100, which enhances dye penetration into fungal nuclei within plant tissue. Following incubation, slides were gently washed three times with 1X PBS buffer to remove excess dye. Fluorescence microscopy was performed using an Olympus BX51 microscope (Olympus Corporation, Japan) equipped with a DAPI filter set (excitation 358 nm, emission 461 nm). Nuclear fusion status was evaluated by observing teliospore morphology and scoring fused versus unfused nuclei at each time point, enabling quantitative assessment of karyogamy progression in both sexual and asexual isolates.

2.2 Heteroecious cycle completion by asexual lineages

2.2.1 Experimental overview

This experiment aims to determine whether asexual lineages of *Melampsora larici-populina* are capable of completing the sexual phase under controlled semi-natural conditions. Sexual

lineages served as a control. The heteroecious life cycle was supported using two hosts: *Populus* spp. as the telial host (asexual phase) and *Larix decidua* as the aecial host (sexual phase). This experimental protocol was based on a previously established method described in Pernaci et al. (2014), with modifications to assess the life cycle completion in asexual lineages.

2.2.2 Poplar host preparation

Poplar plants (*Populus deltoides* × *P. nigra* 'Robusta') were propagated from dormant cuttings in 5-liter containers filled with a 1:1 (v/v) mixture of sand and peat. The substrate was supplemented with calcium carbonate (3.5 g L^{-1}) and a slow-release NPK fertilizer (6.0 g L^{-1}; 13:13:13 formulation) for plant development. Plants were cultivated for 4 months (June - September) in the greenhouse, with natural photoperiod and daily irrigation. All plant cultivation and preparation were conducted at the INRAE facility in Nancy, France.

2.2.3 Poplar inoculation (asexual cycle)

Urediniospore suspensions were prepared at a concentration of 40,000 spores mL^{-1} through mixing 2 mg of spores with 20 mL of water agar (0.1 $g.L^{-1}$). Before inoculation, spore viability was assessed through germination assays on water agar medium (2% agar) in Petri dishes, with germination tube emergence examined under a stereomicroscope after 12–24 hours of incubation at 19°C. Then we selected four to five poplar plants per isolate and inoculated them by spraying the urediniospore suspension on the abaxial surfaces of leaves using a handheld sprayer (Fig. 2A). Then, plants were covered by plastic bags to maintain high humidity and facilitate infection establishment (Fig. 2B). After 24 hours, the plastic bags were replaced by cotton bags (porosity <20 µm) to prevent cross-contamination between isolates and allow sufficient ventilation (Fig. 2C). Plants were kept in unheated greenhouse conditions throughout the autumn period (September-December) to support natural telial development.

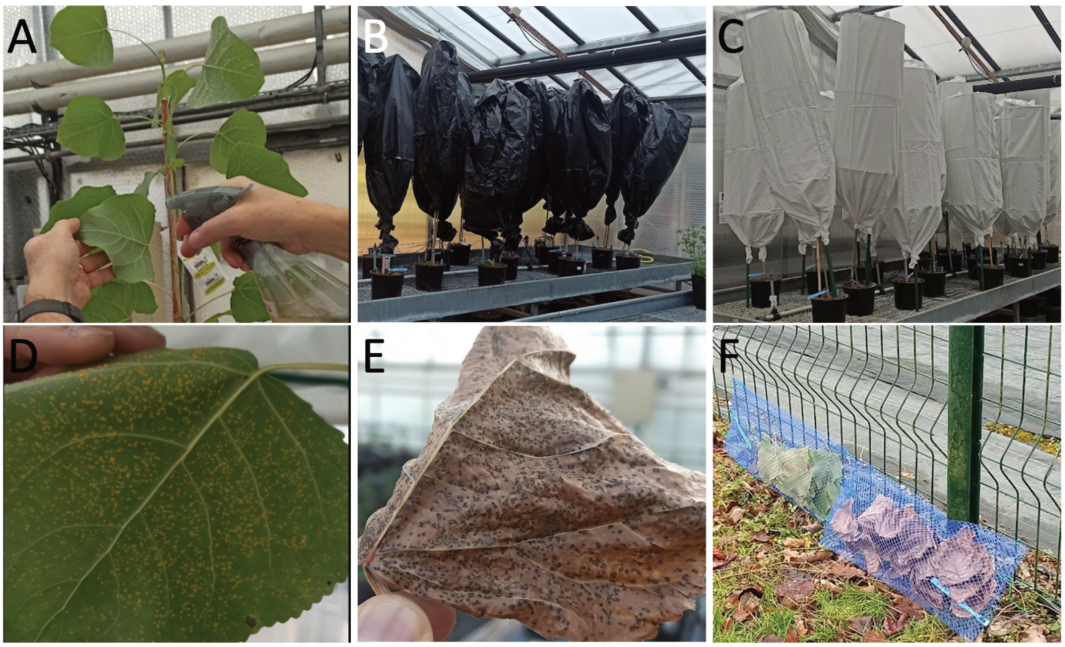


Fig. 2 Experimental protocol for telia production and collection. (A) Spray inoculation of poplar leaves with a urediniospore suspension; (B) Plants are covered with black plastic bags for 24 hours to maintain high humidity; (C) Plants are covered with cotton cloth to prevent cross-contamination; (D) Formation of uredinia on the abaxial surface of poplar leaves; (E) Mature telia developing on senescent poplar leaves; (F) Collection and outdoor placement of telia-bearing leaves in mesh bags.

2.2.4 Telia collection and dormancy breaking

Senescent brown poplar leaves were collected along with green and yellow leaves bearing telia (Fig. 2E). The leaves were placed in plastic mesh containers (1 cm mesh size) and positioned on grassland to undergo natural overwintering treatment. During the winter months (December–February), the leaves were exposed to natural freeze-thaw and wetting-drying cycles to break teliospore dormancy (Fig. 2F). After this treatment, leaves were collected in February and stored at –20 °C until needed for subsequent inoculation of the larch host.

2.2.5 Basidiospore production (sexual cycle)

Poplar leaves with mature telia were soaked in deionized water for 24 hours (Fig. 3A), then placed adaxial side up on moist tissue paper in large plastic dishes and incubated at 19 °C (Fig. 3B). Basidiospore production was monitored during 2-3 days using a stereomicroscope (Fig. 3C).

Leaves exhibiting abundant basidiospore production were assessed for the proportion of the leaf surface producing basidia, and those with sufficient coverage were selected for subsequent inoculation of larch.

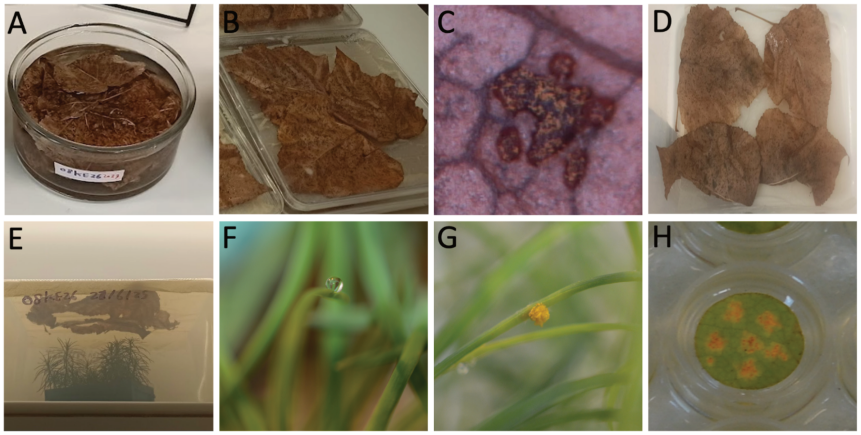


Fig. 3 Experimental protocol for completion of the sexual stage on Larch and subsequent inoculation on Poplar. (A) Senescent poplar leaves soaked in deionized water to induce basidiospore production; (B) Leaves placed adaxial side up on moist tissue paper to facilitate basidiospore formation; (C) Basidiospore production monitored using a stereomicroscope; (D) Leaves bearing basidiospore positioned abaxial side down in semi-transparent boxes, allowing the adaxial surface to face upward for larch infection; (E) Larch seedlings exposed to basidiospore bearing leaves inside boxes (F) Pycnia formation observed on larch needles; (G) Aecia development on larch needles; (H) Poplar leaf discs producing uredinia upon inoculation with aeciospores.

2.2.6 Larch inoculation

Poplar leaves producing basidiospores were positioned 1–5 cm above larch seedlings inside a transparent plastic box for 1-3 days to facilitate basidiospore deposition and infection of the larch needles (Figs. 3D, 3E). After exposure, the poplar leaves were removed, and the larch seedlings were maintained in growth chambers at 19 °C with a 16:8 hour light: dark photoperiod for 4-5 weeks. Larch seedlings were examined every three days for the appearance of pycnia (Fig. 3F), followed by the development of aecia (Fig. 3G). Natural spermatization of pycnia was facilitated either through direct needle contact or by the activity of small dipteran flies (Sciaridae) naturally present around the larch seedlings. As noted here, each *M. larici-populina* isolate was inoculated individually (one per larch box) to ensure selfing. However, because the ability of asexual lineages to self-fertilize was uncertain, we also conducted mixed inoculations.

These involved combining basidiospore sources from both sexual and asexual isolates and applying them to the same larch seedling to assess the potential for outcrossing.

2.2.7 Assessing life cycle completion

Individual aecia were checked and harvested from larch needles every 2-3 days upon maturation, with aeciospores from each single aecium considered as an individual. The aeciospores were used to inoculate 12 mm diameter poplar leaf discs (Fig. 3H). After 8-12 days of incubation at 19 °C, sporulating leaf discs were used to inoculate four additional poplar discs to generate biological replicates. These replicates were assessed through microsatellite genotyping, which was performed following Xhaard et al. (2011) to verify the genetic purity of the offspring, thereby confirming the successful completion of the heteroecious life cycle without contamination. In the case of mixed inoculations, these analyses also served to detect outcrossed progeny resulting from sexual reproduction between asexual and sexual isolates.

3 Results

3.1 Nuclear fusion dynamics

Nuclear fusion dynamics in teliospores of both sexual and asexual isolates were assessed through microscopic analysis of DAPI-stained sections. In most sections, nuclear staining was obscured by either surrounding plant tissues or the teliospore wall, resulting in few or no discernible nuclei. Nevertheless, a few teliospores with distinct nuclei were observed, enabling a limited evaluation of karyogamy status.

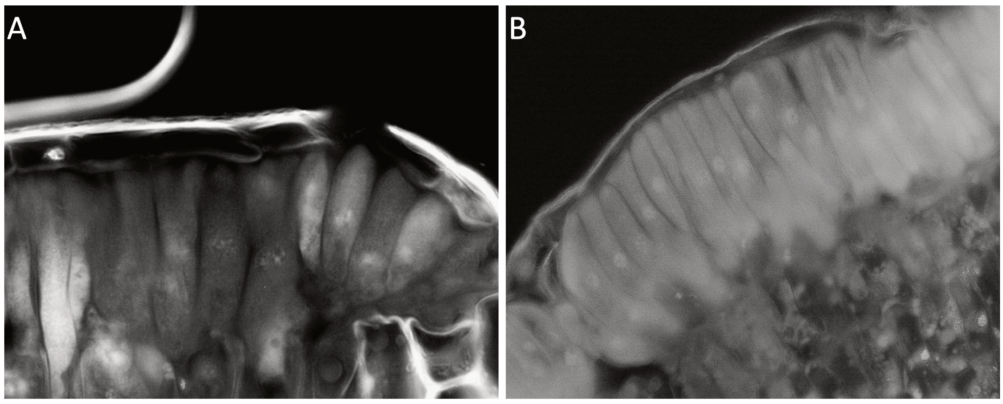


Fig. 4 Microscopic micrographs of teliospores of asexual isolates of *M. larici-populina* at different developmental stages. A) 08KE26 asexual isolate at 18 days post-inoculation (dpi), 40× objective. (B) 09BS78 asexual isolate at 45 dpi, 40× objective. Scale bars = 50 µm.

In the 08KE26 asexual isolate sampled at 18 dpi, some teliospores contained two nuclei positioned close to each other, and with one example appearing as a single, merged DAPI signal suggestive of nuclear fusion (Fig. 4A). In the 09BS78 asexual isolate sampled at 45 dpi, most teliospores were observed with a single, uniform DAPI-stained body, consistent with fused nuclei (Fig. 4B). These observations indicate that asexual isolates can form teliospores and may exhibit nuclear fusion. However, the number of teliospores with clearly visible nuclei was very low.

3.2 Life cycle completion assessment

3.2.1 Isolate selection and poplar inoculation

The experiment results described here were conducted in 2024 to assess life cycle completion in *M. larici-populina* isolates. Six isolates, representing both reproductive modes, were used in this experiment: four isolates from asexual lineages (08KE26, 08KE38, 09BS78, 96RAM7) and two isolates from sexual lineages as controls (98AG31, 93GS3). Twenty-four poplar plants were inoculated, with four replicate plants per isolate. Viability testing indicated high germination rates across all isolates, ranging from 87% to 100% (Table 1). The sexual isolates 98AG31 and 93GS3 exhibited the highest viability (100% and 97%, respectively), whereas asexual isolates showed comparable germination rates (87–98%).

Table 1. Spore viability and inoculation parameters

Isolate	Reproductive mode	Germination rate (%)	Leaves inoculated	Average per plant
08KE26	Asexual	90	83	20.8
08KE38	Asexual	87	83	20.8
09BS78	Asexual	89	97	24.3
96RAM7	Asexual	98	87	21.8
98AG31	Sexual	100	80	20.0
93GS3	Sexual	97	95	23.8

Average inoculation success resulted in 22.5 leaves per plant across all treatments, providing sufficient material for subsequent life cycle assessment.

3.2.2 Telia formation and basidiospore production

During autumn, telia formation was observed across all isolates, although with notable variation in abundance. Isolate 08KE36 exhibited reduced telial development compared to other isolates, with many senescent leaves showing few telia. To break teliospore dormancy, telia-bearing leaves were placed in plastic mesh nets hung on a fence near the grassland. Following these dormancy-breaking treatments, basidiospore production was assessed in both sexual and asexual isolates over a three-day observation period. Stereomicroscope examination revealed variation in basidiospore formation capacity between isolates. Basidiospore production was evaluated using a three-point visual scoring system based on the intensity of basidiospore formation on productive leaves: 1 = sparse, 2 = moderate, and 3 = abundant (Table 2).

Table 2. Basidiospore production assessment across experimental isolates

Isolate	Reproductive mode	Production intensity score	Larch inoculation
08KE26	Asexual	1	Limited success
08KE38	Asexual	0	Not attempted
09BS78	Asexual	2	Limited success
96RAM7	Asexual	3	Successful
98AG31	Sexual	1	Not attempted
93GS3	Sexual	3	Successful

Production intensity: 1 = sparse, 2 = moderate, 3 = abundant basidiospore formation

3.2.3 Sexual stage development on larch and secondary inoculation

Many differences were observed in larch infection success among selfed isolates (Table 2). Selfing capacity varied: the asexual isolate 96RAM7 produced many pycnia followed by well-developed aecia, demonstrating strong self-fertility and high sexual competence. In contrast, the remaining asexual selfed isolates (08KE26, 09BS78, and 93GS3) produced only a few pycnia and no detectable aecial development (Fig. 5).

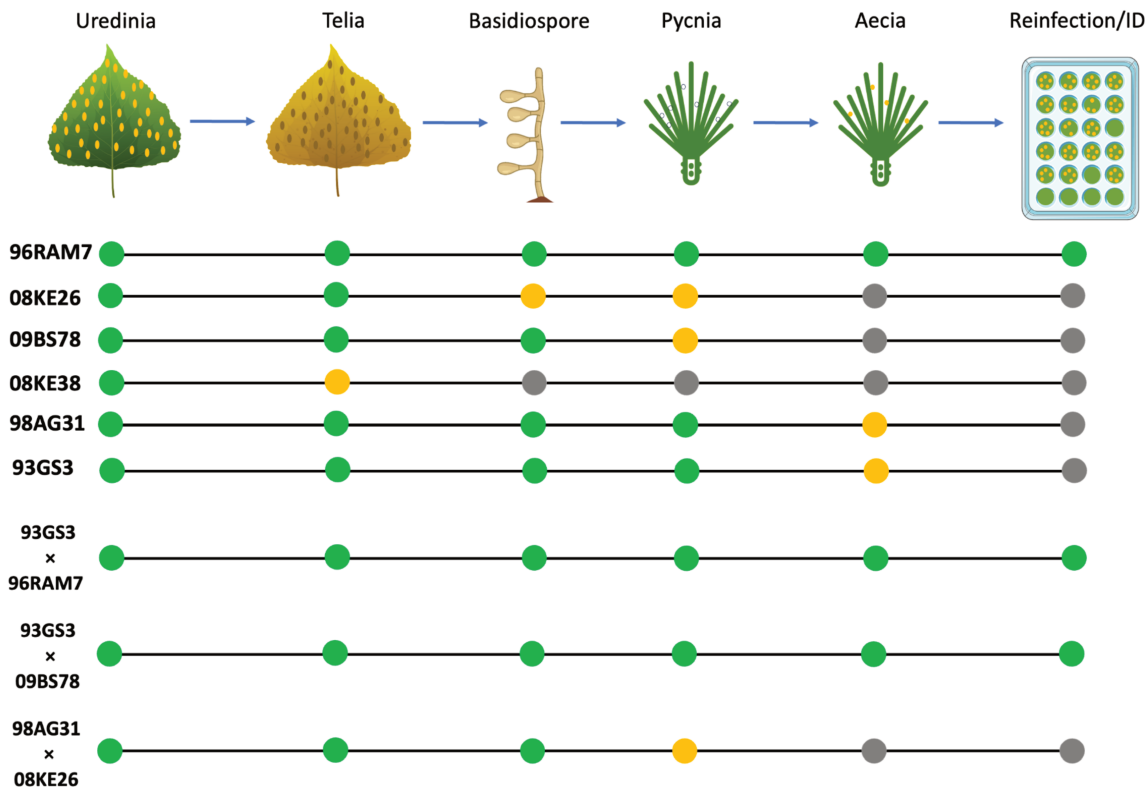


Fig. 5 Summary of results from sexual-cycle completion assays. Summary of experimental results evaluating the ability of asexual isolates to complete each stage of the *M. larici-populina* heteroecious life cycle. The first row depicts the poplar rust developmental stages evaluated (uredinia on poplar leaves → telia on poplar leaves → basidiospores on senescent poplar leaves → pycnia on larch → aecia on larch → reinfection of poplar with identity confirmation to exclude contamination). Colored circles summarize the outcome of each stage for individual isolates (self-fertilization) and pairwise outcrosses between sexual and asexual lineages. Green indicates successful development, yellow indicates reduced or limited development, and gray indicates failure.

Among the outcrosses, combinations involving the sexual isolate 93GS3 were the most successful. Both 93GS3 × 96RAM7 and 93GS3 × 09BS78 formed abundant pycnia, which subsequently developed into many aecial pustules, indicating full completion of the sexual cycle. Conversely, the outcross 98AG31 × 08KE26 produced only a few pycnia and no aecia, suggesting a developmental arrest at the early stages of infection (Fig. 5). Overall, these results show that sexual reproductive success on larch is highly variable among isolates and crosses; outcrosses involving 93GS3 and the selfing of 96RAM7 were particularly successful.

3.2.4 Confirmation of sexual capacity in asexual lineages

In order to verify the genotype of each offspring obtained from selfing or outcrossing experiments, urediniospores that produced secondary infections on poplar in each cross were genotyped with 26 microsatellite markers (Table 3). Offspring from selfed isolates, such as 96RAM7 and 09BS78, displayed recombined genotypes deriving from their parental isolates, confirming successful self-fertilization. In contrast, hybrids resulting from outcrosses, such as 96RAM7 × 93GS3 and 09BS78 × 93GS3, exhibited combined alleles from both parents at each locus, providing evidence of sexual recombination. For asexual isolates 09BS78 and 08KE26, and the sexual isolate 93GS3, which produced only a few pycnia and no aecia under selfing conditions, pycnial tissue was genotyped directly from larch needles. There were no data obtained from their microsatellite markers.

Table 3. Representative microsatellite loci genotypes for parental isolates and progeny

	Isolate	Mlp_100	Mlp_49		Mlp_54		Mlp_73		Mlp_77		Mlp_82b		Mlp_95	
Parents	93GS3	208208	356	365	138	138	431	431	434	434	178	193	456	456
	08KE26	196211	347	356	126	132	435	446	432	432	196	196	453	453
	09BS78	196232	347	356	132	135	438	446	432	434	178	196	453	453
	96RAMR7	196205	347	356	132	135	435	446	432	432	175	196	453	453
Selfed progeny	96RAMR7	205205	347	356	132	135	435	446	432	432	196	196	453	453
	96RAMR7	196205	347	356	135	135	435	446	432	432	175	175	453	453
	09BS78	196232	356	356	135	135	438	446	432	434	178	196	453	453
	09BS78	196232	347	356	135	135	438	446	432	432	178	196	453	453
Outcrossed progeny	96RAM7 X 93GS3	205208	347	365	132	138	431	446	432	434	175	178	453	456
	96RAM7 X 93GS3	205208	356	356	132	138	431	435	432	434	175	193	453	456
	09BS78 X 93GS3	196208	347	365	135	138	431	446	434	434	178	196	453	456
	09BS78 X 93GS3	208232	347	356	132	138	431	438	434	434	178	196	456	456

Only the 7 most informative polymorphic microsatellite loci (out of 26) are displayed; loci with invariant or homozygous alleles across all parent and progeny isolates are omitted for clarity.

These results demonstrate that asexual isolates are quite variable; some are arrested at specific developmental stages, while others retain sexual capacity and can produce genetically diverse progeny through selfing or outcrossing.

4 Discussion

The main result of this study demonstrates that even *M. larici-populina* isolates considered asexual in natural populations, based on population genetic analyses (see Abdalrahem et al., 2025), retain the ability to undergo nuclear fusion and form pycnia and aecia on larch, completing the heteroecious macrocyclic sexual life cycle under controlled conditions. This capacity was observed both in selfing and outcrossing experiments, with microscopic examination indicating that asexuality in *M. larici-populina* is not due to irreversible loss of sexual function, but likely reflects ecological or evolutionary constraints in nature.

Results from selfing experiments showed clear differences in reproductive success among isolates. Isolate 96RAM7 was particularly notable, producing abundant basidiospores, pycnia, and well-developed aecia, completing the sexual cycle. In contrast, other isolates, such as 08KE26 and 09BS78, formed pycnia but showed no evidence of aecia. These differences suggest variation in self-compatibility, possibly linked to mating-type composition (Bakkeren & Kronstad, 1994; Luo et al., 2024), or potentially influenced by the high presence of flies observed in the 96RAM7 larch seedling box, which may have helped pycnial transfer and fertilization and contributed to its selfing process (Bolton et al., 2008). The absence of similar insect activity in other treatments could partially explain their lower success rates.

Crossing experiments revealed additional variation in compatibility. Some outcrosses successfully produced aecia, such as 93GS3 × 96RAM7 and 93GS3 × 09BS78, which formed abundant pycnia, while others stalled after producing only a few pycnia, such as 98AG31 × 08KE26. For example, 08KE26 showed low performance in both selfing and outcrossing, stopping at the pycnial stage with only a few pycnia produced, and 08KE38 produced very few teliospores and almost no basidia. This indicates that asexual isolates vary: some can undergo sexual reproduction with good performance, others have low performance, and others are unable to continue beyond a specific stage. These results support the idea that latent sexual potential persists in many isolates, but its expression may depend on partner compatibility, or

possibly on micro-environmental conditions during infection (Bakkeren & Kronstad, 1994; Freytag et al., 1988; Mojerlou, Schwessinger, et al., 2025).

From an ecological view, the limited expression of sexual reproduction of these isolates in natural populations may be driven by regional climate. Most asexual isolates are prevalent in southern France (Abdalrahem, Andrieux, et al., 2025), where milder Mediterranean winters could facilitate continuous asexual cycling (“green bridge” effect), meaning that poplar trees can retain green leaves throughout winter. This was directly observed in February in Cavalaire-sur-Mer (southern France), where this bridge could allow poplar rust fungi to maintain asexual reproduction through the year, thereby avoiding the need to complete the costly sexual cycle, such as finding a compatible mating partner. In contrast, continental regions with harsh winters may necessitate sexual reproduction for overwintering survival, maintaining this evolutionary trajectory in sexual lineages.

To measure fused nuclei dynamics in teliospores of both isolates, it's necessary to determine their number during microscopic examination. However, because of poor visualization and the few number of nuclei that could be seen, it was not possible to carry out this quantification. Even so, the observations we did make are still valuable and link directly to the life cycle completion assay described earlier. In that experiment, some asexual isolates were shown to form teliospores, produce pycniospores, and, a few, complete the entire life cycle. This microscopy work, although limited, adds another piece to that picture by demonstrating that nuclear fusion also occurs.

For example, in the 08KE26 isolate at 18 dpi, most teliospores contained two nuclei positioned close together, and one teliospore contained one nucleus that looked fused. In the 09BS78 isolate at 45 dpi, most teliospores showed a single nucleus, which was consistent with karyogamy. These findings agree in part with a previous study on poplar rust (Hacquard et al., 2013), where nuclear fusion in teliospores was more clearly observed, likely due to differences in sectioning and imaging methods.

Overall, the results from the field infection experiments, combined with microscopic examination, provide complementary evidence for the ability of asexual isolates to undergo sexual reproduction. The controlled inoculation trials on larch demonstrated that many asexual isolates retain the full developmental potential to progress through pycnial and aecial stages, while the microscopic analysis confirmed the occurrence of nuclear fusion. This dual approach shows that the absence of sexual reproduction in nature is not due to the loss of sexual capacity, but rather to ecological or environmental factors that offer an asexual pathway in nature. These findings are consistent with our hypothesis that asexual isolates of *M. larici-populina* retain the ability to invest in sexual reproduction when conditions are favorable.

5 Challenges and future directions

5.1 Nuclear fusion observation in teliospores

This experiment faced several challenges that prevented me from achieving the main goal, which was to quantify the number of fused nuclei in teliospores of both sexual and asexual isolates. The first major difficulty was section thickness. Cutting very thin and even leaf sections proved challenging. I began at 15 µm, but these sections were too thick to allow clear visualization of nuclei. When I reduced the thickness to 10 µm, the vibratome I used (Leica VT1000S, Leica Microsystems, Germany) was not precise enough to produce consistent cuts at that setting. In the study by Haquard et al. (2013), a different sectioning device and a completely different fixation protocol were used.

Another limitation was staining quality. I tested DAPI at both 1:100 dilution in PBS and 1:10 dilution, but in both cases, it was still difficult to visualize the nuclei. I then modified the staining protocol by using DAPI at 1:100 together with DMSO and Triton X to improve penetration through the teliospore wall. This adjustment produced a slight improvement, but it was not sufficient for reliable quantitative analysis. The thick, melanised teliospore wall continued to act as a strong barrier to the stain, making nuclear structures difficult to observe. I also tried to preserve the fluorescence signal by using an antifade mounting medium (SlowFade™ Gold Antifade Mountant, Thermo Fisher Scientific, USA) and sealing the coverslips with nail polish. While this helped slightly, the improvement was small, and the main issues of section thickness and uneven staining remained.

Looking ahead, the main improvement would be to use a sectioning system that can produce thinner and more uniform sections, as section thickness was the primary factor limiting image clarity in this study. In addition, using a more advanced imaging system could further enhance resolution and contrast, making it easier to visualize nuclei in both sexual and asexual isolates.

5.1 Assessment of life cycle completion

Before establishing the final protocol described in the Methods and Results, two earlier trials (2022 and 2023) were conducted. Each provided important lessons that shaped our successful experiment in 2024.

In the initial trial (2022), four isolates were tested, one putatively asexual (08KE26) and three putatively sexual (98AG31, 98AB07, 93GS3). While germination rates were high across all isolates, differences in canopy position at the time of sampling resulted in varying teliospore maturity, from fully developed telia in senescent lower canopy leaves to immature sporulation in the upper canopy. Although leaves from all leaves from canopy levels were collected and preserved. In the second trial (2023), the design was improved by including four confirmed asexual isolates (08KE26, 09BS12, 09BS78, 96RAMR1) alongside one sexual control (98AG31). However, the overwintering setup presented an unexpected challenge: soil worms consumed many of the leaves placed directly on the grass, causing significant sample loss. This damage limited the ability to assess basidiospore production. While some older stored materials from previous years showed moderate basidial activity, fresh 2023 samples from both sexual and asexual isolates showed little to no activity. A limited number of inoculations using 08KE26 material led to pycniospore production on larch needles, but these infections did not progress to aecial formation.

These early trials highlighted several practical limitations, including variability in teliospore maturity at collection, worm damage, and low rates of life cycle completion under the tested conditions. The final trial incorporated solutions to these issues, leading to the more consistent results described in the main experiment.

Chapter three

Genomic and Transcriptomic Signatures of Reproductive Mode Variation in the Poplar Rust Fungus

This chapter addresses the third objective of my PhD project: Genomic and transcriptomic study of reproductive mode variation (Objective 5.3). The study was developed alongside the population genetic analyses presented in Chapter One and the experiments assessing the commitment to sexual reproduction of asexual isolates described in Chapter Two.

From Chapter One, we identified several distinct asexual lineages. In Chapter Two, we further demonstrated that these lineages differ in their ability to undergo sexual reproduction. While some isolates can complete the sexual life cycle, others are stopped at certain developmental stages. These differences may reflect environmental pressures, but background genetic differences may also support them. At early stages of the thesis project, we did not yet know whether isolates from the asexual lineage may undergo telia differentiation up to sexual-related stages towards successful reproduction. One plausible explanation for the presence of recurrent asexual isolates in the field could have been the presence of a green bridge under favourable environmental conditions for urediniospores from one year to the next, or due to particular features of urediniospores to meet winter conditions in the asexual lineage. Also, obtaining material from sexual stages on larch was particularly complex with the reference sexual isolate 98AG31 (Lorrain et al., 2018), and it could have been a real challenge with isolates from the asexual lineage. Finally, our current knowledge of the telial stage at the molecular level remains rather limited, with only a single stage explored *in planta* (Hacquard et al., 2013) and from isolated teliospores (Tao et al., 2017, 2019; Zhan et al., 2023). Based on these different elements, we envision to explore deeper in this thesis the developmental stages associated with the uredinia to telia transition and telia differentiation.

To explore further reproductive mode variation in the poplar rust fungus *M. larici-populina*, we first conducted a comparative genomic analysis using representative isolates from both sexual and asexual lineages. Whole-genome sequencing was employed to identify structural variants and variation at the mating-type (MAT) loci that could be associated with reproductive modes

(Chapter Three - Section One). In parallel, comparative transcriptomic analysis across key developmental stages before the entry of *M. larici-populina* into dormancy enabled us to assess gene expression programs related to uredinia production, transition to telia differentiation, then entry into overwintering in selected isolates from both sexual and asexual lineages (Chapter Three - Section Two). Together, these complementary genomic and transcriptomic datasets provide insights into both specific loci (e.g., mating-type regions) and the broader regulatory networks that may facilitate or constrain further sexual development ahead of meiosis in the case of sexual reproduction or ahead of uredinia formation and telia differentiation.

For clarity, Section One focuses on the diversity of MAT genes. This part of the work will be later combined with a genomic effort generated concomitantly to the present doctoral project by Ms. Emma Corre, a PhD student of the UMR IAM, under the co-supervision of Dr. Cécile Lorrain at ETH Zurich and Dr. Sébastien Duplessis at INRAE. Emma Corre completed a close to chromosome-level phased telomere-to-telomere genome assembly of the poplar rust fungus (*M. larici-populina* reference genome version 3 of sexual isolate 98AG31: assembly and TE annotation completed in Spring 2025; gene annotation ongoing as of summer 2025 at the Joint Genome Institute). She also assisted in the assembly of a reference genome for the isolate 08KE26 from the asexual lineage, for which I produced spores and isolated high molecular weight genomic DNA. She reached a telomere-to-telomere phased genome assembly through a combination of PacBio HiFi long read sequencing and HiC chromosome conformation capture (Corre E., 2025). Together, some of the analyses conducted in both doctoral projects (*M. larici-populina* isolate 08KE26 genome assembly, MAT loci annotation and analysis of chromosome rearrangements, population genomics survey of MAT genes) will form the foundation of a collaborative manuscript with shared co-authorship. Since the gene annotation of the new version of the phased reference genome was still ongoing in Summer 2025, we did not establish a final draft for this common article. Thereafter are described my personal contributions to this collaborative work in direct relation to my own thesis project.

By integrating population-level diversity, reproductive phenotypes, and molecular mechanisms, this chapter provides a functional perspective on reproductive mode variation in the poplar rust fungus.

Section one

Mating-Type Locus Diversity and Genomic Variation in Sexual and Asexual Lineages

1 Introduction

The evolutionary maintenance of sexual reproduction remains one of biology's paradoxes, particularly given the notable cost of sex. Many comparative studies have contrasted sexual and asexual sister species or genera to study this transition in the evolution of sex reproduction (Bast et al., 2016; Doellman et al., 2024b). However, such designs are limited by divergence between species and confounding ecological differences. Fungi are one of the unique models for addressing these questions because many species exhibit facultative reproductive plasticity, switching between sexual and asexual modes within the same species or population (Abdalrahem, Andrieux, et al., 2025; Billiard et al., 2012). This dual capacity within a single species circumvents the phylogenetic constraints of traditional comparative approaches and enables direct tests of the costs and benefits underlying the evolution and maintenance of sex. The phylum Basidiomycota demonstrates diversity in reproductive strategies. These fungi, which include mushrooms, rusts, smuts, and yeasts, have evolved intricate molecular mechanisms to control sexual compatibility, ensuring outcrossing through highly specialized mating-type loci (Nieuwenhuis et al., 2013; Coelho et al., 2017).

1.1 Basidiomycota mating systems

Sexual reproduction in Basidiomycota is controlled by mating-type (MAT) loci that regulate nuclear pairing and dikaryon formation (Swiezynski & Day, 1960; Raudaskoski & Kothe, 2010). There are two main mating systems: bipolar architectures, in which a single fused MAT locus determines compatibility, and tetrapolar organizations, where two unlinked MAT loci independently govern compatibility (Coelho et al., 2010; Nieuwenhuis et al., 2013; Sun et al., 2019). Although numerous studies have investigated mating-type function and organization, population-level analyses of mating-type diversity in natural fungal communities remain an active field of research in the scientific community. This gap particularly concerns rust fungi, which are important plant pathogens, for which recent genomic research has uncovered substantial mating-type complexity within natural populations (Peris et al., 2022; Holden, Bakkeren, Hubensky, Bamrah, Abbasi, Qutob, de Graaf, et al., 2023; Luo et al., 2024; Henningsen et al., 2024). These findings underscore the critical need for expanded large-scale genomic

surveys to achieve a deeper understanding of mating-type diversity and its implications for fungal evolution and disease management.

In a bipolar system, sexual compatibility is determined by one mating-type locus harbouring both homeodomain and pheromone receptor genes physically linked within the same chromosomal region (Kües et al., 2011). This genetic linkage ensures that the two components of mating-type determination (transcriptional regulation through homeodomain proteins and cell-cell recognition through pheromone signaling) are inherited as a single unit. The result is that individuals can mate only with partners carrying the opposite mating-type allele (Bakkeren & Kronstad, 1994; Nieuwenhuis et al., 2013). In contrast, in a tetrapolar system, the genes controlling sexual compatibility are distributed across two unlinked loci, typically located on different chromosomes. The first locus, known as the pheromone/receptor (P/R) region, harbours genes encoding pheromone precursors and their corresponding G-protein-coupled receptors for pre-fusion recognition and compatibility between potential mating partners (Fowler et al., 1999; Slessareva & Dohlman, 2006). The second locus, the HD region, encodes homeodomain transcription factors that regulate post-fusion nuclear pairing and the initiation of dikaryotic development (Kües, 2000; Raudaskoski & Kothe, 2010; Kües et al., 2011).

The separation of these two mating functions across unlinked loci increases mating-type diversity. Where bipolar systems generate only two mating types, tetrapolar systems can theoretically produce p × h mating types, where p represents the number of variants at the P/R locus and h represents the number of allelic variants at the HD locus (Casselton & Olesnicky, 1998; Coelho et al., 2017). In natural populations, this can result in hundreds or thousands of distinct mating types within a single species, increasing the probability that any two randomly selected individuals will be sexually compatible.

1.2 The tetrapolar homeodomain and pheromone/pheromone receptors mating systems

Phylogenetic analyses across Basidiomycota support the hypothesis that the tetrapolar mating system represents the ancestral condition for this phylum. This conclusion is based on several lines of evidence, the consistent presence of both HD and P/R gene families throughout the phylum, and the apparent evolutionary transitions from tetrapolar to bipolar systems through chromosomal rearrangements and gene loss events (Coelho et al., 2010; Maia et al., 2015; Sun et al., 2019).

The homeodomain loci contain two genes (HD1 and HD2) encoding homeodomain transcription factors with complementary functions in sexual development (Kües & Casselton, 1992). These proteins contain characteristic helix-turn-helix DNA-binding domains that enable them to regulate transcription of genes essential for dikaryotic establishment (Gehring et al., 1994; Casselton & Olesnicky, 1998). Sequence comparisons between different HD alleles within a single species frequently reveal levels of divergence (Holden, Bakkeren, Hubensky, Bamrah, Abbasi, Qutob, de Graaf, et al., 2023; Gomez-Zapata et al., 2025). The pheromone precursors (Mfa) and pheromone receptors (STE3-like) exhibit complementary but distinct evolutionary patterns compared to HD regions. P/R loci contain genes encoding small peptide pheromones that serve as mating signals and G-protein-coupled receptors that detect these chemical cues and initiate intracellular signaling cascades leading to mating responses (Casselton & Olesnicky, 1998; Xue et al., 2008; Coelho et al., 2017). The specificity of pheromone-receptor interactions creates the molecular basis for mating-type discrimination, with compatibility requiring that individuals express non-self combinations of pheromones and receptors.

1.3 Poplar rust as a model system

The European poplar rust fungus, *M. larici-populina*, may offer a potential model for exploring key gaps in our understanding of mating-type diversity at the population level. This fungus follows a heteroecious life cycle, alternating between two distinct hosts: poplar (*Populus* spp.) for clonal multiplication and larch (*Larix* spp.) for the sexual phase. Importantly, *M. larici-populina* was among the first two rust fungi with a reference genome fully sequenced, and the availability of numerous public short-read datasets provides a good foundation for population-scale analyses (Duplessis, Cuomo, et al., 2011; Lorrain et al., 2019).

The coexistence of both sexual and asexual populations of *M. larici-populina* occurs in nature. Some lineages undergo sexual cycles with genetic recombination, whereas recent studies indicate that several lineages persist predominantly through asexual reproduction (Xhaard et al., 2011; Abdalrahem, Andrieux, et al., 2025). This offers a unique framework to compare how reproductive mode influences mating-type evolution and population structure. By conducting comprehensive population genomics on *M. larici-populina* isolates collected from diverse geographic regions and reproductive contexts, we can improve our understanding of how tetrapolar mating systems function and evolve in natural populations. Such insights also contribute to the broader study of evolutionary dynamics underlying sexual and asexual reproduction, particularly in economically important plant pathogens.

2 Material and Methods

2.1 Population genomics datasets and reference isolates

We assembled a population-scale genomic dataset of *M. larici-populina* comprising 213 paired-end Illumina short-read libraries generated from publicly available repositories and newly produced sequencing runs. Public datasets were retrieved from the European Nucleotide Archive (ENA) and include 56 isolates derived from a selfing progeny of the reference isolate 98AG31 (Supplementary Table S1; Duplessis et al., 2011; Pernaci et al., 2014; Persoons et al., 2014). In addition to these public data, we generated 40 previously unpublished Illumina paired-end libraries following a standardized isolation and library preparation protocol, aligned with the parameters used for the ENA collections. Short reads were subjected to quality assessment, adapter removal, and quality trimming with “fastp” v0.23.1 (Chen, 2023). Filtered reads were assembled de novo with “MEGAHIT” v1.2.9 to generate contig-level assemblies for downstream analyses (D. Li et al., 2015, 2016). Population genomics data were all generated from DNA isolated from dikaryotic (n+n nuclei) urediniospores multiplicated on poplar leaves.

The newly sequenced set encompass seven isolates representative of the different asexual lineages previously identified from the population genetic analyses (Abdalrahem, Andrieux, et al., 2025), and the existing public collections with two additional asexual lineage isolates. Together, this sampling captures both sexual and asexual diversity within *M. larici-populina*, providing a framework for downstream analyses of mating-type diversity at the population level.

To enable haplotype-resolved analyses of mating-type (MAT) loci, we analyzed two phased genome assemblies representing sexual (isolate 98AG31) and asexual (isolate 08KE26) reproductive modes (Corre E., 2025). Both assemblies were generated using a hybrid strategy combining PacBio long-read sequencing, Illumina short-read polishing, and Hi-C chromatin conformation capture for scaffolding and phasing. This approach produced four

chromosome-scale haplotypes (two per isolate), which were used to compare the HD1/HD2 homeodomain genes and the STE3 pheromone receptor genes.

2.2 Identification and annotation of MAT genes

MAT protein sequences for *M. larici-populina* were retrieved from the 98AG31 v2 annotation on the USA Department of Energy (DOE) Joint Genome Institute (JGI) MycoCosm website (best gene model IDs for mating-type genes are as follows: HD1, 90168; HD2, 124184; STE3.1, 123975; STE3.2, 73569; STE3.3, 123740; and STE3.4, 86096) (Duplessis, Cuomo, et al., 2011; Grigoriev et al., 2014). All assemblies were screened for MAT candidates by protein-to-genome homology search using “tblastn” from “BLAST+: v2.16.0, with databases constructed by makeblastdb (Camacho et al., 2009), and hit-bearing contigs were retained for refinement using “seqkit” v2.9.0. were then aligned with “exonerate” v2.4.0 (protein2genome) to generate splice-aware gene models and target GFF output, using gene-specific thresholds (HD1/HD2: 30% minimum identity; STE3.1–STE3.4: 45% minimum identity; bestn=1; intron bounds 5–50,000 bp) (Slater & Birney, 2005). The output annotations were transformed into a standardized GFF3 for downstream analyses.

2.3 Comparative analysis of MAT loci across the new reference phased genomes

We compared MAT loci in the four phased haplotypes (98AG31-H0/H1; 08KE26-H0/H1) against the 98AG31 v2 reference. For each gene (HD1, HD2, STE3.1–STE3.4), we: (i) confirmed presence/absence by intersecting curated MAT coordinates with GFF3 features; (ii) extracted CDS features to compute coding length (nt, aa) and exon count; (iii) recorded strand orientation and genomic placement (chromosome/scaffold and start–end coordinates); and (iv) noted locus colocalization and orientation consistency across haplotypes.

2.4 Population-level comparative analysis of MAT genes

We analyzed predicted protein sequences for each MAT gene (HD1, HD2, STE3.1 to STE3.4) across all isolates, including the four phased haplotypes. To eliminate short fragments of predicted protein, we filtered the proteins based on their length compared to the average for each gene, retaining only those within ±20% of the mean length using “seqkit”. To determine the number of allelic variants for each gene, we clustered the filtered proteins with “CD-HIT” v4.8.1 at a threshold of 90% sequence identity and 80% coverage (L. Fu et al., 2012). For detailed analysis, we globally aligned representative alleles of each gene in “Geneious Prime” v2024.0.4 (Biomatters, Auckland, New Zealand) employing the BLOSUM62 substitution matrix for visualization. Additionally, to construct a phylogenetic tree, we aligned the protein sequences with “MAFFT” v7 (Katoh & Standley, 2013) and trimmed the alignment with “trimAl” v1.4.1 (Capella-Gutiérrez et al., 2009). Phylogenetic trees were constructed with “IQ-TREE” v2 using ModelFinder for model selection (Kalyaanamoorthy et al., 2017; Minh et al., 2020). We assessed branch support with 1,000 ultrafast bootstrap replicates and 1,000 SH-aLRT tests. The trees were visualized on the Interactive Tree of Life (iTOL) web-based platform (Letunic & Bork, 2021).

3 Results

3.1 Organization of mating-type (MAT) genes in sexual and asexual phased genomes

Across the four phased haplotypes (98AG31-H0/H1; 08KE26-H0/H1), all loci expected for the tetrapolar MAT system were present (Fig. 1): STE3.1 (386 amino acids (aa)) on chromosome 1 and a pair of homeodomain genes (HD1=453 aa, HD2=393 aa) located on chromosome 4. The remaining pheromone receptor genes (STE3.3 = 386 aa, STE3.4 = 341 aa) were found on chromosome 9. In comparison, STE3.2 (398 aa) was identified on chromosome 9 in 98AG31 and on small RagTag unassembled scaffolds (19_RagTag, 88_RagTag) in 08KE26, which is consistent with incomplete haplotype scaffolding rather than true absence. Gene orientation was conserved within loci across haplotypes (e.g., adjacent, oppositely oriented HD1/HD2 on

chromosome 4). These patterns are summarized in (Fig. 1), with arrows showing gene orientation and chromosome identifiers included for each haplotype.

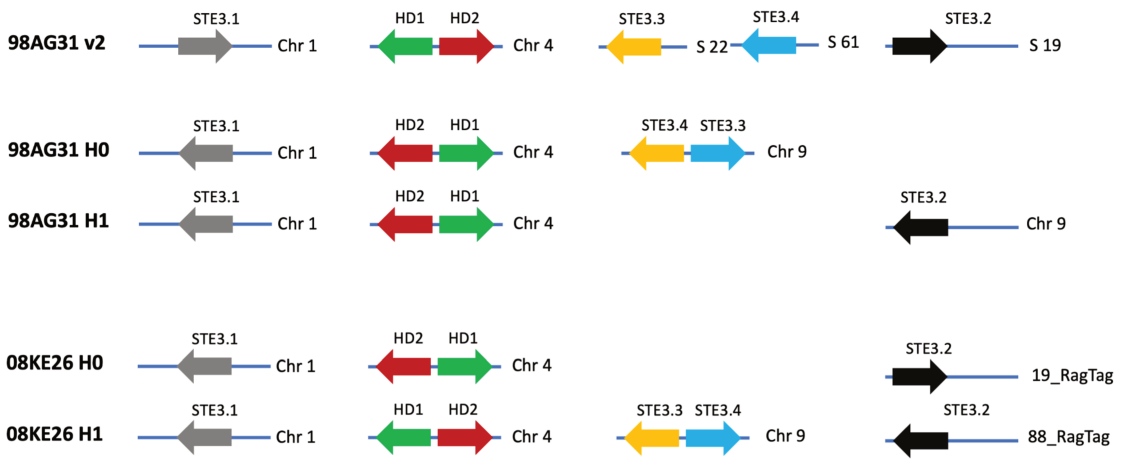


Fig. 1 Chromosomal placement and orientation of MAT genes across sexual (98AG31) and asexual (08KE26) phased haplotypes. Gene arrows indicate the orientation of HD1 (red), HD2 (green), STE3.1 (gray), STE3.2 (black), STE3.3 (yellow), and STE3.4 (blue). Rows correspond to one isolate haplotype phase. HD1/HD2 colocalize on Chr4; STE3.1 on Chr1; STE3.3 and STE3.4 on Chr9; STE3.2 on Chr9 (98AG31) or small RagTag scaffolds (08KE26).

Phasing resolved two alleles per homeodomain gene within each isolate (HD1.1/HD1.2; HD2.1/HD2.2). In comparing the phased genomes of the sexual isolate (98AG31) and the asexual isolate (08KE26), no differences in genomic sequence were detected neither for HD1 nor HD2. The alleles in both haplotypes (H0 and H1) were 100% identical between the two isolates, indicating that, at least in this comparison, reproductive mode variation did not correspond to structural variation at the HD loci.

3.2 Diversity of the HD locus

Clustering analysis of HD1 and HD2 across the collection of 217 (including four phased genomes) *M. larici-populina* isolates revealed a complex allelic polymorphism landscape that challenges traditional views of rust fungal mating systems. Instead of the binary allelic states reported previously in the reference genome of *M. larici-populina* MAT loci, the homeodomain genes exhibit multiallelic states, with seven divergent alleles (Fig. 2, Fig. S1). For the seven representative HD1 alleles, the average length was 461 amino acids, with 235 identical sites (49.8%) and an overall pairwise identity of 74.8%. The seven HD2 alleles had an average length of 402 amino acids, with 197 identical sites (48.5%) and a pairwise identity of 72.2%, indicating notable allelic divergence while preserving conserved gene structure across isolates.

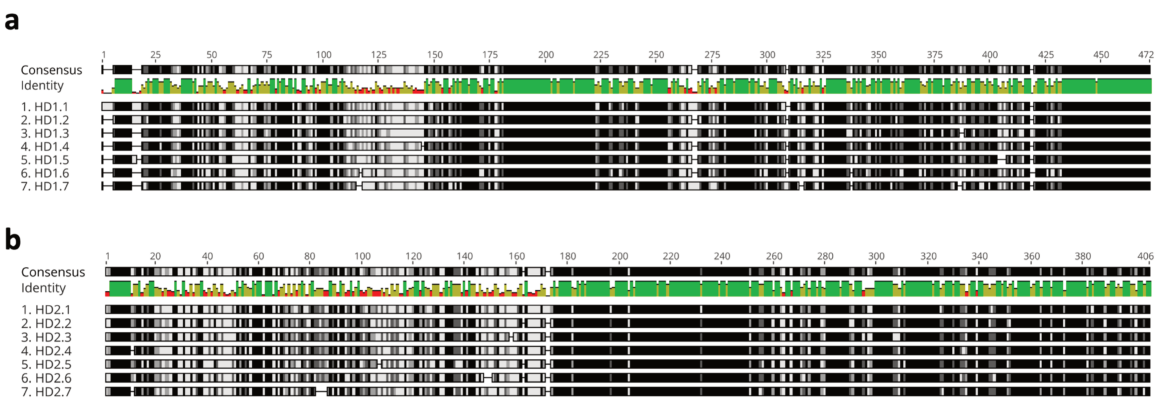


Fig.2 Amino acid diversity of HD1 and HD2 homeodomain genes in *Melampsora larici-populina* isolates. Amino acid sequence alignments of (a) seven HD1 allelic variants and (b) seven HD2 variants derived from 217 *M. larici-populina* isolates. Conserved regions are shown in green, with variable sites highlighted by color-coded bars. Black shading indicates conserved residues, while lighter (gray) shading shows variable positions. Colored bars above the alignment indicate amino acid conservation (green = conserved, yellow = variable, red = divergent). Alignments were performed using global alignment with the BLOSUM62 substitution matrix in Geneious Prime.

This resulted in HD1 and HD2 consisting in seven alleles each, labelled HD1.1–HD1.7 and HD2.1–HD2.7, respectively. This shows that MAT homeodomain diversity extends well beyond a two-allele model, revealing seven distinct allelic states for each gene and underscoring extensive historical divergence

3.3 Pairings of homeodomain alleles (HD1–HD2) across chromosomes, within isolates

Analysis of the collection panel (213 isolates; four phased were excluded) revealed a strongly modal state at two alleles per gene (Fig. 3a-b) consistent with phased sequences within isolates. In total, 144 isolates carried two alleles of both homeodomain genes (HD1, HD2), reflecting the expected organization. A few additional cases with equal copy numbers were also observed, including isolates with a single copy (HD1 = 1, HD2 = 1), three copies (HD1 = 3, HD2 = 3), or four copies (HD1 = 4, HD2 = 4). In addition, some isolates showed unequal copy numbers between HD1 and HD2. These different copy cases likely reflect a mixture of true biological states or technical effects (e.g., assembly fragmentation or genomic contamination), whereas the two-allele state remains the dominant configuration across the population.

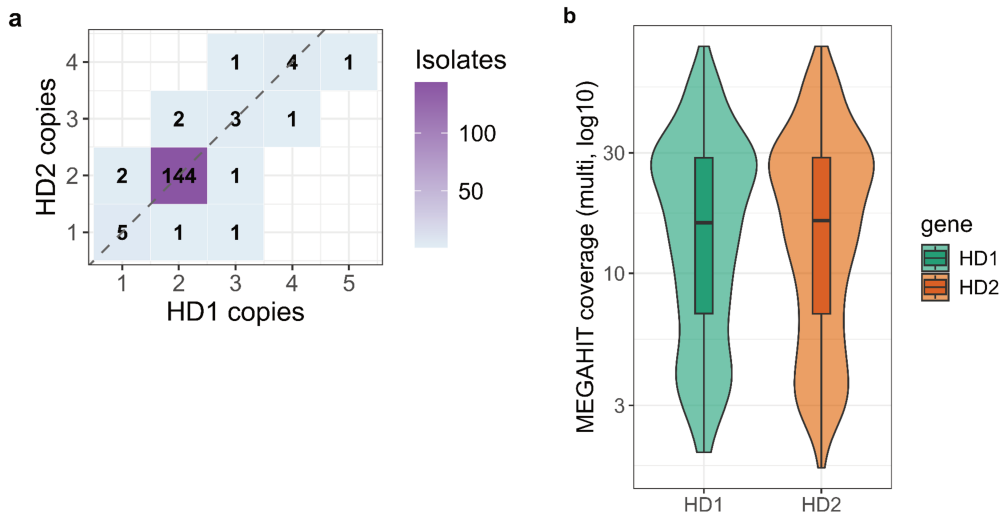


Fig. 3 Copy-number variation of HD1 and HD2 genes across isolates. a. Heatmap of isolate counts across HD1–HD2 copy-number combinations. **c.** Isolate counts by equality status. **b.** Coverage distributions (MEGAHIT, multi log10) for HD1 and HD2 supporting concordant copy-number inference.

Having established that most isolates carry two HD1 and two HD2 alleles, we then investigated whether these alleles assort randomly or form specific pairwise associations across contigs. Analysis of the HD gene alleles has revealed a clear and consistent pattern of one-to-one associations between the clusters of HD1 and HD2 alleles (Fig. 4). Each HD1 allele is only paired with one HD2 allele partner across all surveyed contigs, with no evidence of alternative combinations. For instance, HD1.1 is linked exclusively with HD2.2, HD1.4 pairs with HD2.7, HD1.3 pairs with HD2.5, HD1.2 pairs with HD2.6, HD1.6 pairs with HD2.3, HD1.7 pairs with HD2.4, and HD1.5 pairs with HD2.1. This pattern indicates a strong co-evolutionary association between the HD1 and HD2 alleles, which is expected since they are physically linked (located close together on the genome).

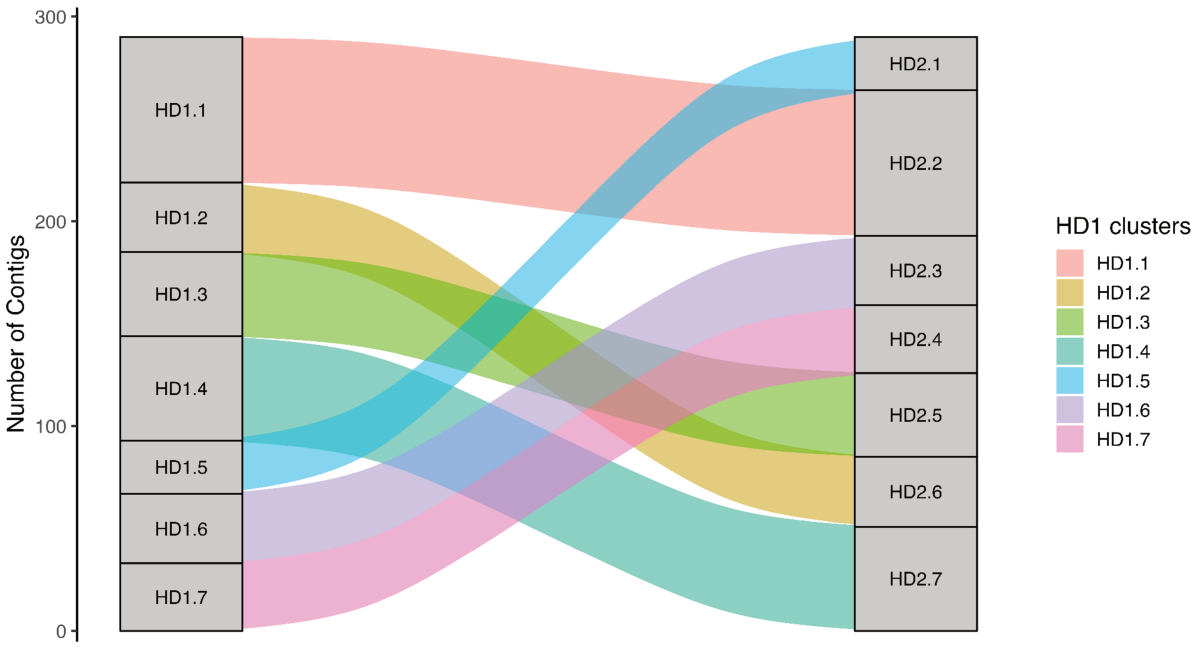


Fig 4. Distribution of HD gene allele associations across contigs. Each HD1 allele is closely paired with a single HD2 allele, with no evidence of alternative pairings. For example, HD1.1 is exclusively linked to HD2.2, while other HD1 alleles show similarly strict one-to-one relationships with their respective HD2 alleles.

We next examined how HD locus pairs are distributed within dikaryotic isolates (n + n). This analysis demonstrated a consistent pattern across all isolates (Fig. S2); all isolates were heterozygous at the HD locus, with each nucleus harboring a distinct HD1–HD2 combination. For example, dikaryons contained complementary allele loci such as HD1.1–HD2.2 in one

nucleus together with HD1.2–HD2.6 in the other nucleus. The absence of homozygous HD combinations between nuclei confirms the mating compatibility requirements essential for dikaryon formation.

3.5 Allelic diversity of STE3 pheromone receptor genes

In our analysis of short-read assemblies at the genome collection level, we found that HD1 and HD2 often assembled on the same contig but in opposite orientations. This finding aligns with their arrangement in both the phased genomes and the v2 reference genome. In contrast, STE3 pheromone receptor genes (STE3.1 to STE3.4) are located on different contigs, separate from the HD locus/contig. This distribution supports the separate genomic locations observed in rust fungi, characteristic of a tetrapolar mating type system. We therefore next examined the sequence diversity of STE3 receptors across isolates to assess how these genes vary within the population.

Across the isolates, STE3.1, with a protein length of 392 amino acids, formed a single cluster without divergent alleles. STE3.3, which is 386 amino acids long, also formed a single cluster across the isolates. These findings indicate that both STE3.1 and STE3.3 are highly conserved genes with no variation. In contrast, the other pheromone receptor genes showed notable allelic diversity. For STE3.2, six distinct sequence clusters were identified (Fig. 5a). The three main clusters correspond to proteins that are 398, 302, and 308 amino acids in length. Additionally, three other clusters were identified, each containing one to three isolates with predicted protein lengths of less than 300 amino acids. Also, STE3.4 resolved into four clusters (Fig. 5b). Two main clusters comprised proteins of 341 and 297 amino acids, whereas the remaining two minor clusters included five isolates and two isolates, respectively, with protein lengths below 270 amino acids.

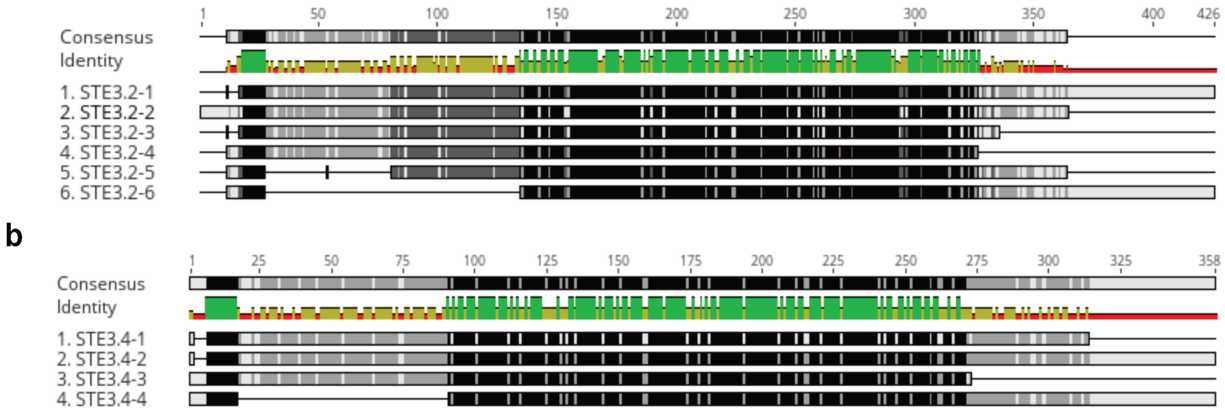


Fig. 5 Sequence diversity of STE3 pheromone receptor genes in poplar rust fungi. (**a**) Alignment of representative sequences from the six STE3.2 allele clusters. The three main clusters correspond to proteins of 398, 302, and 308 amino acids (aa), while three additional minor clusters include one to three isolates each, with protein lengths below 290 aa. (**b**) Alignment of representative sequences from the four STE3.4 allele clusters. Two major clusters correspond to proteins of 341 aa and 297 aa, while two minor clusters have protein lengths below 270 aa. Black shading indicates conserved residues, while lighter (gray) shading shows variable positions. Colored bars above the alignment indicate amino acid conservation (green = conserved, yellow = variable, red = divergent). Alignments were performed using global alignment with the BLOSUM62 substitution matrix in Geneious Prime.

These results show that STE3.2 and STE3.4 retain allelic variation, unlike the highly conserved STE3.1 and STE3.3, highlighting the unequal evolutionary dynamics among STE3 family members in rust fungi.

4 Discussion

At the genomic level, our analyses revealed no differences between sexual and asexual lineages in terms of mating-type gene content. Both groups retain the complete set of mating-type (HD1, HD2) and pheromone receptor (STE3.1 – STE3.4) genes, consistent with reports in other fungi where isolates classified as asexual retain complete mating-type loci (Doellman et al., 2024b). The inability of some isolates to undergo the complete sexual cycle is therefore unlikely to result from the loss of core MAT components. More likely, it reflects differences in gene regulation or the involvement of other genes, for example, genes related to meiosis or sporulation. Resolving this will require a complete analysis of MAT target gene networks, meiotic gene expression profiles, and post-transcriptional regulatory trajectories that correspond to sexual development in rust fungi (Liu et al., 2018; Yeager et al., 2021).

Our findings reveal a more complex organization of homeodomain (HD) genes in *M. larici-populina* than is suggested by the reference genome. Although the reference assembly identifies only a single copy of HD1 and HD2 (Duplessis, et al., 2011), the new phased genome analysis uncovered four distinct variants (HD1.1, HD1.2, HD2.1, HD2.2) located on chromosome four in almost the same genomic region, sharing less than 70% amino acid similarity. This pattern highlights deep allelic variation rather than simple sequence polymorphism. Also, coverage analysis confirmed that these variants are not assembly artifacts; each allele exhibits high read coverage and maps to separate contigs. The same result was consistently reproduced with both long-read phased assemblies and short-read sequence data, confirming robust support for each variant on separate contigs, arguing against assembly artifacts and supporting biological reality (Cuomo et al., 2017). At the collection level, 155 out of 213 isolates carried two distinct HD1 and two distinct HD2 alleles, indicating that most isolates effectively harbor two distinct HD variants. Comparable patterns have also been observed in other rust fungi; for instance, Luo et al. (2024) showed that HD loci are multiallelic across cereal rust species, lending further support to the idea that allelic diversification at the HD loci is a widespread feature within the Pucciniales.

We explored the diversity of mating-type genes in *M. larici-populina*, focusing on both the homeodomain (HD) transcription factors and the STE3 pheromone receptor genes. Using phased genome assemblies together with collection data from 213 isolates, we found more diversity than what is described in the reference genome. The population we analyzed included isolates from both sexual and asexual lineages, which allowed us to compare how different reproductive modes influence genetic variation. Our results show that HD genes are more complex than the copies reported in the reference genome, displaying considerable allelic diversity. In contrast, the STE3 genes exhibit variable patterns of diversification within the pheromone receptor family itself, which may indicate distinct evolutionary trajectories. The STE3 family shows varying levels of conservation among its members. STE3.1 and STE3.3 are highly conserved, while STE3.2 displays diversity with six distinct allelic variants. Similarly,

STE3.4 also exhibits diversity, having four different allelic variants. This pattern of differential diversification within the STE3 family was observed in cereal rust fungi, where certain STE3 genes maintain polymorphisms while others are highly conserved (Luo et al., 2024). These findings provide the basis for discussing how mating-type diversity contributes to population structure and reproductive compatibility in rust fungi, with importance for understanding the dynamics between conservation and diversification in pheromone-receptor systems.

This broader view is consistent with findings in other rust fungi. For example, Holden et al. (2023) reported nine HD alleles in *Puccinia striiformis* when sampling a larger and geographically diverse population, compared to the more limited diversity observed in earlier studies. Such results reinforce the central point of our study: increasing sample size consistently reveals novel HD variants. What might appear, in early or individual-focused analyses, to be "just two alleles" per locus is often only the visible portion of a much richer allelic iceberg. Thus, broad sampling is essential not only to document the extent of HD diversity but also to understand its evolutionary and ecological consequences for mating compatibility and population structure (Kües, 2015).

In contrast to the high allelic variation observed at HD loci, the STE3 pheromone receptor family displays markedly lower diversity. Among the members examined, only STE3.2 formed three clear allelic clusters, consistently recovered through phylogenetic analysis and CD-HIT clustering, while STE3.1, STE3.3, and STE3.4 showed little to no clustering across our isolates. This pattern aligns with the findings in cereal rust fungi, where Luo et al. (2024) reported minimal intraspecies variation for STE3.2-2 and STE3.2-3, suggesting a predominantly biallelic structure for these receptors. Furthermore, earlier work in basidiomycetes has documented the functional importance and relative conservation of pheromone receptor genes, implying they follow distinct evolutionary constraints compared to highly diversified loci like HD (Raudaskoski & Kothe, 2010). Taken together, our data support the idea that pheromone receptors (STE3) display limited allelic diversity, even while some (like STE3.2) maintain enough variation to be detected in population-wide analyses.

It is important to clarify that our collection-level analyses were based on DNA extracted from dikaryotic isolates of *M. larici-populina* (n+n), where two haploid nuclei coexist within the same individual. Because rust fungi are cultivated and collected from many urediniospores, each isolate represents genetic material from both nuclear genomes rather than from a single haploid cell. This biological feature has important implications for interpreting our results, as the apparent presence of multiple copies or uneven coverage across loci in the de novo assembled contigs reflects the contributions of both nuclei within the dikaryon. In this context, in our study, the detection of more than two HD alleles within the same isolate. In some cases, short-read assemblies revealed three, and sometimes even four, versions of HD1 and HD2 within a single dikaryotic individual. This became clear because of the sequencing and assembly method we used, which enabled us to recover signals that might not have been visible with other tools. Importantly, coverage analysis consistently showed two HD1 and two HD2 alleles with high coverage (corresponding to the two nuclei in the dikaryotic state), along with additional alleles with lower coverage.

This indicates that the extra copies require further investigation, particularly to identify recombination signals and assess whether these additional variants could be contamination within the same isolate. We stress that this finding should not be viewed as a technical artifact, but rather as an observation that warrants further analysis. Specifically, examining progeny isolates and employing complementary sequencing or assembly methods will be essential to determine if these extra HD alleles truly represent new recombinants. If confirmed, this within-isolate diversification would add a new layer of complexity to the understanding of how mating-type loci evolve in rust fungi.

Data availability

The public sequencing data analyzed in this study are available from the European Nucleotide Archive (ENA). The isolate metadata (Supplementary Table S1), together with a reproducible Snakemake workflow implementing MAT gene extraction, alignment, and downstream analyses, as well as R scripts for visualization, are available at https://github.com/ammarabdalrahem/MAT_diversity_mlp_rust_fungi/

Supplementary figures

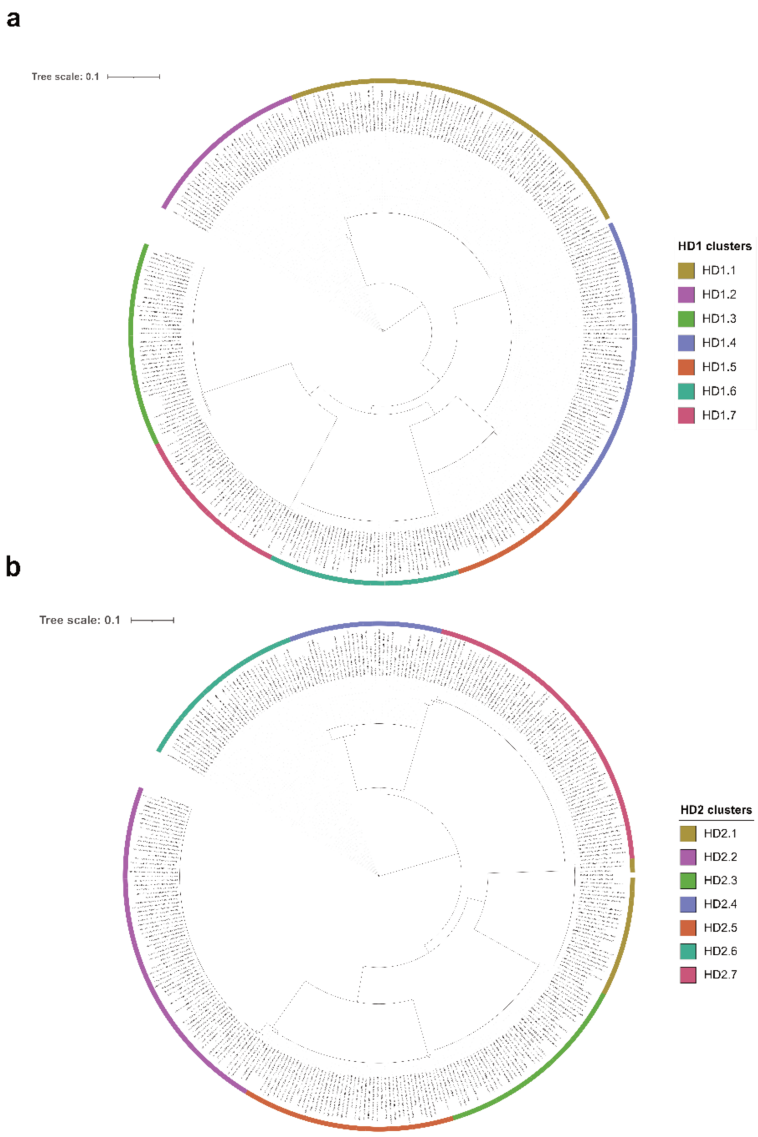


Fig. S1 Allele-specific coverage patterns in isolates with expanded HD copy number. (**a**) Isolates carrying three copies of each homeodomain gene (HD1 = 3, HD2 = 3) show two copies with high coverage and one copy with lower coverage. (**b**) Isolates with four copies of each gene (HD1 = 4, HD2 = 4) display two copies with high coverage and two copies with lower coverage. The consistent patterns observed in both HD1 and HD2 suggest coordinated duplication or recombination events rather than independent gene-specific changes.

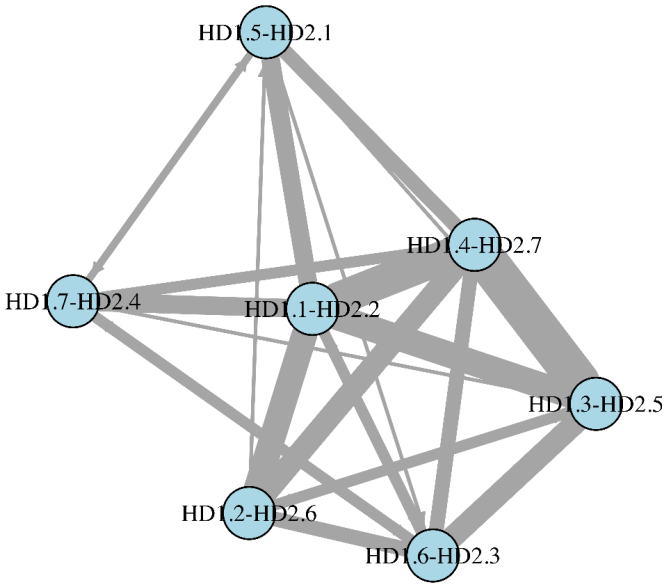


Fig. S2 Network of HD1-HD2 allele pair combinations in dikaryotic *M. larici-populina* isolates. Nodes represent unique combinations of HD locus (HD1.X-HD2.Y) retrieved from dikaryotic (n+n) isolate contigs. Directed edges show combinations across the two contigs within each isolate, with the thickness of the edges indicating the number of isolates that display that particular combination. The absence of self-loops indicates that all analyzed isolates show a heterozygous HD locus.

Section two

Comparative Developmental Transcriptomics of Sexual and Asexual Lineages in the Poplar Rust Fungus *Melampsora larici-populina*

1 Introduction

As seen in the previous chapter, during its life cycle, *Melampsora larici-populina* alternates from a clonal reproductive mode on poplar to a sexual reproductive mode, where two nuclei (n+n) fuse, a process known as karyogamy, during the telial stage. This stage differentiation is a critical but often overlooked step in the rust life cycle. Recent population genetic studies revealed that *M. larici-populina* consists not only of sexual lineages but also of asexual lineages that bypass the sexual cycle and persist on poplar trees (Chapter one), potentially through green-bridge mechanisms (Abdalrahem, Andrieux, et al., 2025). In chapter 2, we showed evidence that asexual lineages vary in their ability to undergo sexual reproduction; some have successfully engaged in sexual reproduction, while others were stopped at specific developmental stages. However, subtler differences may exist in gene expression programs or other phases of the life cycle that differ from those of sexual lineages.

To better understand these distinct developmental stages in the poplar rust life cycle and how the fungus interacts with its plant hosts, our research department has contributed to advancing the understanding of the poplar rust life cycle through transcriptional profiling studies. Earlier work by Duplessis et al. (2011) examined the early infection process on poplar leaves and uncovered dynamic shifts in gene expression patterns related to the virulence process and to infection establishment. Building on this foundation, Hacquard et al. (2013) expanded the research to include the telial stage, demonstrating transcriptional changes associated with overwintering. A recent study focused on the host-specific gene expression on the two host trees poplar and larch (Lorrain et al., 2018). These expression studies were important in revealing stage-specific transcriptional profiles, focusing on the reference sexual isolate 98AG31. However, further work is needed to understand the process of developmental stages occurring within the poplar rust life cycle (Duplessis et al., 2021a), including the study of the variability among sexual and asexual lineages. The overwintering phase appears as an important transitional step between clonal and sexual reproductive modes, yet it remains overlooked in most rust fungal species. The pioneer transcriptomic study from Hacquard et al. (2013) consisted of a single early time point in autumn, during the differentiation of telia on poplar

leaves, and was complemented by a time course of telia differentiation for a handful of selected genes related to meiosis. Other recent transcriptomic studies focused on teliospores in a few rust species (i.e., *Gymnosporangium* spp. and *Puccinia striiformis* f. sp. *tritici*), but always at a single time point in the related life cycles, and none precisely addressed developmental processes (Tao et al., 2017, 2019; Zhan et al., 2023). Additional comprehensive transcriptomic work is needed to complete our view of the genetic processes at play at this stage of the rust life cycle.

Our study builds on earlier transcriptomic work and addresses existing gaps by covering a wider range of developmental stages. We examined stages from early infection of poplar leaves through urediniospore formation and release then, early, mid, and late telia development, providing a broad foundation for understanding how gene expression programs change over time in living, aging then, decaying host tissues. We also add a new perspective by comparing sexual and asexual isolates, offering new insights about how lineage reproductive history shapes gene expression throughout the life-cycle of *M. larici-populina*.

2 Material and Methods

2.1 Experimental Design

A comparative transcriptomic analysis was designed between two isolates selected from sexual and asexual lineages of *M. larici-populina*. The experiment was designed to capture gene expression dynamics across key developmental stages of the infection cycle on poplar leaves. Sampling time points were adapted from previous transcriptomic studies of the fungus (Duplessis, Hacquard, et al., 2011; Hacquard et al., 2013), with modifications to broaden coverage of the infection process, with time points at 6, 10, 18, 25, 32, 39, and 46 days post-inoculation (dpi).

2.2 Plant and fungal material and infection protocol

Detached leaves of hybrid poplar (*Populus deltoides* × *P. nigra* 'Robusta') were used for inoculation. Two isolates of *M. larici-populina* were selected for this study: isolate 98AG31, representing a sexual lineage and serving as a reference strain, and isolate 08KE26, representing an asexual lineage (Duplessis, Cuomo, et al., 2011; Abdalrahem, Andrieux, et al., 2025). An urediniospore suspension was prepared at a final concentration of 2 mg spores in 20 mL sterile water-agar (approximately 1×10^5 spores/mL) and evenly sprayed onto the lower surface of detached poplar leaves following a standard protocol (Rinaldi et al., 2007). Inoculated leaves were incubated at 19 °C under controlled conditions to allow infection, and at 10 dpi, the leaves were transferred in an incubator at 10 °C with an 8-hour artificial photoperiod to promote telia formation. Samples were collected at 6, 10, 18, 25, 32, 39, and 46 dpi. At each time point, three independent leaves were sampled per isolate, and urediniospores were additionally harvested from the leaves at 10 dpi to represent this spore form out of the host. This sampling strategy covered the progression from early infection (6 dpi), through uredinia development and sporulation (10 dpi and isolated urediniospores), initiation of telia (18 dpi), maturation (25-32 dpi), and late telia development (39-46 dpi). In total, eight distinct samples were obtained for each isolate: seven derived from leaf infected tissue (6, 10, 18, 25, 32, 39, and 46 dpi) and one spore sample.

2.3 RNA isolation and quality control

At each time point, leaf tissue samples were collected and immediately flash-frozen in liquid nitrogen to halt cellular activity and preserve RNA integrity. These samples were subsequently stored at –80 °C until RNA extraction. Total RNA was extracted using the RNeasy Plant Mini Kit (Qiagen), with an on-column DNase I treatment (Qiagen) to remove residual genomic DNA, following established procedures (Duplessis, Hacquard, et al., 2011; Hacquard et al., 2013). RNA quality and integrity were assessed by electrophoresis (Fragment analyzer, Agilent Technologies) and spectrophotometric analysis (NanoDrop, Fisher Thermo Scientific). High-quality RNA samples were submitted to Novogene (Munich, Germany), where libraries were prepared and sequenced on an Illumina platform (NovaSeq 6000) according to the manufacturer's protocol.

2.4 Quality filtering and gene quantification

Raw sequencing reads (paired-end, 150 bp) were processed with fastp v0.23.1 (Chen et al., 2018) for adapter trimming and quality filtering. To distinguish host- and pathogen-derived reads, a sequential mapping strategy was applied. Reads were first aligned to the *Populus deltoides* reference genome (assembly v2.0, annotation v2.1; JGI Phytozome14, Goodstein et al., 2012) using STAR v2.7.11a in two-pass mode (Dobin et al., 2013). Reads that did not map to the host were subsequently aligned to the *M. larici-populina* reference genome (version 2; JGI MycoCosm, Grigoriev et al., 2014; Persoons et al., 2022) with optimized parameters for fungal transcriptomes (--alignIntronMin 5, --alignIntronMax 3000, --alignMatesGapMax 3000, --outFilterMultimapNmax 100).

Gene-level read counts were generated with featureCounts v2.0.6 from the Subread package (Liao et al., 2013, 2014), based on exon features summarized by gene identifiers, with the paired-end strand-specific option applied. Analyses were conducted on three biological replicates per condition to ensure robustness.

2.5 Exploratory analyses of isolate- and stage-specific expression

We aimed to provide a preliminary analysis of the complete transcriptional study. We started with an explanatory analysis using a custom R script to understand and explore the expression data. We first evaluated sample clustering using size-normalized log2 counts per million (log2CPM), while advanced normalization employed variance-stabilizing transformation (VST) counts from "DESeq2" v1.44.0 (Love et al., 2014). Hierarchical clustering was performed (distance matrix computed by Pearson correlation distance and average linkage) using "stats" v4.4.1. The similarity between the log2CPM and VST-based dendrograms was assessed using "dendextend" v1.19.1 (Galili, 2015). Then, the presence/absence of gene expression across isolates was examined to identify isolate-specific expression genes. Raw counts were converted to binary data, indicating presence or absence. Genes were considered isolate-specific if they were detected in only one isolate (98AG31 or 08KE26) and were expressed at all stages in that isolate (expressed = total counts ≥ 24 across all replicates per isolate). Then, the top 20 expressed genes were ranked for annotation. Additionally, within each isolate, genes were defined as stage-specific if they were expressed in at least two replicates, specifically at one timepoint and absent from the others.

2.6 Differential expression analysis plans

A differential expression analysis was performed using "DESeq2" on the raw counts. There are two thresholds to define significantly differentially expressed genes: (i) an adjusted p-value (false discovery rate, FDR) < 0.01 for the adjusted p-value, which controls for false positives across multiple tests, and (ii) an absolute log2 fold-change ≥ 1, which is shrunk with "apeglm" (Zhu et al., 2019), indicating at least a two-fold change in expression that ensures biological relevance. Four complementary differential expression analyses were planned: (1) Global baseline comparison: all samples were tested against a single reference condition 98AG31 at 6 dpi, providing a common time point for interpreting changes across isolates and stages; (2) Within-isolate across time points: 98AG31 and 08KE26 were analyzed separately, first to identify genes showing any temporal variation, and then to assess changes between timepoints (e.g., T6/T10, T10/T18), to capture the direction of developmental expression program changes; (3)

Isolate comparisons at the same stages: Both isolates were compared at each timepoint to detect isolate-specific differences; (4) Core reference re-analysis: a union of significantly expressed genes was built across the previous three plans, and differential expression was re-run.

3 Results

3.1 Sequencing quality assessment

After receiving the raw sequencing reads from the provider, data quality was assessed following adapter trimming and filtering with fastp (Fig. 1). Across all infection stages, more than 99% of reads were retained after preprocessing, indicating minimal loss of data. The proportion of Q30 bases consistently exceeded 98%, confirming high base-calling accuracy across all samples. GC content remained stable (~42–44%) and was comparable between isolates, in line with expectations for *M. larici-populina* transcriptomes. Duplication rates showed some variability at later stages (32, 39, and 46 dpi), with the asexual isolate 08KE26 displaying higher levels than the sexual isolate 98AG31. All quality metrics were within acceptable ranges, supported by saturation curves (Supplementary Fig. 1), confirming the suitability of the data for downstream alignment and expression analyses.

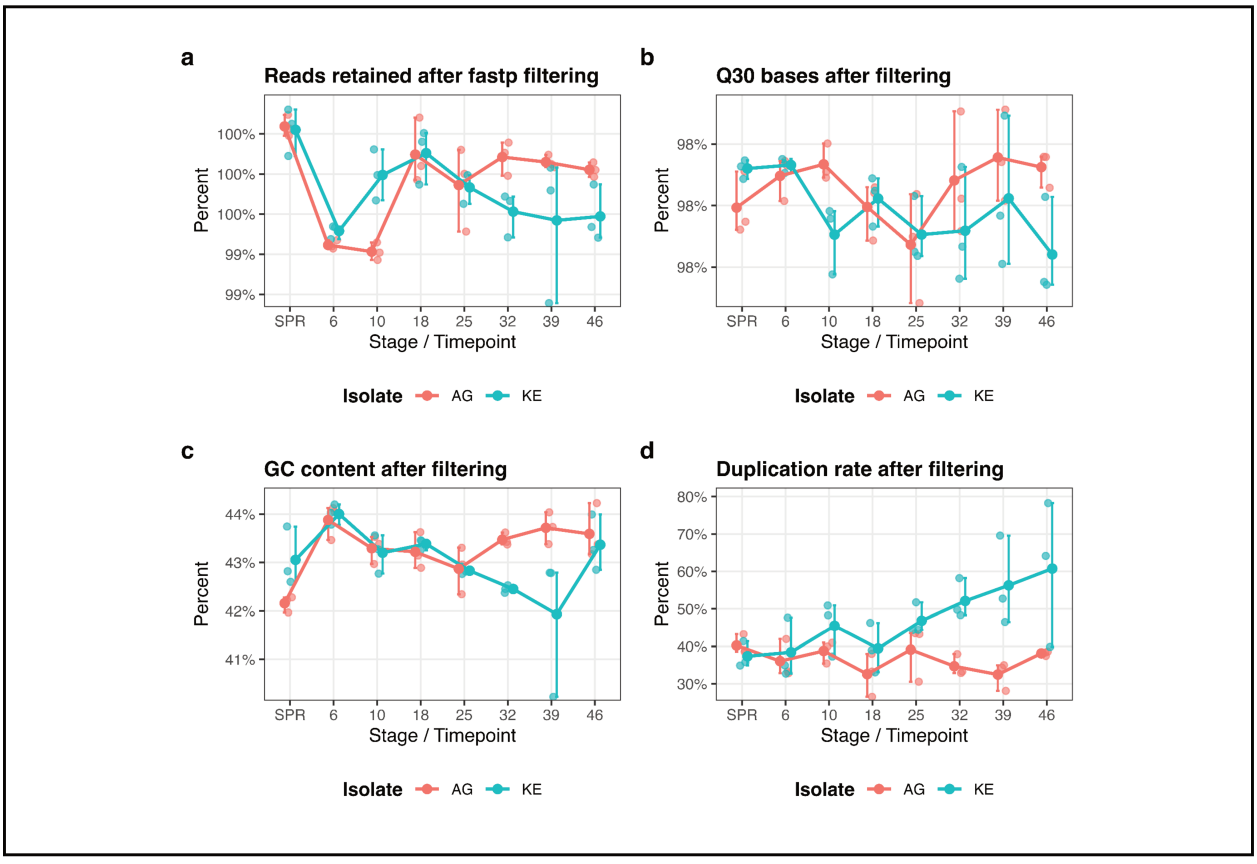


Fig. 1 Sequencing quality metrics after filtering with fastp. Quality control metrics after adapter trimming and quality filtering with fastp are shown for the 98AG31 (sexual, red, termed AG) and 08KE26 (asexual, blue, termed KE) isolates across successive infection stages and developmental time points. (a) Percentage of reads retained after filtering. (b) Percentage of bases with Phred quality scores ≥ 30 (Q30). (c) GC content of filtered reads. (d) Duplication rate after filtering. Points represent biological replicates (n = 3), with error bars indicating the standard error of the mean. These metrics collectively illustrate the overall quality and consistency of sequencing data retained for downstream analyses.

3.2 Read mapping quality metrics and potential reference bias

To determine whether the use of the *M. larici-populina* 98AG31 reference genome introduced mapping bias, we compared alignment statistics for both the reads of the 98AG31 sexual isolate and the reads of the 08KE26 asexual isolate mapped against the CDS of predicted genes in the reference genome (Mellp2_3_GeneCatalog_20151130.gft) across infection stages (Fig. 2). Overall, mapping performance was comparable between isolates. The percentage of uniquely mapped reads was similar across timepoints, with minor fluctuations at specific stages (e.g., 32 and 39 dpi), where 08KE26, despite not being the reference, showed slightly higher values. The per-base mismatch rate was consistently close to zero in both isolates, indicating negligible divergence from the reference at the aligned positions, with reads from isolate 08KE26 being slightly higher than for the reference isolate. Likewise, the proportions of multi-mapped and unmapped reads followed similar trends in 98AG31 and 08KE26. Together, these results demonstrate that using the 98AG31 genome as a reference did not introduce detectable mapping bias with 08KE26 samples and that read alignment was of sufficient quality for downstream expression analyses.

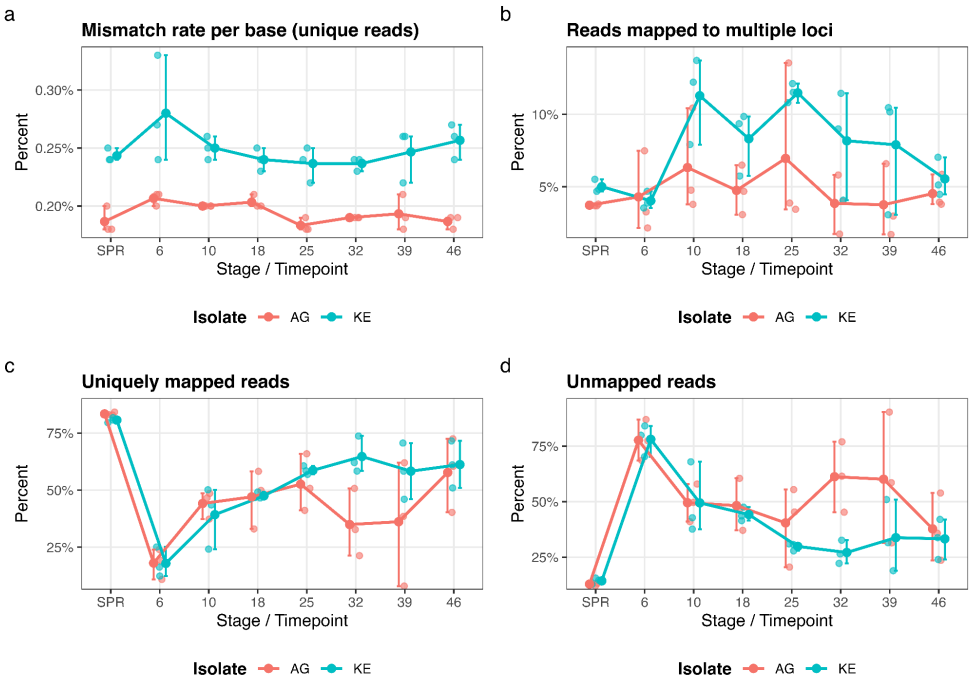


Fig. 2 Read mapping metrics for two poplar rust isolates across infection stages. STAR alignment quality metrics relative to the *M. larici-populina* 98AG31 reference genome v2 are shown for two isolates, 98AG31 (sexual, red, termed AG) and 08KE26 (asexual, blue, termed KE), across successive infection timepoints. (a) Per-base mismatch rate among uniquely mapped reads. (b) Percentage of reads mapping to multiple loci. (c) Percentage of uniquely mapped reads. (d) Percentage of unmapped reads (calculated as 100 − [uniquely mapped + multi-mapped]). Points represent individual biological replicates (n = 3), with error bars indicating the standard error of the mean.

3.3 Samples variation aligns with developmental stage and reproductive strategy

To assess sample relationships, we compared sample cluster dendrograms before and after a variance-stabilizing transformation (VST) built from the read counts attributed to *M. larici-populina* CDS per sample. The tanglegram analysis (Fig. 3a) showed virtually no change in clustering structure, with an entanglement score of 0.0 and Baker's γ of 1.0, indicating that VST normalization did not change the overall sample relationships. Then we explored sample variation using principal component analysis (PCA) on VST normalized counts (Fig. 3b). The first principal component (PC1, 35% variance) captured stage transcriptional expression variation, with samples clustering according to their timepoints for both isolates. Urediniospores, uredinia stages (T6, T10), early teliospores (T18), and mid teliospores (T25) were clearly separated, whereas the late teliospore stages (T32–T46) showed partial overlap. The urediniospore stage (TS) clustered between T18 and T25, suggesting a proper transcriptional pattern distinct from the poplar infection stage at which urediniospores are produced and released (i.e., T10). The second component (PC2, 29% variance) distinguished the two isolates, with 98AG31 (sexual) and 08KE26 (asexual).

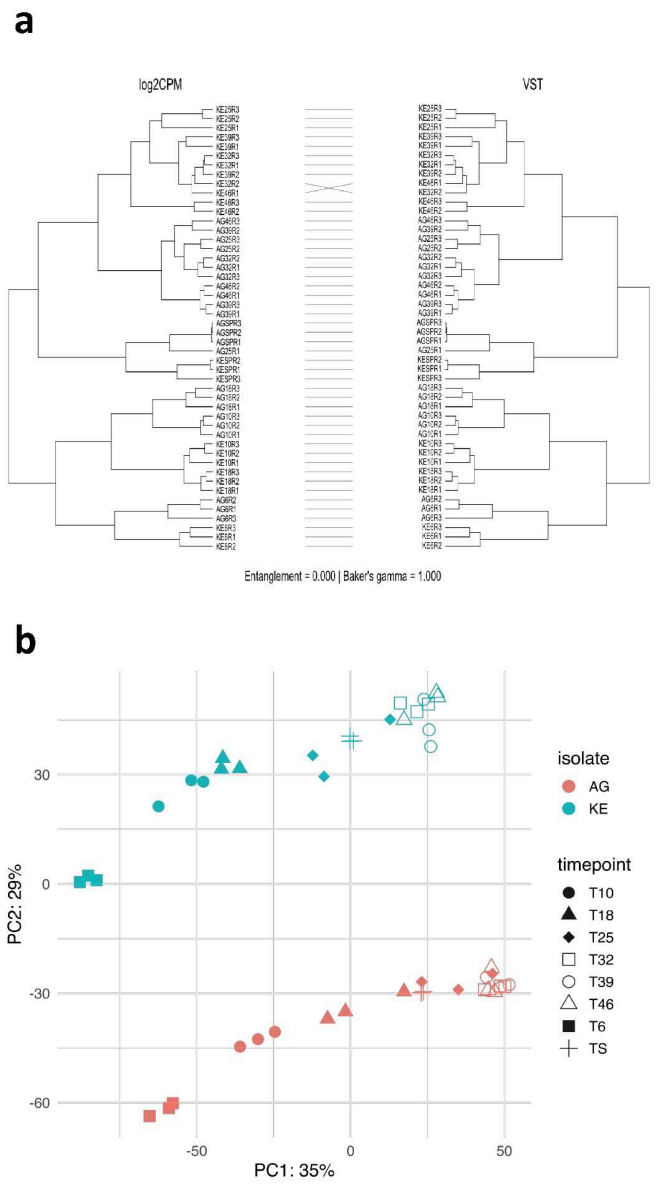


Fig. 3. Sample clustering reflects developmental progression and reproductive mode. (a) Hierarchical clustering of samples transformed by log2CPM (log2 counts per million) and normalized by VST variance-stabilizing transformation (VST). **(b)** Principal component analysis (PCA) of VST-normalized counts. PC1 (35% variance) captured developmental progression. PC2 (29% variance) distinguished the two isolates, separating 98AG31 (sexual, red, termed AG) and 08KE26 (asexual, blue, termed KE).

These analyses show a strong clustering by both time point and isolate, confirming data quality and supporting the presence of biologically meaningful differences across fungal developmental stages and between the two lineages with distinct reproductive modes.

3.4 Isolate- and timepoint-specific gene expression

The number of expressed genes in the catalog initially contains 19,550 genes. An exploratory analysis of raw counts identified large sets of genes expressed exclusively in one isolate at all developmental stages. A total of 843 genes were uniquely detected in the *M. larici-populina* asexual isolate 08KE26, while 732 genes were specific to the sexual isolate 98AG31. Within each set, the 20 most highly expressed genes were examined in more detail; however, none of these candidates had a clear interpretable functional annotation.

In addition to isolate-level differences, we assessed expression restricted to individual developmental stages. The number of timepoint-specific genes varied across stages and between isolates (Table 1). For example, the sexual isolate 98AG31 exhibited the largest set of stage-specific transcripts at T10 (38 genes), whereas the asexual isolate 08KE26 showed its peak at TS (25 genes). These stage-restricted expression patterns suggest the occurrence of transient transcriptional programs during the fungal life cycle.

Table 1. Number of timepoint-specific genes detected in each isolate.

Isolate	Timepoint	Number of specific genes
Sexual (98AG31)	6	24
	10	38
	18	24
	25	5
	32	0
	39	0
	46	1
	SPR	22
Asexual (08KE26)	6	15
	10	12
	18	8
	25	12
	32	0
	39	1
	46	2
	SPR	25

4 Discussion

The preliminary transcriptomic analysis reveals that reproductive mode in *M. larici-populina* may represent a constitutive rather than purely environmental trait, with implications for understanding the evolution and maintenance of asexual lineages in rust fungi. To our knowledge, this study provides the first RNA-seq dataset covering advanced developmental stages of *M. larici-populina* with transition from uredinia to telia then overwintering under controlled laboratory conditions. Previous studies, such as Hacquard et al. (2013) and Duplessis et al. (2011), offered valuable insights into gene expression in this species. However, our dataset extends this knowledge by capturing a broader range of developmental stages with next-generation sequencing in a standardized experimental setting.

4.1 The impact of developmental stage on transcriptional variation

The principal clustering of samples was driven by the developmental stage, which aligns with established patterns in rust fungi, where stage-specific transcriptional programs seem to be conserved features. The clear separation of temporal coordination observed in our study, particularly the transition from early infection to telia formation, might mirror the tightly regulated karyogamy and meiosis processes documented in poplar rust telia, where nuclear fusion and meiotic gene expression occur in precise temporal windows. This is shown by Hacquard et al. (2013) on *M. larici-populina,* using a custom oligoarray that included 13,093 genes, of which 2,035 were significantly regulated between telia and uredinia. Therefore, approximately 15% of the gene catalog was differentially expressed. A similar perspective on the wheat stripe rust fungus, *Puccinia striiformis* f. sp. *tritici,* as revealed by transcriptomic studies, shows a distinct expression profile across all five spore forms, with each stage exhibiting many unique expressed genes (Zhan et al., 2023). The shift from photosynthetically active to senescent leaf tissue during our time course may explain some of the dramatic transcriptional changes observed. Previous transcriptomic study of the Japanese apple rust fungus *Gymnosporangium yamadae* on apple trees revealed similar host-dependent transcriptional reprogramming (Tao et al., 2020), with detoxification pathways activated early and secondary metabolism pathways induced later, suggesting that rust fungi adapt their genetic programs to

changing host physiological conditions. Our controlled laboratory conditions likely simplified these host-pathogen dynamics compared to natural conditions, where additional environmental stressors and host genetic diversity would contribute to transcriptional complexity. These studies indicate that stage-specific gene expression is a fundamental feature of rust fungi, shaped by both their developmental programs and the host environments.

4.2 Reproductive mode imprints on gene expression

The transcriptomic separation between sexual and asexual isolates under similar conditions represents a previously undocumented phenomenon in rust fungi. However, in *Neurospora crassa,* isolates with different mating types during asexual development show some differences in gene expression between the two mating types (Z. Wang et al., 2012). The magnitude of differences observed between sexual and asexual *M. larici-populina* isolates suggests deeper evolutionary divergence than simple variation. Comparative transcriptomic studies of *M. larici-populina* infection on different hosts (larch *versus* poplar) revealed that 20% of expressed genes were host-specific (Lorrain et al., 2018). The isolate-specific transcripts identified in our study may represent a similar phenomenon, where reproductive mode has driven specialization comparable to host adaptation. The fact that these differences persist across developmental stages suggests constitutive rather than conditional regulatory divergence, which could have evolved through adaptive optimization for different life history strategies.

For the moment, we have not explored the biological function of these unique genes. Future analyses, such as functional annotation, will be necessary to determine whether these genes are associated with a specific reproductive mode or other functions.

Our study continued previous transcriptome surveys of the *M. larici-populina* life cycle by expanding a comprehensive view of genetic programs expressed across developmental stages. We add a new dimension that has not been addressed before in rust fungi: a direct comparison between a sexual and an asexual isolate. The transcriptomic patterns observed here also inform understanding of rust fungal evolution more broadly. Recent studies suggest that sexual reproduction is the ancestral state in rust fungi, with asexual reproduction representing derived

evolutionary trajectories (McTaggart et al., 2022). The constitutive transcriptional differences we observed could represent molecular signatures of such evolutionary transitions, potentially involving coordinated changes in multiple regulatory networks rather than simple loss of sex.

5 Limitations and future directions

The main limitation of our current study is that we relied on version 2 of the *M. larici-populina* genome. Although this version provides a useful reference and annotation, it is not sufficient for functional interpretation and will be soon outdated. Several pieces of evidence gathered in the past years through various genomics studies (e.g. gene family re-annotation, identification of avirulence genes) highlighted issues in the quality of the version 2 genome annotation. Although the annotation was overall of adequate quality for setting comparative studies, improvements are still awaited, regarding both gene identification and functional annotation. As a result, many genes with stage- or isolate-specific expression profiles lack consistent annotation information, hindering our ability to explore their biological significance. This limitation is now being addressed, as the IAM (Interactions Arbres – Microorganismes) Department is working on version 3 of the reference genome, which is haplotype phased and much better assembled (i.e., close to gapless telomere-to-telomere assembly). The improvement of the assembly and the provision of newly generated RNA-seq data will undoubtedly impact the quality of gene annotation, such as Gene Ontology (GO) enrichment, KEGG pathway mapping, and other functional genomics approaches. These analyses will provide a biological meaning needed to understand whether the identified transcripts are linked to processes such as sexual/asexual reproduction, stage, or host interaction.

In the present study, we compared only one sexual isolate and one asexual isolate, so it remains unclear whether the observed differences reflect reproductive strategy itself or simply lineage-specific variation. Expanding the analysis to a broader panel of sexual and asexual

isolates could help to clarify whether the transcriptomic signatures we identified are consistent features of reproductive mode or limited to particular genetic backgrounds.

Finally, while our experiments were performed on detached leaves under controlled laboratory settings, they provided insights into MAT-pathway expression, simplifying the environmental and host interactions that shape poplar rust development. Future work should expand these experiments to live poplar and larch hosts in natural environments (particularly since we now know that asexual isolates can complete sexual cycles and infect larch needles). Integrating field-based infection trials with comprehensive time-series transcriptomics and broader sampling of both poplar and larch populations will ground our findings more ecologically relevant. Along with more advanced genomic resources, such as high-quality phased assemblies and population-scale resequencing, this approach will transition the study of M. larici-populina from merely cataloging gene presence and expression to developing mechanistic models of how developmental pathways, host signals, and reproductive strategies work together to drive rust life cycles.

Supplementary figures

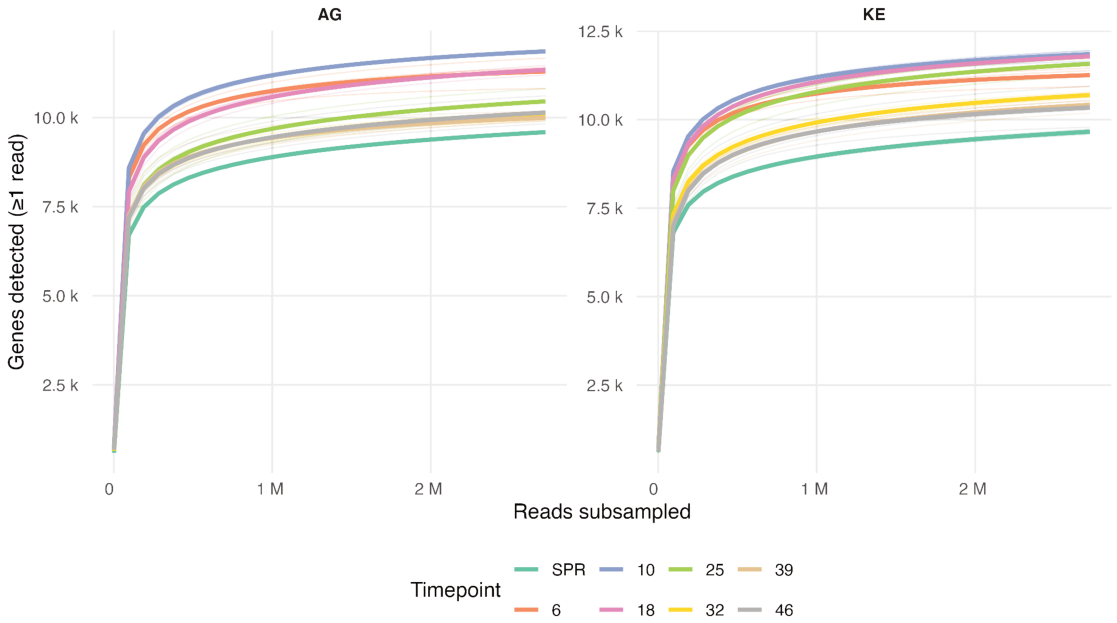


Fig. 1. Saturation curves per isolate and timepoint. Saturation curves of detected genes (≥1 read) as a function of subsampled RNA-seq read depth. Thin lines represent individual replicates, and thick lines represent the mean per timepoint. Curves are shown for the two isolates: 98AG31 (sexual, red, termed AG) and 08KE26 (asexual, blue, termed KE).

General Discussion & Perspectives

Discussion

In this thesis, I investigated the variability of reproductive modes in the poplar rust fungus *Melampsora larici-populina* with the aim of revisiting the classical view of its life cycle. The coexistence of sexual and asexual lineages in natural populations raises broader questions about how reproductive strategies contribute to pathogen diversity, host adaptation, and the evolutionary maintenance of sex. This study provides a comprehensive overview of rust fungi reproduction by combining population surveys, experimental comparisons between sexual and asexual lineages, and molecular analyses of mating-type gene diversity and gene expression during key developmental stages. The findings highlight the implications of these reproductive strategies on evolutionary paths and pathogen dynamics.

At the population level (**Chapter One**), we benefited from a comprehensive dataset of temporal samples of *M. larici-populina* to perform an in-depth population genetic analysis. We identified the coexistence of sexual and asexual lineages in natural populations. Additionally, we carried out a theoretical simulation study to better understand the dynamics of key population genetic parameters used to study partially clonal populations (e.g., poplar rust) across generations and how they can be reliably interpreted. This framework also provided the basis for selecting and comparing lineages in **Chapter Two**. At the individual lineage level, we compared the completeness of the sexual life cycle of asexual lineages using greenhouse experiments and microscopic examinations of nuclear fusion dynamics. These analyses revealed that some asexual lineages maintain the ability to re-invest in sexual reproduction, while others are halted at certain stages. This variability was also investigated in parallel using molecular analyses of genomic and transcriptomic data (**Chapter Three**). We moved to the organism and functional level and revealed the diversity of homeodomain genes of mating-type (MAT) loci, the genetic control center of sexual reproduction. Additionally, we highlighted marked differences in gene expression programs between sexual and asexual isolates. We found unique expressed genes during different developmental stages of the rust life cycle, from early infection to the late overwintering stage. Together with the previous chapters, these results have allowed us to build

a better view of reproductive strategies, from the population level down to the molecular level. This enhances our ability to understand their differences and provides a broader perspective on the evolutionary development of plant pathogenic fungi.

Our findings confirm that many asexual lineages of *M. larici-populina* persist for decades alongside sexually reproducing populations and are reliably detected in the temporal data. This observation supports our hypothesis that asexuality alone can maintain the continuity of lineages over extended periods. Yet, our experiments challenge this notion by showing that not all asexual lineages are entirely restricted from sex. Some retain a hidden potential to reinvest in sexual reproduction, while others appear to have broken this possibility. This contrast raises many questions: what separates those lineages that remain capable of returning to sex from those that cannot? And is there a genetic determinism for asexuality, or do they survive only by chance thanks to favorable environmental conditions?

Taken together, these findings suggest a more dialectical interpretation. On the one hand, asexuality provides stability and continuity, allowing lineages to persist, skipping the sexual cycle. On the other hand, the persistence of sexual potential in certain asexual lineages reveals that asexuality is not always irreversible but may coexist with the possibility of recombination. Our analyses of gene expression support this view: beyond the conserved mating-type genes, other loci appear to govern the contrasting abilities of lineages, indicating an underlying genetic basis that mediates the balance between the two strategies. According to this logic, the reproductive system of these rust fungi represents a paradox: a dynamic interplay in which stability (asexual) and flexibility (sexual/recombination) are not mutually exclusive but rather interplay within the same evolutionary framework.

Revisiting the classical view of poplar rust life cycle

The classical view of the poplar rust lifecycle has long been that it is an annual sexual life cycle between two alternate plant hosts. However, recent evidence, including our work, challenges this assumption (Abdalrahem et al., 2025; Xhaard et al., 2011). The identification of asexual lineages in *M. larici-populina* (chapter 1 first study) initially relied on examining standard

population genetic indices, such as negative inbreeding coefficient F_{IS} values and reduced clonal richness (*R*), which are used as indicators of long-lasting asexual reproduction (Halkett et al., 2005; Arnaud-Haond et al., 2007; Stoeckel et al., 2021; Abdalrahem et al., 2025). To further rule out the possibility that such patterns arose by chance due to sexual recombination and allelic shuffling, we performed an analysis to verify the discrimination power of our marker set and excluded random recurrence as a probable explanation (Stoeckel et al., 2024). While these indices are informative, they remain indirect indications. To strengthen this interpretation, we complemented them with a biological approach: the fungus has an annual life cycle (Lorrain et al., 2019), so the recurrence of the same multilocus lineages (MLLs) over multiple years directly indicates long-term asexual persistence. This hierarchy of evidence, from indirect to direct, underscores the importance of combining population genetic indices with biological observations when revisiting assumptions about reproductive modes. The key question is how a fungus with an obligate sexual annual cycle can maintain asexual lineages across decades.

Therefore, the persistence of asexual lineages cannot be seen apart from their ecological environment; the environment itself plays an active role in influencing reproductive strategies (Tilquin & Kokko, 2016; Liegeois et al., 2023). Our plausible explanation is the existence of a “green bridge” in milder winters (Mboup et al., 2009), such as the Mediterranean climate in southern France, where poplar trees may retain green leaves throughout the winter. Observations of rust-infected poplar leaves in February 2024 in Cavalaire-sur-Mer (J. Pinon, pers. comm.), followed by identification through genotyping of asexual lineages, supports this “green bridge” hypothesis. Similar observations of overwintering rust-infected poplar leaves in Paris in November 2024 (F. Halkett, pers. comm.) support the idea that environmental conditions could facilitate the asexual lineages' persistence, allowing the pathogen to bypass the sexual phase. In the context of climate change, what was once an exceptional overwintering strategy could become a more common outcome, thereby altering the evolutionary trajectory of the pathogen. Milder winters may contribute to the persistence of asexual lineages by extending the duration of the growing season, thus increasing the probability of green bridges, which facilitates asexual long-lasting. This environmental influence on plant pathogenic fungi

could affect not only their reproductive behavior, as seen in the wheat stripe rust fungus, *Puccinia striiformis* f. sp. *tritici* (Ali, Gladieux, Leconte, et al., 2014), where the fungus reproduces sexually in the Himalayan and a few other regions, primarily globally, but also its adaptability and pathogenicity (Velásquez et al., 2018; Boixel et al., 2022). The situation might increase the risk of epidemic spread, as many invasive fungal species depend more on asexuality for rapid growth (Drenth et al., 2019).

To quantify developmental dynamics objectively, a deep-learning pipeline could be developed to detect and track the development of telia over time in image series, enabling direct comparison between sexual and asexual lineages. The analysis of temporal phenotypes (counts, area, growth rates, transitions from uredinia to telia) would allow us to statistically test whether asexual lineages (e.g., 08KE38) exhibit a slower kinetics of telia development relative to sexual lineages.

Our results showed that *M. larici-populina* can maintain long-term asexual lineages. This observation, along with recent evidence of sexual reproduction in *Austropuccinia psidii* (McTaggart et al., 2020), shows that *M. larici-populina* has the ability to spread not only through sexual reproduction but also asexually, as is found for wheat stripe rust (Ali et al., 2014; Hovmøller et al., 2016). Typically, this may account for the establishment of poplar rust in North Africa, South Africa and Chile, where the alternate host, larch, is absent (Barrès et al., 2008) . These examples highlight a broader paradox across rust fungi: reproductive modes are neither fixed nor uniform but instead could change within a given species depending on ecological or evolutionary contexts, which adds to the evolutionary potential of rust fungi (McDonald & Linde, 2002).

Imperfect fungi or imperfect observation?

The label “Fungi Imperfecti” (also known as Deuteromycota) has traditionally been used for species for which sexual reproduction was never observed. However, this terminology may reflect the limitations of past observations rather than a strict biological reality (Dyer & Kück, 2017). Although rust fungi (Pucciniales) maintained their ability to undergo both sexual and

asexual reproduction (Duplessis et al., 2021b; Talhinhas et al., 2023), sexuality is considered the null hypothesis for this group (McTaggart et al., 2022). From this perspective, assuming asexuality as one of the dominant modes without comprehensive evidence undermines the logic of the baseline hypothesis. However, what appears as reproductive strategy exclusivity may be an artefact of the observational window, where time, space, and host availability shape our visibility. In some rust fungi species, sexual reproduction remained undetected for an extended period, as seen in the case of coffee rust caused by *Hemileia vastatrix* (Carvalho et al., 2011). Alternatively, in some populations, the contribution of asexuality might be underestimated due to only a few individuals being sampled. The chance of detecting asexuality in a natural population increases with a greater number of isolates analyzed through population genetic parameters (Halkett et al., 2005; Arnaud-Haond et al., 2007). In such cases, the likelihood of detecting a signal of asexuality might increase, as we observed with poplar rust (Abdalrahem et al., 2025). These contrasting cases demonstrate that both sexuality and asexuality can remain hidden, surfacing only when ecological conditions analysis or sampling strategies allow them to be seen.

This broader perspective became especially important in our experiments. When we studied asexual lineages of *M. larici-populina* under controlled laboratory conditions, we observed that some of them were able to complete the full sexual cycle once the appropriate host and conditions were provided. This suggests that what we perceive as asexual poplar rust in the natural population may not represent an irreversible loss, but rather a reservoir of frozen genotypes saved by favorable ecological conditions that can commit to sexual reproduction in the future. Nevertheless, when an asexual isolate undergoes sexual reproduction with a sexual isolate, its track is lost since it is recombined. Thus, only the fraction of the asexual lineage that overwinters asexually can be retrieved as an asexual lineage in the following season. Therefore, the asexual lineages of poplar rust that are stopped at specific stages and do not complete the full sexual cycle require more investigation. This plasticity in reproductive modes can be interpreted as a form of bet-hedging rather than mixed strategies because both modes coexist (Seger & Brockmann, 1987; X.-Y. Li et al., 2016). On one side, asexual reproduction allows the

fungus to preserve and “freeze” well-adapted gene combinations over time in more stable environmental conditions. On the other side, maintaining sexual potential provides opportunities for allele reshuffling and increasing genetic diversity. In this way, fungal species that can switch between two modes (e.g., poplar rust) distribute their reproductive investment across both strategies, ensuring continuity while maintaining the ability to adapt when conditions change. With this respect, reproduction mode polymorphism can have some analogy with dormancy as a bet-hedging strategy (Menu et al., 2000).

An additional layer of complexity arises not only from observation but also from interpretation. In population genetics, indices such as inbreeding coefficients are often used as indirect indicators of clonal reproduction within natural populations (Arnaud-Haond et al., 2007; Stoeckel et al., 2021). However, our theoretical simulation study of partially clonal populations across generations showed that the clonality rate and the population's evolutionary dynamics can drastically impact the variation of these indices. For example, when the clonality rate is ≥ 0.98, the values of genetic indices become highly sensitive and vary randomly depending on the mixture of young and old lineages within the same population (Becheler et al., 2017). This creates a diffusion point: depending on which lineages are sampled, the indices can show very different signals, sometimes indicating strong clonality (e.g., negative F_{IS} values) because of many old, long-lasting asexual lineages, and other results suggesting sexual recombination might occur due to the young asexual lineages. In such cases, the signal of reproductive mode is less a reflection of the true nature of the system and more a consequence of how lineage age and clonality rate influence population structure. What appears as strong clonality under one analysis scheme may, under another, suggest sexual recombination. The paradox may lie not in the fungus but in the way we read the data. From this perspective, imperfect observation is worsened by imperfect interpretation, highlighting the need for caution when inferring reproductive mode from comprehensive population genetic analysis.

Future studies on partially clonal populations (e.g., poplar rust) should not rely solely on population genetic indices to determine reproductive mode. Instead, these indices should be

supplemented with estimates of the clonality rate (Becheler et al., 2017), which offer an essential context for understanding the genetic signals. Currently, ClonEstiMate employs changes in allelic frequencies over time to infer clonality rates. However, incorporating changes in genotype frequencies as well would enhance the detection of mixed reproductive systems where sexual reproduction and long-lasting asexual lineages coexist (Becheler et al., 2017). This approach could better capture the admixture signatures characteristic of populations where clonal persistence occurs alongside occasional sexual events.

Beyond indices: The importance of experimental validation

Population genetic indices can indicate asexuality within the population, while bioinformatic analyses reveal the completeness of the genetic sexual machinery in fungi, suggesting that sexual reproduction still remains theoretically possible (Bakkeren & Kronstad, 1994; Doellman et al., 2024b). Yet what appears viable in silico does not always translate to in vivo situations, and this difference between potential and actual expression is exactly where caution is needed. Even when a genome contains the complete sexual toolkit, experimental validation can determine whether such fungi are truly capable of sex. In our work on poplar rust, for instance, some lineages reinvested in sexual reproduction under controlled conditions despite appearing asexual in nature. Similar work was done on wheat rust to examine the ability to undergo sexual reproduction in the lab because it was known as asexual (Rodriguez-Algaba et al., 2014). In contrast, others did not, underscoring the limits of genetic inference on its own. A recent study on *Lepraria* (lichenized fungal genus), a genus long regarded as strictly asexual, reached a similar conclusion: although functional mating-type loci were preserved, no sexual structures have ever been observed in nature (Doellman et al., 2024b). Together, these findings emphasize the importance of molecular analysis potential being grounded in experiments.

The presence of MAT loci in asexual fungi should not be taken as proof that they reproduce sexually. Recent evidence shows that these genes might also serve functions beyond mating, such as in spore formation or cell cycle regulation (Q. Wang et al., 2017). From this viewpoint, their retention remains meaningful even if a sexual process is not observed in nature.

Therefore, keeping MAT loci does not necessarily indicate hidden sexuality, but may instead signify other roles that are equally vital for fungal survival and evolution.

MAT expression equivalence in sexual and asexual isolates

In comparing sexual isolates with their asexual counterparts, we found that the MAT genes were largely conserved (synteny) in both representative isolates. The complete set of genes, including the homeodomain genes (HD1 and HD2) and the pheromone/pheromone receptor genes, was present in asexual isolates as in the sexual ones. This pattern was consistent across all seven asexual isolates we examined, pointing to a simple but important conclusion: at the level of genetic reproductive machinery, sexual and asexual representatives are not fundamentally different (Luo et al., 2024; Mojerlou, Moeller, et al., 2025). The lack of observed sex in natural populations may not be caused by the loss of this genetic structure (Rodriguez-Algaba et al., 2014). Instead, ecological conditions or regulatory mechanisms may determine whether sexual reproduction occurs.

Beyond conserved locus structure, across the poplar infection stages, we detected comparable expression profiles for HD1/HD2 and STE3-family receptors in sexual isolate 98AG31 and asexual isolate 08KE26, with similar patterns between both isolates (Fig. 1). This indicates that conserved MAT-pathway transcription is maintained irrespective of different evolutionary history of reproductive mode, suggesting that the absence of observed sex in asexual lineages may not originate from loss of the function of MAT genes under these conditions. Concordant patterns between both isolates support stage-associated regulation that is conserved across lineages, rather than being lineage-specific.

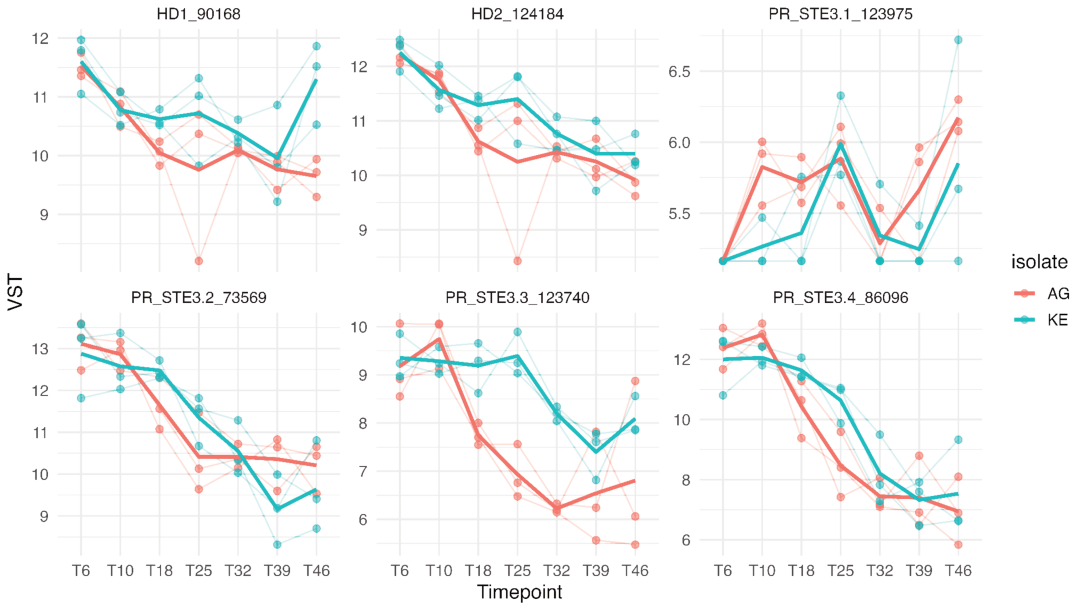


Fig. 1 MAT expression parity across sexual vs asexual isolates at different developmental stages. VST-normalized expression trajectories for MAT-pathway genes HD1, HD2, and STE3-family receptors across infection timepoints (T6–T46) on poplar leaves, comparing 98AG31 (sexual, red, termed AG) and 08KE26 (asexual, blue, termed KE); bold lines show group means with lighter lines representing biological replicates.

The expression of MAT genes can be observed from the early stages of infection through the late development of telia in poplar, which occurs beyond the typical mating functions on the alternate host (Fig. 1). These expression profiles suggest that the MAT pathway remains active during clonal multiplication stages. These findings align with previous studies where some MAT gene expression was detected during asexual stages in telia of *M. larici-populina* and *Puccinia triticina*, and where silencing HD genes in *P. triticina* impaired urediniospore production during infection, collectively supporting pleiotropic developmental roles beyond mating (Duplessis, Hacquard, et al., 2011; Hacquard et al., 2013; Cuomo et al., 2017; Luo et al., 2024). Additionally, expression network analyses of the smut fungus Ustilago maydis have revealed that MAT genes emerged as drivers of plant surface recognition and biotrophy establishment modules, suggesting a regulatory role that extends to effector expression and nutrient acquisition (Lanver et al., 2018). For future work with the new version of annotation of *M. larici-populina*, we will be able to build expression pathways from transcriptome data obtained across different infection stages. This could help position MAT genes within a broader regulatory role,

potentially revealing whether asexual lineages retain the full MAT-associated transcriptional network.

Considering the conserved structure and expression of the MAT gene across both sexual and asexual lineages in the examined stages, I propose that environmental triggers, particularly nutritional limitation stress, may be the key trigger for the transition from clonal multiplication to sexual reproduction in *M. larici-populina*. This hypothesis draws support from numerous reported studies of nutrient limitation as a key inducer of sexual development across many fungal taxa. For instance, nitrogen deficit triggers MAT gene expression and mating in *Saccharomyces cerevisiae*, while combined nutrient starvation and pheromone availability induce sexual development in *Schizosaccharomyces pombe*. Recently, nutritional stress was shown to enable mating in previously 'sterile' white cells of *Candida albicans* (Fukui et al., 1986; Wallen & Perlin, 2018b; Zhong et al., 2025). In *M. larici-populina*, specifically, the transition from uredinial to telial stages, where sexual processes initiate through nuclear fusion forming karyogamy, often co-occurs with host senescence and declining nutrient availability. I therefore propose that controlled starvation experiments, to induce sexual development across fungal taxa, could test whether dikaryotic stages in the poplar rust uredia stage in isolates retain the capacity to transition to the karyogamy stage as an initial step for sexual reproduction.

Despite the obligate biotrophy of *M. larici-populina* (Duplessis et al., 2011; Lorrain et al., 2019), previous reports on axenic media-based cultivation of rust fungi, including attempts with *Melampsora* spp., suggest that growth and even sporulation are sometimes possible under carefully optimized conditions (Wolf, 1974; Krzan, 1981). However, results vary across species and media. In line with this, a preliminary plant-based cultivation study, conducted according to the protocol (Nemr et al., 2020) during my PhD, did not establish growth, highlighting the experimental difficulty. Future experiments could systematically test this hypothesis by cultivating rust fungi under graduated nutrient limitation protocols. These methods could help determine whether the sexual ability observed in the tests is due to nutritional gating of conserved sexual machinery or if it is due to specific limitations within certain lineages.

Diversity in Mating-Type genes

What our results highlight is the importance of moving beyond single reference genomes toward population-level analyses. When we compared the sexual reference isolate 98AG31 with the representative asexual isolate 08KE26 using phased genomes, we found striking similarities in the organization of the mating-type (MAT) genes. Each haplotype contained two copies of the homeodomain transcription factors, which we designate as HD1.1 and HD1.2 for the bW-HD1 locus and HD2.1 and HD2.2 for the bE-HD2 locus. Remarkably, these copies were highly conserved between sexual and asexual isolates; for example, HD1.1 in the sexual isolate closely matched HD1.1 in the asexual one, and the same was true for HD2 variants. This pattern is not unique to *M. larici-populina*: other phased genomes of Pucciniales show similar results, typically reporting two HD1 and two HD2 genes per dikaryotic isolate (e.g., *P. triticina*, *P. striiformis* f. sp. *tritici, A. psidii, M. lini*) (Cuomo et al., 2017; Edwards et al., 2022; Luo et al., 2024; Sperschneider et al., 2025). Such findings reflect the inherent structure of dikaryotic genomes, where each haplotype contributes one allele per HD locus. However, as more isolates are examined at the population level, it becomes increasingly clear that this simple “two-alleles” view underestimates the true diversity present across natural populations, which is shown in recent studies to reveal many HD alleles as multiallelic within the four cereal rust fungi (Luo et al., 2024).

When we began this study, comprehensive population-level studies of HD diversity in rust fungi were scarce (Holden, Bakkeren, Hubensky, Bamrah, Abbasi, Qutob, Graaf, et al., 2023). At that time, the prevailing view, mainly based on a few reference genomes, was that each species carried only two allelic versions of the HD1 and HD2 genes. Our analyses, however, revealed a broader allelic diversity, with multiple HD1 and HD2 variants circulating within natural populations of *M. larici-populina*. Meanwhile, other groups also reported similar findings in other rust fungi, noting allelic variation at the HD loci. (Luo et al., 2024; Gomez-Zapata et al., 2025). The convergence of these independent studies strengthens the conclusion that multiallelism at mating-type loci is a common and evolutionarily important characteristic of rust fungi.

It is important to note that our contribution was developed in parallel with these efforts and relied on independent datasets and methods. Additionally, our work offered direct population-level evidence in *M. larici-populina*. The fact that different research groups studying various rust fungi species have now reached similar conclusions highlights the robustness of this finding. Overall, these results represent a shift from the simplistic two-allele view of reference genomes toward a more detailed understanding of mating-type diversity across populations.

Expression programs reflect reproductive history

Reproductive mode polymorphism represents a fundamental evolutionary strategy across the Pucciniales. Our comparative transcriptomic analyses of *M. larici-populina* revealed differences in gene expression program between sexual and asexual lineages, positioning these findings within a growing understanding of how rust fungi balance sexual recombination with clonal multiplication to gain evolutionary success. Our demonstration of constitutive transcriptional differences in *M. larici-populina* adds a new dimension to this understanding, indicating that the history of reproductive evolution creates molecular imprints as distinct regulatory programs, as shown between sexual and asexual lineages. This regulatory difference may help explain why asexual lineages persist across rust species.

Comparative transcriptome analyses across developmental stages between sexual and asexual lineages revealed that sexual lineages consistently expressed a broad set of genes across all stages, a pattern absent in the asexual lineage. Conversely, specific genes were active throughout the cycle in the asexual isolate but not in the sexual one. Similar studies have also observed divergent expression programs across developmental stages in early research on poplar rust (Duplessis, Hacquard, et al., 2011; Hacquard et al., 2013) and other rust fungi, such as wheat stripe rust (*Puccinia striiformis* f. sp. *tritici*), which exhibit different programs for different spore stages (Zhan et al., 2023).

A recent study in rust genomics, with comparative genomics of cereal rusts, reveals that mating-type (MAT) loci are largely conserved across species, with HD loci exhibiting multiallelic structures (which aligns with our result on poplar rust) (Luo et al., 2024). However, our transcriptomic findings complement these genomic insights by indicating that reproductive mode differences extend beyond MAT loci alone. This interpretation aligns with recent observations in *P. striiformis* f. sp. *tritici*, in which asexual lineages show reduced ability to produce teliospores compared to sexually reproducing populations, potentially indicating evolved constraints on sexual competence (Schwessinger et al., 2020).

Another limitation of our study relates to the resources available in the reference genome annotation. Our analyses were based on version 2 of the *M. larici-populina* reference genome released by JGI. Whereas this resource has been invaluable, it is still incomplete and in some cases presents uncertain annotations. This has limited our ability to confidently interpret certain genes that appeared specific to sexual or asexual lineages, as well as genes expressed at particular developmental stages. The release of version 3, with improved annotation and functional details, will be an essential step toward overcoming this limitation and is greatly awaited. Incorporating these enhanced resources in future work will allow for more accurate functional assignments to determine the genetic programs that shape reproductive diversity in poplar rust.

A more resolved time series, especially during the early hours of host infection, will be essential to capture transient transcriptional shifts that our seven-day sampling intervals may have missed. We therefore propose continuous cycle infection trials on both poplar and larch, with dense sampling at early infection (e.g., 0, 6, 12, 24, 48, 72 hours post-inoculation) followed by sampling every two days. These high-resolution time courses between hosts can reveal host-specific regulatory checkpoints and clarify whether expression shifts that trigger sexual commitment on larch are influenced by events during poplar infection.

Opportunities and challenges

A key challenge that emerges from this work is explaining why, within the same population and geographic setting, some isolates reproduce sexually while others persist clonally. In the poplar rust population (mainly in Southern France), both lineages retain the complete set of examined MAT genes required for sexual reproduction, yet their reproductive strategies differ. This raises an evolutionary question related to the core question of sex evolution: Why do some lineages maintain sex, while others are asexual, even when they coexist in the same environment? Addressing this will require careful analysis of ecological conditions, gene expression dynamics, and evolutionary history.

Our results concluded that asexuality in poplar rust may be due to the loss of the ability to reproduce sexually. Instead, it might reflect a coexistence of both reproductive strategies in the same natural populations; therefore, this offers a good model to study the evolution of sex in a system where both strategies are present under comparable conditions. Insights from this model are likely to contribute not only to the understanding of rust fungi but also to broader discussions about the maintenance of sexual reproduction.

Another challenge is due to our interpretation of the differences in gene expression programs we observed between sexual and asexual isolates. At present, it is uncertain whether these patterns reflect differences in the evolutionary trajectories of reproductive modes or lineage-specific variations unrelated to reproductive mode polymorphism. In other words, the variation we detected may be influenced as much by population history as by the biology of sex itself. Addressing this issue will require a more comprehensive sampling of both kinds of lineages from different geographic regions. Such comparative analyses would make it possible to separate the effects of evolutionary history from those of local adaptation and clarify whether divergent transcriptional programs are linked to reproductive strategy or are instead shaped by demographic and geographic structures within rust fungi populations.

During my PhD work, we assembled and analyzed numerous fungal mitochondrial genomes, with a particular focus on poplar rust fungi (see Appendix I). This resource provides a good foundation for population genetic studies at the species level. The mitochondrial genome was overlooked, especially in studies of rust species. It is worth trying to use it for tracking lineage, assessing evolutionary history, and distinguishing genetic structure among isolates within *M. larici-populina* and related Pucciniales. With the new work that increases the availability of complete fungal mitochondrial genomes, future studies are well positioned to leverage this "mitochondrial treasure" for lineage tracking and testing of evolutionary hypotheses about closely related species.

Bibliography

Abdalrahem, A., Andrieux, A., Becheler, R., Duplessis, S., Frey, P., Marçais, B., Schiffer-Forsyth, K., Stoeckel, S., & Halkett, F. (2025). *Long-lasting coexistence of multiple asexual lineages alongside their sexual counterparts in a fungal plant pathogen* (p. 2025.03.28.645883). bioRxiv. https://doi.org/10.1101/2025.03.28.645883

Abdalrahem, A., Duplessis, S., Frey, P., Stoeckel, S., & Halkett, F. (2025). Behind the scenes of a shadow play: Illuminating the population genetics of partially clonal organisms through the prism of clone age distribution. *Preprint*.

Agrawal, A. F. (2006). Evolution of Sex: Why Do Organisms Shuffle Their Genotypes? *Current Biology*, *16*(17), R696–R704. https://doi.org/10.1016/j.cub.2006.07.063

Aime, M. C., Bell, C. D., & Wilson, A. W. (2018). Deconstructing the evolutionary complexity between rust fungi (Pucciniales) and their plant hosts. *Studies in Mycology*, *89*, 143–152. https://doi.org/10.1016/j.simyco.2018.02.002

Aime, M. C., & McTaggart, A. R. (2021). A higher-rank classification for rust fungi, with notes on genera. *Fungal Systematics and Evolution*, *7*, 21–47. https://doi.org/10.3114/fuse.2021.07.02

Ali, S., Gladieux, P., Leconte, M., Gautier, A., Justesen, A. F., Hovmøller, M. S., Enjalbert, J., & Vallavieille-Pope, C. de. (2014). Origin, Migration Routes and Worldwide Population Genetic Structure of the Wheat Yellow Rust Pathogen Puccinia striiformis f.sp. Tritici. *PLOS Pathogens*, *10*(1), e1003903. https://doi.org/10.1371/journal.ppat.1003903

Ali, S., Gladieux, P., Rahman, H., Saqib, M. S., Fiaz, M., Ahmad, H., Leconte, M., Gautier, A., Justesen, A. F., Hovmøller, M. S., Enjalbert, J., & de Vallavieille-Pope, C. (2014). Inferring the contribution of sexual reproduction, migration and off-season survival to the temporal maintenance of microbial populations: A case study on the wheat fungal pathogen uccinia striiformis f.sp. tritici. *Molecular Ecology*, *23*(3), 603–617. https://doi.org/10.1111/mec.12629

Arnaud-Haond, S., Duarte, C. M., Alberto, F., & Serrão, E. A. (2007). Standardizing methods to address clonality in population studies. *Molecular Ecology*, *16*(24), 5115–5139. https://doi.org/10.1111/j.1365-294X.2007.03535.x

Bakkeren, G., & Kronstad, J. W. (1994). Linkage of mating-type loci distinguishes bipolar from tetrapolar mating in basidiomycetous smut fungi. *Proceedings of the National Academy of Sciences*, *91*(15), 7085–7089. https://doi.org/10.1073/pnas.91.15.7085

Balloux, F., Lehmann, L., & de Meeûs, T. (2003). The Population Genetics of Clonal and Partially Clonal Diploids. *Genetics*, *164*(4), 1635–1644. https://doi.org/10.1093/genetics/164.4.1635

Barrès, B., Halkett, F., Dutech, C., Andrieux, A., Pinon, J., & Frey, P. (2008). Genetic structure of the poplar rust fungus Melampsora larici-populina: Evidence for isolation by distance in Europe and recent founder effects overseas. *Infection, Genetics and Evolution*, *8*(5), 577–587. https://doi.org/10.1016/j.meegid.2008.04.005

Barton, N. H., & Charlesworth, B. (1998). Why Sex and Recombination? *Science*, *281*(5385), 1986–1990. https://doi.org/10.1126/science.281.5385.1986

Bast, J., Parker, D. J., Dumas, Z., Jalvingh, K. M., Tran Van, P., Jaron, K. S., Figuet, E., Brandt, A., Galtier, N., & Schwander, T. (2018). Consequences of Asexuality in Natural Populations: Insights from Stick Insects. *Molecular Biology and Evolution*, *35*(7), 1668–1677. https://doi.org/10.1093/molbev/msy058

Bast, J., Schaefer, I., Schwander, T., Maraun, M., Scheu, S., & Kraaijeveld, K. (2016). No Accumulation of Transposable Elements in Asexual Arthropods. *Molecular Biology and Evolution*, *33*(3), 697–706. https://doi.org/10.1093/molbev/msv261

Beadle, G. W., & Coonradt, V. L. (1944). HETEROCARYOSIS IN NEUROSPORA CRASSA. *Genetics*, *29*(3), 291–308. https://doi.org/10.1093/genetics/29.3.291

Becheler, R., Masson, J.-P., Arnaud-Haond, S., Halkett, F., Mariette, S., Guillemin, M.-L., Valero, M., Destombe, C., & Stoeckel, S. (2017). ClonEstiMate, a Bayesian method for quantifying rates of clonality of populations genotyped at two-time steps. *Molecular Ecology Resources*, *17*(6), e251–e267. https://doi.org/10.1111/1755-0998.12698

Bell, G. (1983). *The Evolution of Sex:The Masterpiece of Nature: The Evolution and Genetics of Sexuality* (Vol. 58).

Bian, Z., Xu, Z., Peer, A., Choi, Y., Priest, S. J., Akritidou, K., Dasgupta, A., Dahlmann, T. A., Kück, U., Nowrousian, M., Sachs, M. S., Sun, S., & Heitman, J. (2025). Essential genes encoded by the mating-type locus of the human fungal pathogen Cryptococcus neoformans. *mBio*, *16*(4), e00223-25. https://doi.org/10.1128/mbio.00223-25

Billiard, S., López-Villavicencio, M., Devier, B., Hood, M. E., Fairhead, C., & Giraud, T. (2011). Having sex, yes, but with whom? Inferences from fungi on the evolution of anisogamy

and mating types. *Biological Reviews*, *86*(2), 421–442. https://doi.org/10.1111/j.1469-185X.2010.00153.x

Billiard, S., López-Villavicencio, M., Hood, M. E., & Giraud, T. (2012). Sex, outcrossing and mating types: Unsolved questions in fungi and beyond. *Journal of Evolutionary Biology*, *25*(6), 1020–1038. https://doi.org/10.1111/j.1420-9101.2012.02495.x

Boixel, A., Chelle, M., & Suffert, F. (2022). Patterns of thermal adaptation in a globally distributed plant pathogen: Local diversity and plasticity reveal two-tier dynamics. *Ecology and Evolution*, *12*(1), e8515. https://doi.org/10.1002/ece3.8515

Bolton, M. D., Kolmer, J. A., & Garvin, D. F. (2008). Wheat leaf rust caused by Puccinia triticina. *Molecular Plant Pathology*, *9*(5), 563–575. https://doi.org/10.1111/j.1364-3703.2008.00487.x

Calo, S., Billmyre, R. B., & Heitman, J. (2013). Generators of Phenotypic Diversity in the Evolution of Pathogenic Microorganisms. *PLOS Pathogens*, *9*(3), e1003181. https://doi.org/10.1371/journal.ppat.1003181

Camacho, C., Coulouris, G., Avagyan, V., Ma, N., Papadopoulos, J., Bealer, K., & Madden, T. L. (2009). BLAST+: Architecture and applications. *BMC Bioinformatics*, *10*(1), 421. https://doi.org/10.1186/1471-2105-10-421

Capella-Gutiérrez, S., Silla-Martínez, J. M., & Gabaldón, T. (2009). trimAl: A tool for automated alignment trimming in large-scale phylogenetic analyses. *Bioinformatics*, *25*(15), 1972–1973. https://doi.org/10.1093/bioinformatics/btp348

Carvalho, C. R., Fernandes, R. C., Carvalho, G. M. A., Barreto, R. W., & Evans, H. C. (2011). Cryptosexuality and the Genetic Diversity Paradox in Coffee Rust, Hemileia vastatrix. *PLOS ONE*, *6*(11), e26387. https://doi.org/10.1371/journal.pone.0026387

Casselton, L. A., & Olesnicky, N. S. (1998). Molecular Genetics of Mating Recognition in Basidiomycete Fungi. *Microbiology and Molecular Biology Reviews*, *62*(1), 55–70. https://doi.org/10.1128/mmbr.62.1.55-70.1998

Chen, S. (2023). Ultrafast one-pass FASTQ data preprocessing, quality control, and deduplication using fastp. *iMeta*, *2*(2), e107. https://doi.org/10.1002/imt2.107

Chen, S., Zhou, Y., Chen, Y., & Gu, J. (2018). fastp: An ultra-fast all-in-one FASTQ preprocessor. *Bioinformatics*, *34*(17), i884–i890. https://doi.org/10.1093/bioinformatics/bty560

Coelho, M. A., Bakkeren, G., Sun, S., Hood, M. E., & Giraud, T. (2017). FUNGAL SEX: THE BASIDIOMYCOTA. *Microbiology Spectrum*, *5*(3), 10.1128/microbiolspec.FUNK-0046–2016. https://doi.org/10.1128/microbiolspec.FUNK-0046-2016

Coelho, M. A., Sampaio, J. P., & Gonçalves, P. (2010). A Deviation from the Bipolar-Tetrapolar Mating Paradigm in an Early Diverged Basidiomycete. *PLOS Genetics*, *6*(8), e1001052. https://doi.org/10.1371/journal.pgen.1001052

Combosch, D. J., & Vollmer, S. V. (2013). Mixed asexual and sexual reproduction in the Indo-Pacific reef coral Pocillopora damicornis. *Ecology and Evolution*, *3*(10), 3379–3387. https://doi.org/10.1002/ece3.721

Corre, E. (2025). *Invasion of Pucciniales genomes by transposable elements: How transposition events shape the evolution and the genomic architecture in rust fungi*.

Cuomo, C. A., Bakkeren, G., Khalil, H. B., Panwar, V., Joly, D., Linning, R., Sakthikumar, S., Song, X., Adiconis, X., Fan, L., Goldberg, J. M., Levin, J. Z., Young, S., Zeng, Q., Anikster, Y., Bruce, M., Wang, M., Yin, C., McCallum, B., … Fellers, J. P. (2017). Comparative Analysis Highlights Variable Genome Content of Wheat Rusts and Divergence of the Mating Loci. *G3 Genes|Genomes|Genetics*, *7*(2), 361–376. https://doi.org/10.1534/g3.116.032797

de Meeûs, T., Prugnolle, F., & Agnew, P. (2007). Asexual reproduction: Genetics and evolutionary aspects. *Cellular and Molecular Life Sciences*, *64*(11), 1355–1372. https://doi.org/10.1007/s00018-007-6515-2

Dobin, A., Davis, C. A., Schlesinger, F., Drenkow, J., Zaleski, C., Jha, S., Batut, P., Chaisson, M., & Gingeras, T. R. (2013). STAR: Ultrafast universal RNA-seq aligner. *Bioinformatics*, *29*(1), 15–21. https://doi.org/10.1093/bioinformatics/bts635

Dodge, B. O. (1936). Facultative and Obligate Heterothallism in Ascomycetes. *Mycologia*, *28*(5), 399–409. https://doi.org/10.2307/3754113

Doellman, M. M., Sun, Y., Barcenas-Peña, A., Lumbsch, H. T., & Grewe, F. (2024a). *Rethinking asexuality: The enigmatic case of functional sexual genes in Lepraria (Stereocaulaceae)* (p. 2024.06.11.598483). bioRxiv. https://doi.org/10.1101/2024.06.11.598483

Doellman, M. M., Sun, Y., Barcenas-Peña, A., Lumbsch, H. T., & Grewe, F. (2024b). Rethinking asexuality: The enigmatic case of functional sexual genes in Lepraria (Stereocaulaceae). *BMC Genomics*, *25*(1), 1003. https://doi.org/10.1186/s12864-024-10898-8

Drenth, A., McTaggart, A. R., & Wingfield, B. D. (2019). Fungal clones win the battle, but recombination wins the war. *IMA Fungus*, *10*(1), 18. https://doi.org/10.1186/s43008-019-0020-8

Du, Z., Peng, Y., Zhang, G., Chen, L., Jiang, S., Kang, Z., & Zhao, J. (2023). Direct Evidence Demonstrates that Puccinia striiformis f. Sp. Tritici Infects Susceptible Barberry to Complete Sexual Cycle in Autumn. *Plant Disease*, *107*(3), 771–783. https://doi.org/10.1094/PDIS-08-22-1750-RE

Duplessis, S., Cuomo, C. A., Lin, Y.-C., Aerts, A., Tisserant, E., Veneault-Fourrey, C., Joly, D. L., Hacquard, S., Amselem, J., Cantarel, B. L., Chiu, R., Coutinho, P. M., Feau, N., Field, M., Frey, P., Gelhaye, E., Goldberg, J., Grabherr, M. G., Kodira, C. D., … Martin, F. (2011). Obligate biotrophy features unraveled by the genomic analysis of rust fungi. *Proceedings of the National Academy of Sciences*, *108*(22), 9166–9171. https://doi.org/10.1073/pnas.1019315108

Duplessis, S., Hacquard, S., Delaruelle, C., Tisserant, E., Frey, P., Martin, F., & Kohler, A. (2011). Melampsora larici-populina Transcript Profiling During Germination and Timecourse Infection of Poplar Leaves Reveals Dynamic Expression Patterns Associated with Virulence and Biotrophy. *Molecular Plant-Microbe Interactions®*, *24*(7), 808–818. https://doi.org/10.1094/MPMI-01-11-0006

Duplessis, S., Lorrain, C., Petre, B., Figueroa, M., Dodds, P. N., & Aime, M. C. (2021a). Host Adaptation and Virulence in Heteroecious Rust Fungi. *Annual Review of Phytopathology*, *59*(1), 403–422. https://doi.org/10.1146/annurev-phyto-020620-121149

Duplessis, S., Lorrain, C., Petre, B., Figueroa, M., Dodds, P. N., & Aime, M. C. (2021b). Host Adaptation and Virulence in Heteroecious Rust Fungi. *Annual Review of Phytopathology*, *59*(1), 403–422. https://doi.org/10.1146/annurev-phyto-020620-121149

Dyer, P. S., & Kück, U. (2017). Sex and the Imperfect Fungi. *Microbiology Spectrum*, *5*(3), 10.1128/microbiolspec.funk-0043–2017. https://doi.org/10.1128/microbiolspec.funk-0043-2017

Edwards, R. J., Dong, C., Park, R. F., & Tobias, P. A. (2022). *A phased chromosome-level genome and full mitochondrial sequence for the dikaryotic myrtle rust pathogen, Austropuccinia psidii* (p. 2022.04.22.489119). bioRxiv. https://doi.org/10.1101/2022.04.22.489119

Fabre, B., Bastien, C., Husson, C., Marcais, B., Frey, P., & Halkett, F. (2021). *Un effet papillon dans les peupleraies françaises: Les répercussions d'un contournement de résistance sur les méthodes de sélection variétale* (pp. 329–339).

Fowler, T. J., DeSimone, S. M., Mitton, M. F., Kurjan, J., & Raper, C. A. (1999). Multiple Sex Pheromones and Receptors of a Mushroom-producing Fungus Elicit Mating in Yeast. *Molecular Biology of the Cell*, *10*(8), 2559–2572. https://doi.org/10.1091/mbc.10.8.2559

Freytag, S., Bruscaglioni, L., Gold, R. E., & Mendgen, K. (1988). Basidiospores of rust fungi (*Uromyces* species) differentiate infection structures*in vitro*. *Experimental Mycology*, *12*(3), 275–283. https://doi.org/10.1016/0147-5975(88)90042-4

Fu, C., Coelho, M. A., David-Palma, M., Priest, S. J., & Heitman, J. (2019). Genetic and genomic evolution of sexual reproduction: Echoes from LECA to the fungal kingdom. *Current Opinion in Genetics & Development*, *58–59*, 70–75. https://doi.org/10.1016/j.gde.2019.07.008

Fu, L., Niu, B., Zhu, Z., Wu, S., & Li, W. (2012). CD-HIT: Accelerated for clustering the next-generation sequencing data. *Bioinformatics*, *28*(23), 3150–3152. https://doi.org/10.1093/bioinformatics/bts565

Fukui, Y., Kaziro, Y., & Yamamoto, M. (1986). Mating pheromone-like diffusible factor released by Schizosaccharomyces pombe. *The EMBO Journal*, *5*(8), 1991–1993. https://doi.org/10.1002/j.1460-2075.1986.tb04454.x

Galili, T. (2015). dendextend: An R package for visualizing, adjusting and comparing trees of hierarchical clustering. *Bioinformatics*, *31*(22), 3718–3720. https://doi.org/10.1093/bioinformatics/btv428

Gehring, W. J., Qian, Y. Q., Billeter, M., Furukubo-Tokunaga, K., Schier, A. F., Resendez-Perez, D., Affolter, M., Otting, G., & Wüthrich, K. (1994). Homeodomain-DNA recognition. *Cell*, *78*(2), 211–223. https://doi.org/10.1016/0092-8674(94)90292-5

Gioti, A., Mushegian, A. A., Strandberg, R., Stajich, J. E., & Johannesson, H. (2012). Unidirectional Evolutionary Transitions in Fungal Mating Systems and the Role of Transposable Elements. *Molecular Biology and Evolution*, *29*(10), 3215–3226. https://doi.org/10.1093/molbev/mss132

Giraud, T., Yockteng, R., López-Villavicencio, M., Refrégier, G., & Hood, M. E. (2008). Mating System of the Anther Smut Fungus Microbotryum violaceum: Selfing under Heterothallism. *Eukaryotic Cell*, *7*(5), 765–775. https://doi.org/10.1128/EC.00440-07

Glass, N. L., Vollmer, S. J., Staben, C., Grotelueschen, J., Metzenberg, R. L., & Yanofsky, C. (1988). DNAs of the Two Mating-Type Alleles of Neurospora crassa Are Highly Dissimilar. *Science*, *241*(4865), 570–573. https://doi.org/10.1126/science.2840740

Gomez-Zapata, P. A., Tellgren-Roth, C., Samils, B., Stenlid, J., Kaitera, J., Durling, M. B., Seidl, M. F., & Olson, Å. (2025). *Mating Type Gene Divergence is Associated with Life Cycle*

Differentiation in Scots Pine Blister Rust (p. 2025.05.02.651906). bioRxiv. https://doi.org/10.1101/2025.05.02.651906

Goodenough, U., & Heitman, J. (2014). Origins of Eukaryotic Sexual Reproduction. *Cold Spring Harbor Perspectives in Biology*, *6*(3), a016154. https://doi.org/10.1101/cshperspect.a016154

Goodstein, D. M., Shu, S., Howson, R., Neupane, R., Hayes, R. D., Fazo, J., Mitros, T., Dirks, W., Hellsten, U., Putnam, N., & Rokhsar, D. S. (2012). Phytozome: A comparative platform for green plant genomics. *Nucleic Acids Research*, *40*(D1), D1178–D1186. https://doi.org/10.1093/nar/gkr944

Grigoriev, I. V., Nikitin, R., Haridas, S., Kuo, A., Ohm, R., Otillar, R., Riley, R., Salamov, A., Zhao, X., Korzeniewski, F., Smirnova, T., Nordberg, H., Dubchak, I., & Shabalov, I. (2014). MycoCosm portal: Gearing up for 1000 fungal genomes. *Nucleic Acids Research*, *42*(Database issue), D699–D704. https://doi.org/10.1093/nar/gkt1183

Grognet, P., & Silar, P. (2015). Maintaining heterokaryosis in pseudo-homothallic fungi. *Communicative & Integrative Biology*, *8*(4), e994382. https://doi.org/10.4161/19420889.2014.994382

Hacquard, S., Delaruelle, C., Frey, P., Tisserant, E., Kohler, A., & Duplessis, S. (2013). Transcriptome analysis of poplar rust telia reveals overwintering adaptation and tightly coordinated karyogamy and meiosis processes. *Frontiers in Plant Science*, *4*. https://www.frontiersin.org/articles/10.3389/fpls.2013.00456

Hacquard, S., Petre, B., Frey, P., Hecker, A., Rouhier, N., & Duplessis, S. (2011). The Poplar-Poplar Rust Interaction: Insights from Genomics and Transcriptomics. *Journal of Pathogens*, *2011*, 716041. https://doi.org/10.4061/2011/716041

Halkett, F., Simon, J.-C., & Balloux, F. (2005). Tackling the population genetics of clonal and partially clonal organisms. *Trends in Ecology & Evolution*, *20*(4), 194–201. https://doi.org/10.1016/j.tree.2005.01.001

Hartfield, M., & Keightley, P. D. (2012). Current hypotheses for the evolution of sex and recombination. *Integrative Zoology*, *7*(2), 192–209. https://doi.org/10.1111/j.1749-4877.2012.00284.x

Hartmann, F. E., Duhamel, M., Carpentier, F., Hood, M. E., Foulongne-Oriol, M., Silar, P., Malagnac, F., Grognet, P., & Giraud, T. (2021). Recombination suppression and evolutionary strata around mating-type loci in fungi: Documenting patterns and understanding evolutionary and mechanistic causes. *New Phytologist*, *229*(5), 2470–2491. https://doi.org/10.1111/nph.17039

Heitman, J. (2010). Evolution of Eukaryotic Microbial Pathogens via Covert Sexual Reproduction. *Cell Host & Microbe*, *8*(1), 86–99. https://doi.org/10.1016/j.chom.2010.06.011

Heitman, J. (2015). Evolution of sexual reproduction: A view from the fungal kingdom supports an evolutionary epoch with sex before sexes. *Fungal Biology Reviews*, *29*(3), 108–117. https://doi.org/10.1016/j.fbr.2015.08.002

Henningsen, E. C., Lewis, D., Nazareno, E., Huang, Y.-F., Steffenson, B. J., Boesen, B., Kianian, S. F., Stone, E., Dodds, P. N., Sperschneider, J., & Figueroa, M. (2024). *A high-resolution haplotype pangenome uncovers somatic hybridization, recombination and*

intercontinental migration in oat crown rust (p. 2024.03.27.583983). bioRxiv. https://doi.org/10.1101/2024.03.27.583983

Hojsgaard, D., & Hörandl, E. (2015). A little bit of sex matters for genome evolution in asexual plants. *Frontiers in Plant Science*, *6*. https://doi.org/10.3389/fpls.2015.00082

Holden, S., Bakkeren, G., Hubensky, J., Bamrah, R., Abbasi, M., Qutob, D., de Graaf, M.-L., Kim, S. H., Kutcher, H. R., McCallum, B. D., Randhawa, H. S., Iqbal, M., Uloth, K., Burlakoti, R. R., & Brar, G. S. (2023). Uncovering the history of recombination and population structure in western Canadian stripe rust populations through mating type alleles. *BMC Biology*, *21*(1), 233. https://doi.org/10.1186/s12915-023-01717-9

Holden, S., Bakkeren, G., Hubensky, J., Bamrah, R., Abbasi, M., Qutob, D., Graaf, M.-L. de, Kim, S. H., Kutcher, H. R., McCallum, B. D., Randhawa, H. S., Iqbal, M., Uloth, K., Burlakoti, R., & Brar, G. S. (2023). *Uncovering the history of recombination and population structure in western Canadian stripe rust populations through mating-type alleles* (p. 2023.03.30.534825). bioRxiv. https://doi.org/10.1101/2023.03.30.534825

Hörandl, E. (2009). A combinational theory for maintenance of sex. *Heredity*, *103*(6), 445–457. https://doi.org/10.1038/hdy.2009.85

Hoshino, M., Cossard, G., Haas, F. B., Kane, E. I., Kogame, K., Jomori, T., Wakimoto, T., Glemin, S., & Coelho, S. M. (2024). Parallel loss of sexual reproduction in field populations of a brown alga sheds light on the mechanisms underlying the emergence of asexuality. *Nature Ecology & Evolution*, *8*(10), 1916–1932. https://doi.org/10.1038/s41559-024-02490-w

Hovmøller, M. S., Walter, S., Bayles, R. A., Hubbard, A., Flath, K., Sommerfeldt, N., Leconte, M., Czembor, P., Rodriguez-Algaba, J., Thach, T., Hansen, J. G., Lassen, P., Justesen, A. F., Ali, S., & de Vallavieille-Pope, C. (2016). Replacement of the European wheat yellow rust population by new races from the centre of diversity in the near-Himalayan region. *Plant Pathology*, *65*(3), 402–411. https://doi.org/10.1111/ppa.12433

Hsueh, Y.-P., Xue, C., & Heitman, J. (2007). G protein signaling governing cell fate decisions involves opposing Gα subunits in Cryptococcus neoformans. *Molecular Biology of the Cell*, *18*(9), 3237–3249. https://doi.org/10.1091/mbc.E07-02-0133

Jiang, X., Xu, Z., Li, J., Shi, Y., Wu, W., & Tao, S. (2011). The influence of deleterious mutations on adaptation in asexual populations. *PloS One*, *6*(11), e27757. https://doi.org/10.1371/journal.pone.0027757

Ju, J.-Q., Fissore, R. A., & Sun, S.-C. (2023). Editorial: Editors' showcase 2022: insights in molecular and cellular reproduction. *Frontiers in Cell and Developmental Biology*, *11*. https://doi.org/10.3389/fcell.2023.1321358

Kaffarnik, F., Müller, P., Leibundgut, M., Kahmann, R., & Feldbrügge, M. (2003). PKA and MAPK phosphorylation of Prf1 allows promoter discrimination in Ustilago maydis. *The EMBO Journal*, *22*(21), 5817–5826. https://doi.org/10.1093/emboj/cdg554

Kaishian, P., Layug, C. R. K., Anderson, M., Berg, D. R., & Aime, M. C. (2024). Rust HUBB: DNA barcode-based identification of Pucciniales. *IMA Fungus*, *15*(1), 3. https://doi.org/10.1186/s43008-023-00132-7

Kalyaanamoorthy, S., Minh, B. Q., Wong, T. K. F., von Haeseler, A., & Jermiin, L. S. (2017). ModelFinder: Fast model selection for accurate phylogenetic estimates. *Nature Methods*, *14*(6), 587–589. https://doi.org/10.1038/nmeth.4285

Kämper, J., Reichmann, M., Romeis, T., Bölker, M., & Kahmann, R. (1995). Multiallelic recognition: Nonself-dependent dimerization of the bE and bW homeodomain proteins in ustilago maydis. *Cell*, *81*(1), 73–83. https://doi.org/10.1016/0092-8674(95)90372-0

Katoh, K., & Standley, D. M. (2013). MAFFT Multiple Sequence Alignment Software Version 7: Improvements in Performance and Usability. *Molecular Biology and Evolution*, *30*(4), 772–780. https://doi.org/10.1093/molbev/mst010

Kobayashi, K., & Hoshi, M. (2011). Sex-inducing effect of a hydrophilic fraction on reproductive switching in the planarian Dugesia ryukyuensis (Seriata, Tricladida). *Frontiers in Zoology*, *8*(1), 23. https://doi.org/10.1186/1742-9994-8-23

Kolmer, J. A., Ordoñez, M. E., & Groth, J. V. (2018). The Rust Fungi. In *eLS* (pp. 1–9). John Wiley & Sons, Ltd. https://doi.org/10.1002/9780470015902.a0021264.pub2

Krobanan, K., & Shen, W.-C. (2018). Nuclear dynamics in the homothallic ascomycete Sordaria fimicola. *Botanical Studies*, *59*, 17. https://doi.org/10.1186/s40529-018-0233-y

Krzan, Z. I. D. (1981). Trials with axenic cultures of the rust fungus Melampsora larici-populina Kleb. *Arboretum Kornickie (Poland)*, *26*. https://agris.fao.org/search/en/providers/122651/records/6471258e804410faf61eede5

Kües, U. (2000). Life History and Developmental Processes in the Basidiomycete Coprinus cinereus. *Microbiology and Molecular Biology Reviews*, *64*(2), 316–353. https://doi.org/10.1128/mmbr.64.2.316-353.2000

Kües, U. (2015). From two to many: Multiple mating types in Basidiomycetes. *Fungal Biology Reviews*, *29*(3), 126–166. https://doi.org/10.1016/j.fbr.2015.11.001

Kües, U., & Casselton, L. A. (1992). Homeodomains and regulation of sexual development in basidiomycetes. *Trends in Genetics: TIG*, *8*(5), 154–155. https://doi.org/10.1016/0168-9525(92)90207-k

Kües, U., James, T. Y., & Heitman, J. (2011). 6 Mating Type in Basidiomycetes: Unipolar, Bipolar, and Tetrapolar Patterns of Sexuality. In S. Pöggeler & J. Wöstemeyer (Eds.), *Evolution of Fungi and Fungal-Like Organisms* (pp. 97–160). Springer. https://doi.org/10.1007/978-3-642-19974-5_6

Lamelza, P., Young, J. M., Noble, L. M., Caro, L., Isakharov, A., Palanisamy, M., Rockman, M. V., Malik, H. S., & Ailion, M. (2019). Hybridization promotes asexual reproduction in Caenorhabditis nematodes. *PLOS Genetics*, *15*(12), e1008520. https://doi.org/10.1371/journal.pgen.1008520

Langton, F. A., & Elliott, T. J. (1980). Genetics of secondarily homothallic basidiomycetes. *Heredity*, *45*, 99–106. https://doi.org/10.1038/hdy.1980.53

Lanver, D., Müller, A. N., Happel, P., Schweizer, G., Haas, F. B., Franitza, M., Pellegrin, C., Reissmann, S., Altmüller, J., Rensing, S. A., & Kahmann, R. (2018). The Biotrophic Development of Ustilago maydis Studied by RNA-Seq Analysis[OPEN]. *The Plant Cell*, *30*(2), 300–323. https://doi.org/10.1105/tpc.17.00764

Laurans, F., & Pilate, G. (1999). Histological Aspects of a Hypersensitive Response in Poplar to Melampsora larici-populina. *Phytopathology®*, *89*(3), 233–238. https://doi.org/10.1094/PHYTO.1999.89.3.233

Lawrence, G., Dodds, P., & Ellis, J. (2007). Rust of flax and linseed caused by Melampsora lini. *Molecular Plant Pathology*, *8*, 349–364. https://doi.org/10.1111/j.1364-3703.2007.00405.x

Lee, S. C., Ni, M., Li, W., Shertz, C., & Heitman, J. (2010). The Evolution of Sex: A Perspective from the Fungal Kingdom. *Microbiology and Molecular Biology Reviews : MMBR*, *74*(2), 298–340. https://doi.org/10.1128/MMBR.00005-10

Lehtonen, J., Jennions, M. D., & Kokko, H. (2012). The many costs of sex. *Trends in Ecology & Evolution*, *27*(3), 172–178. https://doi.org/10.1016/j.tree.2011.09.016

Letunic, I., & Bork, P. (2021). Interactive Tree Of Life (iTOL) v5: An online tool for phylogenetic tree display and annotation. *Nucleic Acids Research*, *49*(W1), W293–W296. https://doi.org/10.1093/nar/gkab301

Li, D., Liu, C.-M., Luo, R., Sadakane, K., & Lam, T.-W. (2015). MEGAHIT: An ultra-fast single-node solution for large and complex metagenomics assembly via succinct de Bruijn graph. *Bioinformatics*, *31*(10), 1674–1676. https://doi.org/10.1093/bioinformatics/btv033

Li, D., Luo, R., Liu, C.-M., Leung, C.-M., Ting, H.-F., Sadakane, K., Yamashita, H., & Lam, T.-W. (2016). MEGAHIT v1.0: A fast and scalable metagenome assembler driven by advanced methodologies and community practices. *Methods*, *102*, 3–11. https://doi.org/10.1016/j.ymeth.2016.02.020

Li, X.-Y., Lehtonen, J., & Kokko, H. (2016). Sexual reproduction as bet-hedging. *Annals of the International Society of Dynamic Games*. https://doi.org/10.1007/978-3-319-70619-1_10

Liao, Y., Smyth, G. K., & Shi, W. (2013). The Subread aligner: Fast, accurate and scalable read mapping by seed-and-vote. *Nucleic Acids Research*, *41*(10), e108–e108. https://doi.org/10.1093/nar/gkt214

Liao, Y., Smyth, G. K., & Shi, W. (2014). featureCounts: An efficient general purpose program for assigning sequence reads to genomic features. *Bioinformatics*, *30*(7), 923–930. https://doi.org/10.1093/bioinformatics/btt656

Liegeois, M., Sartori, M., & Schwander, T. (2023). What Ecological Factors Favor Parthenogenesis over Sexual Reproduction? A Study on the Facultatively Parthenogenetic Mayfly Alainites muticus in Natural Populations. *The American Naturalist*, *201*(2), 229–240. https://doi.org/10.1086/722515

Littlefield, L. J., & Heath, M. (1979). *Ultrastructure of Rust Fungi*. Elsevier Science.

Liu, L., He, G.-J., Chen, L., Zheng, J., Chen, Y., Shen, L., Tian, X., Li, E., Yang, E., Liao, G., & Wang, L. (2018). Genetic basis for coordination of meiosis and sexual structure maturation in Cryptococcus neoformans. *eLife*, *7*, e38683. https://doi.org/10.7554/eLife.38683

Lively, C. M., & Morran, L. T. (2014). The ecology of sexual reproduction. *Journal of Evolutionary Biology*, *27*(7), 1292–1303. https://doi.org/10.1111/jeb.12354

Lorrain, C., Gonçalves dos Santos, K. C., Germain, H., Hecker, A., & Duplessis, S. (2019). Advances in understanding obligate biotrophy in rust fungi. *New Phytologist*, *222*(3), 1190–1206. https://doi.org/10.1111/nph.15641

Lorrain, C., Marchal, C., Hacquard, S., Delaruelle, C., Pétrowski, J., Petre, B., Hecker, A., Frey, P., & Duplessis, S. (2018). The Rust Fungus Melampsora larici-populina Expresses a Conserved Genetic Program and Distinct Sets of Secreted Protein Genes During Infection of Its Two

Host Plants, Larch and Poplar. *Molecular Plant-Microbe Interactions®*, *31*(7), 695–706. https://doi.org/10.1094/MPMI-12-17-0319-R

Louet, C., Saubin, M., Andrieux, A., Persoons, A., Gorse, M., Pétrowski, J., Fabre, B., De Mita, S., Duplessis, S., Frey, P., & Halkett, F. (2023). A point mutation and large deletion at the candidate avirulence locus AvrMlp7 in the poplar rust fungus correlate with poplar RMlp7 resistance breakdown. *Molecular Ecology*, *23*(10), 2472–2483. https://doi.org/10.1111/mec.16294

Love, M. I., Huber, W., & Anders, S. (2014). Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biology*, *15*(12), 550. https://doi.org/10.1186/s13059-014-0550-8

Luo, Z., McTaggart, A., & Schwessinger, B. (2024). Genome biology and evolution of mating-type loci in four cereal rust fungi. *PLOS Genetics*, *20*(3), e1011207. https://doi.org/10.1371/journal.pgen.1011207

Lynch, M., Bürger, R., Butcher, D., & Gabriel, W. (1993). The Mutational Meltdown in Asexual Populations. *Journal of Heredity*, *84*(5), 339–344. https://doi.org/10.1093/oxfordjournals.jhered.a111354

Maia, T. M., Lopes, S. T., Almeida, J. M. G. C. F., Rosa, L. H., Sampaio, J. P., Gonçalves, P., & Coelho, M. A. (2015). Evolution of Mating Systems in Basidiomycetes and the Genetic Architecture Underlying Mating-Type Determination in the Yeast Leucosporidium scottii. *Genetics*, *201*(1), 75–89. https://doi.org/10.1534/genetics.115.177717

Maynard Smith, J. (1978). *The Evolution of Sex*. https://cir.nii.ac.jp/crid/1370285712589795084

Mboup, M., Leconte, M., Gautier, A., Wan, A. M., Chen, W., De Vallavieille-Pope, C., & Enjalbert, J. (2009). Evidence of genetic recombination in wheat yellow rust populations of a Chinese oversummering area. *Fungal Genetics and Biology*, *46*(4), 299–307. https://doi.org/10.1016/j.fgb.2008.12.007

McDonald, B. A., & Linde, C. (2002). Pathogen Population Genetics, Evolutionary Potential, and Durable Resistance. *Annual Review of Phytopathology*, *40*(1), 349–379. https://doi.org/10.1146/annurev.phyto.40.120501.101443

McTaggart, A. R., du Plessis, E., Roux, J., Barnes, I., Fraser, S., Granados, G. M., Ho, W. W. H., Shuey, L. S., & Drenth, A. (2020). Sexual reproduction in populations of Austropuccinia psidii. *European Journal of Plant Pathology*, *156*(2), 537–545. https://doi.org/10.1007/s10658-019-01903-y

McTaggart, A. R., James, T. Y., Idnurm, A., Park, R. F., Shuey, L. S., Demers, M. N. K., & Aime, M. C. (2022). Sexual reproduction is the null hypothesis for life cycles of rust fungi. *PLOS Pathogens*, *18*(5), e1010439. https://doi.org/10.1371/journal.ppat.1010439

Meng, J.-W., Zhu, W., He, M.-H., Wu, E.-J., Duan, G.-H., Xie, Y.-K., Jin, Y.-J., Yang, L.-N., Shang, L.-P., & Zhan, J. (2015). Population genetic analysis reveals cryptic sex in the phytopathogenic fungus Alternaria alternata. *Scientific Reports*, *5*, 18250. https://doi.org/10.1038/srep18250

Menu, F., Roebuck, J., & Viala, M. (2000). Bet-Hedging Diapause Strategies in Stochastic Environments. *The American Naturalist*, *155*(6), 724–734. https://doi.org/10.1086/303355

Meunier, C., Hosseini, S., Heidari, N., Maryush, Z., & Johannesson, H. (2018). Multilevel Selection in the Filamentous Ascomycete Neurospora tetrasperma. *The American Naturalist*, *191*(3), 290–305. https://doi.org/10.1086/695803

Minh, B. Q., Schmidt, H. A., Chernomor, O., Schrempf, D., Woodhams, M. D., von Haeseler, A., & Lanfear, R. (2020). IQ-TREE 2: New Models and Efficient Methods for Phylogenetic Inference in the Genomic Era. *Molecular Biology and Evolution*, *37*(5), 1530–1534. https://doi.org/10.1093/molbev/msaa015

Mojerlou, S., Moeller, M., Schwessinger, B., & Rodriguez-Algaba, J. (2025). Beyond Asexual: Genomics-Driven Progress in Unveiling Sexual Reproduction in Cereal Rust Fungi. *Molecular Plant-Microbe Interactions®*, *38*(2), 206–212. https://doi.org/10.1094/MPMI-10-24-0122-FI

Mojerlou, S., Schwessinger, B., & Rodriguez-Algaba, J. (2025). Sexual recombination in cereal rust fungi: Knowns and unknowns of pathogen evolution and adaptation. *PLOS Pathogens*, *21*(2), e1012908. https://doi.org/10.1371/journal.ppat.1012908

Morrow, C. A., & Fraser, J. A. (2009). Sexual reproduction and dimorphism in the pathogenic basidiomycetes. *FEMS Yeast Research*, *9*(2), 161–177. https://doi.org/10.1111/j.1567-1364.2008.00475.x

Muller, H. J. (1964). The relation of recombination to mutational advance. *Mutation Research/Fundamental and Molecular Mechanisms of Mutagenesis*, *1*(1), 2–9. https://doi.org/10.1016/0027-5107(64)90047-8

Nauta, M. J., & Hoekstra, R. F. (1992). Evolution of reproductive systems in filamentous ascomycetes. I. Evolution of mating types. *Heredity*, *68*(5), 405–410. https://doi.org/10.1038/hdy.1992.60

Neiman, M., Meirmans, S., & Meirmans, P. G. (2009). What Can Asexual Lineage Age Tell Us about the Maintenance of Sex? *Annals of the New York Academy of Sciences*, *1168*(1), 185–200. https://doi.org/10.1111/j.1749-6632.2009.04572.x

Nemr, R. A., Khalil, M., Sarhan, M. S., Abbas, M., Elsawey, H., Youssef, H. H., Hamza, M. A., Morsi, A. T., El-Tahan, M., Fayez, M., Patz, S., Witzel, K., Ruppel, S., El-Sahhar, K. F., & Hegazi, N. A. (2020). "In situ similis" Culturing of Plant Microbiota: A Novel Simulated Environmental Method Based on Plant Leaf Blades as Nutritional Pads. *Frontiers in Microbiology*, *11*. https://doi.org/10.3389/fmicb.2020.00454

Nieuwenhuis, B. P. S., Billiard, S., Vuilleumier, S., Petit, E., Hood, M. E., & Giraud, T. (2013). Evolution of uni- and bifactorial sexual compatibility systems in fungi. *Heredity*, *111*(6), Article 6. https://doi.org/10.1038/hdy.2013.67

Nieuwenhuis, B. P. S., & James, T. Y. (2016). The frequency of sex in fungi. *Philosophical Transactions of the Royal Society B: Biological Sciences*, *371*(1706), 20150540. https://doi.org/10.1098/rstb.2015.0540

Otto, S. P. (2009). The Evolutionary Enigma of Sex. *The American Naturalist*, *174*(S1), S1–S14. https://doi.org/10.1086/599084

Otto, S. P., & Lenormand, T. (2002). Resolving the paradox of sex and recombination. *Nature Reviews Genetics*, *3*(4), 252–261. https://doi.org/10.1038/nrg761

Passer, A. R., Clancey, S. A., Shea, T., David-Palma, M., Averette, A. F., Boekhout, T., Porcel, B. M., Nowrousian, M., Cuomo, C. A., Sun, S., Heitman, J., & Coelho, M. A. (2022). Obligate sexual reproduction of a homothallic fungus closely related to the Cryptococcus pathogenic species complex. *eLife*, *11*, e79114. https://doi.org/10.7554/eLife.79114

Peris, D., Lu, D. S., Kinneberg, V. B., Methlie, I.-S., Dahl, M. S., James, T. Y., Kauserud, H., & Skrede, I. (2022). Large-scale fungal strain sequencing unravels the molecular diversity in mating loci maintained by long-term balancing selection. *PLOS Genetics*, *18*(3), e1010097. https://doi.org/10.1371/journal.pgen.1010097

Pernaci, M., De Mita, S., Andrieux, A., Pétrowski, J., Halkett, F., Duplessis, S., & Frey, P. (2014). Genome-wide patterns of segregation and linkage disequilibrium: The construction of a linkage genetic map of the poplar rust fungus Melampsora larici-populina. *Frontiers in Plant Science*, *5*. https://www.frontiersin.org/articles/10.3389/fpls.2014.00454

Persoons, A., Hayden, K. J., Fabre, B., Frey, P., De Mita, S., Tellier, A., & Halkett, F. (2017). The escalatory Red Queen: Population extinction and replacement following arms race dynamics in poplar rust. *Molecular Ecology*, *26*(7), 1902–1918. https://doi.org/10.1111/mec.13980

Persoons, A., Maupetit, A., Louet, C., Andrieux, A., Lipzen, A., Barry, K. W., Na, H., Adam, C., Grigoriev, I. V., Segura, V., Duplessis, S., Frey, P., Halkett, F., & De Mita, S. (2022). Genomic Signatures of a Major Adaptive Event in the Pathogenic Fungus Melampsora larici-populina. *Genome Biology and Evolution*, *14*(1), evab279. https://doi.org/10.1093/gbe/evab279

Persoons, A., Morin, E., Delaruelle, C., Payen, T., Halkett, F., Frey, P., De Mita, S., & Duplessis, S. (2014). Patterns of genomic variation in the poplar rust fungus Melampsora larici-populina identify pathogenesis-related factors. *Frontiers in Plant Science*, *5*. https://www.frontiersin.org/articles/10.3389/fpls.2014.00450

Raudaskoski, M., & Kothe, E. (2010). Basidiomycete Mating Type Genes and Pheromone Signaling. *Eukaryotic Cell*, *9*(6), 847–859. https://doi.org/10.1128/EC.00319-09

Rinaldi, C., Kohler, A., Frey, P., Duchaussoy, F., Ningre, N., Couloux, A., Wincker, P., Le Thiec, D., Fluch, S., Martin, F., & Duplessis, S. (2007). Transcript Profiling of Poplar Leaves upon Infection with Compatible and Incompatible Strains of the Foliar Rust Melampsora larici-populina. *Plant Physiology*, *144*(1), 347–366. https://doi.org/10.1104/pp.106.094987

Rodriguez-Algaba, J., Walter, S., Sørensen, C. K., Hovmøller, M. S., & Justesen, A. F. (2014). Sexual structures and recombination of the wheat rust fungus *Puccinia striiformis* on *Berberis vulgaris*. *Fungal Genetics and Biology*, *70*, 77–85. https://doi.org/10.1016/j.fgb.2014.07.005

Ross, L., Hardy, N. B., Okusu, A., & Normark, B. B. (2013). LARGE POPULATION SIZE PREDICTS THE DISTRIBUTION OF ASEXUALITY IN SCALE INSECTS. *Evolution*, *67*(1), 196–206. https://doi.org/10.1111/j.1558-5646.2012.01784.x

Rushworth, C. A., Windham, M. D., Keith, R. A., & Mitchell-Olds, T. (2018). Ecological differentiation facilitates fine-scale coexistence of sexual and asexual Boechera. *American Journal of Botany*, *105*(12), 2051–2064. https://doi.org/10.1002/ajb2.1201

Schwessinger, B., Chen, Y.-J., Tien, R., Vogt, J. K., Sperschneider, J., Nagar, R., McMullan, M., Sicheritz-Ponten, T., Sørensen, C. K., Hovmøller, M. S., Rathjen, J. P., & Justesen, A. F. (2020). Distinct Life Histories Impact Dikaryotic Genome Evolution in the Rust Fungus Puccinia striiformis Causing Stripe Rust in Wheat. *Genome Biology and Evolution*, *12*(5), 597–617. https://doi.org/10.1093/gbe/evaa071

Seger, J., & Brockmann, H. (1987). What is Bet-Hedging. *Oxf. Surv. Evol. Biol.*, *4*.

Slater, G. S. C., & Birney, E. (2005). Automated generation of heuristics for biological sequence comparison. *BMC Bioinformatics*, *6*(1), 31. https://doi.org/10.1186/1471-2105-6-31

Slessareva, J. E., & Dohlman, H. G. (2006). G Protein Signaling in Yeast: New Components, New Connections, New Compartments. *Science*, *314*(5804), 1412–1413. https://doi.org/10.1126/science.1134041

Smith, J. M. (1964). Group Selection and Kin Selection. *Nature*, *201*(4924), 1145–1147. https://doi.org/10.1038/2011145a0

Sperschneider, J., Chen, J., Anderson, C., Morin, E., Zhang, X., Lewis, D., Henningsen, E., Grigoriev, I. V., Rathjen, J. P., Jones, D. A., Duplessis, S., & Dodds, P. N. (2025). A Chromosome-Scale Genome Assembly of the Flax Rust Fungus Reveals the Two Unusually Large Effector Proteins, AvrM3 and AvrN. *Molecular Plant-Microbe Interactions®*. https://doi.org/10.1094/MPMI-04-25-0047-R

Stoeckel, S., Becheler, R., Bocharova, E., & Barloy, D. (2024). GenAPoPop 1.0: A user-friendly software to analyse genetic diversity and structure from partially clonal and selfed autopolyploid organisms. *Molecular Ecology Resources*, *24*(1), e13886. https://doi.org/10.1111/1755-0998.13886

Stoeckel, S., & Masson, J.-P. (2014). The Exact Distributions of FIS under Partial Asexuality in Small Finite Populations with Mutation. *PLOS ONE*, *9*(1), e85228. https://doi.org/10.1371/journal.pone.0085228

Stoeckel, S., Porro, B., & Arnaud-Haond, S. (2021). The discernible and hidden effects of clonality on the genotypic and genetic states of populations: Improving our estimation of clonal rates. *Molecular Ecology Resources*, *21*(4), 1068–1084. https://doi.org/10.1111/1755-0998.13316

Sun, S., Coelho, M. A., Heitman, J., & Nowrousian, M. (2019). Convergent evolution of linked mating-type loci in basidiomycete fungi. *PLOS Genetics*, *15*(9), e1008365. https://doi.org/10.1371/journal.pgen.1008365

Swiezynski, K. M., & Day, P. R. (1960). Heterokaryon formation in Coprinus lagopus. *Genetics Research*, *1*(1), 114–128. https://doi.org/10.1017/S0016672300000112

Talhinhas, P., Carvalho, R., Tavares, S., Ribeiro, T., Azinheira, H., Ramos, A. P., Silva, M. do C., Monteiro, M., Loureiro, J., & Morais-Cecílio, L. (2023). Diploid Nuclei Occur throughout the Life Cycles of Pucciniales Fungi. *Microbiology Spectrum*, *11*(4), e01532-23. https://doi.org/10.1128/spectrum.01532-23

Tao, S.-Q., Auer, L., Morin, E., Liang, Y.-M., & Duplessis, S. (2020). Transcriptome Analysis of Apple Leaves Infected by the Rust Fungus Gymnosporangium yamadae at Two Sporulation Stages. *Molecular Plant-Microbe Interactions®*, *33*(3), 444–461. https://doi.org/10.1094/MPMI-07-19-0208-R

Tao, S.-Q., Cao, B., Morin, E., Liang, Y.-M., & Duplessis, S. (2019). Comparative transcriptomics of Gymnosporangium spp. Teliospores reveals a conserved genetic program at this specific

stage of the rust fungal life cycle. *BMC Genomics*, *20*(1), 723. https://doi.org/10.1186/s12864-019-6099-x

Tao, S.-Q., Cao, B., Tian, C.-M., & Liang, Y.-M. (2017). Comparative transcriptome analysis and identification of candidate effectors in two related rust species (Gymnosporangium yamadae and Gymnosporangium asiaticum). *BMC Genomics*, *18*(1), 651. https://doi.org/10.1186/s12864-017-4059-x

Tilquin, A., & Kokko, H. (2016). What does the geography of parthenogenesis teach us about sex? *Philosophical Transactions of the Royal Society B: Biological Sciences*, *371*(1706), 20150538. https://doi.org/10.1098/rstb.2015.0538

Tripp, E. A. (2016). Is asexual reproduction an evolutionary dead end in lichens? *The Lichenologist*, *48*(5), 559–580. https://doi.org/10.1017/S0024282916000335

van der Kooi, C. J., & Schwander, T. (2014). On the fate of sexual traits under asexuality. *Biological Reviews*, *89*(4), 805–819. https://doi.org/10.1111/brv.12078

Velásquez, A. C., Castroverde, C. D. M., & He, S. Y. (2018). Plant and pathogen warfare under changing climate conditions. *Current Biology : CB*, *28*(10), R619–R634. https://doi.org/10.1016/j.cub.2018.03.054

Vialle, A., Feau, N., Frey, P., Bernier, L., & Hamelin, R. C. (2013). Phylogenetic species recognition reveals host-specific lineages among poplar rust fungi. *Molecular Phylogenetics and Evolution*, *66*(3), 628–644. https://doi.org/10.1016/j.ympev.2012.10.021

Villegas, D., Bartaula, R., Cantero-Martínez, C., Luster, D., Szabo, L., Olivera, P., Berlin, A., Rodriguez-Algaba, J., Hovmøller, M. S., McIntosh, R., & Jin, Y. (2022). Barberry plays an

active role as an alternate host of Puccinia graminis in Spain. *Plant Pathology*, *71*(5), 1174–1184. https://doi.org/10.1111/ppa.13540

Wallen, R. M., & Perlin, M. H. (2018a). An Overview of the Function and Maintenance of Sexual Reproduction in Dikaryotic Fungi. *Frontiers in Microbiology*, *9*. https://doi.org/10.3389/fmicb.2018.00503

Wallen, R. M., & Perlin, M. H. (2018b). An Overview of the Function and Maintenance of Sexual Reproduction in Dikaryotic Fungi. *Frontiers in Microbiology*, *9*, 503. https://doi.org/10.3389/fmicb.2018.00503

Wang, Q., Wang, S., Xiong, C. L., James, T. Y., & Zhang, X. G. (2017). Mating-type genes of the anamorphic fungus Ulocladium botrytis affect both asexual sporulation and sexual reproduction. *Scientific Reports*, *7*(1), 7932. https://doi.org/10.1038/s41598-017-08471-3

Wang, Z., Kin, K., Lopez-Giraldez, F., Johannesson, H., & Townsend, J. P. (2012). Sex-specific gene expression during asexual development of Neurospora crassa. *Fungal Genetics and Biology*, *49*(7), 533–543. https://doi.org/10.1016/j.fgb.2012.05.004

Warren, R. J., & Mokadam, C. (2025). Asexuality and species invasion. *Biodiversity and Conservation*, *34*(1), 29–43. https://doi.org/10.1007/s10531-024-02976-w

Williams, G. C. (1975). *Sex and Evolution*. Princeton University Press.

Wilson, A. M., Wilken, P. M., van der Nest, M. A., Steenkamp, E. T., Wingfield, M. J., & Wingfield, B. D. (2015). Homothallism: An umbrella term for describing diverse sexual behaviours. *IMA Fungus*, *6*(1), 207–214. https://doi.org/10.5598/imafungus.2015.06.01.13

Wolf, F. T. (1974). The cultivation of rust fungi upon artificial media. *Canadian Journal of Botany*, *52*(4), 767–772. https://doi.org/10.1139/b74-099

Xhaard, C., Fabre, B., Andrieux, A., Gladieux, P., Barrès, B., Frey, P., & Halkett, F. (2011). The genetic structure of the plant pathogenic fungus Melampsora larici-populina on its wild host is extensively impacted by host domestication. *Molecular Ecology*, *20*(13), 2739–2755. https://doi.org/10.1111/j.1365-294X.2011.05138.x

Xue, C., Hsueh, Y.-P., & Heitman, J. (2008). Magnificent seven: Roles of G protein-coupled receptors in extracellular sensing in fungi. *FEMS Microbiology Reviews*, *32*(6), 1010–1032. https://doi.org/10.1111/j.1574-6976.2008.00131.x

Yan, S., Wang, W., & Shen, J. (2020). Reproductive polyphenism and its advantages in aphids: Switching between sexual and asexual reproduction. *Journal of Integrative Agriculture*, *19*(6), 1447–1457. https://doi.org/10.1016/S2095-3119(19)62767-X

Yeager, R., Bushkin, G. G., Singer, E., Fu, R., Cooperman, B., & McMurray, M. (2021). Post-Transcriptional Control of Mating-Type Gene Expression during Gametogenesis in Saccharomyces cerevisiae. *Biomolecules*, *11*(8), 1223. https://doi.org/10.3390/biom11081223

Yun, S.-H., Berbee, M. L., Yoder, O. C., & Turgeon, B. G. (1999). Evolution of the fungal self-fertile reproductive life style from self-sterile ancestors. *Proceedings of the National Academy of Sciences*, *96*(10), 5592–5597. https://doi.org/10.1073/pnas.96.10.5592

Zhan, G., Guo, J., Tian, Y., Ji, F., Bai, X., Zhao, J., Guo, J., & Kang, Z. (2023). High-throughput RNA sequencing reveals differences between the transcriptomes of the five spore forms of

Puccinia striiformis f. Sp. Tritici, the wheat stripe rust pathogen. *Stress Biology*, *3*(1), 29. https://doi.org/10.1007/s44154-023-00107-z

Zhong, C., Guan, S., Xu, M., Ma, M., Li, C., Tao, L., Huang, G., & Guan, M. (2025). An extremely poor nutritional condition enables efficient white cell mating in Candida albicans. *mSphere*, *10*(6), e0029125. https://doi.org/10.1128/msphere.00291-25

Zhu, A., Ibrahim, J. G., & Love, M. I. (2019). Heavy-tailed prior distributions for sequence count data: Removing the noise and preserving large differences. *Bioinformatics*, *35*(12), 2084–2092. https://doi.org/10.1093/bioinformatics/bty895

Appendices

List of Appendices

Appendix I

Accepted manuscript “De novo assembly of complete circular mitochondrial genomes from 2,695 fungal species” in journal of *Scientific Data*.

Appendix II

Published manuscript “A roadmap for equitable reuse of public microbiome data” in journal of *Nature Microbiology.*

Appendix I:

The manuscript, titled “De novo assembly of complete circular mitochondrial genomes from 2,695 fungal species,” has been accepted for publication in the Journal of *Scientific Data*.

The story of this work started when I became curious about the mitochondrial genome of *Melampsora larici-populina* and mitochondria inheritance in fungi. I realized that the complete resolution and annotation of the mitochondrial genome for this rust fungus had not yet been published. This observation overlapped with the interests of my Egyptian former colleague and his colleagues at Eurac Research, Institute for Biomedicine, and Institute for Mummy Studies, who were pursuing a broader project aimed at retrieving the complete mitochondrial genomes across the fungal kingdom. This overlap in interests led to an independent collaboration outside the main scope of my PhD, resulting in the study presented in this appendix.

My contribution involved designing the computational analysis, curating the data for public deposition, writing, reviewing, and approving the final version of the manuscript.

Scientific Data

https://doi.org/10.1038/s41597-025-06447-x

Article in Press

De novo assembly of complete circular mitochondrial genomes from 2,695 fungal species

Received: 10 April 2025

Accepted: 9 December 2025

Cite this article as: Sarhan, M.S., Abdalrahem, A., Maixner, F. *et al. De novo* assembly of complete circular mitochondrial genomes from 2,695 fungal species. *Sci Data* (2025). https://doi.org/10.1038/s41597-025-06447-x

Mohamed S. Sarhan, Ammar Abdalrahem, Frank Maixner & Christian Fuchsberger

We are providing an unedited version of this manuscript to give early access to its findings. Before final publication, the manuscript will undergo further editing. Please note there may be errors present which affect the content, and all legal disclaimers apply.

If this paper is publishing under a Transparent Peer Review model then Peer Review reports will publish with the final article.

ARTICLE IN PRESS

© The Author(s) 2025. **Open Access** This article is licensed under a Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International License, which permits any non-commercial use, sharing, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if you modified the licensed material. You do not have permission under this licence to share adapted material derived from this article or parts of it. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit http://creativecommons.org/licenses/by-nc-nd/4.0/.

ARTICLE IN PRESS

Title

De novo assembly of complete circular mitochondrial genomes from 2,695 fungal species

Authors

Mohamed S. Sarhan[1, 2, *], Ammar Abdalrahem[3,4], Frank Maixner[5], Christian Fuchsberger[1]

Affiliations

[1] Institute for Biomedicine, Eurac Research, Bolzano 39100, Italy

[2] Department CIBIO, University of Trento, Trento, Italy

[3] Université de Lorraine, INRAE, IAM, F-54000 Nancy, France

[4] PHIM, Université de Montpellier, IRD, CIRAD, INRAE, Institut Agro, Montpellier, France

[5] Institute for Mummy Studies, Eurac Research, Bolzano 39100, Italy

[*]Corresponding author: Mohamed S. Sarhan (mohamed.sarhan@eurac.edu)

ARTICLE IN PRESS

ARTICLE IN PRESS

Abstract

Fungal mitochondrial genomes are critical for understanding phylogenetics, evolution, and ecology of the Kingdom Fungi, yet they remain underrepresented in public databases. To address this, we developed a workflow to recover mitochondrial genomes from 12,902 fungal short read sequencing data housed in the Sequence Read Archive (SRA) records, assembling complete circular genomes from 2,695 species. This effort expanded fungal mitochondrial genome diversity by nearly 2.3X particularly in understudied phyla such as Mucoromycota (11X increase) and Zoopagomycota (8X increase). The new dataset contains novel yet undescribed mitochondrial genomes at numerous taxonomic levels, including 15 classes, 64 orders, 178 families, and 544 genera. Taxonomic analysis revealed broad ecological representation among the top-assembled species, including human pathogens (e.g., *Cryptococcus tetragattii*), plant pathogens (e.g., *Melampsora larici-populina*), edible mushrooms (e.g., *Suillus luteus*), and industrial fungi. By leveraging the not yet fully exploited SRA sequencing data, this study fills critical gaps in fungal mitochondrial genomics, tripling the currently known mitochondrial genome diversity of the Kingdom Fungi, and provides an extensive resource for phylogenetic and evolutionary research.

Background & Summary

Fungi represent one of the most diverse and ecologically significant kingdoms of life, with an estimated diversity of 2 to 5 million species globally [1,2]. However, only approximately 150,000 species are described representing less than 5% of the total estimated diversity [3]. Fungal species definition has been debated for several decades as there is no consensus on how to delineate species [4]. With the introduction of molecular typing and DNA barcoding techniques, genomic regions like internal transcribed spacers (ITS) and 18S rRNA gene have been used extensively to molecularly type fungi and to assist in species definition. While this approach has been useful for decades, with increasing number of described species, single gene barcoding became inefficient to delineate species at high resolution. Therefore, molecular mycologists started to employ multi-locus typing and even whole genome sequencing to achieve higher taxonomic resolution, which can exceed the species-level to the sub-species level [4]. This is reflected currently by the increasing number of complete/draft fungal genomes deposited in different public databases [5].

Fungi, like most eukaryotes, carry extrachromosomal DNA molecules, such as mitochondrial DNA or plasmids [6]. Those extrachromosomal DNA are present in multiple copies exceeding with order of magnitude the copy numbers of nuclear DNA. Mitochondrial genomes, with their compact size, high copy number, and conserved gene content, have been used in elucidating fungal phylogeny, population genetics, and evolutionary dynamics [7-10]. However, they are often ignored or advertently removed from

ARTICLE IN PRESS

the downstream analysis when performing whole genome sequencing of fungi [11] or erroneously assembled as part of the nuclear genome [12].

As of June 2024, the nucleotide database of the NCBI hosted 3,774 complete fungal mitochondrial genomes representing 1,114 species, with the majority belonging to the phyla Ascomycota and Basidiomycota, with fewer representatives from the other phyla. In contrast, the SRA database identified 99,472 fungal genomic sequencing records, from 4,994 species, representing under-utilized resources that could potentially be used for mitochondrial genome assembly. Leveraging these gaps between the fungal mitochondrial genomes and the available public SRA records, we undertook a large-scale effort to *de novo* assemble complete fungal mitochondrial genomes aiming to extend the current mitochondrial diversity to new groups. To achieve this, we employed a fungal mitochondrial genomes recovery workflow that involves quality control check, followed by *de novo* assembly from paired end (PE) short reads, and extraction of mitochondrial genomes (**Methods**).

In conclusion, this dataset significantly expanded the known diversity of fungal mitochondrial genomes by leveraging underutilized public sequencing data, successfully assembling complete mitochondrial genomes from 2,695 species—nearly tripling the previously available records. The dataset introduced the first mitochondrial genomes for numerous taxonomic groups, including 15 classes, 64 orders, 178 families, and 544 genera, with particularly notable expansions in understudied phyla such as Mucoromycota (11X) and Zoopagomycota (8X). By filling critical gaps in fungal mitochondrial genomics, this resource enables deeper phylogenetic, ecological, and evolutionary studies, while also providing a foundation for future research on fungal biodiversity, host-pathogen interactions, and mitochondrial evolution. The workflow and findings underscore the value of mining public sequencing data to unlock hidden genomic diversity, presenting a model for similar large-scale efforts in other eukaryotic lineages.

Methods

Data collection

Fungal mitochondrial genomes – To retrieve the already available complete mitochondrial genomes of the Kingdom Fungi, we used the NCBI online platform querying for “mitochondrion[Title] AND ("complete genome" OR "complete sequence")” then applied the following filters: i) species == “fungi”; ii) molecule type == “genomic DNA”; iii) sequence type == “Nucleotide”; and iv) genetic compartment == “Mitochondrion”. The resulting records were exported to a sequence file in FASTA format.

Fungal SRA records – To check the available fungal short reads SRA records, we queried for “fungi” in the SRA database, then applied the following filters: i) source == “DNA”; ii) library layout == “paired”; iii)

platform == “Illumina”; iv) strategy == “genome”; and v) file type == “fastq”. The resulting records were exported as a TSV file including all associated metadata.

To exclude the SRA records from species with already existing complete mitochondrial genomes, we mapped the accession numbers of the fungal mitochondrial genomes and the SRA accessions to the NCBI taxonomy database [13], using the R package “taxonomizr” v0.10.7 (https://github.com/sherrillmix/taxonomizr). For each mitochondrial genome accession number, we used the function “accessionToTaxa” to get its taxonomic ID (taxID), then we used the function “getTaxonomy” to get the taxonomic lineage. We considered only the major taxonomic ranks i.e., kingdom, phylum, class, order, family, genus, and species. For the SRA records, we applied the function “getTaxonomy” directly on the taxID associated with each BioSample.

Then, all SRA records with representative mitochondrial genomes available, either in RefSeq or in the International Nucleotide Sequence Database Collaboration (INSDC) were excluded and the remaining SRA accessions were used for further *de novo* assembly.

Mitochondrial genome assembly workflow

For each SRA record, we downloaded the raw sequencing data in fastq format using the tool “fasterq-dump” v3.0.1 (https://github.com/ncbi/sra-tools), then trimmed the adapters and low-quality sequences using “fastp” v0.23.4 [14] applying the default parameters. The resulting files were rarefied to 5 million paired end reads using “seqtk” v1.4 (r122) (https://github.com/lh3/seqtk); followed by *de novo* assembly using “SPAdes” v4.0.1 [15] with the options “--gfa11 -k 21,33,55,77,89 --isolate”. To extract the mitochondrial genomes from the resulting assembly graphs, we used the tool “GetOrganelle” v1.7.7.1 [16], specifying the search database “fungus_mt” to confine the search to fungal mitochondrial genomes. The tool labels the extracted mitochondrial genomes as complete and circular or as scaffolded genomes. For the scaffolded genomes, we repeat the process of rarefaction and the assembly but using 2 million paired end reads (**Fig. 1**).

Mitochondrial genome annotation and phylogeny

The assembled complete mitochondrial genomes as well as those retrieved from the RefSeq and the INSDC were annotated using “MFannot” v1.37 [17], using their corresponding genetic code.

Phylogenetic analysis was conducted using the amino acid sequences of the mitochondrial protein-coding genes (PCGs) *cox1*, *cox2*, cox3, *cob*, *nad1*, *nad2*, *nad3*, *nad4*, *nad4L*, *nad5*, *nad6*, *atp6*, *atp8*, and *atp9*. The sequences for each gene were aligned using “MAFFT” v7.505 [18], and poorly aligned regions were trimmed using “trimAl” v1.4.rev22 [19], followed by concatenation and alignment assessment using “AMAS” (https://github.com/marekborowiec/AMAS)[20]. A maximum-likelihood phylogeny was inferred with

ARTICLE IN PRESS

"FastTreeMP" v2.1.11 [21] under the LG+Γ model (Le & Gascuel amino acid substitution matrix with gamma-distributed rate heterogeneity). Branch lengths represent the number of substitutions per site, and node support values were estimated using FastTree's approximate likelihood ratio test (SH-like local support). The tree was rooted with the phylum Rozellomycota (including 2 genomes), using the R package "castor" and visualized and annotated using the R packages "ggtree" and "ggtreeExtra".

Data Records

Nucleotide sequence data reported are available in the Third Party Annotation Section of the DDBJ/ENA/GenBank databases under the accession numbers TPA: BK072095-BK074789[22].

Additionally, the following data files are available in Figshare [23]:

1- "Assembled_fungal_mitochondrial_genomes.fasta": Combined FASTA file including all newly assembled mitochondrial genomes.
2- "fungal_mitochondrial_genomes_phylogeny.tre": Maximum-likelihood phylogeny of all fungal mitochondrial genomes. The analysis includes all fungal complete mitochondrial genomes assembled in this study as well as genomes from the RefSeq and the INSDC databases.
3- "All_fungal_mitochondrial_genomes_accessions.tsv": Table including NCBI accession numbers and taxonomic information of the newly assembled mitochondrial genomes as well as their corresponding BioProject accession, BioSample accession, and PubMed references (when available). The table additionally includes the accession numbers of the already existing mitochondrial genomes from the RefSeq and INSDC databases. For all records, genome sizes and GC content is provided.

ARTICLE IN PRESS

Technical Validation

To validate the newly assembled mitochondrial genomes, we followed three strategies. First, we considered the genomes complete if they were identified by GetOrganelle [16] as circular and complete. Second, after annotation with MFannot [17], we set an empirical threshold for completeness based on the presence of a minimum number of core protein-coding genes (PCGs) at certain taxonomic ranks. While fungal mitochondrial genomes have no known set of core genes [24], a typical fungus will contain 4 gene complexes, i.e., complex I (*nad1*, *nad2*, *nad3*, *nad4*, *nad4L*, *nad5*, and *nad6*), complex III (*cob*), complex IV (*cox1*, *cox2*, and *cox3*), and complex V (*atp6*, *atp8*, *atp9*) [6]. Based on the RefSeq mitochondrial genome annotations, we used different completeness thresholds for different taxonomic ranks (down to the order-level) as follows: 1) The orders of Saccharomycetales, Saccharomycodales, and Schizosaccharomycetales of Ascomycota should include at least 7 genes; 2) The order Pezizales of

Ascomycota, to include at least 8 genes; 3) The phylum Chytridiomycota (orders Caulochytriales and Cladochytriales), to include at least 11 genes; 4) The order Lobulomycetales of Chytridiomycota to include 12 genes; 5) The orders Acrospermales, Aulographales, Eremomycetales, Jahnulales, Microthyriales, Mytilinidiales, Phaeotrichales, Pleosporales, and Venturiales of Ascomycota to include at least 12 genes; 6) The orders Pucciniales, Septobasidiales, and Urocystidales of Basidiomycota to include at least 13 genes; 7) and finally the remaining orders of Fungi to include 13 genes at least.

Finally, we used only the genomes that passed the previous filters and used them for phylogenetic overview, including genomes from the NCBI RefSeq and the INSDC databases (**Fig. 2**). The constructed phylogenetic tree included a total of 4,012 genomes providing a comprehensive overview of the Kingdom Fungi. The newly assembly genomes fell within the expected clades compared with the RefSeq and INSDC genomes.

Code Availability

The assembly workflow was implemented in a python script (assembly_workflow.py) passing SRA run accession as input and outputting the assembly contigs and graphs, which are used by GetOrganelle for mitochondrial genome extraction (**Methods**). The script uses already published tools and explained in the Methods section. The script is available on GitHub at https://github.com/msabrysarhan/fungal_mtDNA.

Data Availability

Nucleotide sequence data reported are available in the Third Party Annotation Section of the DDBJ/ENA/GenBank databases under the accession numbers TPA: BK072095-BK074789, and the metadata is available at https://doi.org/10.6084/m9.figshare.28750034.

Figure legends

Fig. 1. Overview of fungal mitochondrial genome *de novo* assembly workflow. Workflow for assembling fungal mitochondrial genomes from publicly available paired-end short-read whole-genome sequencing (WGS) data (For details, please refer to the Methods).

Fig. 2. Validation of the newly assembled fungal mitochondrial genome in comparison with the RefSeq and INSDC genomes. (**a**) Proportion of fungal mitochondrial genomes from different sources. (**b**) Maximum-likelihood phylogeny of fungal mitochondrial genomes. The analysis includes all fungal complete mitochondrial genomes assembled in this study as well as species representatives from the RefSeq and the INSDC databases. The phylogeny was inferred with FastTreeMP from concatenated mitochondrial protein sequences (*cox1*, *cox2*, *cox3*, *atp6*, *atp8*, *atp9*, *cob*, *nad1*, *nad2*, *nad3*, *nad4*, *nad4L*, *nad5* and *nad6*) under the LG+Γ model. Branch lengths represent the number of substitutions per site. Node support values were estimated using FastTree's approximate likelihood ratio test (SH-like local

ARTICLE IN PRESS

support) but are not shown (79% of the nodes are supported with > 70%, for details please refer to the dataset file “fungal_mitochondrial_genomes_phylogeny.tre”). The tree was rooted with the phylum Rozellomycota (including 2 genomes). The colors of the tree tips refer to the phylum of origin while the colors of the outer ring refer to the source database. (**c**) The phylogenetic tree of the panel (b) collapsed at the phylum level to provide a comprehensive overview on the Kingdom Fungi, using the latest taxonomic proposal [25]. The red dots refer to the branches supported by > 70%.

Acknowledgements

This work was supported by the "MOC - MultiOmics Centre for Food and Health" project. The MOC project is co-funded by the European Union (European Regional Development Fund - EFRE). Ammar Abdalrahem was supported by a PhD fellowship from the French Ministry of Education and Research (MESR) and by the French Plan Investissement d’Avenir (PIA) Lab of Excellence ARBRE [ANR-11-LABX-0002- 01]. The authors thank the Department of Innovation, Research and University of the Autonomous Province of Bozen/Bolzano, Italy for covering the Open Access publication costs.

Author contributions

M.S.S. conceived the original idea. M.S.S. and A.A. designed and performed the computational analysis. M.S.S. performed the data visualization and wrote the first draft of the manuscript. M.S.S. and A.A. curated the data for public deposition. F.M. and C.F. edited and revised the manuscript. All authors read and approved the final version of the manuscript.

Competing interests

The authors declare no competing interests.

ARTICLE IN PRESS

ARTICLE IN PRESS

References

1 Hawksworth, D. L. & Lücking, R. J. M. s. Fungal diversity revisited: 2.2 to 3.8 million species. **5**, 10.1128/microbiolspec. funk-0052-2016 (2017).

2 Paterson, R. R. M., Solaiman, Z. & Santamaria, O. J. S. R. Guest edited collection: fungal evolution and diversity. **13**, 21438 (2023).

3 James, T. Y., Stajich, J. E., Hittinger, C. T. & Rokas, A. J. A. R. o. M. Toward a fully resolved fungal tree of life. **74**, 291-313 (2020).

4 Chethana, K. T. *et al.* What are fungal species and how to delineate them? **109**, 1-25 (2021).

5 Li, Y. *et al.* A genome-scale phylogeny of the kingdom Fungi. **31**, 1653-1665. e1655 (2021).

6 Kouvelis, V. N., Kortsinoglou, A. M., James, T. Y. J. E. o. F. & Organisms, F.-L. The evolution of mitochondrial genomes in fungi. 65-90 (2023).

7 Kulik, T., Van Diepeningen, A. D. & Hausner, G. J. F. i. M. Vol. 11 628579 (Frontiers Media SA, 2021).

8 Song, N., Geng, Y. & Li, X. J. F. i. M. The mitochondrial genome of the phytopathogenic fungus Bipolaris sorokiniana and the utility of mitochondrial genome to infer phylogeny of Dothideomycetes. **11**, 863 (2020).

9 Zhang, S. *et al.* Dynamic evolution of eukaryotic mitochondrial and nuclear genomes: a case study in the gourmet pine mushroom Tricholoma matsutake. **23**, 7214-7230 (2021).

10 Sauters, T. J. & Rokas, A. J. C. B. Patterns and mechanisms of fungal genome plasticity. **35**, R527-R544 (2025).

11 Jung, H. *et al.* Twelve quick steps for genome assembly and annotation in the classroom. **16**, e1008325 (2020).

12 Persoons, A. *et al.* Patterns of genomic variation in the poplar rust fungus Melampsora larici-populina identify pathogenesis-related factors. **5**, 450 (2014).

13 Schoch, C. L. *et al.* NCBI Taxonomy: a comprehensive update on curation, resources and tools. **2020**, baaa062 (2020).

14 Chen, S., Zhou, Y., Chen, Y. & Gu, J. fastp: an ultra-fast all-in-one FASTQ preprocessor. *Bioinformatics* **34**, i884-i890, doi:10.1093/bioinformatics/bty560 (2018).

15 Nurk, S., Meleshko, D., Korobeynikov, A. & Pevzner, P. A. metaSPAdes: a new versatile metagenomic assembler. *Genome research* **27**, 824-834 (2017).

16 Jin, J.-J. *et al.* GetOrganelle: a fast and versatile toolkit for accurate de novo assembly of organelle genomes. *Genome Biology* **21**, 241, doi:10.1186/s13059-020-02154-5 (2020).

17 Lang, B. F. *et al.* Mitochondrial genome annotation with MFannot: a critical analysis of gene identification and gene model prediction. **14**, 1222186 (2023).

18 Katoh, K. & Standley, D. M. MAFFT Multiple Sequence Alignment Software Version 7: Improvements in Performance and Usability. *Molecular Biology and Evolution* **30**, 772-780, doi:10.1093/molbev/mst010 (2013).

19 Capella-Gutiérrez, S., Silla-Martínez, J. M. & Gabaldón, T. J. B. trimAl: a tool for automated alignment trimming in large-scale phylogenetic analyses. **25**, 1972-1973 (2009).

20 Borowiec, M. L. J. P. AMAS: a fast tool for alignment manipulation and computing of summary statistics. **4**, e1660 (2016).

21 Price, M. N., Dehal, P. S. & Arkin, A. P. J. P. o. FastTree 2–approximately maximum-likelihood trees for large alignments. **5**, e9490 (2010).

22 Sarhan, M. S., Abdalrahem, A., Maixner, F. & Fuchsberger, C. Third Party Assembly (TPA) for Accession Range BK072095-BK074789 https://www.ncbi.nlm.nih.gov/nuccore/?term=BK072095%3ABK074789%5BACCN%5D. *NCBI GenBank* (2025).

ARTICLE IN PRESS

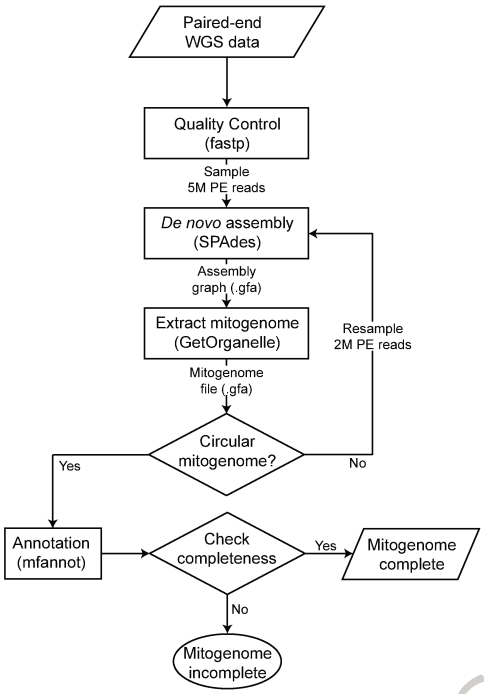
23 Sarhan, M. S., Abdalrahem, A., Maixner, F. & Fuchsberger, C. De novo assembly of complete circular mitochondrial genomes from 2,695 fungal species. *figshare*, doi:https://doi.org/10.6084/m9.figshare.28750034 (2025).

24 Fonseca, P. L. *et al.* Global characterization of fungal mitogenomes: new insights on genomic diversity and dynamism of coding genes and accessory elements. **12**, 787283 (2021).

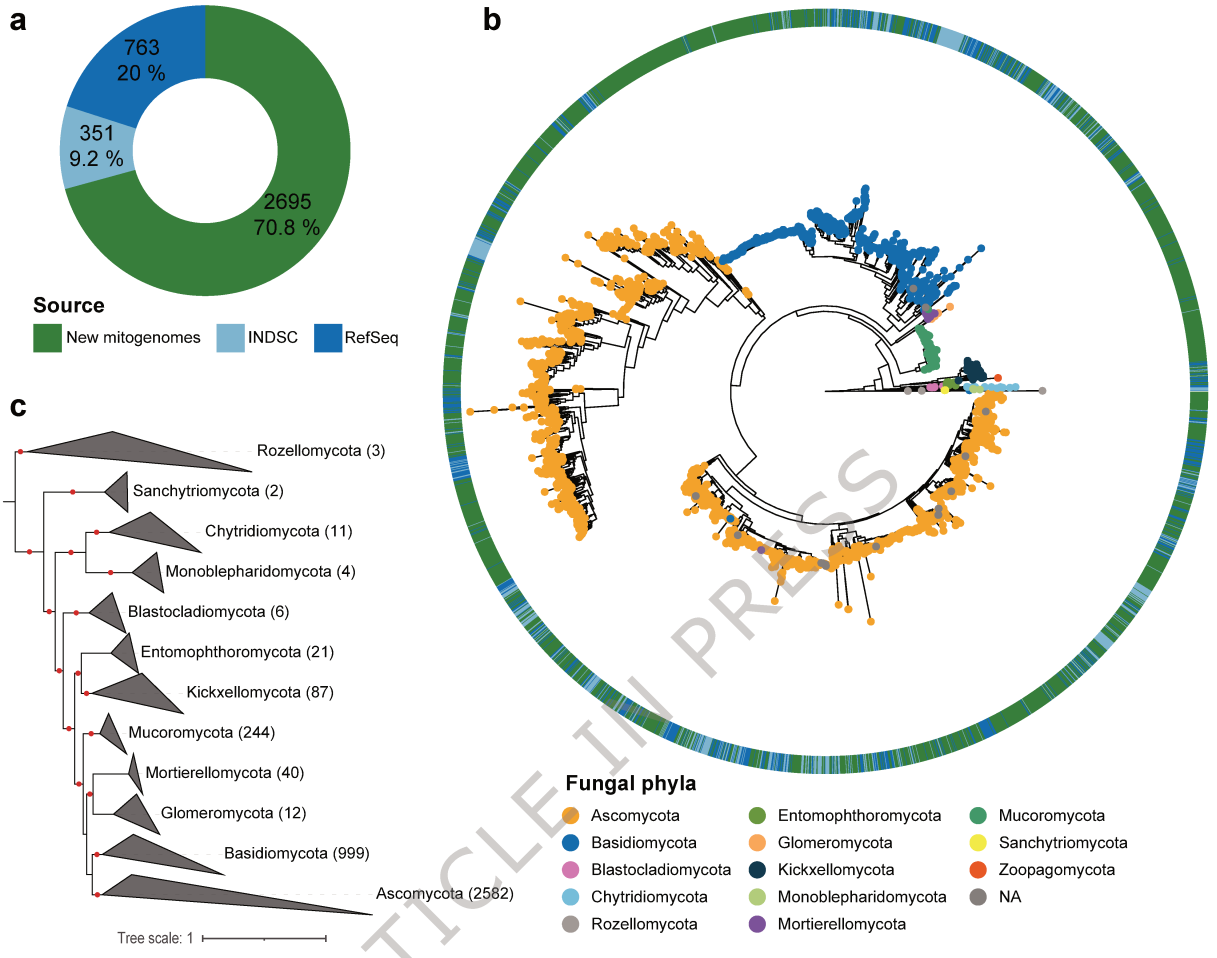
25 Wijayawardene, N. N. *et al.* Classes and phyla of the kingdom Fungi. **128**, 1-165 (2024).

ARTICLE IN PRESS

ARTICLE IN PRESS



ARTICLE IN PRESS



ARTICLE IN PRESS

Appendix II:

The manuscript, titled “A roadmap for equitable reuse of public microbiome data” was published in journal of *Nature Microbiology*.

While I was involved in the project aiming at the reconstruction of mitochondrial genomes from open international Short Reads Archives, I participated in an international scientific consortium called "the data reuse consortium". As part of this consortium, I contributed to an article entitled "A roadmap for fair reuse of public microbiome data" that was first published on the BioRxiv website (doi: 10.1101/2024.06.21.599698) then published for publication in Nature Microbiology (doi: 10.1038/s41564-025-02116-2).

nature microbiology

Consensus Statement

https://doi.org/10.1038/s41564-025-02116-2

A roadmap for equitable reuse of public microbiome data

Received: 27 March 2023

Accepted: 13 August 2025

Published online: 26 September 2025

Check for updates

Laura A. Hug[1,217], Roland Hatzenpichler[2,3,4,217], Cristina Moraru[5,217], André R. Soares[5,6,217], Folker Meyer[7,8,217], Anke Heyder[9], The Data Reuse Consortium* & Alexander J. Probst[5,6,217]

Science benefits from rapid open data sharing, but current guidelines for data reuse were established two decades ago, when databases were several million times smaller than they are today. These guidelines are largely unfamiliar to the scientific community, and, owing to the rapid increase in biological data generated in the past decade, they are also outdated. As a result, there is a lack of community standards suited to the current landscape and inconsistent implementation of data sharing policies across institutions. Here we discuss current sequence data sharing policies and their benefits and drawbacks, and present a roadmap to establish guidelines for equitable sequence data reuse, developed in consultation with a data consortium of 167 microbiome scientists. We propose the use of a Data Reuse Information (DRI) tag for public sequence data, which will be associated with at least one Open Researcher and Contributor ID (ORCID) account. The machine-readable DRI tag indicates that the data creators prefer to be contacted before data reuse, and simultaneously provides data consumers with a mechanism to get in touch with the data creators. The DRI aims to facilitate and foster collaborations, and serve as a guideline that can be expanded to other data types.

Sequence data reuse has been an evolving topic over the past two decades. The Fort Lauderdale Agreement (FLA), a public declaration by biomedicine scientists supporting the free and unrestricted use of genome sequencing data, was coined in 2003, before the advent of metagenomics and during a time when sequencing was still too costly to be performed by individual laboratories[1]. The FLA concluded that large genome projects should be released before publication to allow unrestricted and immediate reuse, which would accelerate the advancement of science. The FLA strengthened the Bermuda Principles defined in 1996[2], which advocated for the release of sequence data 24 h after generation and before publication of research papers. In 2009, after the Human Genome Project highlighted the advantages of sharing data early and widely, the Toronto Statement (TOR)[3] advocated for the prepublication release of other biological data types beyond genomics data. Finally, in 2014, 141 United Nations member states and the European Union entered into the Nagoya Protocol[4] (Regulation (EU) No 511/2014), which calls on data creators and data users to develop, update and use voluntary codes of conduct, guidelines and

[1]Department of Biology, University of Waterloo, Waterloo, Ontario, Canada. [2]Department of Microbiology and Cell Biology, Thermal Biology Institute, Montana State University, Bozeman, MT, USA. [3]Department of Chemistry and Biochemistry, Thermal Biology Institute, Montana State University, Bozeman, MO, USA. [4]Center for Biofilm Engineering, Montana State University, Bozeman, MT, USA. [5]Environmental Metagenomics, Research Center One Health Ruhr, University Alliance Ruhr, Faculty of Chemistry, University of Duisburg-Essen, Essen, Germany. [6]Centre for Water and Environmental Research (ZWU), University of Duisburg-Essen, Essen, Germany. [7]Institute for AI in Medicine, University Hospital Essen, University of Duisburg-Essen, Essen, Germany. [8]Department of Computer Science, University of Duisburg-Essen, Essen, Germany. [9]Department of Psychology, Ruhr University Bochum, Bochum, Germany. [217]These authors contributed equally: Laura A. Hug, Roland Hatzenpichler, Cristina Moraru, André R. Soares, Folker Meyer, Alexander J. Probst. *A list of authors and their affiliations appears at the end of the paper. e-mail: alexander.probst@uni-due.de

https://doi.org/10.1038/s41564-025-02116-2

BOX 1

Definitions of roles of scientists and legal entities related to microbiome and sequence data

Data consumer: any legal entity interested in using public data

Data creator: entity or entities, that is, individual or multiple researchers, who designed the study, obtained the samples and intended to publish an analysis; typically assumed to have priority on analysis

Data distributor: public databases that provide access to the digital data (for example, GenBank, ENA, DDBJ; Box 2)

Data generator: entity that renders the sample into a digital object (for example, a sequencing facility that processes biological material and produces sequence-related files to transmit to another entity)

Data owner: legal entity who owns the data by rights; this can be different from 'data creator' (for example, an institute at which the data creator is employed or a nation or state)

best practices in relation to access and benefit-sharing of genetic data (see Box 1 for descriptions of the different roles of researchers working with sequence datasets).

Large-scale sequence data analysis has become mainstream with a wide array of tools available, making data mining accessible to many labs. Now, ~20 years after the FLA, GenBank holds an estimated ~5.09 terabase pairs (Tbp)[5] of biological sequence data, and the Sequence Read Archive (SRA) holds 90.89 petabase pairs (Pbp) as of February 2024 (Supplementary Fig. 1 and Box 2). These databases are several million times larger than the available sequence data at the time that the FLA or TOR was formulated. With the rapid, continuous increase in public sequence data (projected to reach ~500 Pbp in 2030, Supplementary Fig. 1), data mining projects (or those requiring large public datasets for artificial intelligence training) have increased in both frequency and scope, necessitating a revisit and potential overhaul of the 20-year-old guidelines depicted in the FLA and TOR[1,3].

In 2016, the FAIR principles for data management were defined, which place an emphasis on submitted data being machine actionable (that is, computational systems should be able to Find, Access, Interoperate and Reuse data with minimal human intervention)[6]. These principles were designed to promote good scientific practice and to serve as a guideline for those wishing to enhance the reusability of their data. The FAIR principles have since been adopted as recommendations or requirements by major funding bodies, including the US National Institute of Health and the European Commission[7]. The FAIR data principles prioritize data reuse and computer-driven data mining, and include a specific requirement for (meta)data to be released with a clear and accessible data reuse licence (principle R.1). To date, this aspect of the FAIR principles has not been implemented in a straightforward or machine-readable way, lacking a coordinated implementation between databases and the community.

Biological sciences, and particularly their subdisciplines associated with generating sequence data, have been at the forefront of data availability compared with the fields of earth sciences, mathematics, physics and chemistry[8]. For instance, astrophysics is a subdiscipline that traditionally relies on data sharing and reuse owing to the exorbitant costs of research data. A study investigating the motivational factors behind data sharing and reuse within this field identified several demotivating factors[9]. Among them were the lack of data standards, the lack of facilitating platforms, inconsistency between datasets, limited documentation, difficulties finding and reusing data, and last but not least, competition and fear of being 'scooped' (accidentally or purposefully)[9]. The latter point is considered in the FLA. While the FLA recommends swift prepublication of data generated by large sequencing consortia, it also states that "[…] the contributions and interests of the large-scale data producers should be recognized and respected by the users of the data, and the ability of the production centers to analyse and publish their own data should be supported by their funding agencies […]"[1]. This highlights one of the significant and enduring tensions between data creators and data consumers. Both data creators and data consumers are indispensable to advancing biological sciences, particularly in the realm of sequence data analysis, in which many data creators also act as data consumers. Unrestricted public use of microbiome data, on which data creators have not yet published, does not always align with the interests of data creators.

How to achieve unrestricted data reuse and, at the same time, give due credit to data creators has been discussed by the scientific community[10]. Data reuse in this work refers specifically to those cases in which the sequence data will be featured in a publication prepared by a data consumer, whether in figures, tables or text, or as an important aspect of the workflow that leads to new insights or conclusions (see Supplementary Table 2 for some Data Reuse Information (DRI) usage scenarios). In this spirit, a recent study thoroughly analysed the pros and cons of early data release and considered both the needs of data creators and consumers. The authors proposed immediate, unrestricted release of sequence data before publication, in parallel with the adoption of a reward system (for example, separate promotion and tenure tracks) for acknowledgement of data creators by universities and research institutions[11]. In addition, the authors proposed making the datasets and the protocols used for their generation citable through Digital Object Identifiers (DOIs). If implemented, these measures would create a safer environment for data sharing, benefiting all parties involved and, most importantly, supporting the advancement of science.

Mechanisms for crediting data creators beyond citing associated publications are not yet widespread in the scientific community. Creating separate tenure tracks or other incentives for data creators and data consumers requires sizable changes in evaluation criteria, and would require substantial time to propagate through institutions. DOIs for datasets, on the other hand, seem relatively easy to implement and would provide data creators with a reportable impact metric. However, their use has not yet been widely adopted, possibly owing to associated costs with purchasing and maintaining DOIs, which can be prohibitive for many publicly funded research institutions. Potential measures to lower DOI costs could include large-scale agreements between research institutions and DOI providers. Other mechanisms of data citation have been discussed in the community but have also not been widely adopted[12,13]. Currently, data creators do not have any incentive or reward for releasing sequence data before an associated publication.

It is crucial to implement methodological and ethical guidelines that are based on the principles of good scientific practice and which are driven by the scientific community to facilitate appropriate use of public data. This need has been highlighted by recent conflicts between data creators and data consumers that played out over social media. Implementing and following guidelines for unpublished data usage by all scientists would create 'safe spaces' for data creators to publish their first analyses of data—particularly if they are delayed by resource, time or personnel constraints. The research topic also affects the expectations for open data. Research related to public health necessitates swift data release to counteract pandemics or identify zoonotic diseases. For example, in the event of a pandemic, there should be no data restriction

https://doi.org/10.1038/s41564-025-02116-2

BOX 2

Abbreviations extensively used in microbiome data research

COGs (clusters of orthologous groups): represents a collection of proteins from complete bacterial and archaeal genomes, grouped into clusters of orthologues, and associated with functional annotations

DDBJ (DNA Data Bank of Japan): a database of nucleotide sequence data maintained by the National Institute of Genetics (NIG) in Japan

EMBL (European Molecular Biology Laboratory): a research organization that conducts basic research in molecular biology and offers a range of scientific resources

EMBL-EBI (European Bioinformatics Institute): a bioinformatics research centre belonging to the EMBL, which maintains and provides access to several sequence-related databases (for example, ENA, Interpro, PDBe, UniProt)

ENA (European Nucleotide Archive): a database of raw sequence data and annotated sequence data, from a wide range of organisms, maintained by EMBL-EBI

GenBank (Genetic Sequence Database): a database of DNA and RNA sequences from a wide range of organisms, along with associated annotations and metadata

GSC (Genomic Standards Consortium): an international organization dedicated to the development and implementation of standards and best practices in genomics and related fields

IMG/M (Integrated Microbial Genomes and Metagenomes): a data management and analysis system for microbial genomes and metagenomes maintained by the Department of Energy's Joint Genome Institute (JGI) in the USA

IMG/VR (Integrated Microbial Genomes with Virus-related Datasets): a specialized database storing virus genomic and metagenomic sequences, annotations and metadata

InterPro (integrated resource of protein domains and functional sites): a database that integrates information on protein domains, motifs and functional sites from a variety of sources

INSDC (International Nucleotide Sequence Database Collaboration): a data-sharing initiative between DDBJ, EMBL-EBI and NCBI

KEGG (Kyoto Encyclopedia of Genes and Genomes): a comprehensive database and knowledge base of biological systems, including genes, proteins and biochemical pathways; maintained by Kanehisa Laboratories in Japan

KOG (Eukaryotic Orthologous Groups): a collection of proteins from eukaryotic genomes, grouped into clusters of orthologues, and associated with functional annotations

NCBI (National Center for Biotechnology Information): part of NIH; a central repository for molecular sequence data, including several databases (for example, GenBank, RefSeq, SRA, COG, KOG and so on)

NIH (National Institutes of Health): a biomedical research agency of the federal government of the USA

PDBe (Protein Data Bank in Europe): a comprehensive collection of 3D structures of proteins and other macromolecules

PFAM (protein family database): a database of protein families, domains and functional sites

RefSeq (Reference Sequence Database): a comprehensive, non-redundant database of reference genomic, transcriptomic and proteomic sequences, for a wide range of organisms

SRA (Sequence Read Archive): a public repository for raw sequence data generated by platforms such as Sanger, Illumina, Ion Torrent and Pacific Biosciences

UniProt (Universal Protein Resource): a comprehensive protein sequence database, including additional field-specific contextual information (for example, protein domain structure and known interactions)

InterPro (Integrated resource of protein domains and functional sites): a database that integrates information on protein domains, motifs and functional sites from a variety of sources

INSDC (International Nucleotide Sequence Database Collaboration): a data sharing initiative between DDBJ, EMBL-EBI and NCBI

KEGG (Kyoto Encyclopedia of Genes and Genomes): a comprehensive database and knowledge base of biological systems, including genes, proteins and biochemical pathways; maintained by Kanehisa Laboratories in Japan

KOG (euKaryotic Orthologous Groups): a collection of proteins from eukaryotic genomes, grouped into clusters of orthologues, and associated with functional annotations

NCBI (National Center for Biotechnology Information): part of NIH; a central repository for molecular sequence data, including several databases (for example, GenBank, RefSeq, SRA, COG, KOG and so on)

NIH (National Institutes of Health): a biomedical research agency of the federal government of the USA

PDBe (Protein Data Bank in Europe): a comprehensive collection of 3D structures of proteins and other macromolecules

PFAM (Protein family database): a database of protein families, domains and functional sites

RefSeq (Reference Sequence Database): a comprehensive, non-redundant database of reference genomic, transcriptomic and proteomic sequences, for a wide range of organisms

SRA (Sequence Read Archive): a public repository for raw sequence data generated by platforms such as Sanger, Illumina, Ion Torrent and Pacific Biosciences

UniProt (Universal Protein Resource): a comprehensive protein sequence database, including additional field-specific contextual information (as, for example, protein domain structure and known interactions)

on research related to the pandemic[14]. The general goal should be to promote open sharing of complete datasets as early and as widely as possible, across all institutions and individuals. This necessitates a technical framework that enhances the communication between data creators and data consumers regarding data reuse.

Here we propose a roadmap to enable equitable reuse of public microbiome data. This roadmap (1) addresses the lack of consensus in the field of microbiome research regarding public microbiome data use and reuse, (2) promotes communication between data consumers and data creators and (3) facilitates the rapid advancement of the microbiome field, including supporting the continued increases in data mining. To achieve the goal of this roadmap, we propose the introduction of a new machine-readable metadata tag, named DRI, containing Open Researcher and Contributor IDs (ORCIDs) of the data creators associated with data in public databases. The DRI will clearly indicate the point of contact for communication and if communication is desired by data creators. The ability to provide a point of contact for data reuse will lead to more rapid and complete data deposition. Following adoption by databases, authors and scientific journals would ideally integrate statements confirming that the best practices governed by DRI use were used in manuscripts and submission processes.

The roadmap is directly in line with the FAIR data principles, specifically contributing to FAIR principle R.1 in providing a machine-readable licence for data usage. This roadmap and its adoption by the scientific community (222 scientists as part of the Data Reuse Consortium—Supplementary Table 1—totalling 229 supporters, including the co-authors of this paper) will provide a citable resource regarding guidelines for public data reuse, will enable appropriate data reuse by data consumers and will reduce tension for data creators when submitting data. Ultimately, this roadmap outlines the expected practices for open data use for sequence data and represents a model for other biological data such as metabolomics or proteomics data.

Survey on data reuse

We created an anonymous survey with Google Forms, which was distributed to the international scientific community on 15 January 2024, to accumulate opinion data on a number of key topics related to this manuscript (Supplementary Box 1). Participation was voluntary and anonymous, and participants gave informed consent before participation. To ensure participant confidentiality, it was emphasized that signing the consensus statement could not be linked to survey responses. The University of Duisburg-Essen ethics committee evaluated the study and declared that the need for ethics approval has been waived. Questions included in the survey were formulated in a neutral fashion with the intent of not biasing responses, which were anonymized to ensure openness and transparency. This survey was online and open for responses over a total of 21 days. Efforts to ensure widespread awareness of this survey included actively advertising it across multiple social media platforms (X.com, LinkedIn, Bluesky.app) over the duration of the survey. To achieve this, 39 authors made use of their accounts across these platforms while also leveraging their working group and other institutional accounts. A blog advertising this survey was additionally hosted at the Springer Nature Research Communities blog[15]. Finally, a total of 78 microbiology institutions across the world were contacted to increase participation from the Global South and underrepresented segments of the global scientific community related to this survey. Raw data pertaining to the anonymous responses to this survey in TSV format are available at the Open Science Foundation[16].

This resulted in responses from 306 scientists representing all continents (except Antarctica) with feedback on community interest in and likelihood of adopting the proposed roadmap (Supplementary Figs. 2–10 and Supplementary Tables 3–7). Survey questions were designed to enable quantitative analysis, namely, to appreciate the fraction of the community that agreed or disagreed with proposed aspects of the roadmap delineated in this manuscript. Raw data for this survey were imported into R 4.3.1 running in RStudio 2023.12.0 (Build 369) using the tidyverse ecosystem of data analysis packages to read TSV inputs (readr, version 2.1.5), filter (dplyr, version 1.1.4) and generate plots (ggplot2, version 3.5.1)[17,18]. For Fig. 1, survey data were processed using R, and visualizations in Fig. 1a–c were generated using ggplot2. Figure 1d was generated with a copyright-free image, coloured by normalizing height (in pixels) to percentage categories. Colours and formatting were manually edited. Sankey diagrams were generated via the ggsankey package (version 0.0.99999). Tables were generated with the kableExtra package (version 1.3.4)[19]. R code as well as system and package versions used for all data analysis in this manuscript are publicly available at a GitHub repository: https://github.com/GeoMicroSoares/DataUsage_Data_Analysis.

The initial DRI tag and the strategy for its implementation were defined by the Data Reuse Core Team before the survey. After the survey, the Data Reuse Core Team refined the roadmap and DRI strategy through conversations with members of the Data Reuse Consortium, of the Joint Genome Institute, of the European Nucleotide Archive and of the Genomic Standards Consortium, and the reviewers of this manuscript. Throughout the refinement process, the Data Reuse Core Team identified compromises between differing priorities and ensured that the DRI strategy was actionable within current database structures.

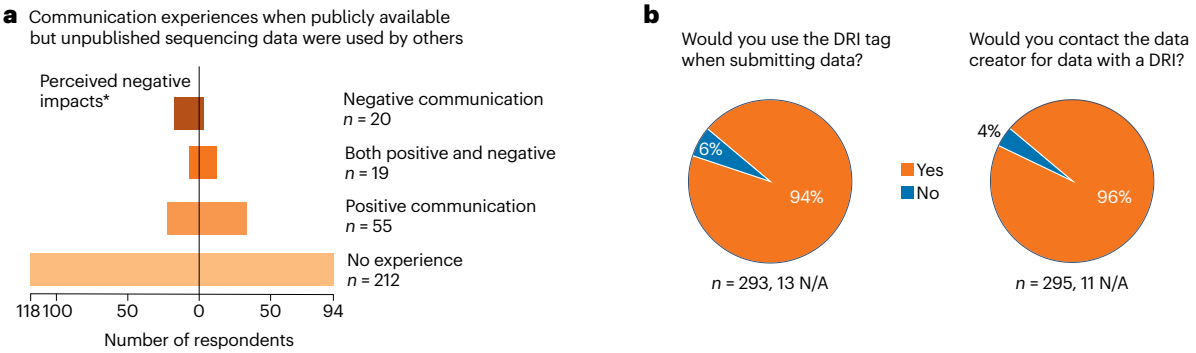
Microbiome data types most frequently reused include amplicon datasets of single genes, such as 16S rRNA or internal transcribed spacer (ITS) genes, as well as reads and assemblies of genomes, metagenomes and metatranscriptomes. These sequence data, while typically structured as machine-readable files, can come in various formats depending on the hosting repository.

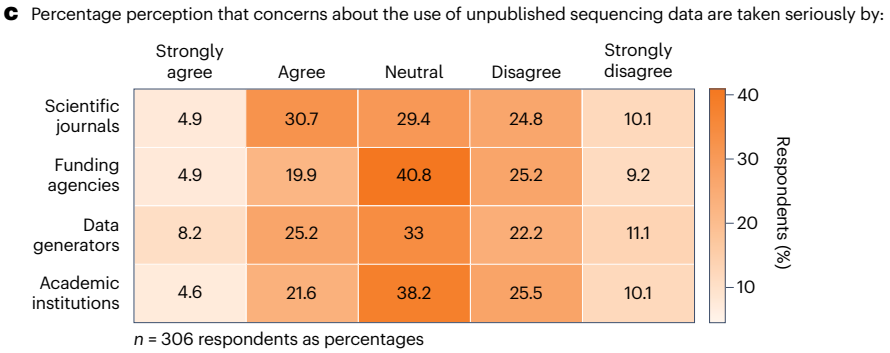
Organizations hosting data usually have their own legal framework governing data download. In the case of the EMBL-EBI ENA database, for example, while records are typically available with no restrictions on reuse, the Terms of Use[20] recognize that third parties may assert restrictions on reuse for a variety of reasons. The database is working towards more systematic implementation of categorizing reusability, probably under Creative Commons (CC0) licensing[21]. This puts the responsibility on the data consumer to ensure their action is indeed legal and covered by the copyrights that may be associated with the data they are accessing, even if the licensing information is not machine readable.

The current state of data policies across online repositories of biological data, including EMBL-EBI, is largely reflective of a now outdated statement in the TOR. Here data users are advised to contact data creators to 'discuss publication plans', a demanding and often unfeasible approach given today's rate of data production and the availability of new data analysis pipelines. The TOR, from 2009, and the FLA, from 2003, are the agreements guiding the field to date, although their contents are often not well known by microbiome scientists across all career stages (Supplementary Figs. 4–6). The FLA specifies, among other things, that "sequence assemblies of 2 kb or greater by large-scale sequencing efforts" must be rapidly released. The language around the scale of data reflects the outdated guidance, and the emphasis on large-scale data creators is no longer in line with the prevalence of independent labs as data creators today. Interestingly, the conflict between data sharing and publishing a first analysis was acknowledged in the FLA, but it does not provide guidelines for navigating this concern. The TOR elaborated further on the conflict between "data [creators] and data users", stating that, in the author's experience, conflicts have rarely arisen.

The TOR lists a set of conditions (scale, utility, reference data, community acceptance) to consider for prepublication data sharing that are of limited relevance for today's data landscape[3]. Currently, individual research groups can contribute substantial datasets, both in terms of size and scientific value, and the idea of individualized, private agreements on data sharing, as suggested by the TOR, is no longer viable. In response to this need, repositories have developed their own policies (for example, the EMBL-EBI licence information). Data distributors

https://doi.org/10.1038/s41564-025-02116-2





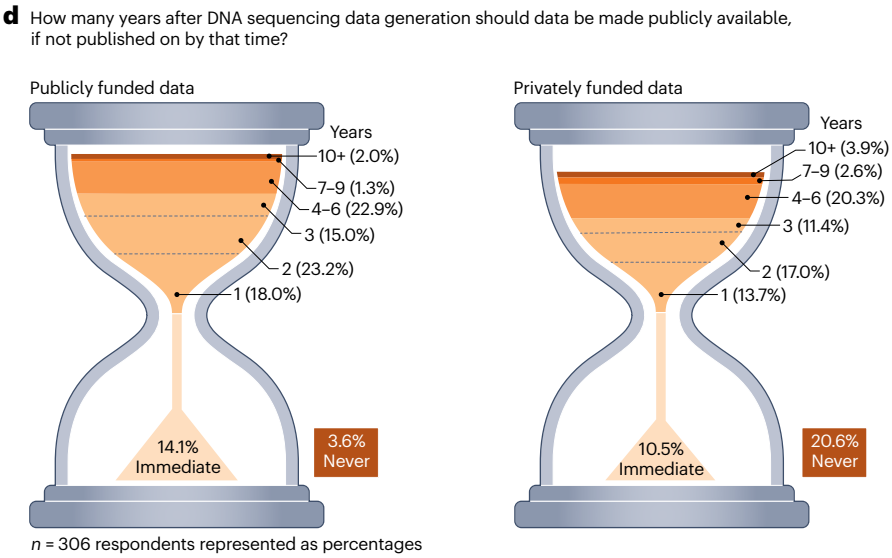


Fig. 1 | Summary results from a survey of 306 scientists on data reuse. A 21-question survey examining the scientific community's perceptions of data reuse was conducted in January 2024, with the survey distributed through social media, a blog and email to hundreds of scientists and two dozen scientific societies. A total of 306 respondents contributed to the survey. **a**–**d**, Responses were summarized for key questions and visualized using Adobe Illustrator, in which the first panel corresponds to questions 15–17 (**a**), the second panel to questions 20–21 (**b**), the third panel to questions 9–12 (**c**) and the fourth panel to questions 7–8 (**d**) (a descriptive analysis of responses to this survey and anonymous raw data to all questions are available in Supplementary Figs. 2–10 and Supplementary Tables 3–7). N/A indicates the question was left blank. Positive communication was defined as being contacted before data analysis or publication and asked for collaboration and opinion, or a positive answer when you requested data removal from a manuscript. Negative communication was defined as no contact before publication, or refusal to remove data from a manuscript upon request. The asterisk in **a** indicates that respondents agree or strongly agree that 'Unauthorized use of my sequence data by other authors has had (or will have) negative impacts on my research programme and/or my mentees'. Respondents selected single-year intervals for the data presented in the hourglass image; years 4–6 and 7–9 were combined given very low proportions for all years except year 5. Notably, respondents did not agree on a time interval after which data should be made available in the absence of an available publication (Supplementary Fig. 7). As a result, our roadmap does not include a recommendation as to when publicly available data with a DRI but no publication can be reused without contacting the data creators (Fig. 2).

recognize that some public datasets, although hosted by their respective services, carry additional restrictions that are currently neither easily visible nor machine readable. Given the ease of accessibility and sheer volume of sequence data, it has become impractical and, in some cases, impossible for a data consumer to verify and to comply with recommendations and restrictions.

Identifying conflicts of interest between data consumers and data creators

The current scale of open sequence data, including massive open datasets such as the *Tara* Oceans (7.2 Tbp of metagenomic data) and Integrative Human Microbiome (1.3 Tbp of metagenomic data as of 2019) projects, has made meta-analyses drawing on public data a powerful avenue to explore microbial systems[22,23]. Use of public data is now routine; close to 80% of respondents to our poll on data usage identified themselves as both data creators and data consumers (Box 1 and Supplementary Table 7). Access to public data has unequivocally improved the depth of the science conducted. However, it is often difficult to assess whether data reuse follows the expectations of communication and collaboration outlined in the TOR. There are many widely used software tools that include use of secondarily accessed data for which no primary publication exists (for example, the Genome Taxonomy Database, GToTree)[24,25]. Identification of the publication status associated with specific data in many repositories is not straightforward. As more governments begin to require data deposition on short or immediate timelines, there is a growing tension between data creators and data consumers around public data use. Clarifying and facilitating data reuse is therefore in the best interest of the community.

The first step is to identify roots of the potential conflict of interest between data creators and data consumers, as established following discussions between the authors of this manuscript as well as within their academic networks. These are discussed in detail below.

Disconnect between efforts of data creation and ease of reuse

One source of conflict is a disconnect between the efforts expended by data creators in generating sequence data and associated metadata, and the ease of reuse and limited or inconsistent acknowledgement of data origins by consumers. Creators sometimes feel that their monetary, time and intellectual investments to design and conduct sample collection and experiments; obtain permission for, plan, fund and carry out research expeditions; process samples; and deposit data and metadata are not adequately acknowledged or are potentially ignored by data consumers. Data creators must obtain legal documents (for example, sampling permits, visas) and follow international agreements (for example, the Convention on Biological Diversity, Nagoya Protocol), and manage the risks that come with certain fieldwork (for example, treacherous terrain, wilderness areas, areas with high criminal activity). In addition, they must secure funding for custom-design vehicles and instrumentation needed for sample collection (for example, drill ships, research vessels, submarines, buoys, remote samplers) and maintain research sites in hard-to-access areas (for example, polar regions or the Amazon). Unbeknownst to data consumers, the original data creators may be bound by restrictive agreements on appropriate or ethical data use if research was conducted in a national park, on private land or land owned by Indigenous nations, and/or for samples obtained from human specimens or biobanks[26]. There are currently limited rewards for data creators when their data are reused, and data creators have little incentive to make detailed metadata available. Systems for reporting and incentivizing data deposition have been proposed but are not yet the norm[11,27].

Timely deposition contrasts with lengthy multi-omics analyses

Both data creators and consumers generally share ideals of open science and rapid advancement of science. However, conflicts arise from disagreements in the timing of sharing data and prioritization of access. Data creators must balance long trainee timelines with publication of datasets intended for multiple research questions. Publishing a first paper on a large dataset and depositing the full data may make additional research projects associated with that dataset vulnerable to scooping. Results from our poll suggest that 53% of researchers are concerned about negative impacts on their research programme and/or mentees from unauthorized data reuse (Fig. 1 and Supplementary Figs. 7–9). As a result, partial or raw datasets or datasets lacking key metadata are deposited in place of more polished, complete datasets with full metadata to guide interpretation of genome data. In the absence of open data, data consumers are frequently unable to access contextual information (for example, physicochemical parameters, geolocations) of the field site that are essential to accurately interpret the data, to the detriment of downstream analyses. These issues are exacerbated by public sequence databases lacking (links to) the associated metadata and a general lack of familiarity of many data consumers with the literature on, or the environmental context of, a specific system.

Perceived threats to research and career goals

Duplication of effort and the potential for lowered impact or difficulties publishing replicated results are a loss for both creators and consumers. For data creators, raw data underlying published scientific results must be made public to meet expectations for reproducibility. However, unrestricted access to public data can compromise permits, site access agreements and research ethics board approvals, all of which can negatively impact the data creators' and their mentees' ongoing research. For data consumers, even unintentional reuse of restricted data can slow research progress while appropriate permissions are sought, delay publications while data are removed and, in extreme cases, lead to paper retractions. The perceived threat is that unauthorized data reuse can also negatively impact planned research directions, funded research goals, acquisition of new funds or career perspectives of early career researchers for both creators and consumers. A lack of formal structure for data reuse causes tension for data creators and consumers alike.

A roadmap to reduce tension between data creators and data consumers

There are multiple potential avenues for mitigating the three conflicts of interest discussed above, yet not every approach is suitable or can be realized. For instance, funding agencies have the power to set rules for data release and data reuse in principle. However, besides the differentiation between private and taxpayer-funded agencies, funders usually have diverging agendas that not only differ across political borders but are also heterogeneous within a single country. To address the current tension(s) between data creators and data consumers and to update the existing agreements from more than 15 years ago, we propose a comprehensive roadmap for data reuse (Fig. 2). We recommend following this roadmap except in cases in which institutions or funding agencies have a different policy for data reuse in place or there is a restricting licence associated with the dataset itself. This roadmap was developed with the aim of minimizing friction between data creators and data consumers, while promoting open science, and involves the introduction of a new machine-readable DRI metadata tag for facilitating communication between data creators, generators and consumers.

Transparent, equitable and ethical use of public data necessitates clear labelling of its usability by data distributors. Although we initially considered a system that places limitations on free data use, following consultations with the Genomic Standards Consortium and polling 306 microbiome scientists, we have converged on an approach that focuses on both simplicity and openness while achieving nearly all the desired effects. The DRI tag would attach ORCIDs of the data creators to deposited data, signalling that data creators wish to be contacted (for example, via email) before data use[28]. This way, the DRI tag also provides stable contact information to allow data consumers to easily reach data creators. ORCIDs are both free and ubiquitous and, most importantly, are already used internally by the INSDC community. The absence of a DRI tag would signal that the authors of the dataset agree to its reuse without the need to be further involved or contacted.

https://doi.org/10.1038/s41564-025-02116-2

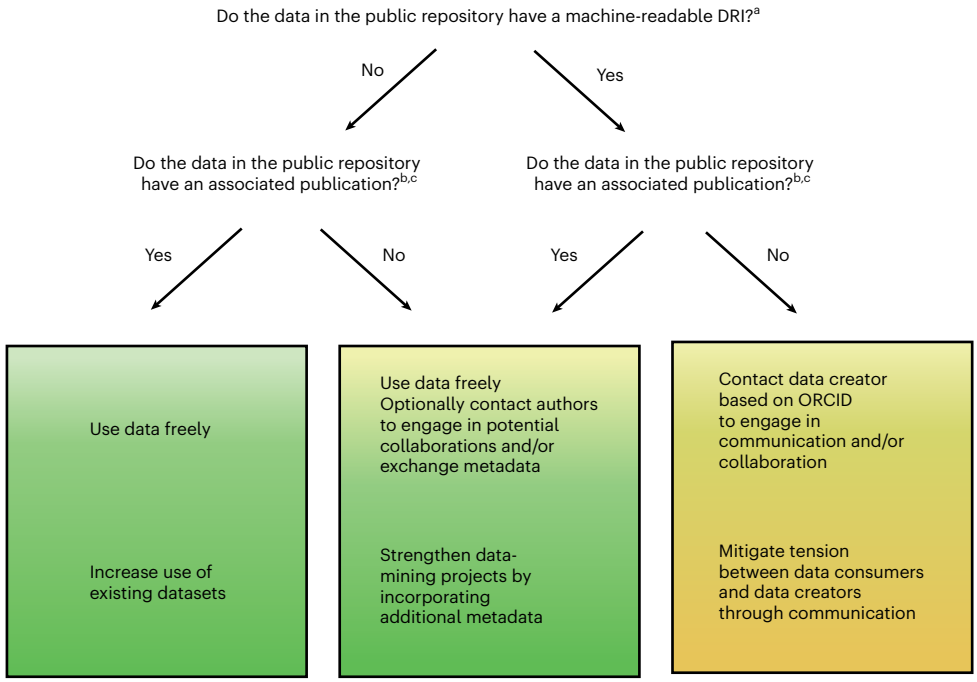


Fig. 2 | Recommendations for equitable reuse of public microbiome data. This flow diagram is applicable unless there is a differing policy in place by the institution using the data consumer or there is a restricting licence associated with the dataset itself. [a]DRI can be updated or added. [b]Manuscripts on a preprint server or in a peer-reviewed journal are included in these recommendations. [c]Repositories could more easily update publication status.

The DRI will consist of a tag with one or more associated ORCIDs identifying the data creators. In computer science notation, the DRI tag will have the following structure:

DRI = {ORCID1, ORCID2, …}

We note that, implicitly, data consumers are expected to acknowledge or cite any data they use for their scientific work. For the new ethical use of data with DRI, we expect data consumers to follow the approach summarized in Fig. 2.

Sequencing datasets published in public databases (for example, in Genbank) have tags, attributes and fields that indicate which publications are connected to the respective datasets. Examples of such tags are the following: (1) for GenBank entries, 'REFERENCE', 'REFERENCE/AUTHORS', 'REFERENCE/TITLE' and 'REFERENCE/JOURNAL'; (2) for BioSample, 'reference for biomaterial'; and (3) for BioProject, 'Publications'. The content of these tags is input initially and/or updated by the data submitter. In theory, they can be updated later either manually by the submitter or automatically by the system. Large scientific literature databases such as PubMed (https://pubmed.ncbi.nlm.nih.gov/) and Europe PMC (https://europepmc.org/) actively monitor published scientific articles and index listings of sequencing accession numbers, generating crucial linkage information that can help address updating such information in public sequence databases. However, as of 6 November 2024, only 147,632 and 78,430 publications for PubMed and Europe PMC, respectively, have had 196,632 and 91,440 sequence accession numbers assigned. Linkage information on literature and sequence data stored in these databases contrasts poorly with the scale of exponentially growing data hosted in the NCBI SRA, amounting to 9 million accession numbers as of 28 March 2023 (12 petabytes of sequence data). Coordination between publication and sequencing databases could in theory be improved if both database types use ORCID associated with their entries. In our roadmap, the DRI, which will contain the ORCID of at least one of the data creators (typically the corresponding creators, that is, the project leader), could address these issues. The content of this tag would be input or updated solely by the data submitter, preferably during the initial data deposition.

The presence of a DRI tag indicates that data creators prefer to be contacted if a data consumer reuses their data, especially if the respective data have no associated publication. The intentions behind this preference can be manifold, including a willingness to share additional metadata or datasets, or the preference to collaborate to help protect early career researchers' (for example, PhD students) ability to finish their studies and graduate. Including one or several ORCID(s) with the DRI will provide a stable point of contact, bolster transparency in science and adherence to the FAIR principles (findable, accessible, interoperable and reusable)[6] and also facilitate science through the exchange of metadata and increased collaboration. There have been instances in which such collaboration with data creators, for example, provision of additional metadata that are not publicly available, has strengthened the content of research studies[29,30]. This should be the norm rather than the exception. In other cases, communication with data creators has allowed proper citation of datasets used, and acknowledgement of funding that supported critical datasets, thus providing some benefit to the data creator for their data reuse[29,31].

In the long run, the DRI tag would help facilitate automatic updates of the publication-associated tags. For example, a GenBank sequence entry could be updated as follows: if the GenBank entry or its corresponding BioProject and BioSample entries are mentioned in a PubMed-indexed publication together with the ORCID(s) found in the DRI, then the new publication will be automatically added to the respective publication tags. The proposed metadata tag (that is, the DRI) will substantially reduce the amount of time needed to clarify the status of any public dataset, enabling automated rules for dataset screening and reducing tension between data creators and data consumers in the future. The DRI thus bridges a gap generated by the FAIR principles, for those who are invested in making their data open access, but where it is equitable for the creator to maintain some control over its reuse.

Ideally, DRIs would propagate automatically within databases (for example, from datasets of sequencing reads to the assembled genomes) and across other online databases of -omics sequence data for a given data creator. In the absence of automatic connections, data

https://doi.org/10.1038/s41564-025-02116-2

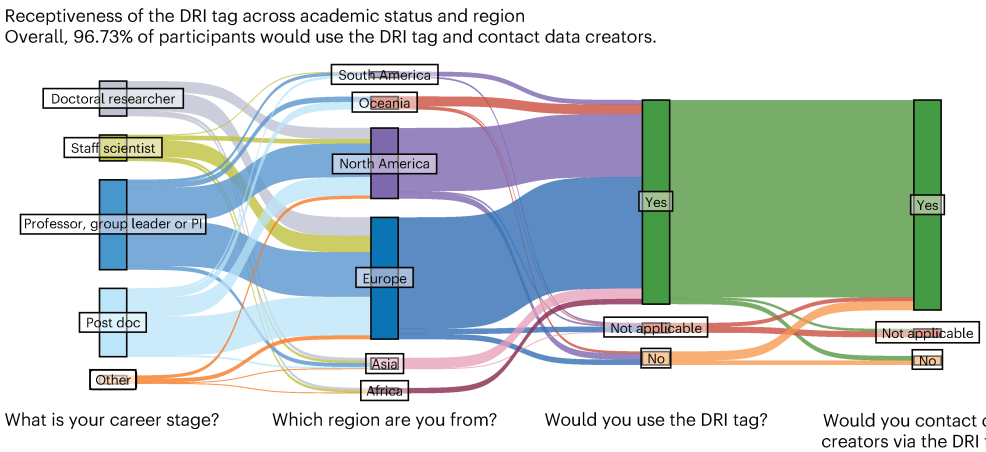


Fig. 3 | The receptiveness of survey participants towards implementing a DRI tag. Flow height (response category, in colour) per stage (survey question, *x*-axis) is proportional to number of responses. See Supplementary Table 2 for respondent numbers in each category. PI, principal investigator.

consumers can manually add DRIs to downstream data depositions (for example, metagenome assembled genomes) via a custom metadata field. These can be the original DRIs from the original data creator or, following conversation with the data creator, new DRIs connected to the data consumer.

Outlook

In this Consensus Statement, we propose a roadmap to facilitate equitable data reuse in the microbiome field. The implementation of machine-readable labels reflecting contact information of data creators will permit efficient reuse of data, accelerate scientific discoveries and hopefully lead data creators to more readily share their valuable datasets with the scientific public at an earlier stage. Using the standards devised by the GSC[32], complete metadata are expected to be made available by data creators. Here we propose the addition of a DRI tag to GSC metadata. Enabling equitable use of public microbiome data will rely on close collaborative work between the data creators, data distributors and data consumers. We envision that the GSC will play a pivotal role in helping all parties (that is, data distributors) implement a DRI tag within submission systems for microbiome data to public repositories, as well as in socializing the new approach to data mining (that is, the flowchart and algorithm given above). We note that even tangible incentives (for example, MG-RAST granting priority access to computing for data made public[33]) have not alleviated data creators' hesitancy to make their data available. However, while we do not anticipate the DRI will achieve a complete resolution to this challenge, we do think it is an essential first step in the right direction.

The full adoption of a DRI tag for microbiome data in public repositories will ultimately require broad support from the scientific community, data distributors, and journals and publishing houses. As an encouraging first step, the ENA has independently implemented an ORCID metadata category, allowing data creators to attach identifying information to their submissions. Propagation of this practice to other major databases will set the stage for the DRI to be used to screen for data availability. In the meantime, data creators are encouraged to apply DRI metadata tags to their datasets. This will allow data consumers to connect with data creators to discuss data reuse. We, the Data Reuse Core Team and Data Reuse Consortium, propose that scientific manuscripts partially or exclusively making use of public data should include a written statement by the authors confirming that they have complied with these guidelines for public data use. This statement would include protocols for data download and use of analysis tools, or reproducible workflows describing how the tag was incorporated in the workflow or, in case of a missing DRI, how authors adhered to the roadmap outlined in Fig. 2.

We are aware that the implementation of the DRI could substantially impact the timeline from analysis to publication by increasing the workload (for example, needing to identify email addresses of data creators). This is especially true for projects accessing many datasets (for example, mining single genes for phylogenetic trees). We are confident that, with time, automated and standardized informatic tools will become available that will lower the administrative burden of following the DRI guidelines.

At the same time, the high proportion of participants who stated that they would respect the DRI when reusing data (266 participants (96.73%); Fig. 3) suggests that data creators will be able to more freely and more frequently share their data in the future. This would facilitate adoption of the FAIR principles in science and hugely benefit the scientific community in the long run. Moreover, by fostering collaborations, the scientific best practices, and the roadmap for data sharing in microbiome research as introduced here, will enable research by lower-resourced laboratories, reducing financial bias in scientific progress. However, we do not expect that the recommendations in this roadmap will be applied retroactively to datasets already deposited in public databases before the implementation of the DRI.

The scale of nucleic acid sequence data required early pioneers to establish public databases across political borders and to confront data sharing considerations early. Other omics technologies are maturing (for example, proteomics), and scientists are recognizing the need to establish data mining approaches[34]. We propose that this roadmap for equitable reuse of public sequence data should be expanded to other fields including but not limited to proteomics, lipidomics, metabolomics, phenomics, microscopy and spectroscopy as data mining becomes routine with these types of data.

References

1. The Wellcome Trust *Sharing Data from Large-Scale Biological Research Projects: A System of Tripartite Responsibility* (National Human Genome Research Institute, 2003).
2. *Report of the International Strategy Meeting on Human Genome Sequencing held at the Princess Hotel, Southampton, Bermuda, on 25th–28th February 1996* (unpublished manuscript, 1996); http://hdl.handle.net/10161/7715
3. Toronto International Data Release Workshop Authors. Prepublication data sharing. *Nature* **461**, 168–170 (2009).
4. Parties to the Convention on Biological Diversity. *Nagoya Protocol on Access to Genetic Resources and the Fair and Equitable Sharing of Benefits Arising from their Utilization to the Convention on Biological Diversity* 234–249 (Official Journal of the European Union, 2014); https://eur-lex.europa.eu/eli/agree_prot/2014/283/oj

5. *GenBank and WGS Statistics* (National Center for Biotechnology Information, accessed February 2024); https://www.ncbi.nlm.nih.gov/genbank/statistics/
6. Wilkinson, M. D. et al. The FAIR Guiding Principles for scientific data management and stewardship. *Sci. Data* **3**, 160018 (2016).
7. Data Management. *NIH Grants & Funding* https://sharing.nih.gov/data-management-and-sharing-policy/data-management (2025).
8. Womack, R. P. Research data in core journals in biology, chemistry, mathematics, and physics. *PLoS ONE* **10**, e0143460 (2015).
9. Zuiderwijk, A. & Spiers, H. Sharing and re-using open data: a case study of motivations in astrophysics. *Int. J. Inf. Manag.* **49**, 228–241 (2019).
10. Uhlir, P. *Developing Data Attribution and Citation Practices and Standards: Summary of an International Workshop* (National Academies, 2012).
11. Amann, R. I. et al. Toward unrestricted use of public genomic data. *Science* **363**, 350–352 (2019).
12. Borgman, C. L. in *Theories of Informetrics and Scholarly Communication* (ed. Sugimoto, C. R.) 93–116 (De Gruyter, 2016).
13. Cousijn, H., Feeney, P., Lowenberg, D., Presani, E. & Simons, N. *Bringing citations and usage metrics together make data count* **18**, 9 (2019).
14. Rourke, M., Eccleston-Turner, M., Phelan, A. & Gostin, L. Policy opportunities to enhance sharing for pandemic research. *Science* **368**, 716–718 (2020).
15. Hug, L. Contribution needed for developing a new community standard for reusing sequencing data. *Springer Nature Research Communities* https://communities.springernature.com/posts/contribution-needed-for-developing-a-new-community-standard-for-reusing-sequencing-data (2024).
16. Soares, A. R. *Data Usage Manuscript - Data Deposition* (OSFHOME, 2025); https://osf.io/skw4a/
17. Wickham, H. et al. Welcome to the Tidyverse. *J. Open Source Softw.* **4**, 1686 (2019).
18. Wickham, H. et al. *dplyr: a grammar of data manipulation*. R package version 3.1 (2023).
19. Zhu, H. et al. kableExtra: construct complex table with 'kable' and pipe syntax. R package version 3.1 (2024).
20. EMBL-EBI Terms of Use. *EMBL-EBI* https://www.ebi.ac.uk/about/terms-of-use/ (2025).
21. Licensing of EMBL-EBI data resources. *EMBL-EBI* https://www.ebi.ac.uk/licencing/ (2025).
22. Sunagawa, S. et al. Tara Oceans: towards global ocean ecosystems biology. *Nat. Rev. Microbiol.* **18**, 428–445 (2020).
23. Proctor, L. M. et al. The Integrative Human Microbiome Project. *Nature* **569**, 641–648 (2019).
24. Lee, M. D. GToTree: a user-friendly workflow for phylogenomics. *Bioinformatics* **35**, 4162–4164 (2019).
25. Chaumeil, P.-A., Mussig, A. J., Hugenholtz, P. & Parks, D. H. GTDB-Tk: a toolkit to classify genomes with the Genome Taxonomy Database. *Bioinformatics* **36**, 1925–1927 (2020).
26. Jennings, L. et al. Applying the 'CARE Principles for Indigenous Data Governance' to ecology and biodiversity research. *Nat. Ecol. Evol.* **7**, 1547–1551 (2023).
27. Westoby, M., Falster, D. S. & Schrader, J. Motivating data contributions via a distinct career currency. *Proc. R. Soc. B* **288**, 20202830 (2021).
28. Credit where credit is due. *Nature* **462**, 825–825 (2009).
29. Buessecker, S. et al. An essential role for tungsten in the ecology and evolution of a previously uncultivated lineage of anaerobic, thermophilic Archaea. *Nat. Commun.* **13**, 3773 (2022).
30. McKay, L. J. et al. Co-occurring genomic capacity for anaerobic methane and dissimilatory sulfur metabolisms discovered in the Korarchaeota. *Nat. Microbiol.* **4**, 614–622 (2019).
31. Viljakainen, V. R. & Hug, L. A. The phylogenetic and global distribution of bacterial polyhydroxyalkanoate bioplastic-degrading genes. *Environ. Microbiol.* **23**, 1717–1731 (2021).
32. Field, D. et al. The Genomic Standards Consortium. *PLoS Biol.* **9**, e1001088 (2011).
33. Meyer, F. et al. The metagenomics RAST server—a public resource for the automatic phylogenetic and functional analysis of metagenomes. *BMC Bioinform.* **9**, 386 (2008).
34. Vaudel, M. et al. Exploring the potential of public proteomics data. *Proteomics* **16**, 214–225 (2016).

Acknowledgements

The Data Reuse Core Team thanks the numerous participants of our survey who chose not to be listed as authors on this manuscript for their valuable perspectives and the many people who discussed this endeavour with us on multiple occasions. We thank L. Rothe for consultations on data visualizations. A.J.P. acknowledges funding by the German Research Foundation (DFG), CRC 1439/1 and CRC 1439/2, project number 426547801 (project INF). L.A.H. acknowledges support from the Canada Research Chairs. R.H. acknowledges support from the US National Science Foundation (OCE-2049445). C. Moraru acknowledges funding by the German Research Foundation (DFG), Priority Program SPP 2330, project number MO 3498/2-1. F.M. acknowledges support from the German Federal Ministry of Education and Research (BMBF project number 01ZZ2013).

Author contributions

A.J.P. was invited to write this manuscript and formed the 'Data Reuse Core Team' via invitation. All members of this team (A.J.P., L.A.H., A.R.S., C. Moraru, F.M. and R.H.) contributed equally to this manuscript. A.J.P., F.M. and A.R.S. led discussions with JGI, ENA and GSC. L.A.H. and A.R.S. performed the analysis and visualization of survey response data. A.H. supervised the construction of the scientific survey, ensuring the quality and impartiality of questions. The 'Data Reuse Consortium' provided support and feedback to this manuscript.

Competing interests

R.H. was a member (2021–2024) of the User Executive Committee of the US Department of Energy's (DOE) Joint Genome Institute. A.P. is an affiliate scientist at JGI and sits on the Prokaryotic Advisory Committee. All opinions expressed in this paper are the authors' and do not necessarily reflect the policies and views of the DOE. R.K. is a scientific advisory board member and consultant for BiomeSense, Inc., has equity and receives income. He is a scientific advisory board member and has equity in GenCirq. He has equity in and acts as a consultant for Cybele. He is a co-founder of Biota, Inc. and has equity. He is a co-founder of Micronoma and has equity and is a scientific advisory board member. He is a board member of Microbiota Vault, Inc. He is a board member of N=1 IBS advisory board and receives income. He is a Senior Visiting Fellow of HKUST Jockey Club Institute for Advanced Study. The terms of these arrangements have been reviewed and approved by the University of California, San Diego in accordance with its conflict of interest policies.

Additional information

Supplementary information The online version contains supplementary material available at https://doi.org/10.1038/s41564-025-02116-2.

Correspondence and requests for materials should be addressed to Alexander J. Probst.

Peer review information *Nature Microbiology* thanks Marnix Medema and the other, anonymous, reviewer(s) for their contribution to the peer review of this work.

Reprints and permissions information is available at www.nature.com/reprints.

Springer Nature or its licensor (e.g. a society or other partner) holds exclusive rights to this article under a publishing agreement with the author(s) or other rightsholder(s); author self-archiving of the accepted manuscript version of this article is solely governed by the terms of such publishing agreement and applicable law.

© The Author(s), under exclusive licence to Springer Nature Limited 2025, corrected publication 2025

The Data Reuse Consortium

R. Z. Abdallah[10], A. Abdalrahem[11], N. Abdulkadir[12], I. M. Adesiyan[13], L. Alteio[14], K. Anantharaman[15], R. Anderson[16], A-S. Andrei[17], J. A. Baeza[18,19], F. Bak[20], B. Baker[21], A. Bartholomäus[22], N. Bejerman[23], J. Biddle[24], A. Bissett[25], J. A. Blakeley-Ruiz[26], K. Block[27], J. Boldt[28], G. Bonilla-Rosso[29], T. L. Bornemann[5,6], V. S. Brauer[30], W. Brazelton[31], A. Bremges[32], E. Buelow[33], Z. M. Burcham[34], A. Cansdale[35], J. G. Caporaso[36], T. Cernava[37], I. Chatzigiannidou[38], R. Costa[39], C. R. Currie[40], A. Daebeler[41], V. De Anda[42], A. De Santiago[43], L. M. Arake de Tacca[44], J. Debelius[45], S. M. Dittami[46], X. Dong[47], M. Džunková[48], A. Edwards[49], R. Edwards[50], S. Egbert[51], J. C. Engelmann[52], S. P. Esser[5], T. J. G. Ettema[53], C. L. Ettinger[54], A. Petrovic Fabijan[55,56], R. M. W. Ferguson[57], P. Ferretti[58], P. Foucault[59], J. A. Fuhrman[60], A. M. Gada[61], P. Geesink[53], I. R. Gerhardt[62], M. O. Gessner[12,63], D. Giovannelli[64], D. Gittins[65], G. B. Gloor[66], R. A. González-Pech[67], C. Gopalakrishnappa[68], C. Greening[69], R. Gregor[68], A. C. Gregory[70], H.-P. Grossart[12,71], M. Groussin[72], B. Valenzuela Guerrero[73], M. Guzel[74], N. Hamamura[75], T. L. Hamilton[76], J. N. Hamm[77], L. Hart[78], C. Hassenrück[79], M. Hay[80], R. M. Hechler[81], P. Hellwig[82], M. Henson[83], M. Herold[84], P. J. Hesketh-Best[76], M. Hess[85], L. Hillary[85], T. C. Hitch[86], S. S. Hivarkar[87], K. J. Hoff[88], E. F. Hom[89], S. Hou[90], L. W. Hugerth[91], Y. Hwang[92], N. Ilott[93], Z. J. Jay[3], S. P. Jungbluth[94], E. Karimi[95], Y. M. Kaspareit[96], C. Keating[97], M. Kellom[98], E. A. Kiledal[99], I. Klarenberg[100], R. Knight[101,102,103,104], A. K. Koech[105], E. V. Koonin[106], K. Kormas[107], K. Kujala[108], N. C. Kyrpides[98], S. L. La Rosa[109], C. C. Laczny[110], K. Lahmers[111], X. Lan[112], A. A. Lateef[113,114], S. H. Lau[115], F. Leese[6,116], M. Á. Lezcano[117], S. S. Li[118], R. N. Lima[119], S. Lücker[120], A. Mahnert[121], S. Majidian[122], L. Malfertheiner[123], A. Marshall[124], S. Meaden[35], C. J. Meehan[125], D. V. Meier[126], C. Melkonian[127,128], D. R. Mende[129], J. L. Meyer[130], G. Michoud[131], V. Mikryukov[132], S. Miravet-Verde[133], J. Muschiol[134], M. K. Nata'ala[135,136], J. D. Neufeld[1], S. Neuhauser[137], O. Osuolale[138], J. Osvatic[139,140], K. M. Pappas[141], D. H. Parks[142], R. H. Parry[143], P. V. Pascoal[44], C. Pavloudi[144], B. Peyton[145], J. Plewka[5], M. Poyet[146], T. Priest[133], E. K. Quaye[146], S. Ramganesh[147], T. Rattei[148], P. Rausch[72], E. Rech[149], C. Rinke[150], C. Robinson[151], A. Rodríguez-Gijón[152], L. M. Rodriguez-R[153,216], R. R. Rohwer[154], T. Roloff[155], J. A. Rothman[156], S. Rückert[6,157], S. E. Ruff[158], J. S. Saini[159], M. G. Santiago-Martínez[160], L. Santoferrara[161], M. S. Sarhan[162], J. H. Saw[163], T. Sbaffi[164], R. B. Schäfer[165], G. Schaible[145], M. Schloter[166], R. A. Schmitz[167], C. Schubert[168], O. Schwengers[169], L. Sehnal[170], A. Sekar[171], J. Sekar[172], M. M. Seyoum[173], M. B. Shah[5], I. Sharon[174], B. Siebers[6,175], E. T. Sieradzki[176], D. Skliros[177], O. L. Snoeyenbos-West[178], A. Sorbie[179], D. R. Speth[148], C. G. Sprehn[98], P. Srivastava[180], T. L. Stach[5,6], J. E. Stajich[156], J. Starke[5], A. D. Steen[181], R. Stöckl[182], T. Stoikidou[183], N. Stopnisek[184], R. Sukumaran[185], B. Sures[6,186], S. Suzuki[187], D. Tamarit[127], P. Thieringer[188], R. Y. Tito[189,190], C. B. Trivedi[191], G. Trubl[192], J. Truu[193], M. Tsiknia[194], J. Ugalde[195], L. E. Valentin-Alvarado[196], X. Vázquez-Campos[197], J. Vierheilig[198,199], F. A. B. von Meijenfeldt[200], M. Wagner[148], C. J. Walsh[201], S. Wang[202], Y. Wang[203], C.-E. Wegner[204], T. Weir[205], L. C. Weiss[206], J. L. Weissman[207], A. Wichels[208], C. L. Williams[49], T. A. Williams[209], A. Z. Worden[158], T. Woyke[98], M. Wu[210], W. Xiu[211], Y. Zhang[212], J. Zhu[213], R. M. Ziels[214] & B. Zwirzitz[215]

[10]Department of Bioinformatics and Genomics, College of Biotechnology, Misr University for Science and Technology, Giza, Egypt. [11]Université de Lorraine, INRAE, IAM, Nancy, France. [12]Department of Plankton and Microbial Ecology, Leibniz Institute of Freshwater Ecology and Inland Fisheries (IGB), Berlin, Germany. [13]Department of Environmental and Occupational Health, School of Public Health, University of Medical Sciences, Ondo, Nigeria. [14]Austrian Competence Centre for Feed and Food Quality, Safety and Innovation, FFoQSI GmbH, Tulln an der Donau, Austria. [15]Department of Bacteriology, University of Wisconsin-Madison, Madison, WI, USA. [16]Carleton College, Northfield, MN, USA. [17]Department of Plant and Microbial Biology, University of Zurich, Zurich, Switzerland. [18]Department of Biological Sciences, Clemson University, Clemson, SC, USA. [19]Departamento de Biologia Marina, Universidad Catolica del Norte, Coquimbo, Chile. [20]Austrian Institute of Technology, Wien, Austria. [21]The University of Texas at Austin, Austin, TX, USA. [22]GFZ Helmholtz Centre for Geosciences, Potsdam, Germany. [23]CONICET, Buenos Aires, Argentina. [24]University of Delaware, Newark, DE, USA. [25]CSIRO, Canberra, Australian Capital Territory, Australia. [26]Department of Plant and Microbial Biology, North Carolina State University, Raleigh, NC, USA. [27]Universitätsklinikum Essen—IKIM, Essen, Germany. [28]Leibniz Institute DSMZ - German Collection of Microorganisms and Cell Cultures, Braunschweig, Germany. [29]Agroscope, Zurich, Switzerland. [30]Aquatic Microbiology, Environmental Microbiology and Biotechnology, Faculty of Chemistry, University of Duisburg-Essen, Essen, Germany. [31]School of Biological Sciences, University of Utah, Salt Lake City, UT, USA. [32]Evonik Operations GmbH, RD&I Biotechnology, Halle, Germany. [33]CNRS, UMR 5525, VetAgro Sup, Grenoble INP, TIMC, Univ. Grenoble Alpes, Grenoble, France. [34]University of Tennessee, Knoxville, TN, USA. [35]Department of Biology, University of York, York, UK. [36]Department of Biological Sciences, Northern Arizona University, Flagstaff, AZ, USA. [37]School of Biological Sciences, Faculty of Environmental and Life Sciences, University of Southampton, Southampton, UK. [38]Department of Biotechnology and Biomedicine, Technical University of Denmark, Kongens Lyngby, Denmark. [39]Department of Bioengineering, Instituto Superior Técnico, University of Lisbon, Lisbon, Portugal. [40]Department of Biochemistry and Biomedical Sciences, M.G. DeGroote Institute for Infectious Disease Research, McMaster University, Hamilton, Ontario, Canada. [41]Institute of Soil Biology and Biogeochemistry, Biology Centre CAS, České Budějovice, Czechia. [42]Department of Microbiology and Cell Sciences, Fort Lauderdale Research and Education Center, University of Florida, Fort Lauderdale, FL, USA.

[43]Institute of Bioinformatics, University of Georgia, Athens, GA, USA. [44]Embrapa Cenargen, Brasília, Brazil. [45]Johns Hopkins Bloomberg School of Public Health, Baltimore, MD, USA. [46]UMR8228, Station Biologique de Roscoff, Roscoff, France. [47]Third Institute of Oceanography, Ministry of Natural Resources, Xiamen, China. [48]Institute for Integrative Systems Biology (I2SysBio), University of Valencia and Spanish National Research Council, Valencia, Spain. [49]Aberystwyth University, Aberystwyth, UK. [50]Flinders University, Adelaide, South Australia, Australia. [51]Washington University at St. Louis, St. Louis, MO, USA. [52]NIOZ—Royal Netherlands Institute for Sea Research, Den Burg, The Netherlands. [53]The Laboratory of Microbiology, Wageningen University and Research, Wageningen, The Netherlands. [54]University of California, Riverside, CA, USA. [55]Westmead Institute for Medical Research, Sydney, New South Wales, Australia. [56]Sydney Medical School, The University of Sydney, Sydney, New South Wales, Australia. [57]The University of Essex, Colchester, UK. [58]Section of Genetic Medicine, Department of Medicine, University of Chicago, Chicago, IL, USA. [59]Biology of Intracellular Bacteria Unit, Pasteur Institute, Paris, France. [60]University of Southern California, Los Angeles, CA, USA. [61]Sokoto State University, Sokoto, Nigeria. [62]Embrapa Agricultura Digital, Campinas, Brazil. [63]Institute of Ecology, Berlin Institute of Technology (TUB), Berlin, Germany. [64]Department of Biology, University of Naples Federico II, Naples, Italy. [65]University of California, Berkeley, Berkeley, CA, USA. [66]Western University, London, Ontario, Canada. [67]Department of Biology, The Pennsylvania State University, University Park, PA, USA. [68]Department of Civil and Environmental Engineering, Massachusetts Institute of Technology, Cambridge, MA, USA. [69]Monash University, Melbourne, Victoria, Australia. [70]University of Calgary, Calgary, Alberta, Canada. [71]Institute of Biochemistry and Biology, Potsdam University, Potsdam, Germany. [72]Institute of Clinical Molecular Biology, Kiel University, Kiel, Germany. [73]Departamento de Educación, Facultad de Educación, Universidad de Antofagasta, Antofagasta, Chile. [74]Department of Microbiology, University of Tennessee, Knoxville, TN, USA. [75]Department of Biology, Faculty of Science, Kyushu University, Fukuoka, Japan. [76]Plant and Microbial Biology, University of Minnesota, Minneapolis, MN, USA. [77]Department of Marine Microbiology and Biogeochemistry, Royal Netherlands Institute for Sea Research (NIOZ), Texel, The Netherlands. [78]Life Sciences Institute, University of Michigan, Ann Arbor, MI, USA. [79]Leibniz Institute for Baltic Sea Research Warnemünde (IOW), Rostock, Germany. [80]Royal Veterinary College, London, UK. [81]Department of Ecology & Evolutionary Biology, University of Toronto, Toronto, Ontario, Canada. [82]Otto-von-Guericke-University Magdeburg, Magdeburg, Germany. [83]Northern Illinois University, DeKalb, IL, USA. [84]Luxembourg Institute of Science and Technology, Esch-sur-Alzette, Luxembourg. [85]University of California, Davis, CA, USA. [86]University Hospital of RWTH Aachen, Aachen, Germany. [87]Agharkar Research Institute, Pune, India. [88]University of Greifswald, Greifswald, Germany. [89]University of Mississippi, Oxford, MS, USA. [90]Southern University of Science and Technology, Shenzhen, China. [91]Uppsala University, Uppsala, Sweden. [92]Harvard University, Cambridge, MA, USA. [93]University of Oxford, Oxford, UK. [94]Lawrence Berkeley National Laboratory, United States Department of Energy, Berkeley, CA, USA. [95]Université Paris-Saclay, INRAE, AgroParisTech, Micalis Institute, Jouy-en-Josas, France. [96]Wageningen University & Research, Wageningen, The Netherlands. [97]Durham University, Durham, UK. [98]US Department of Energy Joint Genome Institute, Lawrence Berkeley National Laboratory, Berkeley, CA, USA. [99]University of Michigan, Ann Arbor, MI, USA. [100]Systems Ecology, Amsterdam Institute for Life and Environment (A-LIFE), Faculty of Science, Vrije Universiteit, Amsterdam, The Netherlands. [101]Department of Pediatrics, University of California, San Diego, San Diego, CA, USA. [102]Department of Computer Science & Engineering, University of California, San Diego, San Diego, CA, USA. [103]Shu Chien-Gene Lay Department of Bioengineering, University of California, San Diego, San Diego, CA, USA. [104]Halıcıoğlu Data Science Institute, University of California, San Diego, San Diego, CA, USA. [105]Masinde Muliro University of Science and Technology, Kakamega, Kenya. [106]Computational Biology Branch, Division of Intramural Research, National Library of Medicine, National Institutes of Health, Bethesda, MD, USA. [107]Department of Ichthyology and Aquatic Environment, University of Thessaly, Volos, Greece. [108]Natural Resources Institute Finland, Helsinki, Finland. [109]Faculty of Chemistry, Biotechnology and Food Science, Norwegian University of Life Sciences, Aas, Norway. [110]Luxembourg Centre for Systems Biomedicine, University of Luxembourg, Esch-sur-Alzette, Luxembourg. [111]Virginia Tech, Blacksburg, VA, USA. [112]Nankai University, Tianjin, China. [113]University of Ilorin, Ilorin, Nigeria. [114]Department of Forest Sciences, University of Helsinki, Helsinki, Finland. [115]Department of Microbiology, The University of Hong Kong, Hong Kong, China. [116]Faculty of Biology, University of Duisburg-Essen, Essen, Germany. [117]IMDEA Water Institute, Alcalá de Henares, Madrid, Spain. [118]Microbiome Systems Laboratory, Biomedicine Discovery Institute, Monash University, Melbourne, Victoria, Australia. [119]EMBRAPA/INCT_BioSyn, Brasília, Brazil. [120]Department of Microbiology, RIBES, Radboud University, Nijmegen, The Netherlands. [121]Medical University of Graz, Graz, Austria. [122]Department of Computational Biology, University of Lausanne, Lausanne, Switzerland. [123]Department of Molecular Life Sciences, University of Zurich, Zurich, Switzerland. [124]Department of Biology, Royal Melbourne Institute of Technology, Melbourne, Victoria, Australia. [125]Nottingham Trent University, Nottingham, UK. [126]University of Bayreuth, Bayreuth Center of Ecology and Environmental Research, Bayreuth, Germany. [127]Utrecht University, Theoretical Biology and Bioinformatics, Utrecht, The Netherlands. [128]Wageningen University & Research, Bioinformatics Group, Wageningen, The Netherlands. [129]Amsterdam University Medical Center, Amsterdam, The Netherlands. [130]Department of Soil, Water, and Ecosystem Sciences, University of Florida, Gainesville, FL, USA. [131]River Ecosystems Laboratory, Alpine and Polar Environmental Research Center, Ecole Polytechnique Fédérale de Lausanne EPFL, Lausanne, Switzerland. [132]Institute of Ecology & Earth Sciences, University of Tartu, Tartu, Estonia. [133]Department of Biology, Institute of Microbiology and Swiss Institute of Bioinformatics, ETH Zurich, Zurich, Switzerland. [134]Ocean EcoSystems Biology Unit, Marine Ecology Research Division, GEOMAR Helmholtz Centre for Ocean Research, Kiel, Germany. [135]Department of Computer Science and Interdisciplinary Centre of Bioinformatics, University of Leipzig, Leipzig, Germany. [136]Department of Data Science in Bioeconomy, Leibniz Institute for Agricultural Engineering and Bioeconomy (ATB), Potsdam, Germany. [137]Institute of Microbiology, University of Innsbruck, Innsbruck, Austria. [138]Elizade University, Ilara-Mokin, Nigeria. [139]Division of Clinical Microbiology, Department of Laboratory Medicine, Medical University of Vienna, Vienna, Austria. [140]Joint Microbiome Facility of the Medical University of Vienna and the University of Vienna, Vienna, Austria. [141]Department of Biology, National and Kapodistrian University of Athens, Athens, Greece. [142]The University of Queensland, Brisbane, Queensland, Australia. [143]School of Chemistry and Molecular Biosciences, University of Queensland, Brisbane, Queensland, Australia. [144]European Marine Biological Resource Centre-European Research Infrastructure Consortium (EMBRC-ERIC), Paris, France. [145]Montana State University, Bozeman, MT, USA. [146]Institute of Experimental Medicine, Kiel University, Kiel, Germany. [147]Department of Biochemistry, J.J College of Arts and Science (Autonomous), Pudukkottai, India. [148]Centre for Microbiology and Environmental Systems Science, University of Vienna, Vienna, Austria. [149]National Institute of S&T in Synthetic Biology/EMBRAPA, Brasília, Brazil. [150]University of Queensland, Brisbane, Queensland, Australia. [151]Indiana University, Bloomington, IN, USA. [152]Department of Ecology, Environment, and Plant Sciences, Science for Life Laboratory, Stockholm University, Stockholm, Sweden. [153]Department of Microbiology and Digital Science Center (DiSC), University of Innsbruck, Innsbruck, Austria. [154]UT Austin, Austin, TX, USA. [155]Institute for Medical Microbiology, University of Zurich, Zurich, Switzerland. [156]Department of Microbiology and Plant Pathology, University of California-Riverside, Riverside, CA, USA. [157]Department of Eukaryotic Microbiology, Faculty of Biology, University of Duisburg-Essen, Essen, Germany. [158]Marine Biological Laboratory, Woods Hole, MA, USA. [159]EPFL, Lausanne, Switzerland. [160]Department of Molecular and Cell Biology, The University of Connecticut (UConn), Storrs, CT, USA. [161]Department of Biology, Hofstra University, Hempstead, NY, USA. [162]Institute for Biomedicine, Eurac Research, Bolzano, Italy. [163]The George Washington University, Washington DC, USA. [164]Water Research Institute (IRSA), National Research Council of Italy (CNR), Verbania, Italy. [165]RC One Health Ruhr, Research Alliance Ruhr and Faculty of Biology, University of Duisburg-Essen, Essen, Germany. [166]Helmholtz Munich, Research Unit for Comparative Microbiome Analysis, Munich, Germany. [167]Kiel University, Kiel, Germany.

[168]ETH Zürich, Institute of Microbiology, Zurich, Switzerland. [169]Bioinformatics and Systems Biology, Justus Liebig University Giessen, Giessen, Germany. [170]RECETOX, Faculty of Science, Masaryk University, Brno, Czech Republic. [171]Institute of Medical Genetics and Applied Genomics, Universitätsklinikum Tübingen, Tübingen, Germany. [172]M. S. Swaminathan Research Foundation, Chennai, India. [173]Department of Poultry Science, University of Arkansas, Fayetteville, AR, USA. [174]Tel Hai Academic College, Qiryat Shemona, Israel. [175]Molecular Enzyme Technology and Biochemistry (MEB), Environmental Microbiology and Biotechnology (EMB), Department of Chemistry, University of Duisburg-Essen, Essen, Germany. [176]Department of Agroecology, Aarhus University, Denmark. [177]Laboratory of Environmental Biotechnology, Department of Biotechnology, School of Applied Biology and Biotechnology, Agricultural University of Athens, Athens, Greece. [178]Department of Environmental Science, University of Arizona, Tucson, AZ, USA. [179]Institute for Stroke and Dementia Research, LMU Klinikum, Munich, Germany. [180]School of Engineering, Cardiff University, Cardiff, UK. [181]University of Tennessee - Knoxville, Knoxville, TN, USA. [182]Institute of Microbiology and Archaea Centre, University of Regensburg, Regensburg, Germany. [183]School of Biological Sciences, Institute for Global Food Security, Queen's University Belfast, Belfast, UK. [184]National Laboratory of Health Environment and Food, Maribor, Slovenia. [185]University of Kerala, Kerala, India. [186]Department of Aquatic Ecology, University of Duisburg-Essen, Essen, Germany and Research Center One Health Ruhr, Research Alliance Ruhr, University of Duisburg-Essen, Essen, Germany. [187]RIKEN, Wakō, Japan. [188]University of California Santa Barbara, Santa Barbara, CA, USA. [189]Department of Microbiology, Immunology and Transplantation, KU Leuven, Leuven, Belgium. [190]Center for Microbiology, VIB, Leuven, Belgium. [191]LGC Biosearch Technology, Petaluma, CA, USA. [192]Lawrence Livermore National Laboratory, Livermore, CA, USA. [193]Institute of Molecular and Cell Biology, University of Tartu, Tartu, Estonia. [194]Soil Science and Agricultural Chemistry Lab, Dept. of Natural Resources and Agricultural Engineering, Agricultural University of Athens, Athens, Greece. [195]Universidad Andres Bello, Santiago, Chile. [196]Plant and Microbial Biology, University of California, Berkeley, Berkeley, CA, USA. [197]School of Biotechnology and Biomolecular Sciences, University of New South Wales, Sydney, New South Wales, Australia. [198]Institute of Water Quality and Resource Management, TU Wien, Vienna, Austria. [199]Interuniversity Cooperation Centre Water & Health, Vienna, Austria. [200]NIOZ Royal Netherlands Institute for Sea Research, Yerseke, The Netherlands. [201]Department of Microbiology and Immunology, University of Melbourne at the Peter Doherty Institute for Infection and Immunity, Melbourne, Victoria, Australia. [202]Department of Microbiology, The Chinese University of Hong Kong, Hong Kong, China. [203]School of Life Sciences and Biotechnology, Shanghai Jiao Tong University, Shanghai, China. [204]Heinrich Heine University Düsseldorf, Düsseldorf, Germany. [205]Colorado State University, Fort Collins, CO, USA. [206]Department of Animal Ecology, Evolution and Biodiversity, Ruhr University Bochum, Bochum, Germany. [207]Department of Biology, The City College of New York, New York, NY, USA. [208]Alfred-Wegener-Institut Helmholtz Zentrum für Polar-und Meeresforschung, Helgoland, Germany. [209]University of Bath, Bath, UK. [210]Australian Centre for Water and Environmental Biotechnology, University of Queensland, Brisbane, Queensland, Australia. [211]State Key Laboratory of Geomicrobiology and Environmental Changes (GMEC), China University of Geosciences, Beijing, China. [212]Department of Cell and Molecular Biology, College of the Environment and Life Sciences, University of Rhode Island, Kingston, RI, USA. [213]Li Ka Shing Institute of Health Sciences, Faculty of Medicine, The Chinese University of Hong Kong, Hong Kong, China. [214]Department of Civil Engineering, University of British Columbia, Vancouver, British Columbia, Canada. [215]Institute of Food Science, BOKU University, Vienna, Austria. [216]Present address: Department of Chemistry and Biosciences, Aalborg University, Aalborg, Denmark. A full list of members and their affiliations appears in the Supplementary Information.

2023: Jacques Monod conference "Sex unfolded: sex, asex, sexes"

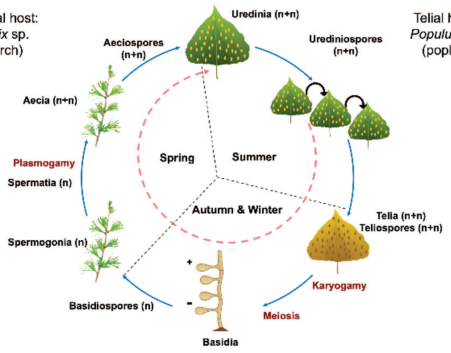
Evolutionary and functional studies of the variability of reproductive modes in the poplar rust fungus, *Melampsora larici-populina*

Ammar ABDALRAHEM, Pascal FREY, Fabien HALKETT, Sébastien DUPLESSIS
INRAE, Université de Lorraine, Interactions Arbres-Microorganismes (IAM), 54000 Nancy, France

Context

Melampsora larici-populina, causing poplar rust exhibits a complex heteroecious life cycle, typical of rust fungi (Pucciniales), alternating between a sexual reproduction phase on larch trees (*Larix* sp.) and an asexual multiplication phase on poplar trees (*Populus* sp.).

Previous population genetic study provided evidence for mixed reproductive mode, with a genetic group that shows hallmarks of long lasting asexual reproduction (Xhaard et al., 2011).



The *Melampsora larici-populina* life cycle

This alternation of sexual and asexual reproduction exhibits some similarities with the cyclical parthenogenesis of aphids and daphnia.

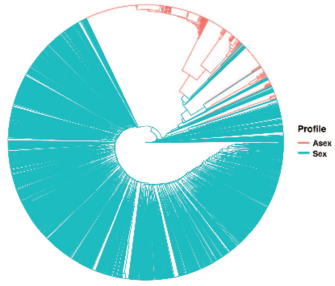
My PhD project proposes to address the issue of variability in reproductive modes and the survival through years of asexual lineages. along **three complementary axes:**

Axis 1: Evolutionary ecology

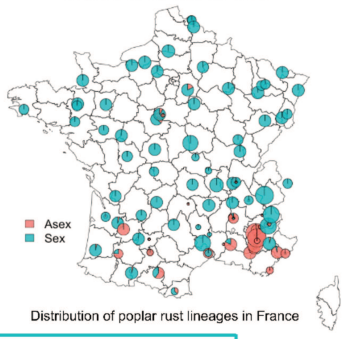
This approach will allow us to determine the geographical distribution of the asexual lineages of *M. larici-populina* using population genetics.

Table 2. Details of asexual lineages

MLL	Isolate No.	MLG No.	Year range
1	255	41	1992:2020
2	23	8	1996:2020
3	43	5	1996:2011
4	38	14	1996:2020
5	13	5	1993:2014
6	49	4	2014:2020
7	14	3	2009:2011



Neighbor Joining tree (SSR data)



Distribution of poplar rust lineages in France

Results

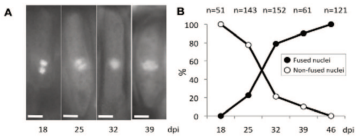
Sexually reproducing lineages are common throughout France, while asexual lineages are more prevalent in southern France. This difference may be influenced by various environmental and/or ecological factors, such as milder winters.

Table 1. Population genetic characteristics

	N	G/N	Ar	Fis	(sd)	rbarD
Asex	438	0.185	4.390	-0.361	0.374	0.361
Sex	1307	0.886	8.771	0.092	0.111	0.018

Axis 3: Phenotypic

To evaluate whether asexual lineages can undergo karyogamy in teliospores similar to sexual lineages, using DAPI staining and a fluorescence microscope



(A) DAPI-stained teliospores nuclei during telia formation, (B) Dynamics of karyogamy. (Hacquard et al., 2013).

Axis 2: Evolutionary transcriptomic

This approach that will aim to elucidate the functional basis of the cycle and to find the determinants of sexual vs. asexual reproductive modes.

Conclusion and perspectives

- The existence of long lasting asexual lineages, largely found in southern France, was validated using a population genetic analysis: some of asexual lineages where repetitively sampled across a time period of 30 years.
- Overall, this thesis aims to contribute to our understanding of the complexity of reproductive mode variability in *M. larici-populina* and how it affects the evolution and adaptation of this important phytopathogenic fungus.

References

Xhaard, C., Fabre, B., Andrieux, A., Gladieux, P., Barrès, B., Frey, P., & Halkett, F. (2011). The genetic structure of the plant pathogenic fungus *Melampsora larici-populina* on its wild host is extensively impacted by host domestication. Molecular Ecology, 20(13), 2739–2755.
Hacquard, S., Delaruelle, C., Frey, P., Tisserant, E., Kohler, A., & Duplessis, S. (2013). Transcriptome analysis of poplar rust telia reveals overwintering adaptation and tightly coordinated karyogamy and meiosis processes. Frontiers in Plant Science, 4: 456.