

DISS. ETH NO. 27662

INTO THE WILD
-
**ADAPTATION GENOMICS OF *EPICHLÖE* GRASS PATHOGENS
IN NATURAL ECOSYSTEMS**

A thesis submitted to attain the degree of
DOCTOR OF SCIENCES of ETH ZURICH
(Dr. sc. ETH Zurich)

presented by

Artemis Diana Treindl

MSc Biology, ETH Zurich
born on February 22nd 1991
citizen of Meilen (ZH)

accepted on the recommendation of

Prof. Dr. Adrian Leuchtman
Prof. Dr. Bruce McDonald
Prof. Dr. Daniel Croll

Contents

SUMMARY	5
ZUSAMMENFASSUNG	8
GENERAL INTRODUCTION	12
CHAPTER I	24
Chromosome-level genomes provide insights into genome evolution, organization and size in <i>Epichloe</i> fungi	
CHAPTER II	61
Genetic diversity and population structure of <i>Epichloe</i> – Fungal pathogens of grasses in natural ecosystems	
CHAPTER III	120
Genome-wide signatures of selection in natural plant pathogen populations	
SUPPLEMENTARY CHAPTER	177
Resolving the phylogeny of the <i>Epichloe typhina</i> species complex using genome-wide SNPs	
SYNTHESIS AND OUTLOOK	194
ACKNOWLEDGEMENTS	204
CURRICULUM VITAE	208

Summary

Pathogenic fungi that are associated with plants compose an integral part of natural biodiversity and present fascinating biological systems to study the genomic basis of adaptation. The lifestyle of these eukaryotic microbes drives rapid and ongoing adaptation by antagonistic selection leaving detectable patterns on relevant genes in the pathogens genomes. Furthermore, some fungal pathogens exhibit particular patterns of genome organization that are thought to further promote rapid evolutionary change. This thesis presents the comparative and population genomic analyses of two pathogenic sibling species of the genus *Epichloe* specialized on different host grasses: *Epichloe typhina* infecting *Dactylis glomerata* and *Epichloe clarkii* infecting *Holcus lanatus*. Using chromosome-level genome assemblies and an unprecedented sampling of seven sympatric population pairs, I investigated the genomic mechanisms, selected genes and dynamic processes underlying adaptive evolution of plant pathogenic fungi in natural ecosystems.

In the first chapter, the completed assemblies and annotations of reference genomes of the two fungal species are presented and the mechanisms that may have differentially shaped the genomes in their evolutionary past are discussed. I show that both genomes display a remarkable bipartite organization, so called two-speed genomes. A gene-rich core genome highly conserved between species was interspersed with blocks of repeat-rich sequence exhibiting an extensive signature of repeat-induced point mutation. This fungal specific defense against the proliferation of transposable elements in the genome generated divergent regions with a high AT-nucleotide content in the *Epichloe* genomes. While the core genomes were of similar size and gene content, these AT-rich regions made up half of the whole genome in *E. clarkii* but only a third in *E. typhina*, resulting in a large difference in genome size between the two species. The results from this chapter demonstrate how the rapid evolution of repeat structure can drive divergence between closely related taxa and highlight the evolutionary role of dynamic compartments in fungal genomes.

The second chapter investigates the genetic structure and diversity of *Epichloe* populations. I sampled seven population pairs of the two sibling species *E. typhina* and *E. clarkii* occurring sympatrically in natural grassland ecosystems across Western Europe. Whole-genome sequencing of 480 haploid fungal isolates was performed and two genome-wide SNP datasets generated, one for each of the study species. Clustering analyses clearly showed that in both species geographically separated populations also formed genetically distinct groups which could result from local adaptation generating and maintaining genetic structure. The amount of gene-flow, however, differed between species with higher levels of genetic differentiation between populations and a stronger effect of isolation by distance in *E. clarkii* compared to *E. typhina*. This pattern may be a result of different dispersal abilities of the two pathogens linked to their spore morphology and is likely also influenced by the genetic structure of host populations. Furthermore, the overall higher genetic diversity within *E. typhina* populations compared to *E. clarkii* suggests highly different effective population sizes. The results from this chapter provide insights into the evolutionary processes and genomic features contributing to population divergence of plant pathogens in natural systems and the strong geographic structure observed in this system is an ideal prerequisite to investigate the genetic basis of adaptation which is the focus of the third chapter.

Here the population genomic datasets were used to detect genome-wide signatures of selective sweeps in *Epichloe* genomes and identify genes putatively involved in host adaptation. Two different selection scan methods were applied to detect signatures of selective sweeps and uncover loci under recent selection at the population level: haplotype-based scans and SFS-based scans. I examined the possibility that the dynamic genome compartments (AT-rich regions) contribute to adaptive evolution in *Epichloe* by testing for an association between these regions and genes under positive selection. Selective sweeps were widespread across the genomes of both species, however, selective sweeps were more abundant in *E. typhina*. This difference may be explained by the larger effective population size of *E. typhina* compared to *E. clarkii* and a therefore a stronger relative effect of selection versus genetic drift. Sweep regions often overlapped with AT-rich regions suggesting a significant role of these regions in adaptive evolution of *Epichloe*. Many different loci were under selection in different populations suggesting that heterogenous selection pressures generate locally adapted populations, but some

selected loci were shared across multiple populations suggesting shared selection regimes and a possible role of these loci in adaptation to the host at the species-level. Genes located in shared sweeps and putatively involved in host-adaptation encoded known effector functions such as small-secreted proteins and cell-wall-degrading enzymes. While the analyses focused on observing evolutionary patterns within populations, the haplotype-structure at candidate loci could also be compared among populations and the observed patterns suggested that different loci are governed by different evolutionary dynamics. This last main chapter delivers insights into the impact and targets of selection in natural populations of fungal plant pathogens and investigates the link between genome organization and adaptation.

Finally, in a short supplementary chapter the phylogenetic and taxonomic ambiguity of the *E. typhina* species complex is addressed, in which *E. typhina* and *E. clarkii* reside. Phylogenetic analysis was performed on genome-wide SNP datasets of 144 isolates collected from 14 different host grass species across Europe prior to and throughout this project and additional genomic data from public data bases. The new phylogeny presented here resolves seven genetically differentiated clades representing closely related cryptic species within in the previously obscure *E. typhina* complex and thus provides a basis for future taxonomic reconsiderations.

The results of this thesis provide insights into the population biology and adaptive evolution of an intriguing and unique biological system, the *Epichloe* fungi. A solid foundation is laid in the form of both genomic resources and the empirical analysis of natural populations to establish a new fungal model system for evolutionary research and to stimulate future scientific investigations focusing on, but not limited to, the adaptation of fungal plant pathogens, genome evolution in eukaryotes and coevolutionary dynamics between pathogens and hosts.

Zusammenfassung

Pathogene Pilze bilden einen allgegenwärtigen Bestandteil der natürlichen Biodiversität und stellen faszinierende biologische Systeme dar, um die genomischen Grundlagen der Adaptation zu untersuchen. Einerseits begünstigt der antagonistische Lebensstil dieser eukaryotischen Mikroben starken Selektionsdruck und treibt damit eine schnelle und fortlaufende Anpassung an, deren genetischer Signatur in den Genomen der Pathogene untersucht werden können. Andererseits besitzen manche pilzliche Pathogene besondere Formen der Genomorganisation, von denen man annimmt, dass sie den schnellen evolutionären Wandel weiter begünstigen. In dieser Dissertation werden vergleichende und populationsgenomische Analysen von zwei pathogenen Geschwisterarten der Gattung *Epichloe* vorgelegt, die auf unterschiedliche Wirtsgräser spezialisiert sind: *Epichloe typhina*, die *Dactylis glomerata* infiziert und *Epichloe clarkii*, die *Holcus lanatus* infiziert. Mit Hilfe von neuen Referenzgenomen auf Chromosomenebene und einer beispiellosen Stichprobe von sieben sympatrischen Populationspaaren untersuchte ich die genomischen Mechanismen, genetischen Ziele und dynamischen Prozesse, die der adaptiven Evolution von pflanzenpathogenen Pilzen in natürlichen Ökosystemen zugrunde liegen.

Im ersten Kapitel werden die fertig assemblierten und annotierten Referenzgenome der beiden Fokusarten präsentiert und die Mechanismen diskutiert, welche die Genome in ihrer evolutionären Vergangenheit unterschiedlich geformt haben könnten. Ich zeige, dass beide Genome eine bemerkenswerte zweiteilige Organisation aufweisen, so genannte «Two-Speed»-Genome. Der genreiche und konservierte Teil des Genoms, das Kerngenom, wies einen bemerkenswerten Grad an Syntänie zwischen den Arten auf, war jedoch mit Blöcken von repetitiven Sequenzen voller Transposons durchsetzt. Diese repetitiven Sequenzen wurden durch «repeat-induced point mutations» hypermutiert, ein pilzspezifischer Abwehrmechanismus, der das Genom vor der Proliferation der Transposons schützen soll. Dieser Vorgang generierte divergente Regionen in den *Epichloe*-Genomen mit einem hohen AT-Nukleotidgehalt. Während die Kerngenome beider Arten von ähnlicher Größe und ähnlichem Gengehalt waren, machten diese AT-

reichen Regionen bei *E. clarkii* die Hälfte des gesamten Genoms aus, bei *E. typhina* jedoch nur einen Drittel, was zu einem großen Unterschied in der Genomgröße zwischen den beiden Arten führte. Die Ergebnisse dieses Kapitels geben Hinweise darauf, wie die schnelle Evolution von repetitiven Teilen des Genoms die Divergenz zwischen eng verwandten Taxa vorantreiben kann und heben die evolutionäre Rolle von dynamischen Kompartimenten in Pilzgenomen hervor.

Das zweite Kapitel untersucht die genetische Struktur und Diversität von *Epichloe*-Populationen. Ich analysierte sieben Populationspaare der beiden Geschwisterarten *E. typhina* und *E. clarkii*, die sympatrisch in natürlichen Grasland-Ökosystemen in Westeuropa vorkommen. Die ganzen Genome von 480 haploiden Pilzisolaten wurden sequenziert und zwei SNP-Datensätze generiert, einer für jede der untersuchten Arten. Clustering-Analysen zeigten deutlich, dass bei beiden Arten geographisch getrennte Populationen auch genetisch differenzierte Gruppen bildeten, was auf lokale Anpassung/Adaptation zurückzuführen sein könnte. Die Menge an Genfluss unterschied sich jedoch zwischen den Arten: Obwohl die Populationen dieselbe geographische Entfernung aufwiesen, war die genetische Differenzierung zwischen den Populationen bei *E. clarkii* grösser im Vergleich zu *E. typhina*. Diesem Muster könnten die unterschiedlichen Ausbreitungsfähigkeiten der beiden Pathogene in Verbindung mit ihrer Sporenmorphologie zugrunde liegen, wobei es ausserdem von der genetischen Struktur der Wirtspopulationen abhängt. Darüber hinaus lässt die insgesamt höhere genetische Diversität innerhalb von *E. typhina*-Populationen im Vergleich zu *E. clarkii* auf sehr unterschiedliche effektive Populationsgrößen schließen. Die Ergebnisse dieses Kapitels geben Einblicke in die evolutionären Prozesse, die zur Populationsdivergenz von Pflanzenpathogenen in natürlichen Systemen beitragen. Die in diesem System beobachtete starke geografische Struktur ist eine ideale Voraussetzung für die Untersuchung der genetischen Grundlagen der Adaptation, die im Mittelpunkt des dritten Kapitels steht.

Hier wurden die populationsgenomischen Datensätze verwendet, um die *Epichloe*-Genome nach Signaturen zu durchsuchen, die auf neue Anpassungen hindeuten, sogenannten selektiven «Sweeps». Diese Analyse sollte dazu dienen die Gene zu identifizieren, die unter anderem an der Wirtsanpassung beteiligt sind. Zwei verschiedene Selektions-Scan-Methoden wurden angewandt, um Signaturen von

selektiven Sweeps zu erkennen und Loci zu identifizieren, die auf Populationsebene unter rezenter Selektion stehen: Haplotyp-basierte Scans und SFS-basierte Scans. Ich untersuchte die Möglichkeit, dass die dynamischen Regionen der Genome, die AT-reichen Sequenzen, zur adaptiven Evolution in *Epichloe* beitragen, indem ich auf einen Zusammenhang zwischen diesen Regionen und Genen unter Selektion testete. Selektive Sweeps waren in den Genomen beider Arten weit verbreitet, jedoch waren sie in *E. typhina* zahlreicher vorhanden. Dieser Unterschied kann mit der größeren effektiven Populationsgröße von *E. typhina* im Vergleich zu *E. clarkii* erklärt werden und somit mit einem stärkeren relativen Effekt von Selektion versus genetischer Drift. Die Selektionsregionen überlappten oft mit AT-reichen Regionen, was auf eine bedeutende Rolle dieser Regionen in der adaptiven Evolution von *Epichloe* hinweist. Viele unterschiedliche Loci standen in verschiedenen Populationen unter Selektion, was darauf hindeutet, dass heterogene Selektionsdrücke lokal angepasste Populationen erzeugen, aber einige Loci standen in mehreren Populationen gleichzeitig unter Selektion, was auf eine mögliche Rolle dieser Loci bei der Wirtsanpassung hindeuten könnte. Gene, die sich in diesen «gemeinsamen» Sweeps befanden, zeigten Ähnlichkeit zu bekannten Effektoren in anderen pflanzenpathogenen Pilzen, wie z. B. kleine sekretierte Proteine und Zellwand-abbauende Enzyme. Während sich die Analysen auf die Beobachtung von Evolutionsmustern innerhalb von Populationen konzentrierten, konnte die Haplotyp-Struktur an Kandidaten-Loci auch zwischen Populationen verglichen werden. Die beobachteten Muster legten nahe, dass verschiedene Loci einer unterschiedlichen Evolutionsdynamik unterworfen sind. Dieses letzte Hauptkapitel liefert Einblicke in die Auswirkungen der Selektion in natürlichen Populationen von pilzlichen Pflanzenpathogenen auf der Ebene der gesamten Genomorganisation und auf der Ebene einzelner Gene.

Schließlich werden in einem kurzen ergänzenden Kapitel die phylogenetischen und taxonomischen Unklarheiten des *E. typhina*-Artenkomplexes angesprochen, in dem *E. typhina* und *E. clarkii* angesiedelt sind. Die phylogenetische Analyse wurde mit genomischen SNP-Datensätzen von 144 Isolaten durchgeführt, die vor und während dieses Projekts von 14 verschiedenen Wirtsgrasarten in ganz Europa gesammelt wurden, sowie mit zusätzlichen genomischen Daten aus öffentlichen Datenbanken. Die hier vorgestellte Phylogenie trennt sieben genetisch differenzierte Gruppen auf, die

möglicherweise eng verwandten kryptischen Arten innerhalb des bisher obskuren *E. typhina*-Komplexes entsprechen. Damit bietet der neue Stammbaum eine Grundlage für zukünftige taxonomische Neubetrachtungen.

Die Ergebnisse dieser Arbeit geben Einblicke in die Populationsbiologie und adaptive Evolution eines faszinierenden und einzigartigen biologischen Systems, den endophytischen *Epichloe*-Pilzen. In Form von genomischen Ressourcen und der empirischen Analyse natürlicher Populationen wird ein solides Fundament gelegt, um ein neues Pilzmodellsystem für die Evolutionsforschung zu etablieren und zukünftige wissenschaftliche Untersuchungen anzuregen mit Fokus auf die Mechanismen der Genomevolution in Eukaryoten und die coevolutionäre Dynamik zwischen Pathogenen und Wirten.

General Introduction

THE SYMBIOTIC CONTINUUM OF PLANT-FUNGAL INTERACTIONS

The kingdom of fungi comprises one of the most diverse and adaptable group of organisms in the natural world. Many fungi colonize living plants and form symbiotic association with them. Such ancient symbioses are thought to date back to over 450 million years ago and have been critical in plant colonization of land and their continued diversification eventually forming the terrestrial ecosystems we know today (Lutzoni et al. 2018, Feijen et al. 2018). The interactions between fungi and plants span the whole symbiotic continuum ranging from mutualistic interactions where both partners benefit, to parasitism where one partner, usually the fungus, exploits the other and is therefore designated a pathogen (Redman et al. 2001, Kogel et al. 2006, but see Leake 2005). This distinction of symbiotic categories based on costs and benefits may imply strict delimitation between them, however, the outcome of the biotic interaction between many species pairs is in fact not fixed. All types of symbiosis are the result of a sophisticated trade-off between the nutrient demand of the fungal symbiont and the defense response of the plant. Mutualism represents a state of tightly regulated balance where the benefits of the interaction outweigh the costs. Conversely, pathogenic interactions indicate a state of disequilibrium in which the cost of harboring the fungal symbiont becomes visible as disease symptoms in the host plant. Indeed, fungal pathogenicity, that is the ability to cause disease and the severeness of symptoms in the host, is a dynamic phenotype and depends on a range of different factors including environmental conditions and genetic compatibility. For example, *Fusarium oxysporum* is a dreaded fungal pathogen of many crop species but this species complex comprises strains that exhibit the whole symbiotic continuum, from pathogenicity through commensalism to mutualism, depending on the fungal strain and host species involved (Ma et al. 2010). Sometimes even small genetic modifications can tip the scale on the symbiotic outcome, for example Eaton et al. (2010) show how deletion of a single gene in *Epichloë festucae* encoding a stress-activated MAP-kinase can switch the symbiotic interaction with its host grass *Lolium perenne* from mutualistic to pathogenic. There are commonalities to different types of symbioses between plants and the fungi colonizing

them: the host plant constitutes the distinct ecological niche that the associated fungus evolves in as its reproduction and dispersal strongly depends on successful infection and establishment in the host tissue. It is therefore likely that the deployment of similar strategies and mechanisms govern plant-fungal interactions across the symbiotic continuum (Kogel et al. 2006, Gladieux 2018), however, fungal plant pathogens make up the most prominent and extensively studied systems among them all.

FUNGAL PATHOGENS ARE FASCINATING MODEL SYSTEMS TO STUDY ADAPTATION

From an evolutionary biologist's perspective, highly specialized and highly detrimental pathogens provide particularly interesting biological systems to study adaptation over short evolutionary time-scales as they are expected to experience strong and ongoing selection pressures imposed by a single most important selective agent, the host (Tellier et al. 2014, Brockhurst et al. 2014, Papkou et al. 2019). As stated above, the biotic interaction the fungal pathogen and its host plant are entangled in is of an antagonistic nature. The pathogen needs to persist in the host environment and receive nutrients and therefore it constantly evolves to overcome host defenses and establish infection while the host plant evolves to reduce the negative fitness effects imposed by the pathogen and develop resistance (Gladieux et al. 2014). This battle plays out in ongoing cycles of adaptation and counteradaptation, also called the coevolutionary arms-race, in which antagonistic selection drives rapid evolutionary change. In the context of this constant struggle to stay adapted, becoming specialized is a successful strategy (if not the often inevitable evolutionary outcome), because biotic heterogeneity, competition with co-infecting genotypes and species, and adaptation to multiple hosts is thought to be costly (Barrett et al. 2009). Not surprisingly, the concept of host adaptation is considered a key driver of ecological divergence in plant-associated fungi, as it can inherently act as a reproductive barrier when closely related fungi are specialized on different host species (Fournier and Giraud 2008, Giraud et al. 2008).

GENOME SCANS REVEAL WHICH GENES UNDERLY ADAPTATION

Population genomic analysis is today's state-of-the-art approach to identify adaptive traits and understand the forces driving divergence of populations and species of fungal pathogens (Grünwald et al. 2016, Everhart et al. 2021). Population genomics employs methods from classical population genetics to a large genome-wide set of genetic markers (Stinchcombe and Hoekstra 2008). Adaptive loci are detected using genome scans based on the local depletion of genetic variation in the genomic region that surrounds the locus under selection, the signature of a selective sweep. Sophisticated genome scan methods have been and are continued to be developed to detect adaptive signals from an ever growing amount of genomic data (Oleksyk et al. 2010, Vitti et al. 2013, Horscroft et al. 2019). Three general classes are differentiated, each identifying statistical outliers relative to the average genome-wide pattern of polymorphism across populations: Linkage-disequilibrium-based methods detecting long, homozygous haplotypes, site-frequency-spectrum-based methods detecting the surplus of rare new mutations at the locus under selection, and population-differentiation-based methods detecting pairwise differences in allele frequencies between populations that emerge when an allele is selected in one population but not the other (e.g. F_{ST}). Choosing the appropriate selection scan method for a given dataset and understanding their different implications is crucial as the sweep signature in genomic data is transient and only revealed at a certain timepoint relative to the onset of selection (Weigand and Leese 2018). Over time, recombination and new mutations will restore genetic variation at the locus and thus remove the signature of selection. How quickly this happens depends on the rate of recombination, for example, high recombination rates in large populations of randomly mating species will break down linkage between the beneficial allele and neighboring variants within a few generations, causing the size of the haplotype affected by a selective sweep to decrease quickly. The effect of selection may therefore be very localized and only evident when the selective sweep is recent enough which is why informative selection scans require a large number of genetic markers distributed along the whole genome.

INTO THE WILD – INVESTIGATING ADAPTATION OF FUNGAL PATHOGENS IN NATURAL ENVIRONMENTS

Fungal pathogens can inflict large economic damage by causing major losses in crop plants (Fisher et al. 2012, McDonald and Stukenbrock 2016), and it is therefore not surprising that most studies on adaptation genomics in plant associated fungi to date have focused on these agricultural systems. Meanwhile, insights concerning the evolutionary biology and adaptive processes in natural populations remain limited to a small number of biological systems such as the anther smut fungi *Microbotryum* (Hartmann et al. 2019). This bias is not due to the lack of fungal plant pathogens in natural environments (Burdon and Laine 2019), but rather because fewer suitable systems for empirical studies have been established so far, which entails the generation of high quality reference genomes. However, with the recent advances in high-throughput sequencing technology, the affordability thereof, and the unceasing development of new computational tools, genomic resources are greatly improved and becoming readily available for an increasing number of non-model species (Bleidorn 2016). The genomics era has therefore paved the way for population genomic research of fungal plant pathogens to be carried into the wild, to study adaptation in natural host-pathogen systems. This research is not only essential to gain a broader understanding of how plants and fungi coevolve and interact in natural systems, but also to be able to predict evolutionary trajectories and assess threats posed by fungal diseases as natural populations may act as reservoirs for emerging pathogens on agricultural crop species (Stukenbrock 2013).

THE *EPICHLOE* STUDY SYSTEM

The work presented in this thesis focuses on the comparative and population genomic analyses of two biotrophic fungal pathogens of the ascomycete genus *Epichloe* infecting different host grasses in natural ecosystems: *Epichloe typhina* infecting *Dactylis glomerata* and *Epichloe clarkii* infecting *Holcus lanatus*. They are closely related yet highly specialized members of the *E. typhina* (Pers.) Brockm. species complex (Craven et al. 2001) and form life-long systemic associations with their host plants (Leuchtman and Clay 1997). *Epichloe* fungi are usually referred to as endophytes in the literature

which by common definition entails 1) the growth of the fungus *within* the host plant tissue and 2) that this infection does not cause symptoms in the plant and is considered to be a mutualistic association (i.e. providing a beneficial effect to both the fungus and its host) (Kogel et al. 2006). There has been, however, considerable disagreement as to this definition of the term endophyte particularly because many plant associated fungi, including devastating pathogens, have asymptomatic life cycle phases or latency periods within host tissue. The distinction between a fungal plant pathogen and a fungal endophyte is therefore not very clear and throughout this thesis I use the term endophyte to refer more broadly to fungi that colonize internal plant parts asymptotically for at least part of their life cycle (see Petrini 1991). This asymptomatic phase may last all of the fungus life-cycle in the case of mutualists or be temporary in the case of some fungal plant pathogens (see Hardoim et al. 2015). The sexually reproducing focal species of this thesis *E. typhina* and *E. clarkii* J.F. White are prime example of such pathogenic endophytes (Figure 1) (Clay and Schardl 2002). During vegetative growth of their host grasses they grow endophytically in the intercellular space of the above-ground plant tissue without causing symptoms. However with the onset of host flowering, the fungus hijacks grass flowering tillers to complete its own reproductive cycle: it forms external fruiting bodies, so called stromata, which enclose undeveloped inflorescences and inhibit flowering and seed production, a phenomenon referred to as choke disease (White Jr. et al. 1991, White Jr. 1997, Leuchtman 2003). Whole plants are usually infected by a single haploid genotype (Leuchtman and Clay 1997), and obligate outcrossing is mediated by a symbiotic fly of the genus *Botanophila* which transfers gametes between two genotypes of opposite mating types (bipolar heterothallic mating system) (White and Bultman 1987, Bultman and Leuchtman 2008, Schardl et al. 2014). After successful fertilization, perithecia with asci develop containing eight haploid ascospores that are then wind-dispersed and mediate horizontal transmission to new hosts (Raynal 1991, Chung and Schardl 1997, Brem and Leuchtman 1999).

Throughout this thesis I refer to *E. typhina* and *E. clarkii* as sibling species, however, it needs to be noted that based on current taxonomy they are assigned the taxonomic rank of subspecies (Leuchtman et al. 2014). Despite their ability to hybridize in laboratory crosses, previous studies have suggested that *E. typhina* and *E. clarkii* are reproductively

isolated evolutionary lineages at an advanced stage of divergence (Schirrmann et al. 2015, 2018, Treindl and Leuchtman 2019), and this is further supported by the results presented here. Furthermore, we refrain from using diaeresis in *Epichloe* to ensure the name does not contain non-ascii characters as suggested by the International Code of Nomenclature (Turland et al. 2018) and adopted by major databases including MycoBank.

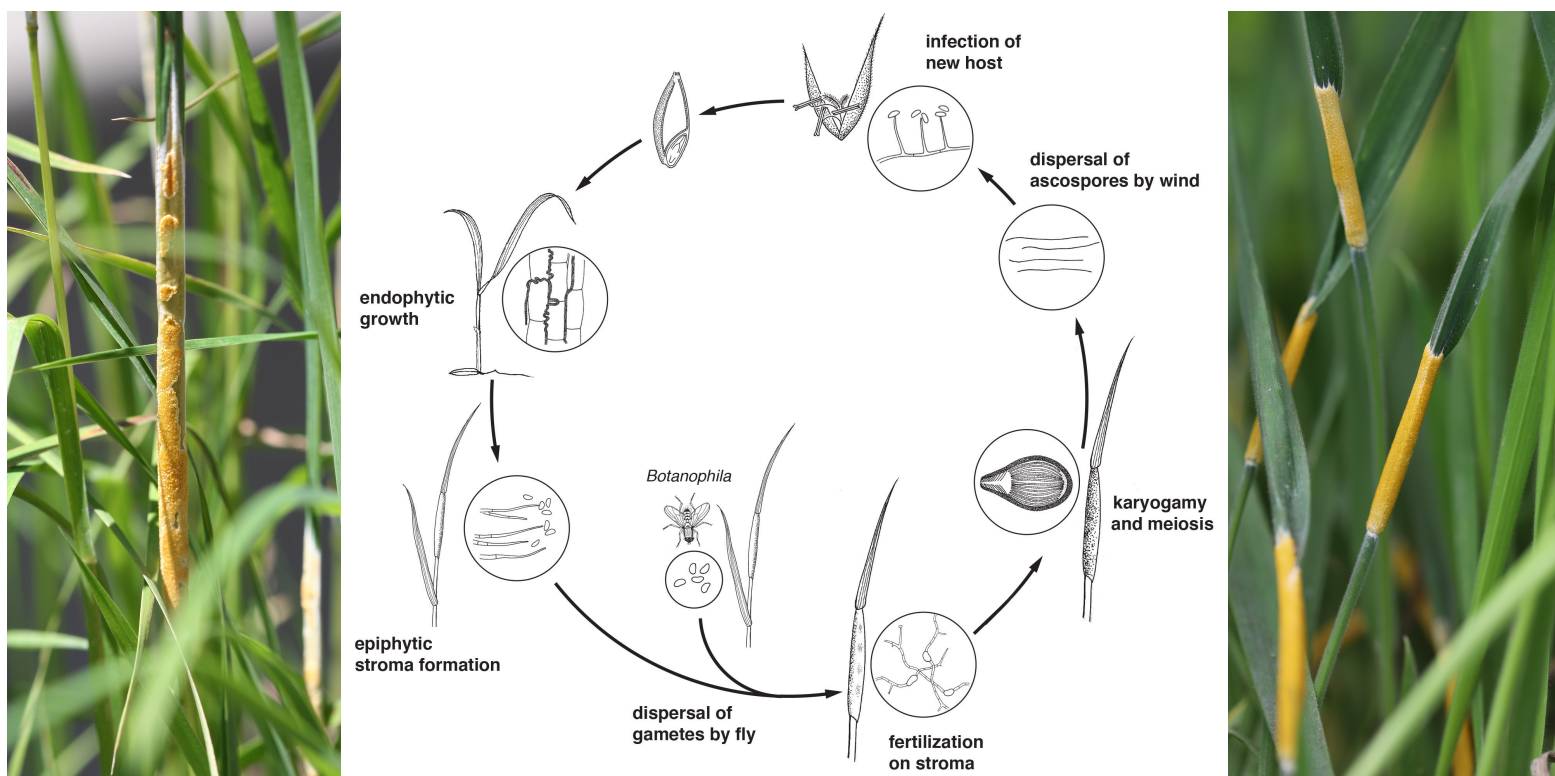


Figure 1: Choked flowering tillers with fungal fruiting bodies (stromata) of study species and sexual life cycle of *Epichloe* fungi. Choke disease caused by *E. typhina* infecting *Dactylis glomerata* (left) and caused by *E. clarkii* infecting *Holcus lanatus* (right). Sexual *Epichloe* possess a bipolar heterothallic mating system and the development of yellow perithecia within stromata indicates successful fertilization with gametes from the opposite mating type. The predominant life cycle state of the fungus is haploid, except for a short diploid phase between karyogamy and meiosis. Figure modified from (Treindl and Leuchtman 2019), photos by A. D. Treindl.

OBJECTIVES AND OUTLINE OF THE PHD THESIS

The pathogenic *Epichloe* endophytes, with their sterilizing lifestyle and high host specificity provide an intriguing and unique biological system for the study of adaptive evolution of fungal plant pathogens in natural ecosystems. I performed a unprecedented parallel population genomic study of two fungal sibling species, including an extensive sampling of seven sympatric populations pairs, which serves the analysis of evolutionary processes within species at the population level but also allows a comparison between species.

The very first objective and precondition for any downstream analysis was the *de novo* assembly of two new reference genomes for the focal species *E. typhina* and *E. clarkii*. Based on the finding that *Epichloe* genomes showed a striking level of structural organization into distinct genomic compartments, I further explored the patterns of genome evolution and repeat content in this system. These analyses are presented in

Chapter I: Chromosome-level genomes provide insights into genome evolution, organization and size in *Epichloe* fungi

The second main objective was the generation of population genomic datasets for *E. typhina* and *E. clarkii* to investigate genetic structure and diversity of these fungal pathogen populations. I sampled seven sympatric population pairs from natural grasslands across Western Europe, I isolated more than 500 fungal strains and sequenced the whole genomes of 480. Based on SNP datasets generated using the new reference genomes for each of the two species, I assessed and compared population structure, population differentiation and genetic diversity of *Epichloe* sibling species. The results are presented in

Chapter II: Genetic diversity and population structure of *Epichloe* - Fungal pathogens of grasses in natural ecosystems

The third objective was to elucidate the genetic basis of adaptation in the *Epichloe* study system and to investigate if dynamic genome compartments play a role in adaptive

evolution. I performed two different genome scans, haplotype-based scans and site-frequency-spectrum-based scans, to detect signatures of recent selective sweeps within each population of each species. I analyzed the gene content of regions under selection and, based on shared sweep loci that were under selection in multiple populations and functionally relevant gene categories, I identified genes that may be relevant for host adaptation. The corresponding analyses are presented in

Chapter III: Genome-wide signatures of selection in natural plant pathogen populations

A last objective, which emerged from the extensive sampling of fungal strains throughout this project and throughout the career of my supervisor Adrian Leuchtman, was to build a new phylogenetic tree based on whole-genome sequencing data for the whole *E. typhina* species complex including *E. typhina* and *E. clarkii* and resolving taxonomic obscurities among its members. The new phylogeny is presented in

Supplementary Chapter: Resolving the phylogeny of the *Epichloe typhina* species complex using genome-wide SNPs

REFERENCES

- Barrett, L. G., J. M. Kniskern, N. Bodenhausen, W. Zhang, and J. Bergelson. 2009. Continua of specificity and virulence in plant host-pathogen interactions: Causes and consequences. *The New Phytologist* 183:513–529.
- Bleidorn, C. 2016. Third generation sequencing: Technology and its potential impact on evolutionary biodiversity research. *Systematics and Biodiversity* 14:1–8.
- Brem, D., and A. Leuchtman. 1999. High prevalence of horizontal transmission of the fungal endophyte *Epichloë sylvatica*. *Bulletin of the Geobotanical Institute ETH* 65:3–12.
- Brockhurst, M. A., T. Chapman, K. C. King, J. E. Mank, S. Paterson, and G. D. D. Hurst. 2014. Running with the Red Queen: The role of biotic conflicts in evolution. *Proceedings of the Royal Society B: Biological Sciences* 281:20141382.
- Bultman, T. L., and A. Leuchtman. 2008. Biology of the *Epichloë*-*Botanophila* interaction: An intriguing association between fungi and insects. *Fungal Biology Reviews* 22:131–138.
- Burdon, J. J., and A.-L. Laine. 2019. The diverse and ubiquitous nature of pathogens. Pages 1–28 *Evolutionary dynamics of plant-pathogen interactions*. Cambridge University Press, Cambridge.
- Chung, K.-R., and C. L. Schardl. 1997. Sexual cycle and horizontal transmission of the grass symbiont, *Epichloë typhina*. *Mycological Research* 101:295–301.
- Clay, K., and C. Schardl. 2002. Evolutionary origins and ecological consequences of endophyte symbiosis with grasses. *The American Naturalist* 160:S99–S127.
- Craven, K. D., P. T. W. Hsiau, A. Leuchtman, W. Hollin, and C. L. Schardl. 2001. Multigene phylogeny of *Epichloë* species, fungal symbionts of grasses. *Annals of the Missouri Botanical Garden* 88:14–34.
- Eaton, C. J., M. P. Cox, B. Ambrose, M. Becker, U. Hesse, C. L. Schardl, and B. Scott. 2010. Disruption of signaling in a fungal-grass symbiosis leads to pathogenesis. *Plant Physiology* 153:1780–1794.
- Everhart, S., N. Gambhir, and R. Stam. 2021. Population genomics of filamentous plant pathogens - A brief overview of research questions, approaches, and pitfalls. *Phytopathology* 111:12–22.
- Feijen, F. A. A., R. A. Vos, J. Nuytinck, and V. S. F. T. Merckx. 2018. Evolutionary dynamics of mycorrhizal symbiosis in land plant diversification. *Scientific Reports* 8:10698.
- Fisher, M. C., D. A. Henk, C. J. Briggs, J. S. Brownstein, L. C. Madoff, S. L. McCraw, and S. J. Gurr. 2012. Emerging fungal threats to animal, plant and ecosystem health. *Nature* 484:186.
- Fournier, E., and T. Giraud. 2008. Sympatric genetic differentiation of a generalist pathogenic fungus, *Botrytis cinerea*, on two different host plants, grapevine and bramble. *Journal of Evolutionary Biology* 21.1:122–132.
- Giraud, T., G. Refrégier, M. Le Gac, D. M. de Vienne, and M. E. Hood. 2008. Speciation in fungi. *Fungal Genetics and Biology* 45:791–802.
- Gladieux, P. 2018. What makes a specialized endophyte special? *Molecular Ecology* 27:3037–3039.
- Gladieux, P., J. Ropars, H. Badouin, A. Branca, G. Aguilera, D. M. De Vienne, R. C. Rodríguez de la Vega, S. Branco, and T. Giraud. 2014. Fungal evolutionary genomics provides insight into the mechanisms of adaptive divergence in eukaryotes. *Molecular Ecology* 23:753–773.

- Grünwald, N. J., B. A. McDonald, and M. G. Milgroom. 2016. Population genomics of fungal and oomycete pathogens. *Annual Review of Phytopathology* 54:323–346.
- Hardoim, P. R., L. S. Van Overbeek, G. Berg, A. M. Pirttilä, S. Compant, A. Campisano, M. Döring, and A. Sessitsch. 2015. The hidden world within plants: Ecological and evolutionary considerations for defining functioning of microbial endophytes. *Microbiology and Molecular Biology Reviews* 79:293–320.
- Hartmann, F. E., R. C. Rodríguez de la Vega, F. Carpentier, P. Gladieux, A. Cornille, M. E. Hood, and T. Giraud. 2019. Understanding adaptation, coevolution, host specialization, and mating system in castrating anther-smut fungi by combining population and comparative genomics. *Annual Review of Phytopathology* 57:431–457.
- Horscroft, C., S. Ennis, R. J. Pengelly, T. J. Sluckin, and A. Collins. 2019. Sequencing era methods for identifying signatures of selection in the genome. *Briefings in Bioinformatics* 20:1997–2008.
- Kogel, K. H., P. Franken, and R. Hückelhoven. 2006. Endophyte or parasite - what decides? *Current Opinion in Plant Biology* 9:358–363.
- Leake, J. R. 2005. Plants parasitic on fungi: Unearthing the fungi in myco-heterotrophs and debunking the “saprophytic” plant myth. *Mycologist* 19:113–122.
- Leuchtmann, A. 2003. Taxonomy and diversity of *Epichloë* endophytes. Pages 169–194 in W. J. J. F., B. C. W., H.-J. N. L., and S. J. W., editors. *Clavicipitalean fungi: Evolutionary biology, chemistry, biocontrol and cultural impacts*. Marcel Dekker, New York.
- Leuchtmann, A., C. W. Bacon, C. L. Schardl, J. F. White, and M. Tadych. 2014. Nomenclatural realignment of *Neotyphodium* species with genus *Epichloë*. *Mycologia* 106:202–215.
- Leuchtmann, A., and K. Clay. 1997. The population biology of grass endophytes. Pages 185–202 in G. C. Carroll and P. Tudzynski, editors. *The Mycota, Vol. V. Part B, Plant Relationships*. Springer-Verlag, Berlin, Heidelberg.
- Lutzoni, F., M. D. Nowak, M. E. Alfaro, V. Reeb, J. Miadlikowska, M. Krug, A. E. Arnold, L. A. Lewis, D. L. Swofford, D. Hibbett, K. Hilu, T. Y. James, D. Quandt, and S. Magallón. 2018. Contemporaneous radiations of fungi and plants linked to symbiosis. *Nature Communications* 9:1–11.
- Ma, L.-J., H. C. Van Der Does, K. A. Borkovich, J. J. Coleman, M.-J. Daboussi, A. Di Pietro, M. Dufresne, M. Freitag, M. Grabherr, and B. Henrissat. 2010. Comparative analysis reveals mobile pathogenicity chromosomes in *Fusarium*. *Nature* 464:367–373.
- McDonald, B. A., and E. H. Stukenbrock. 2016. Rapid emergence of pathogens in agro-ecosystems: Global threats to agricultural sustainability and food security. *Philosophical Transactions of the Royal Society B: Biological Sciences* 371:20160026.
- Oleksyk, T. K., M. W. Smith, and S. J. O’Brien. 2010. Genome-wide scans for footprints of natural selection. *Philosophical Transactions of the Royal Society B: Biological Sciences* 365:185–205.
- Papkou, A., T. Guzella, W. Yang, S. Koepper, B. Pees, R. Schalkowski, M. C. Barg, P. C. Rosenstiel, H. Teotónio, and H. Schulenburg. 2019. The genomic basis of red queen dynamics during rapid reciprocal host–pathogen coevolution. *Proceedings of the National Academy of Sciences of the United States of America* 116:923–928.

- Petrini, O. 1991. Fungal endophytes of tree leaves. Pages 179–197 in J. H. Andrews and S. S. Hirano, editors. *Microbial Ecology of Leaves*. Springer, New York, NY, New York, NY.
- Raynal, G. 1991. Observations françaises sur les *Acremonium*, champignons endophytes des graminées fourragères. *Fourrages* 126:225–237.
- Redman, R. S., D. D. Dunigan, and R. J. Rodriguez. 2001. Fungal symbiosis from mutualism to parasitism: Who controls the outcome, host or invader? *New Phytologist* 151:705–716.
- Schardl, C. L., C. A. Young, N. Moore, N. Krom, P.-Y. Dupont, J. Pan, S. Florea, J. S. Webb, J. Jaromczyk, and J. W. Jaromczyk. 2014. Genomes of plant-associated Clavicipitaceae. *Advances in Botanical Research* 70:291–327.
- Schirrmann, M. K., S. Zoller, D. Croll, E. H. Stukenbrock, A. Leuchtman, and S. Fior. 2018. Genomewide signatures of selection in *Epichloë* reveal candidate genes for host specialization. *Molecular Ecology* 27:3070–3086.
- Schirrmann, M. K., S. Zoller, S. Fior, and A. Leuchtman. 2015. Genetic evidence for reproductive isolation among sympatric *Epichloë* endophytes as inferred from newly developed microsatellite markers. *Microbial Ecology* 70:51–60.
- Stinchcombe, J. R., and H. E. Hoekstra. 2008. Combining population genomics and quantitative genetics: Finding the genes underlying ecologically important traits. *Heredity* 100:158–170.
- Stukenbrock, E. H. 2013. Evolution, selection and isolation: A genomic view of speciation in fungal plant pathogens. *New Phytologist* 199:895–907.
- Tellier, A., S. Moreno-Gámez, and W. Stephan. 2014. Speed of adaptation and genomic footprints of host–parasite coevolution under arms race and trench warfare dynamics. *Evolution* 68:2211–2224.
- Treindl, A. D., and A. Leuchtman. 2019. Assortative mating in sympatric ascomycete fungi revealed by experimental fertilizations. *Fungal Biology* 123:676–686.
- Turland, N. J., J. H. Wiersema, F. R. Barrie, W. Greuter, D. L. Hawksworth, P. S. Herendeen, S. Knapp, W.-H. Kusber, D.-Z. Li, and K. Marhold. 2018. International Code of Nomenclature for algae, fungi, and plants (Shenzhen Code) adopted by the Nineteenth International Botanical Congress Shenzhen, China, July 2017. Koeltz Botanical Books, Glashütten.
- Vitti, J. J., S. R. Grossman, and P. C. Sabeti. 2013. Detecting natural selection in genomic data. *Annual Review of Genetics* 47:97–120.
- Weigand, H., and F. Leese. 2018. Detecting signatures of positive selection in non-model species using genomic data. *Zoological Journal of the Linnean Society* 184:528–583.
- White, J. F., and T. L. Bultman. 1987. Endophyte-host associations in forage grasses. VIII. Heterothallism in *Epichloë typhina*. *American Journal of Botany* 74:1716–1721.
- White Jr., J. F. 1997. Perithecial ontogeny in the fungal genus *Epichloë*: An examination of the clavicipitalean centrum. *American Journal of Botany* 84:170–178.
- White Jr., J. F., A. C. Morrow, G. Morgan-Jones, and D. A. Chambless. 1991. Endophyte-host associations in forage grasses. XIV. Primary stromata formation and seed transmission in *Epichloë typhina*: Developmental and regulatory aspects. *Mycologia* 83:72–81.

Chromosome-level genomes provide insights into genome evolution, organization and size in *Epichloe* fungi

Artemis D. Treindl^a, Jessica Stapley^b, David J. Winter^c, Murray P. Cox^c, Adrian Leuchtman^a

^a Plant Ecological Genetics Group, Institute of Integrative Biology, ETH Zurich, 8092 Zurich, Switzerland

^b Plant Pathology Group, Institute of Integrative Biology, ETH Zurich, 8092 Zurich, Switzerland

^c Statistics and Bioinformatics Group, Massey University, Palmerston North 4442, New Zealand

Manuscript submitted to *Genomics (Elsevier)* pending revisions – this is an edited version

ABSTRACT

Epichloe fungi are common endophytes and pathogens of many cool season grasses, both in wild species and commercially sown cultivars. The *Epichloe*-grass symbiosis is not only of great interest to agricultural research for the fungal bioprotective properties conferred to host grasses but also to evolutionary biologists as a model system for plant-fungal coevolution in natural systems. Here, we assembled and annotated gapless chromosome-level reference genomes of two pathogenic *Epichloe* sibling species. Both genomes have a bipartite genome organization, with blocks of highly syntenic gene-rich core regions separated by blocks of AT-rich DNA. The AT-rich regions show an extensive signature of RIP (repeat-induced point mutation) and the expansion of this compartment accounts for the large difference in genome size between the two species. This study demonstrates how the rapid evolution of repeat structure can drive divergence between closely related taxa and highlights the evolutionary role of dynamic compartments in fungal genomes.

Keywords: Genome compartmentalization, Two-speed genomes, Repeat-induced point mutation, Fungal plant pathogens, Transposable elements, *Epichloë* endophyte

INTRODUCTION

Despite their small size, fungal genomes can provide big insights into eukaryotic genome organization and evolution. Plant-colonizing fungi and oomycetes, in particular, frequently have genomes with a bipartite organization: a conserved “core genome” containing the majority of genes and a highly variable “accessory genome” characterized by a high repeat-content and few genes (Bertazzoni et al. 2018). In some cases these repeat-rich regions are targeted by repeat-induced point mutation (RIP), a defense mechanism against transposable element proliferation, which leads to locally high mutation rates, specifically C-to-T mutations, reduced G/C content and increased genetic diversity (Selker 2002). Regional variation in the levels of structural and sequence polymorphism gives rise to different rates of evolutionary change across the genome, referred to as the “two-speed” genome model (Raffaele and Kamoun 2012). Regions with higher rates of evolution are enriched for repetitive elements and effector genes that are involved in plant-pathogen interactions, and these regions are often important in adaptation (Croll and McDonald 2012, Raffaele and Kamoun 2012, Dong et al. 2015). Despite recent progress, we still lack an understanding of how widespread this bipartite genome organization is, how it varies across taxa and how it relates to evolutionary processes (Torres et al. 2020).

To date, the most evidence for bipartite genomes comes from pathogens of agricultural crops with large economic impacts. Studies of these agriculturally important pathogens have shown that the distribution of repeat rich compartments can vary. They can cover whole accessory or lineage specific chromosomes, for example in *Fusarium oxysporum* (Ma et al. 2010) and *Zymoseptoria tritici* (Plissonneau et al. 2018). In other species, this bipartite architecture is organized in a mosaic-like structure with alternating blocks of genomic compartments distributed along chromosomes; for example, chromosomes of a dothideomycete pathogen (*Leptosphaeria maculans*) contain structural isochores, scattered sequence blocks of distinct base composition that are enriched in transposable elements (Rouxel et al. 2011, Grandaubert et al. 2014). These so called AT-isochores are largely devoid of genes, similar to the genome of the oomycete *Phytophthora infestans*, which also consists of alternating repeat-poor/gene-rich and repeat-rich/gene-poor regions (Raffaele et al. 2010). The genome organization of fungal plant pathogens in natural ecosystems are less well studied, in part because fewer high quality, whole

genome sequences are available for these systems. One exception is the facultative pathogen *Epichloe festucae*. A recent study by Winter et al. (2018) found that the genome of *E. festucae* also displays a two-speed organization and showed that the remarkable patchwork structure affects the three-dimensional configuration of chromosomes creating compartments of distinct gene expression. This suggests a functional role of genome organization in *Epichloe*, although its evolutionary importance remains largely obscure.

To better understand bipartite genome evolution and how this influences adaptive processes in natural fungal plant pathogens, we set out to produce high quality reference genome for two *Epichloe* species. *Epichloe* are obligately biotrophic fungi that form life-long systemic infections of cool-season grasses. The interaction between *Epichloe* and their host grasses serves as a unique model in the study of coevolution and adaptation because they span the whole symbiotic continuum from antagonistic to mutualistic depending on the mode of reproduction. Sexually reproducing *Epichloe* are haploid outcrossers and usually a single genotype infects the whole above-ground tissue of the host grass (Leuchtman and Clay 1997). They grow asymptotically to the point of host flowering and then cause choke disease, whereby the undeveloped host inflorescences are enclosed in the fungal fruiting structure (White Jr. 1997). This pathogenic stage can sterilize the host plant and enables the fungus to horizontally transmit meiotic spores and spread to other plants. Purely asexual *Epichloe*, on the other hand, form infections that do not cause visible symptoms in the host plant and transmit vertically by dissemination via seeds. These mutualistic symbionts can confer various benefits to the host plant including increased drought resistance (West et al. 1993), and protection against herbivory by mammals and invertebrates mediated by the production of bioactive alkaloid compounds (Scharidl et al. 2013b). *Epichloe* endophytes are therefore widely employed in commercial agriculture (Kauppinen et al. 2016, Saikkonen et al. 2016, Lugtenberg et al. 2016). Interestingly, many of these mutualistic symbionts are heteroploid (aneuploid, diploid, or polyploid) interspecific hybrids of two or more ancestral sexual species (Moon et al. 2004). The species studied here are both sexual and possess haploid genomes.

The aim of this study was to sequence DNA and RNA from two closely related taxa in the *E. typhina* species complex (*E. typhina* and *E. clarkii*) that infect different host

grasses, to create finished reference genomes and investigate the evolution of genome organization. We combined Illumina and Pacific Biosciences (PacBio) DNA sequencing to generate complete chromosome-level genome assemblies, and used Illumina RNAseq data to improve gene annotation. Using these new genomic resources, we compared patterns of genome organization, gene content and synteny between the two species.

METHODS & MATERIALS

Taxonomic status and nomenclature

This study focuses on two members of the *E. typhina* species complex: *E. typhina* (Pers.) infecting the grass species *Dactylis glomerata* and *E. clarkii* J.F. White infecting *Holcus lanatus*. The two sibling taxa are currently assigned the taxonomic rank of subspecies based on their sexual compatibility and ability to hybridize in experimental crosses (Leuchtman and Schardl 1998, Leuchtman et al. 2014). Evidence from previous studies suggests, however, that *E. typhina* and *E. clarkii* are reproductively isolated biological entities at an advanced stage of divergence (Schirrmann et al. 2015, 2018, Treindl and Leuchtman 2019), and our ongoing work further supports this, detecting no evidence of gene flow among natural sympatric populations of the two taxa (see Chapter II and Supplementary Chapter of this thesis). For the purpose of this and future studies we therefore consider *E. typhina* and *E. clarkii* as closely related yet distinct species. Further, as now suggested by the International Code of Nomenclature (Turland et al. 2018), we refer to *Epichloe* without its historical diaeresis, aiding to make the data, metadata and analysis open, reusable and reproducible.

Fungal isolates

Epichloe typhina strain Ety_1756 and *Epichloe clarkii* strain Ecl_1605_22 were isolated from infected plant tissue collected from naturally infected grass accessions in Switzerland (Supplementary Table S1 and Figure S1). Note that in subsequent figures we use the abbreviations Ety and Ecl respectively. Pure fungal cultures were obtained from surface sterilized leaf-sheaths of host plants *D. glomerata* and *H. lanatus* following the procedure described by Leuchtman and Clay (1988). Isolates were grown in liquid V8

medium (Christensen et al. 1993) for ten days, strained, washed with sterile deionized water to remove residue medium and subsequently freeze dried.

DNA preparation and sequencing

DNA was extracted according to the simple method for extraction of fungal genomic DNA of Al-Samarrai and Schmid (2000). In short, the protocol involves lysis of 30 mg of freeze dried ground mycelium in fresh sodium dodecyl sulphate buffer, detachment of DNA from polysaccharides by mild shearing, NaCl precipitation of polysaccharides and protein, chloroform extraction and ethanol precipitation. DNA quality and quantity was assessed by automated electrophoresis (2200 TapeStation, Agilent, Santa Clara, CA, USA), spectrophotometric ratios ($A_{260}/A_{280} > 1.8$ and $A_{260}/A_{230} > 2$, Nanodrop spectrophotometer, Thermo Fisher Scientific, Waltham, MA, USA) and fluorometric analysis (Quantus™ Fluorometer, Promega Corporation, Wisconsin, USA). For each isolate we obtained around 20 µg of high molecular weight, high purity genomic DNA. PacBio SMRTbell libraries were prepared at the Functional Genomics Center of Zurich (FGCZ), using 15 µg of DNA. Sequencing was performed on a PacBio RSII instrument at the FGZC (Sequencing and Genomic Technologies Shared Resource, Duke University, NC, USA). For assembly polishing, we generated Illumina paired-end libraries (insert size of 330 bp) using NEBNext® Ultra™ II DNA Library Prep Kits and sequenced these on a HiSeq4000 Illumina sequencer, at a final coverage of approximately 50x for *E. typhina* and 25x for *E. clarkii*.

RNA extraction and sequencing

RNA was isolated from reference *Epichloe* strains Ety_1756 and Ecl_1605_22 grown as axenic culture, as well as in association with their host grasses *D. glomerata* and *H. lanatus*, respectively. For genes expressed *in culture*, RNA was extracted from strains grown on supplemented malt-extract agar (containing 1% malt extract, 1% glucose, 0.25% bacto peptone, and 0.25% yeast extract) for 10 days at room temperature. Three technical replicates per strain were used. For *in planta* RNA extraction, endophyte-free seedlings grown from commercial cultivar seeds of *D. glomerata* (cv. 'Prato', Agroscope Reckenholz-Tänikon), and *H. lanatus* (3624/K01/0.1, Otto Hauenstein seeds) were inoculated with the respective reference strains Ety_1756 (*D. glomerata*) and

Ecl_1605_22 (*H. lanatus*) following the procedure of Latch and Christensen (1985). Inoculated plants were kept for 3-5 days in the dark and then grown in multipot trays in commercial soil (Ökohum Bio Universalerde) at 22°C with a photoperiod of 16 h of light. After 10 weeks, plants were screened for infection by examining aniline blue stained leaf-sheath tissue of three tillers per plant using a light microscope at 400x magnification. Five plants of each species with infection in all three tillers were re-potted into larger pots and moved to the greenhouse. Pseudostems of three plants (biological replicates) with two technical replicates per plant were collected and frozen in liquid nitrogen resulting in a total of six samples per strain. We extracted total RNA using the RNeasy Plant Mini Kit (QIAGEN, Hilden, Germany) and removed residual genomic DNA by DNase treatment (DNase I RNase-Free, New England Biolabs, MA, USA). We assessed RNA concentration by fluorometric analysis (Quantus™ Fluorometer, Promega Corporation, Wisconsin, USA), and quality by automated electrophoresis (2200 TapeStation, Agilent, Santa Clara, CA, USA). Paired-end cDNA libraries were prepared with the Illumina TruSeq RNA kit and sequencing was performed at the FGCZ on an Illumina NovaSeq 6000 (Illumina, San Diego, CA, USA).

Genome assembly

Initial assemblies were produced using three different programs: Canu v1.8 (Koren et al. 2017), FALCON v1.2.2 (Chin et al. 2016) and MECAT2 v20190314 (Xiao et al. 2017). By comparing assemblies using D-GENIES v1.2.0 (Cabanettes and Klopp 2018), we found discontinuities found in one assembly were generally well-resolved in one or both of the other assemblies. As FALCON produced the most contiguous assembly for each of our strains, we used the assemblies produced by this program as the base of our final reference sequence and used overlapping segments present in other assemblies to resolve the remaining gaps. Base-level errors in the final assembly were polished using Pilon v1.23 (Walker et al. 2014), using Illumina DNA reads as input.

Gene annotation

The genome was annotated using v1.4 of the funannotate pipeline (Palmer 2016). This pipeline makes use of Augustus, SNAP, glimmerHMM and CodingQuarry for gene-calling. Step-by-step details and example code are available at

https://github.com/adtreindl/Epichloe_genomes/blob/master/RNASeq_Data_Analysis_predict.md. Briefly, we mapped RNAseq data using STAR v2.5.3 (Dobin et al. 2013), and identified transcripts using stringtie v1.3.3 (Pertea et al. 2015). We then used the `funannotate predict` functions to predict genes, incorporating the transcript information. Gene/protein function was predicted using InterProScan v5.17-56.0 (Jones et al. 2014) and we identified putative effectors using signalP v4.1 (Petersen et al. 2011), and effectorP v2.0 (Sperschneider et al. 2016). We identified orthologous genes shared by both genomes using proteinortho v6.0 beta (Lechner et al. 2011).

Transposable element annotation

To identify transposable elements (TEs) we produced *de novo* TE libraries for each genome using the RepeatModeler v2.0.1 pipeline (Flynn et al. 2020), which uses RepeatScout to identify multicopy regions of the genome and RECON (Bao and Eddy 2002) to discover known TE motifs. We then identified specific elements from each family in our *de novo* libraries using RepeatMasker v4.0.6 (Tarailo-Graovac and Chen 2009).

Identification of genome compartmentalization

We scanned genomes for signatures of RIP using The RIPper (access date 23 October 2020), with default settings (Van Wyk et al. 2019). RIP activity shifts nucleotide composition in affected regions by preferentially inducing C-to-T mutations resulting in a reduced GC content. We also identified AT-rich regions using OcculterCut v1.1 (Testa et al. 2016), which detects bimodal patterns in GC distribution across genome sequences and then segments them into regions of differing nucleotide-content using the Jensen-Shannon divergence (D_{JS}). The RIPper accounts for the overall AT-content of a genome and implements a quantitative measure of RIP activity using the RIP index of Margolin (Margolin et al. 1998), whereas OcculterCut is a qualitative measure assigning genomic regions to defined categories. Given that both methods rely on nucleotide proportions, we expect results from these analyses to be strongly correlated.

Synteny

We assessed synteny, the shared physical orientation of genomic sequence between species, at two levels. For a comparison at the nucleotide level, we performed pairwise alignments of three genomes, the two new genomes and the *E. festucae* Fl1 reference genome, with minimap2 v2.12 (Li 2018) using the preset option for cross-species full genome alignment (asm10 for sequence divergence around 2% and asm20 for sequence divergence up to 10%, giving very similar results). Secondly, synteny was assessed at the gene level by identifying orthologous genes shared by both genomes using proteinortho v6.0beta (Lechner et al. 2011). We used coordinates of single copy orthologous genes to investigate gene order and co-linearity.

Assembly completeness

The completeness of each genome assembly was estimated using BUSCO v3.0.2 (Simão et al. 2015), by searching for conserved genes in the Sordariomycete dataset (library sordariomyceta_odb9).

RESULTS AND DISCUSSION

De novo assembly of two complete chromosome-level genomes

We combined PacBio long-read and Illumina short-read DNA sequencing to generate high quality contiguous reference genomes of *E. typhina* and *E. clarkii*. PacBio sequencing yielded 810,432 and 807,685 subreads with an N50 of 14,350 and 17,455 base pairs, respectively. In total we obtained 6.8 and 6.6 Gb representing a mean coverage of 200x for the *E. typhina* genome and 144x for the *E. clarkii* genome (Table 1). We used a set of different long-read assemblers to generate separate assemblies from these reads (see methods). As FALCON produced the most contiguous assembly for each of our strains, we used the assemblies produced by this program as the base of our final assemblies and used overlapping segments present in other assemblies to confirm resolution of the remaining gaps. We used long-read genome alignments to manually check break points and resolve each gap, giving us strong confidence in our final assemblies.

The final genome assemblies consisted of seven nuclear chromosomes and the complete mitochondrial genome for each species. Nuclear genome size differed between the two sibling species (Table 2), with the *E. typhina* genome spanning 33.83 Mb whereas the *E. clarkii* genome was more than a third larger at 45.62 Mb. Genome sequences were deposited on GenBank (BioProject ID [PRJNA533210](#) for *E. typhina* Ety_1756 and [PRJNA533212](#) for *E. clarkii* Ecl_1605_22).

Table 1: PacBio sequencing statistics

	<i>E. typhina</i>	<i>E. clarkii</i>
total bases	6.8 Gb	6.6 Gb
total subreads	810,432	807,685
mean read length	8,403	11,340
read length N50	14,350	17,455
longest read	67,878	114,135

Table 2: Summary information for AT-rich and gene-rich compartments of *E. typhina* and *E. clarkii* genomes

			AT-rich compartment			Gene-rich compartment			
	Genome size	chr	genes	size	GC-content [%]	genes	size	GC-content [%]	genes
Ety	33.83 Mb	7	8165	10.7 Mb (31.7%)	24.4	10	23.1 Mb (68.3%)	52.7	8155
Ecl	45.62 Mb	7	8736	22.2 Mb (48.6%)	25.7	53	23.4 Mb (51.4%)	52.7	8683

Nuclear genome sizes are reported here, mitochondrial sequence lengths were 98,452 bp (*E. typhina*) and 73,184 bp (*E. clarkii*)

Gene prediction and functional annotation

We identified 8,165 and 8,736 protein coding genes in the genomes of *E. typhina* and *E. clarkii*, respectively. We assigned functional annotations to most of the genes identified; 81.3% and 77.1% of *E. typhina* and *E. clarkii* protein coding genes, respectively. We also identified putative signal peptides in both species (584 in *E. typhina* and 587 in *E. clarkii*), and of these, we identified 138 and 144 putative effectors in *E. typhina* and *E. clarkii*, respectively (see supplementary files here: https://github.com/adtreindl/Epichloe_genomes). We identified 6,923 one-to-one orthologs shared by both species. Duplicated genes were rare with only three duplicated two-to-two orthologs identified. Of all gene pairs, 5,715 (86.1%) formed an orthology group with a single gene from the *E. festucae* Fll genome. The genomes contained 97.5% (*E. typhina*) and 96.7% (*E. clarkii*) of conserved sordariomycete genes as identified by BUSCO (Supplementary Table 2).

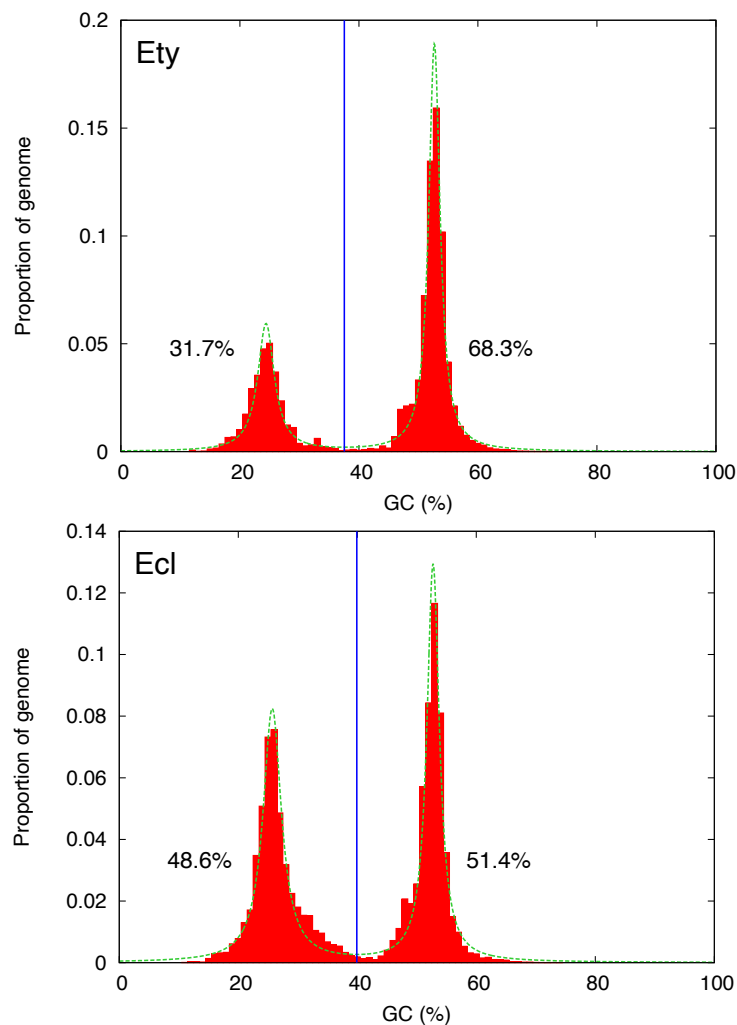


Figure 1: Bimodal distribution of GC content in two *Epichloe* genomes.

E. typhina is shown at the top, *E. clarkii* at the bottom. Vertical blue lines show the GC cut-off chosen by OcculterCut and used to classify genome segments into distinct AT-rich and GC-rich regions. Green dashed lines indicate Cauchy distributions that were fit to the data. The percentage values indicate the percentage of the genome that is AT-rich (left) and GC-rich (right).

Bipartite genome structure and variation in genome size

The distinctive bipartite genome structure seen in some fungal pathogens is evident in *E. typhina* and *E. clarkii* genomes. In line with the previously sequenced *E. festucae*, both new *Epichloe* genomes exhibit a distinct bimodal pattern (distribution) of GC-content (Figure 1). The genomes are compartmentalized into distinct blocks with high AT nucleotide content (AT-rich regions), and alternating blocks of sequence with approximately equal nucleotide content (Figure 2). We found concordance between quantitative signatures of RIP (repeat-induced point mutation) in genomic windows (identified by The RIPper), and qualitative assignment of AT-rich regions identified by

OcculterCut (Supplementary Figures S2 and S3), which was expected given that both methods principally rely on nucleotide distribution along the genome. For subsequent analyses, we refer to the dichotomous categorization into AT-rich and GC-rich regions. The GC-rich compartment contained the vast majority of annotated genes (99.88% and 99.39% in *E. typhina* and *E. clarkii*, respectively), and is henceforth referred to as the gene-rich compartment. In contrast, the AT-rich compartment of these two *Epichloe* genomes was largely devoid of genes (10 and 53 genes, 0.12% and 0.61%, one and two of which were putative effectors), in line with other fungal genomes investigated (Testa et al. 2016). Consistent with this, we found that of 6,923 orthologs shared by both species, most were located in gene-rich regions, whereas only one gene in *E. typhina* and two in *E. clarkii* were located in AT-rich regions. Putative effector genes were significantly less likely to be located in AT-rich regions than expected by chance, however, they were enriched in the regions directly bordering AT-rich regions (Supplementary Figure S16).

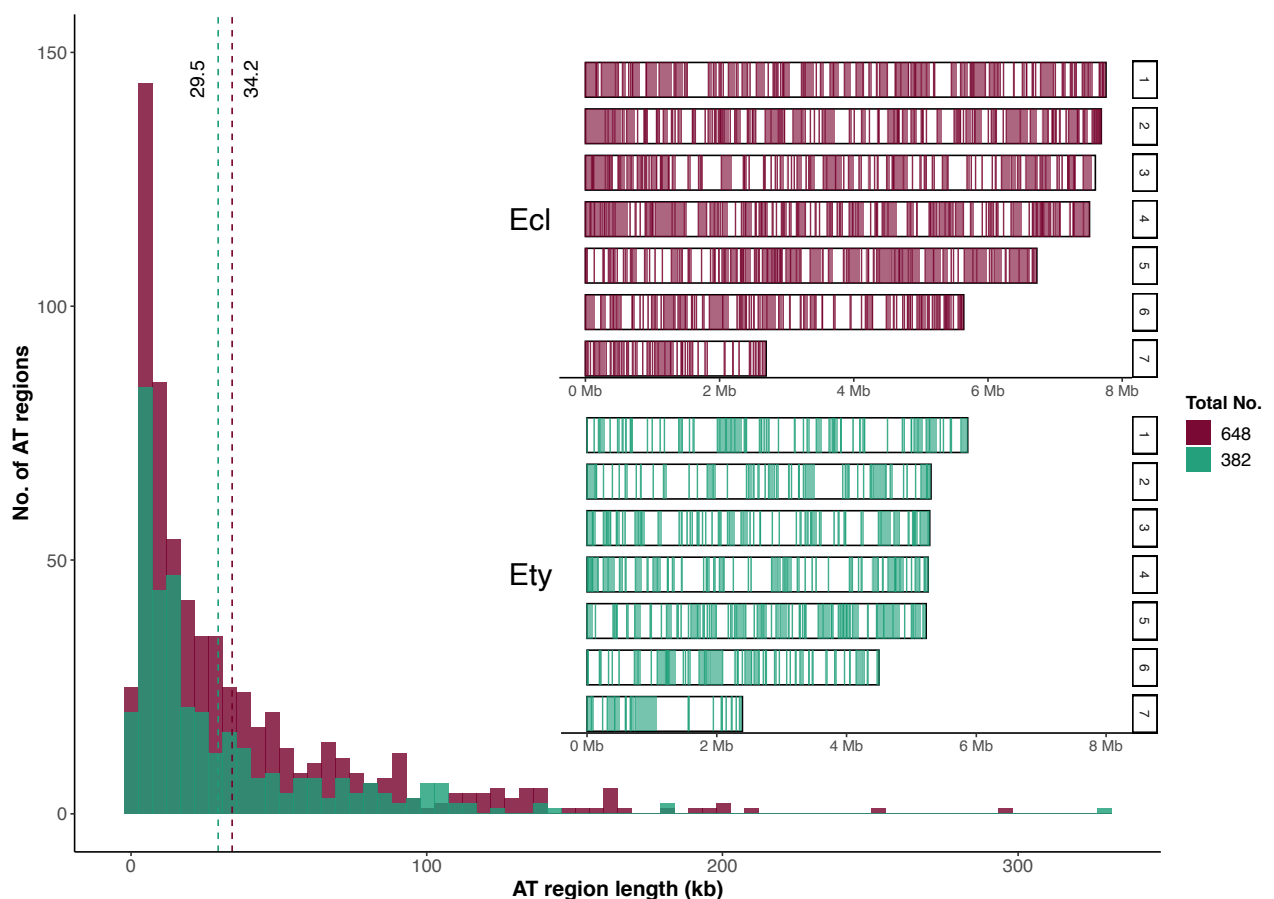


Figure 2: Comparison of AT-rich compartments.

Histogram showing the distribution of AT-rich region length for *E. typhina* (green) and *E. clarkii* (purple). Dashed lines mark mean lengths. Inset displays ideograms with patchwork distribution of AT-rich regions along chromosomes. *E. clarkii* has almost 70% more regions than *E. typhina* (648 vs. 382), with a majority of those regions being relatively short (58% *E. typhina* and 52% *E. clarkii* are ≤ 20 kb).

The genome of *E. clarkii* was a third larger than *E. typhina* and this is largely explained by the repeat content and an expansion of the AT-rich compartment in *E. clarkii*, whereas the gene-rich compartments of the genomes were approximately equal in size (23.1 Mb and 23.4 Mb). The genome of *E. clarkii* contained a larger fraction of AT-rich sequence (10.7 Mb *E. typhina* vs. 22.2 Mb *E. clarkii*), more AT-rich regions (648 vs 382) and had on average slightly longer AT-rich regions (mean 29.5 kb *E. typhina* vs. 34.2 kb *E. clarkii*) (Figure 2). This difference in compartment size is congruent with the different repeat content between genomes: 18.9 Mb (41.4 %) of the *E. clarkii* genome was assigned to a TE family compared to 9.1 Mb (27.0%) TE sequence in *E. typhina*. In both genomes, 95% of all bases annotated as being part of a TE fell into AT-rich regions demonstrating that AT-rich compartments of *E. typhina* and *E. clarkii* are largely comprised of repetitive DNA (Supplementary Table S3 and Figure S4). The composition of TEs in each genome was similar, with the majority of TE sequences (82% *E. typhina* and 86% in *E. clarkii*) being full or partial LTR retrotransposons and the remainder comprised of DNA transposons, LINE elements and unclassified elements. Almost all LTRs were located in AT-rich regions (99% in *E. typhina* and 97% in *E. clarkii*) whereas a relatively high proportion of unclassified elements and DNA transposons occurred in the gene-rich regions of the genome. Unclassified elements likely include short non-autonomous elements such as MITES, which may function as regulators of gene expression in *Epichloe* (Winter et al. 2018).

The genome size difference between the *E. typhina* and *E. clarkii* is substantial and much larger than might be expected given they are closely related members of the same species complex. The difference between these sibling species is almost as large as the differences seen among all *Epichloe* species where genome size has been estimated from genome sequence data (Schardl et al. 2013a). Although no clear association between genome size and phylogeny was detected, closely related strains tended to have more similar genome sizes (Schardl et al. 2014). Although caution is needed when comparing genome size estimates using different sequencing technologies (short-read based genome assemblies may have misassembled or collapsed large repetitive regions), it appears these sibling species have diverged considerably in genome size compared to other sibling taxa. Despite the difference in overall size, gene content was similar in the two genomes; we found a similar number of genes most of which had high sequence

homology between species (84.8% and 79.2% orthologs in *E. typhina* and *E. clarkii* respectively).

Conserved synteny between sibling taxa

We found complete conservation of synteny between *E. typhina* and *E. clarkii* at the chromosome level, with no evidence of major rearrangements between chromosomes (Figure 3). The chromosome numbers do not correspond because *Epichloe* chromosomes are numbered by size with the suspected short arm of the chromosome first; as a result, chromosome 1 in *E. typhina* corresponds to the inverted chromosome 3 in *E. clarkii*. The synteny of these two sibling taxa were compared to the synteny with the more distantly related species *E. festucae* Fll, and, as expected, whole genome alignments with *E. festucae* showed substantially more rearrangements among the seven chromosomes (Supplementary Figure S5-S7) compared to the alignments between *E. typhina* and *E. clarkii*. In gene-rich regions of homologous chromosomes, both macrosynteny (the overall chromosomal gene content) and microsynteny (the order of genes) were highly conserved (Figure 4). The mapping of positions of 6923 single copy orthologous genes showed high gene co-linearity (Supplementary Figure S8, S9, S10B), and in pairwise alignments of homologous chromosomes between *E. typhina* and *E. clarkii*, we found only one large rearrangement affecting more than 47 genes in a gene-rich region (Supplementary Figure S10).

What leads to genome expansion?

RIP is a genomic defense mechanism that counteracts TE activity in fungal genomes (Selker 2002). RIP activity is linked to sexual reproduction (meiosis), and targets duplicated sequences and induces frequent C-to-T mutations, thus causing a strong depletion in GC-content over time (Galagan and Selker 2004). We find clear evidence of RIP in AT-rich regions across both studied *Epichloe* genomes. However, the AT-rich compartment is larger in *E. clarkii* compared to *E. typhina*. This appears to result from both novel AT-rich regions that exist in one species but not the other, as well as expansions of shared AT-rich regions. Genome compartmentalization in eukaryotes is thought to result from uneven rates of TE insertions along the genome. TE insertions in gene-rich regions are more likely to be deleterious by disrupting a coding sequence or

regulatory region of a gene and will be removed from the population by purifying selection. However, in largely non-coding sequence such as gene-sparse, repeat-rich regions, insertions may have little or no effect on fitness and thus they can persist and proliferate (Torres et al. 2020).

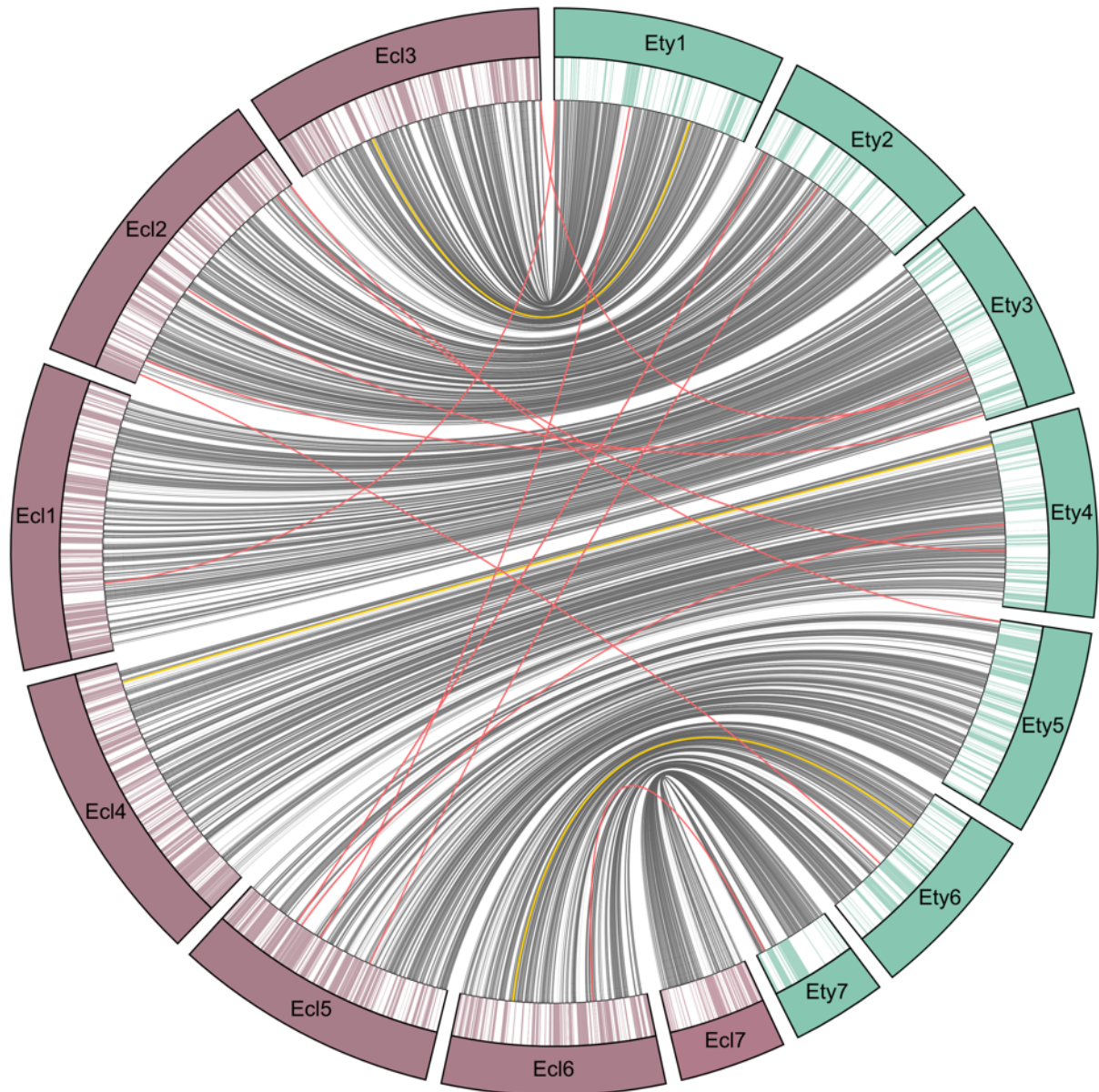


Figure 3: Synteny between *Epichloe* sibling species is conserved in gene-rich regions.

All seven chromosomes of Ety (green) and Ecl (purple) are shown as ideograms of chromosome structure with AT-rich regions in color. Links connect positions of 6,923 single-copy orthologous genes between species, 6,920 of which are located in gene-rich regions in both genomes. Only one gene on chromosome 6 in Ety and two genes on chromosomes 3 and 4 in Ecl are located in an AT-rich region in one species but not the other (yellow). Gene content is highly conserved and gene order strongly co-linear among homologous chromosomes. We found 13 gene pairs that were located on non-homologous chromosomes (red), indicating inter-chromosomal rearrangements.

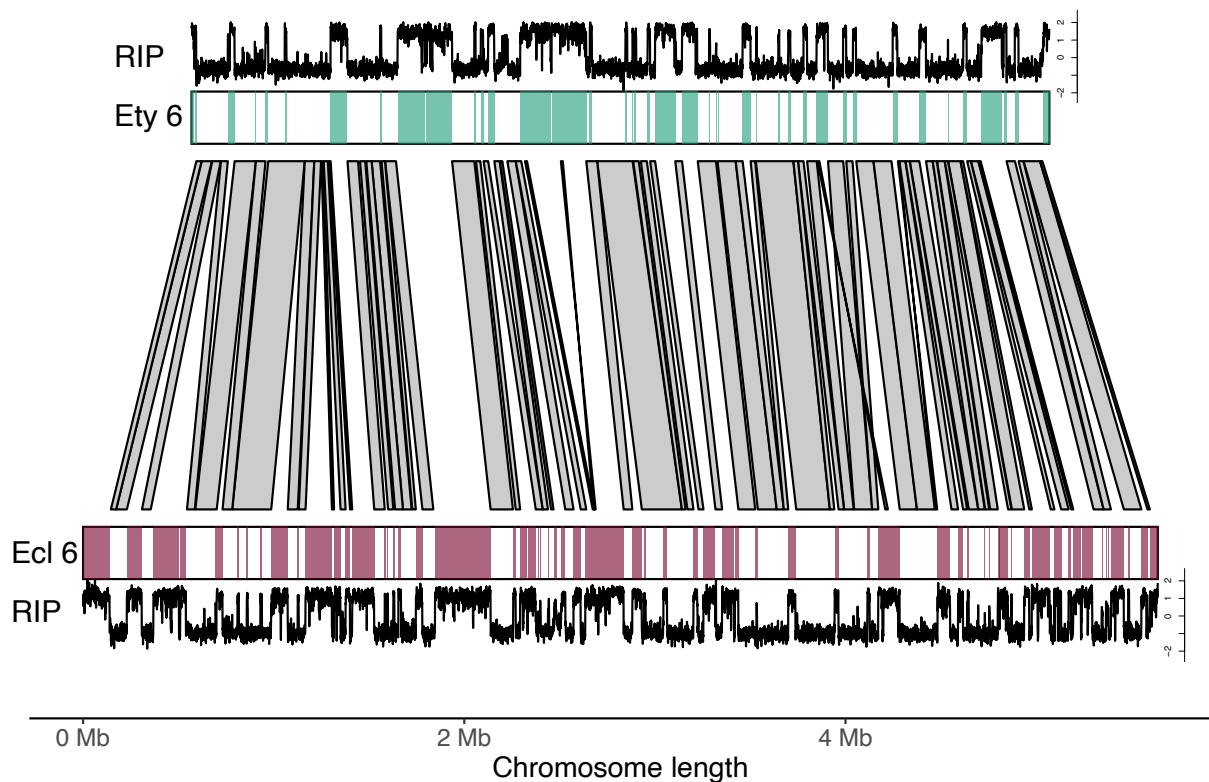


Figure 4: Synteny between *Epichloe* sibling species is conserved in gene-rich regions.

Syntenic regions were identified based on the alignment of homologous chromosomes aligning regions with sequence similarity >90% and a mapping quality >5 (grey polygons). Note that the plots only include alignments >10 kb for aesthetics; comparative figures with unfiltered alignments can be found in the Supplementary (Figure S15). Chromosomes of Ety (green) and Ecl (purple) are shown as ideograms with AT-rich regions identified by OcculterCut in color. Above and below the RIP index calculated with The RIPper in 10 kb windows along chromosomes is plotted with high values corresponding to a strong signature of RIP. Syntenic regions are limited to gene-rich regions with the exception of two short alignments both suggesting rearrangements. These likely correspond to the alignment of repetitive sequence and should therefore be interpreted with caution. Chromosome 6 is shown here as an example, the remaining chromosome pairs can be found in the Supplementary information (Figures S11-S15).

RIP can inactivate the TEs in these non-coding regions, and this generates blocks of AT-rich sequence. We show that the AT-rich compartments of *Epichloe* genomes are largely comprised of repetitive DNA. This is in line with the previously published *E. festucae* FII reference genome, where preexisting repeats were interrupted by new TE-insertions generating nested long repeats, all of which have been heavily targeted by RIP (Winter et al. 2018).

The size difference in the AT-rich compartment, and coincidentally genome size, between *E. typhina* and *E. clarkii* suggests that these genomes likely experienced independent surges of TE activity and subsequent inactivation by RIP. Expansion of the AT-rich

compartment does not appear to simply be the result of elongated AT-rich regions, which could be caused by recombination processes, such as replication slippage increasing the size of existing AT-rich regions. Instead, the lengths of AT-rich regions are similar and each genome contains species specific AT-rich regions supporting new TE invasions. The divergent remodeling of the AT-rich compartment could be explained by differences in the ecological and evolutionary histories of these species, altering the relative strength of selection and genetic drift acting on genomes, via changes in effective population size between species or populations (Möller and Stukenbrock 2017). A reduction in effective population size, for example following a bottleneck, could lead to a stronger effect of genetic drift and a temporarily weaker effect of purifying selection. This in turn can result in TE invasion of previously “protected” parts of the genome, giving rise to new repetitive regions that are targeted by RIP and results in the emergence of novel AT-rich regions (Lynch and Conery 2003). Similarly, phases of asexual reproduction and the absence of recombination may reduce the efficacy of purifying selection and lead to TE proliferation (Kent et al. 2017). Simultaneously, asexual reproduction renders meiotic genome defenses such as RIP inactive or ineffective and may thereby further promote the spread of TEs. Thus, fungal species that rarely undergo meiosis and have reduced RIP activity should exhibit more TE expansion compared to species with frequent sexual cycles and this is indeed supported by a growing number of studies (see Hane et al. 2015). The bottleneck associated with a host jump, invasion and or range expansion is expected to drastically reduce effective population size, and could potentially be followed by a prolonged phase of clonal spread through the new host population, as has been suggested in other fungal pathogens (Goodwin et al. 1994, Dutech et al. 2010, Ali et al. 2014, Grandaubert et al. 2014), for example, when only one mating type is present in a founder population. In such a scenario, neither natural selection nor RIP could prevent the invasion of the genome by TEs, and the consequential genome expansion until sexual reproduction (and RIP) became prevalent again with pressure to adapt and diversify. Given the differentially affected genomic composition in *Epichloe* genomes, we hypothesize that the divergence of *E. typhina* and *E. clarkii* may have been accompanied by a prolonged phase of asexual reproduction, and therefore small effective population size in *E. clarkii*, leading to the observed genome expansion while TE proliferation remained more limited in *E. typhina*.

The similarity of the two *Epichloe* genomes in terms of their chromosome-level synteny in the gene-rich compartment (core genome) indicates recent divergence of the sibling species *E. typhina* and *E. clarkii*. In artificial laboratory settings, strains of these species remain interfertile, supporting only recent divergence (Treindl and Leuchtman 2019). Despite this, the AT-rich compartments have diverged at a much higher rate compared to the core genome, consistent with the two-speed genome model. TE-dynamics and the activity of RIP have played key roles in the genetic divergence of these species, and the genomic environment of AT-rich regions may create hotspots of genetic diversity that serve as cradles for rapid adaptive evolution contributing to local adaptation and divergence at the population level. Here we discuss the mechanisms that may have differentially shaped fungal genomes in their evolutionary past and outline how a bottleneck and relaxed selection could explain the observed patterns. We propose that AT-rich regions may be important for evolution at both the within- and between-species level, however, testing this hypothesis will require genomic analyses of more than a single isolate. Future studies that combine comparative and population genomic approaches in these sibling species, will provide important insight into the role of genome compartmentalization in shaping genomic diversity, and understanding how this influences adaptation and speciation.

CONCLUSIONS

We have assembled and annotated chromosome-level genome sequences for two *Epichloe* sibling species, *E. typhina* and *E. clarkii*. Both genomes display a remarkable bipartite organization, with a gene-rich core genome that is highly conserved between species, interspersed with blocks of divergent AT-rich sequence. AT-rich regions are the result of repeat-induced point mutation (RIP) counteracting TE proliferation in fungal genomes. Effector genes possibly involved in host-pathogen interactions are overrepresented in the AT-bordering regions, where they may experience increased mutation rates due to RIP-leakage. Variation of repeat content and consequently the AT-rich compartment between ecologically similar and evolutionarily close species of these fungi accounts for the large increase in genome size in *E. clarkii* compared to *E. typhina*. These results demonstrate the important role that TE-activity and genome

defense mechanisms play in the genetic divergence of closely related taxa and highlights the role of such dynamic genomic compartments in shaping genome evolution. The genomic resource in the form of complete, high-quality reference genomes of two closely related *Epichloe* species, combined with population data, will serve as an ideal study system for the comparison of genomic architecture and patterns of sequence evolution both within and between species.

DECLARATIONS

Data deposition

Genome assemblies have been uploaded to GenBank (Accession numbers [PRJNA533210](#) for *E. typhina* Ety_1756 and [PRJNA533212](#) for *E. clarkii* Ecl_1605_22). PacBio and Illumina raw DNA reads and RNAseq data have been deposited in the Sequence Read Archive under the same respective accession numbers. A repository providing the scripts and files associated with this manuscript is available at https://github.com/adtreindl/Epichloe_genomes.

Declaration of competing interests

All authors declare that they have no competing interests.

Funding

This research was supported by the Swiss National Science Foundation (grant number 31003A_169269), the New Zealand Tertiary Education Commission via a Bio-Protection Research Centre grant, and the Royal Society Te Apārangi Marsden Fund (grant 18-MAU-02). The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

Author contributions

Artemis D. Treindl: Conceptualization, Methodology, Formal analysis, Data curation, Visualization, Writing; **Jessica Stapley:** Methodology, Software, Formal analysis, Data curation, Writing; **David J. Winter:** Resources, Software, Formal analysis, Writing –

review and editing; **Murray P. Cox**: Resources, Software, Formal analysis, Writing – review and editing; **Adrian Leuchtmann**: Conceptualization, Resources, Writing – review and editing, Supervision, Funding acquisition.

ACKNOWLEDGEMENTS

We would like to kindly thank Claudia Michel and Beatrice Arnold for laboratory assistance, Daniel Berry for advice concerning RNA extractions, and Stefan Zoller and Niklaus Zemp for bioinformatic support. Data presented and analyzed in this paper were generated in collaboration with the Genetic Diversity Centre (GDC), ETH Zurich and the Functional Genomics Center Zurich (FGCZ), UZH.

REFERENCES

- Al-Samarrai, T. H., and J. Schmid. 2000. A simple method for extraction of fungal genomic DNA. *Letters in Applied Microbiology* 30:53–56.
- Ali, S., P. Gladieux, M. Leconte, A. Gautier, A. F. Justesen, M. S. Hovmøller, J. Enjalbert, and C. de Vallavieille-Pope. 2014. Origin, migration routes and worldwide population genetic structure of the wheat yellow rust pathogen *Puccinia striiformis* f. sp. *tritici*. *PLoS Pathogens* 10:e1003903.
- Bao, Z., and S. R. Eddy. 2002. Automated de novo identification of repeat sequence families in sequenced genomes. *Genome research* 12:1269–1276.
- Bertazzoni, S., A. H. Williams, D. A. Jones, R. A. Syme, K.-C. Tan, and J. K. Hane. 2018. Accessories make the outfit: Accessory chromosomes and other dispensable DNA regions in plant-pathogenic fungi. *Molecular Plant-Microbe Interactions* 31:779–788.
- Cabanettes, F., and C. Klopp. 2018. D-GENIES: dot plot large genomes in an interactive, efficient and simple way. *PeerJ* 6:e4958.
- Chin, C.-S., P. Peluso, F. J. Sedlazeck, M. Nattestad, G. T. Concepcion, A. Clum, C. Dunn, R. O'Malley, R. Figueroa-Balderas, A. Morales-Cruz, G. R. Cramer, M. Delledonne, C. Luo, J. R. Ecker, D. Cantu, D. R. Rank, and M. C. Schatz. 2016. Phased diploid genome assembly with single-molecule real-time sequencing. *Nature Methods* 13:1050–1054.
- Christensen, M. J., A. Leuchtmann, D. D. Rowan, and B. A. Tapper. 1993. Taxonomy of *Acremonium* endophytes of tall fescue (*Festuca arundinacea*), meadow fescue (*F. pratensis*) and perennial ryegrass (*Lolium perenne*). *Mycological Research* 97:1083–1092.

- Croll, D., and B. A. McDonald. 2012. The accessory genome as a cradle for adaptive evolution in pathogens. *PLoS Pathogens* 8:8–10.
- Dobin, A., C. A. Davis, F. Schlesinger, J. Drenkow, C. Zaleski, S. Jha, P. Batut, M. Chaisson, and T. R. Gingeras. 2013. STAR: ultrafast universal RNA-seq aligner. *Bioinformatics* 29:15–21.
- Dong, S., S. Raffaele, and S. Kamoun. 2015. The two-speed genomes of filamentous pathogens: Waltz with plants. *Current Opinion in Genetics and Development* 35:57–65.
- Dutech, C., O. Fabreguettes, X. Capdevielle, and C. Robin. 2010. Multiple introductions of divergent genetic lineages in an invasive fungal pathogen, *Cryphonectria parasitica*, in France. *Heredity* 105:220–228.
- Flynn, J. M., R. Hubley, C. Goubert, J. Rosen, A. G. Clark, C. Feschotte, and A. F. Smit. 2020. RepeatModeler2 for automated genomic discovery of transposable element families. *Proceedings of the National Academy of Sciences of the United States of America* 117:9451–9457.
- Galagan, J. E., and E. U. Selker. 2004. RIP: The evolutionary cost of genome defense. *Trends in Genetics* 20:417–423.
- Goodwin, S. B., B. A. Cohen, and W. E. Fry. 1994. Panglobal distribution of a single clonal lineage of the Irish potato famine fungus. *Proceedings of the National Academy of Sciences of the United States of America* 91:11591–11595.
- Grandaubert, J., R. G. T. Lowe, J. L. Soyer, C. L. Schoch, A. P. Van De Wouw, I. Fudal, B. Robbertse, N. Lapalu, M. G. Links, B. Ollivier, J. Linglin, V. Barbe, S. Mangenot, C. Cruaud, H. Borhan, B. J. Howlett, M. H. Balesdent, and T. Rouxel. 2014. Transposable element-assisted evolution and adaptation to host plant within the *Leptosphaeria maculans*-*Leptosphaeria biglobosa* species complex of fungal pathogens. *BMC Genomics* 15:1–27.
- Hane, J. K., A. H. Williams, A. P. Taranto, P. S. Solomon, and R. P. Oliver. 2015. Repeat-induced point mutation: a fungal-specific, endogenous mutagenesis process. Pages 55–68 *Genetic Transformation Systems in Fungi*, Volume 2. Springer.
- Jones, P., D. Binns, H.-Y. Chang, M. Fraser, W. Li, C. McAnulla, H. McWilliam, J. Maslen, A. Mitchell, G. Nuka, S. Pesseat, A. F. Quinn, A. Sangrador-Vegas, M. Scheremetjew, S.-Y. Yong, R. Lopez, and S. Hunter. 2014. InterProScan 5: genome-scale protein function classification. *Bioinformatics (Oxford, England)* 30:1236–1240.
- Kauppinen, M., K. Saikkonen, M. Helander, A. M. Pirttilä, and P. R. Wäli. 2016. *Epichloë* grass endophytes in sustainable agriculture. *Nature Plants* 2:15224.
- Kent, T. V., J. Uzunović, and S. I. Wright. 2017. Coevolution between transposable elements and recombination. *Philosophical Transactions of the Royal Society B: Biological Sciences* 372.
- Koren, S., B. P. Walenz, K. Berlin, J. R. Miller, N. H. Bergman, and A. M. Phillippy. 2017. Canu: scalable and accurate long-read assembly via adaptive k-mer weighting and repeat separation. *Genome research* 27:722–736.
- Latch, G. C. M., and M. J. Christensen. 1985. Artificial infection of grasses with endophytes. *Annals of Applied Biology* 107:17–24.
- Lechner, M., S. Findeiß, L. Steiner, M. Marz, P. F. Stadler, and S. J. Prohaska. 2011. Proteinortho: Detection

- of (Co-)orthologs in large-scale analysis. *BMC Bioinformatics* 12:124.
- Leuchtman, A., C. W. Bacon, C. L. Schardl, J. F. White, and M. Tadych. 2014. Nomenclatural realignment of *Neotyphodium* species with genus *Epichloë*. *Mycologia* 106:202–215.
- Leuchtman, A., and K. Clay. 1988. *Atkinsonella hypoxylon* and *Balansia cyperi*, epiphytic members of the *Balansiae*. *Mycologia* 80:192–199.
- Leuchtman, A., and K. Clay. 1997. The population biology of grass endophytes. Pages 185–202 in G. C. Carroll and P. Tudzynski, editors. *The Mycota, Vol. V. Part B, Plant Relationships*. Springer-Verlag, Berlin, Heidelberg.
- Leuchtman, A., and C. L. Schardl. 1998. Mating compatibility and phylogenetic relationships among two new species of *Epichloë* and other congeneric European species. *Mycological Research* 102:1169–1182.
- Li, H. 2018. *Minimap2*: Pairwise alignment for nucleotide sequences. *Bioinformatics* 34:3094–3100.
- Lugtenberg, B. J. J., J. R. Caradus, and L. J. Johnson. 2016. Fungal endophytes for sustainable crop production. *FEMS Microbiology Ecology* 92.
- Lynch, M., and J. S. Conery. 2003. The origins of genome complexity. *Science* 302:1401–1404.
- Ma, L.-J., H. C. Van Der Does, K. A. Borkovich, J. J. Coleman, M.-J. Daboussi, A. Di Pietro, M. Dufresne, M. Freitag, M. Grabherr, and B. Henrissat. 2010. Comparative analysis reveals mobile pathogenicity chromosomes in *Fusarium*. *Nature* 464:367–373.
- Margolin, B. S., P. W. Garrett-Engele, J. N. Stevens, D. Y. Fritz, C. Garrett-Engele, R. L. Metzberg, and E. U. Selker. 1998. A methylated *Neurospora* 5S rRNA pseudogene contains a transposable element inactivated by repeat-induced point mutation. *Genetics* 149:1787–1797.
- Möller, M., and E. H. Stukenbrock. 2017. Evolution and genome architecture in fungal plant pathogens. *Nature Reviews Microbiology* 15:756–771.
- Moon, C. D., K. D. Craven, A. Leuchtman, S. L. Clement, and C. L. Schardl. 2004. Prevalence of interspecific hybrids amongst asexual fungal endophytes of grasses. *Molecular Ecology* 13:1455–67.
- Palmer, J. M. 2016. *Funannotate*: pipeline for genome annotation.
- Pertea, M., G. M. Pertea, C. M. Antonescu, T.-C. Chang, J. T. Mendell, and S. L. Salzberg. 2015. *StringTie* enables improved reconstruction of a transcriptome from RNA-seq reads. *Nature Biotechnology* 33:290–295.
- Petersen, T. N., S. Brunak, G. von Heijne, and H. Nielsen. 2011. *SignalP 4.0*: discriminating signal peptides from transmembrane regions. *Nature Methods* 8:785–786.
- Plissonneau, C., F. E. Hartmann, and D. Croll. 2018. Pangenome analyses of the wheat pathogen *Zymoseptoria tritici* reveal the structural basis of a highly plastic eukaryotic genome. *BMC Biology* 16:5.
- Raffaele, S., R. A. Farrer, L. M. Cano, D. J. Studholme, D. MacLean, M. Thines, R. H. Y. Jiang, M. C. Zody, S. G. Kunjeti, and N. M. Donofrio. 2010. Genome evolution following host jumps in the Irish potato famine pathogen lineage. *Science* 330:1540–1543.
- Raffaele, S., and S. Kamoun. 2012. Genome evolution in filamentous plant pathogens: Why bigger can be better. *Nature Reviews Microbiology* 10:417–430.
- Rouxel, T., J. Grandaubert, J. K. Hane, C. Hoede, A. P. van de Wouw, A. Couloux, V. Dominguez, V.

- Anthouard, P. Bally, S. Bourras, A. J. Cozijnsen, L. M. Ciuffetti, A. Degrave, A. Dilmaghani, L. Duret, I. Fudal, S. B. Goodwin, L. Gout, N. Glaser, J. Linglin, G. H. J. Kema, N. Lapalu, C. B. Lawrence, K. May, M. Meyer, B. Ollivier, J. Poulain, C. L. Schoch, A. Simon, J. W. Spatafora, A. Stachowiak, B. G. Turgeon, B. M. Tyler, D. Vincent, J. Weissenbach, J. Amselem, H. Quesneville, R. P. Oliver, P. Wincker, M.-H. Balesdent, and B. J. Howlett. 2011. Effector diversification within compartments of the *Leptosphaeria maculans* genome affected by Repeat-Induced Point mutations. *Nature Communications* 2:202.
- Saikkonen, K., C. A. Young, M. Helander, and C. L. Schardl. 2016. Endophytic *Epichloë* species and their grass hosts: from evolution to applications. *Plant Molecular Biology* 90:665–675.
- Schardl, C. L., C. A. Young, U. Hesse, S. G. Amyotte, K. Andreeva, P. J. Calie, D. J. Fleetwood, D. C. Haws, N. Moore, B. Oeser, D. G. Panaccione, K. K. Schweri, C. R. Voisey, M. L. Farman, J. W. Jaromczyk, B. A. Roe, D. M. O’Sullivan, B. Scott, P. Tudzynski, Z. An, E. G. Arnaoudova, C. T. Bullock, N. D. Charlton, L. Chen, M. Cox, R. D. Dinkins, S. Florea, A. E. Glenn, A. Gordon, U. Güldener, D. R. Harris, W. Hollin, J. Jaromczyk, R. D. Johnson, A. K. Khan, E. Leistner, A. Leuchtmann, C. Li, J. Liu, J. Liu, M. Liu, W. Mace, C. Machado, P. Nagabhyru, J. Pan, J. Schmid, K. Sugawara, U. Steiner, J. E. Takach, E. Tanaka, J. S. Webb, E. V. Wilson, J. L. Wiseman, R. Yoshida, and Z. Zeng. 2013a. Plant-Symbiotic Fungi as Chemical Engineers: Multi-Genome Analysis of the Clavicipitaceae Reveals Dynamics of Alkaloid Loci. *PLoS Genetics* 9:e1003323.
- Schardl, C. L., C. A. Young, N. Moore, N. Krom, P.-Y. Dupont, J. Pan, S. Florea, J. S. Webb, J. Jaromczyk, and J. W. Jaromczyk. 2014. Genomes of plant-associated Clavicipitaceae. *Advances in Botanical Research* 70:291–327.
- Schardl, C. L., C. A. Young, J. Pan, S. Florea, J. E. Takach, D. G. Panaccione, M. L. Farman, J. S. Webb, J. Jaromczyk, N. D. Charlton, P. Nagabhyru, L. Chen, C. Shi, and A. Leuchtmann. 2013b. Currencies of mutualisms: Sources of alkaloid genes in vertically transmitted epichloae. *Toxins* 5:1064–1088.
- Schirrmann, M. K., S. Zoller, D. Croll, E. H. Stukenbrock, A. Leuchtmann, and S. Fior. 2018. Genomewide signatures of selection in *Epichloë* reveal candidate genes for host specialization. *Molecular Ecology* 27:3070–3086.
- Schirrmann, M. K., S. Zoller, S. Fior, and A. Leuchtmann. 2015. Genetic evidence for reproductive isolation among sympatric *Epichloë* endophytes as inferred from newly developed microsatellite markers. *Microbial Ecology* 70:51–60.
- Selker, E. U. 2002. Repeat-induced gene silencing in fungi. *Advances in Genetics* 46:439–450.
- Simão, F. A., R. M. Waterhouse, P. Ioannidis, E. V Kriventseva, and E. M. Zdobnov. 2015. BUSCO: assessing genome assembly and annotation completeness with single-copy orthologs. *Bioinformatics* 31:3210–3212.
- Sperschneider, J., D. M. Gardiner, P. N. Dodds, F. Tini, L. Covarelli, K. B. Singh, J. M. Manners, and J. M. Taylor. 2016. EffectorP: predicting fungal effector proteins from secretomes using machine learning. *The New Phytologist* 210:743–761.
- Tarailo-Graovac, M., and N. Chen. 2009. Using RepeatMasker to identify repetitive elements in genomic sequences. *Current Protocols in Bioinformatics* 25:4.10.1-4.10.14.

- Testa, A. C., R. P. Oliver, and J. K. Hane. 2016. OcculterCut: A comprehensive survey of AT-rich regions in fungal genomes. *Genome Biology and Evolution* 8:2044–2064.
- Torres, D. E., U. Oggenfuss, D. Croll, and M. F. Seidl. 2020. Genome evolution in fungal plant pathogens: Looking beyond the two-speed genome model. *Fungal Biology Reviews* 34:136–143.
- Treindl, A. D., and A. Leuchtman. 2019. Assortative mating in sympatric ascomycete fungi revealed by experimental fertilizations. *Fungal Biology* 123:676–686.
- Turland, N. J., J. H. Wiersema, F. R. Barrie, W. Greuter, D. L. Hawksworth, P. S. Herendeen, S. Knapp, W.-H. Kusber, D.-Z. Li, and K. Marhold. 2018. International Code of Nomenclature for algae, fungi, and plants (Shenzhen Code) adopted by the Nineteenth International Botanical Congress Shenzhen, China, July 2017. Koeltz Botanical Books, Glashütten.
- Walker, B. J., T. Abeel, T. Shea, M. Priest, A. Abouelliel, S. Sakhikumar, C. A. Cuomo, Q. Zeng, J. Wortman, S. K. Young, and A. M. Earl. 2014. Pilon: An integrated tool for comprehensive microbial variant detection and genome assembly improvement. *PLoS ONE* 9:e112963.
- West, C. P., E. Izekor, K. E. Turner, and A. A. Elmi. 1993. Endophyte effects on growth and persistence of tall fescue along a water-supply gradient. *Agronomy Journal* 85:264–270.
- White Jr., J. F. 1997. Perithecial ontogeny in the fungal genus *Epichloë*: An examination of the clavicipitalean centrum. *American Journal of Botany* 84:170–178.
- Winter, D. J., A. R. D. Ganley, C. A. Young, I. Liachko, C. L. Schardl, P. Y. Dupont, D. Berry, A. Ram, B. Scott, and M. P. Cox. 2018. Repeat elements organise 3D genome structure and mediate transcription in the filamentous fungus *Epichloë festucae*. *PLoS Genetics* 14:1–29.
- Van Wyk, S., C. H. Harrison, B. D. Wingfield, L. De Vos, N. A. van Der Merwe, and E. T. Steenkamp. 2019. The RIPper, a web-based tool for genome-wide quantification of Repeat-Induced Point (RIP) mutations. *PeerJ* 7:e7447.
- Xiao, C.-L., Y. Chen, S.-Q. Xie, K.-N. Chen, Y. Wang, Y. Han, F. Luo, and Z. Xie. 2017. MECAT: fast mapping, error correction, and de novo assembly for single-molecule sequencing reads. *Nature Methods* 14:1072–1074.

SUPPLEMENTARY MATERIALS

Table S1: *Epichloe* reference strain information

	Strain ID	Origin	Latitude	Longitude	Host plant
<i>Epichloe typhina</i>	Ety_1756	Aubonne VD	46.51232	6.363653	<i>Dactylis glomerata</i>
<i>Epichloe clarkii</i>	Ecl_1605_22	Soglio GR	46.34289	9.537262	<i>Holcus lanatus</i>

Strain IDs, sampling locations, sampling coordinates and host species of *Epichloe* strains for reference genome assemblies.

Table S2: BUSCO statistics of genome assemblies

	<i>Epichloe typhina</i>	<i>Epichloe clarkii</i>
Complete BUSCOs (C)	3632	3604
Complete and single-copy BUSCOs (S)	3626	3600
Complete and duplicated BUSCOs (D)	6	4
Fragmented BUSCOs (F)	31	46
Missing BUSCOs (M)	62	75
Total BUSCO groups searched	3725	3725

Table S3: Transposable element content in *Epichloe* genomes

TE class	<i>Epichloe typhina</i>		<i>Epichloe clarkii</i>	
	total	proportion AT	total	proportion AT
DNA/hAT-Ac	90815	1.00	0	--
DNA/MuLE-MuDR	35061	0.65	171004	0.8
DNA/PIF-Harbinger	0	--	83447	0.67
DNA/PiggyBac	0	--	17200	0.74
DNA/TcMar-Mariner	8221	0.80	28817	0.74
LINE/Tad1	225523	0.99	459845	0.97
LTR	142462	1.00	1298656	0.99
LTR/Copia	3217404	0.98	7842902	0.96
LTR/Gypsy	4096227	0.99	7141460	0.98
Unclassified	1315346	0.70	1890478	0.83
Total	9131059	0.95	18933809	0.95

TE class lists the class of transposable element using Dfam classification with the total number of bp annotated to that class in the second column (total); proportion AT indicates what proportion of bases annotated to this class, is also part of an AT-rich region.

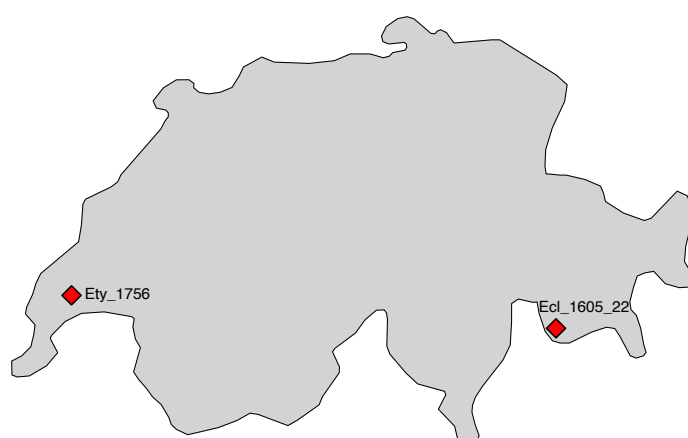


Figure S1: Sampling locations of *Epichloe* reference strains.

Sampling locations in Switzerland indicated by red diamonds in Aubonne (*E. typhina*) and Soglio (*E. clarkii*).

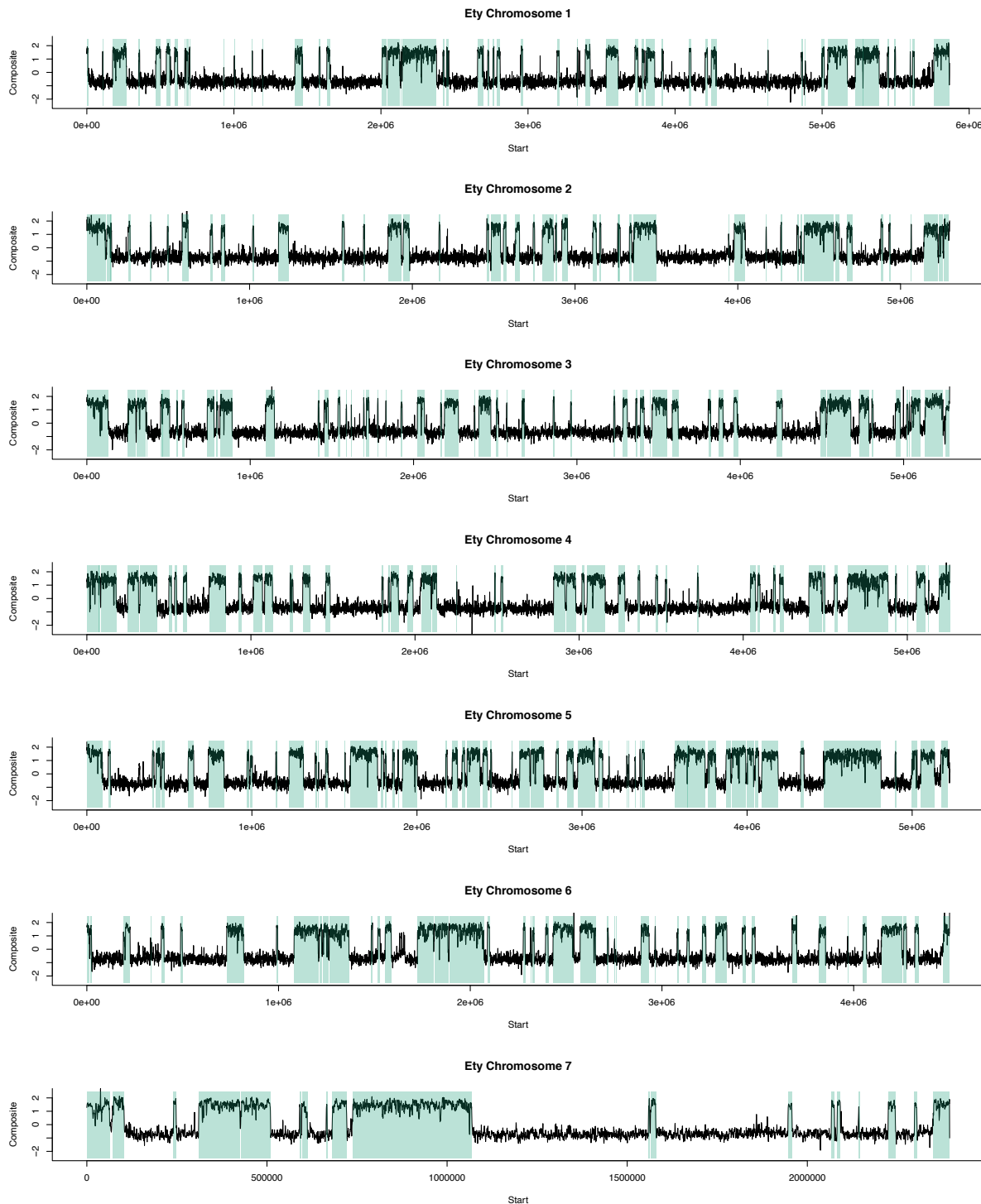


Figure S2: Concordance of RIP signatures along chromosomes of *E. typhina*.

Regions with high RIP index value corresponding to a strong signature of RIP identified by The RIPper match the qualitative assignment of AT-rich regions identified by OcculterCut (colored rectangles). We additionally tested the significance of the overlap between regions with a strong signature of RIP (Large-RIP-Affected-Regions, LRARs) and AT-rich regions using a permutation test (1000 iterations, $p < 0.001$, $Z = 12.19$). The overlap is significantly higher than would be expected by chance.

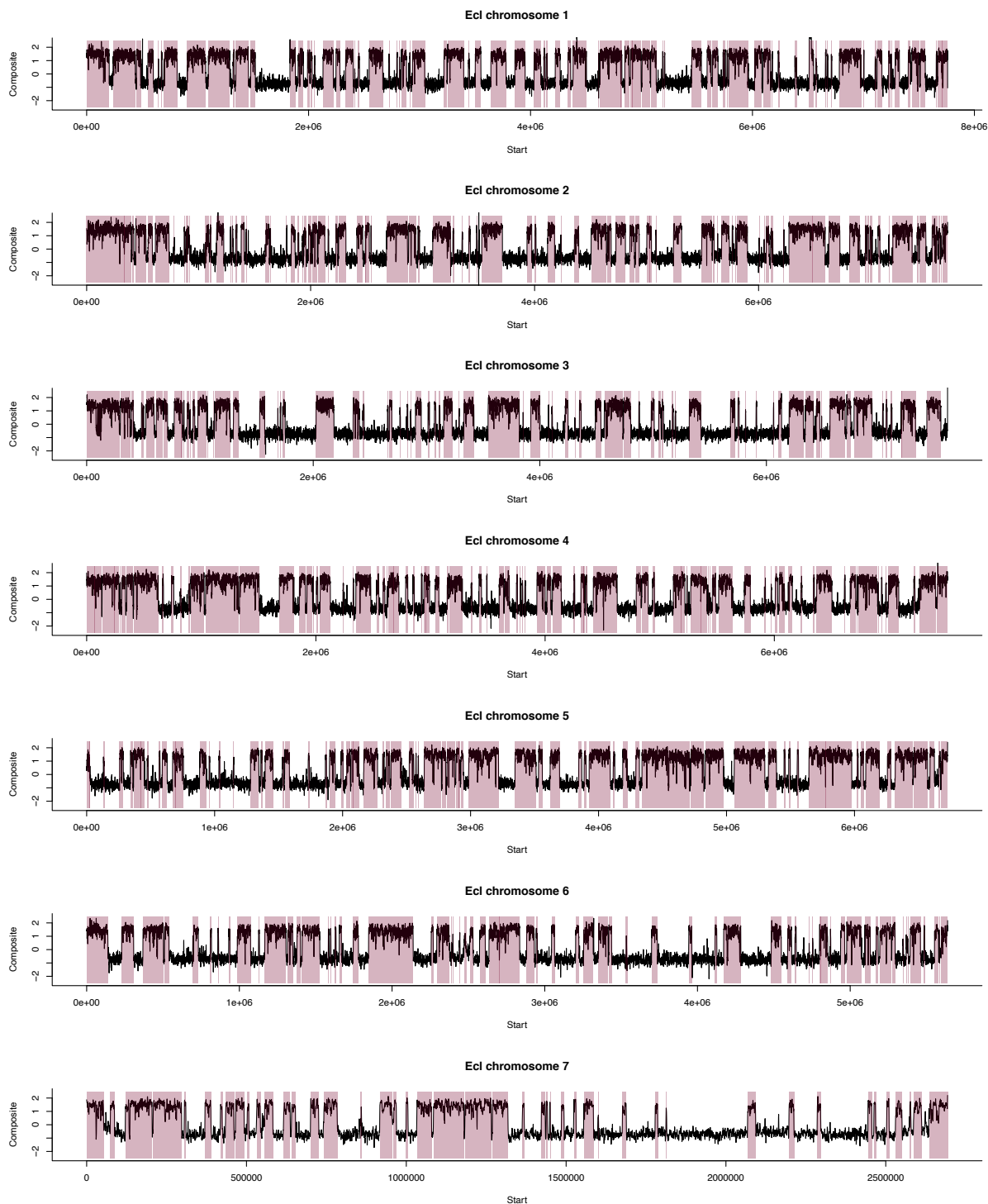


Figure S3: Concordance of RIP signatures along chromosomes of *E. clarkii*.

Regions with high RIP index value corresponding to a strong signature of RIP identified by The RIPper match the qualitative assignment of AT-rich regions identified by OcculterCut (colored rectangles). We additionally tested the significance of the overlap between regions with a strong signature of RIP (Large-RIP-Affected-Regions, LRARs) and AT-rich regions using a permutation test (1000 iterations, $p < 0.001$, $Z = 8.225$). The overlap is significantly higher than would be expected by chance.

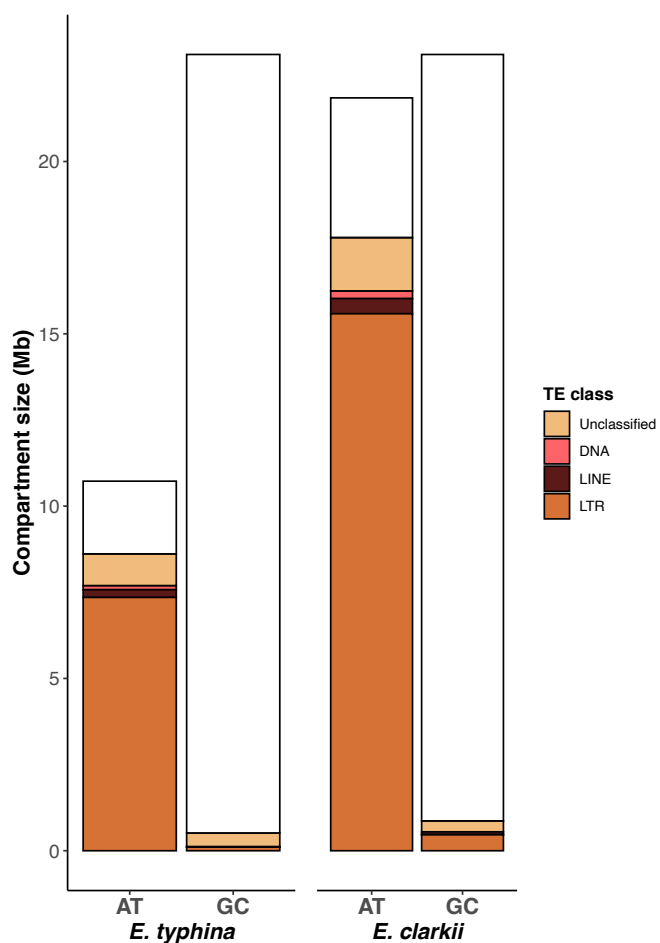


Figure S4: The AT-rich compartments are largely comprised of repetitive DNA.

TEs were classified following Dfam classification (<https://dfam.org/classification>) and the number of basepairs belonging to a given class plotted according to which genomic compartment they were annotated in. Compartment size is represented by the total bar height whereby AT refers to the AT-rich compartment while GC refers to the gene-rich compartment. The AT-rich compartment consists largely of TEs, with 81% of bases in both the Ety and Ecl genomes showing similarity to a known repeat (see Table S3). The composition of TEs in each genome is similar, with the majority of sequences included in a TE being part of an LTR retrotransposon (82% of all TE bases in Ety and 86% in Ecl), and DNA transposons, LINE elements and currently unclassified elements present in both genomes.

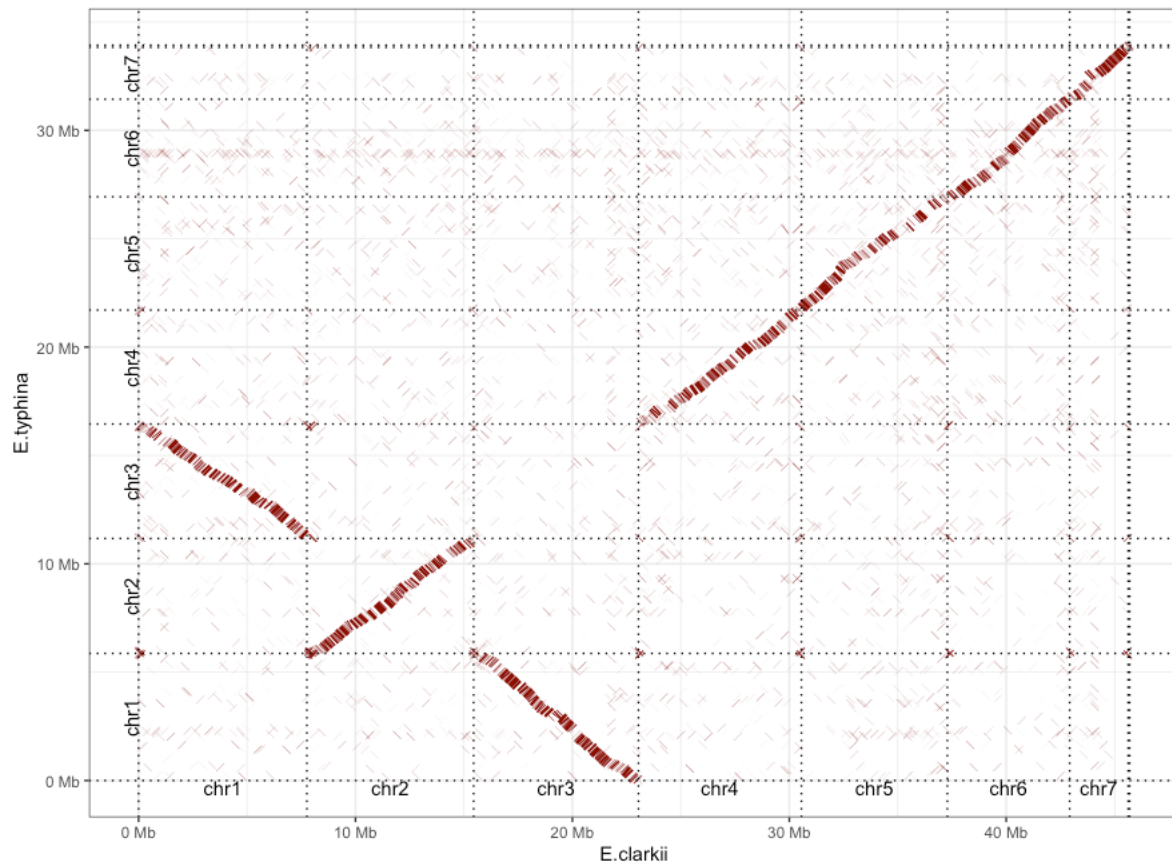


Figure S5: Genome alignment reveals high degree of synteny between *E. typhina* and *E. clarkii*.

Whole genome dotplot of alignment of *E. typhina* (Ety, y-axis) and *E. clarkii* (Ecl, x-axis) genomes show no major rearrangements between chromosomes. Due to naming conventions chromosome 1 in Ety corresponds to the inverted chromosome 3 in Ecl and vice versa.

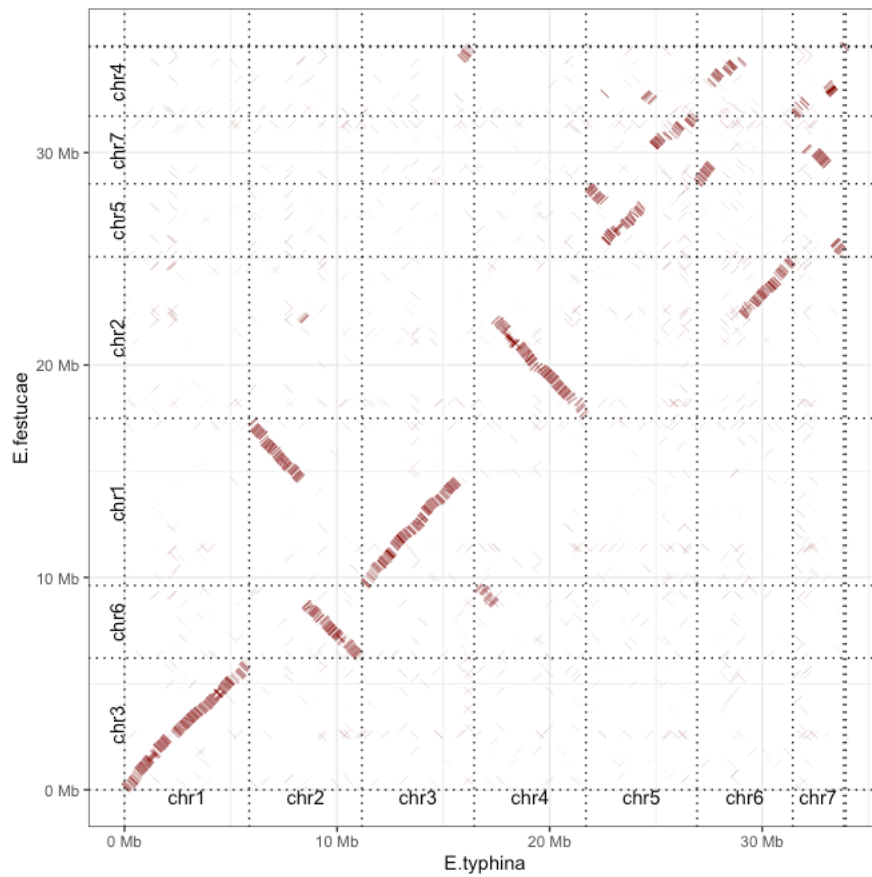


Figure S6: Whole genome alignment between *E. typhina* and *E. festucae*.

Whole genome dotplot of alignment of *E. typhina* (Ety, x-axis) and *E. festucae* (F11, y-axis) genomes show rearrangements between chromosomes. Synteny is retained in some parts of the genomes, for example between chromosome 1 in Ety and the homologous chromosome 3 in F11.

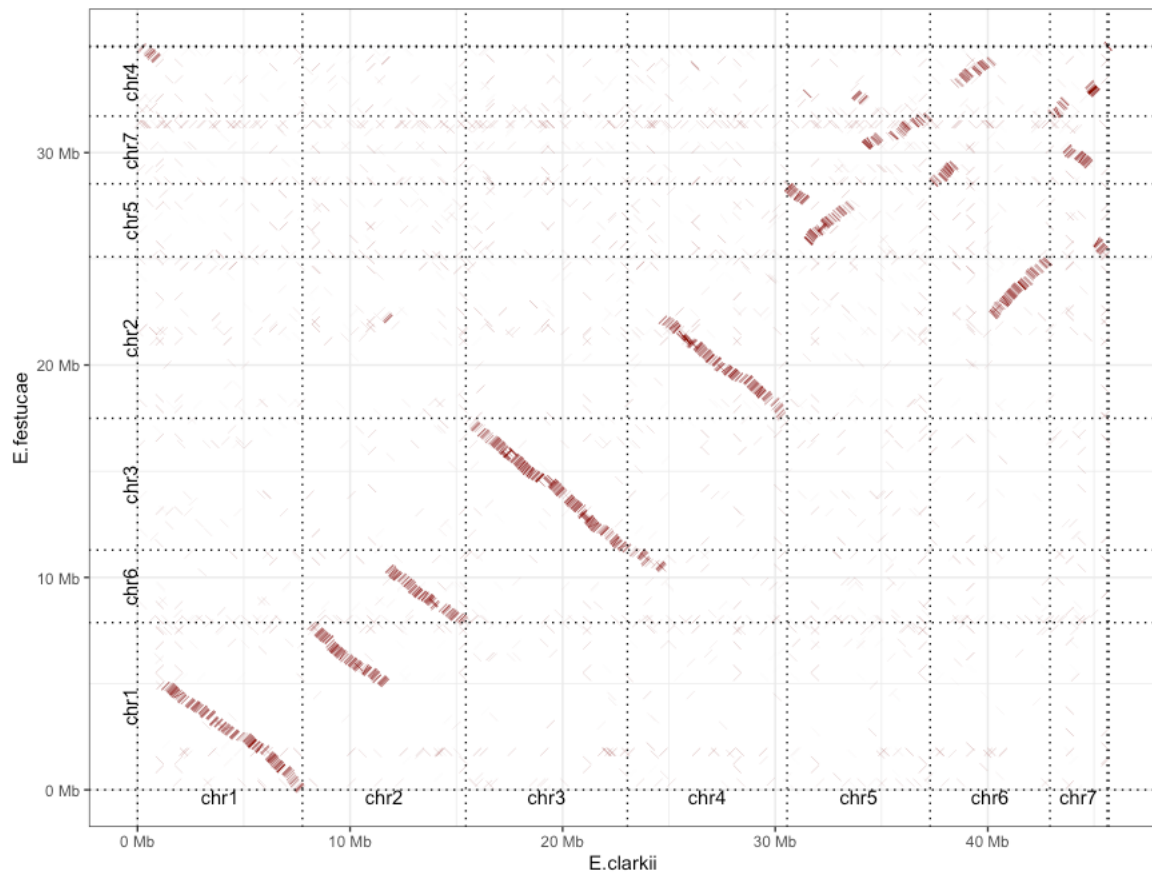
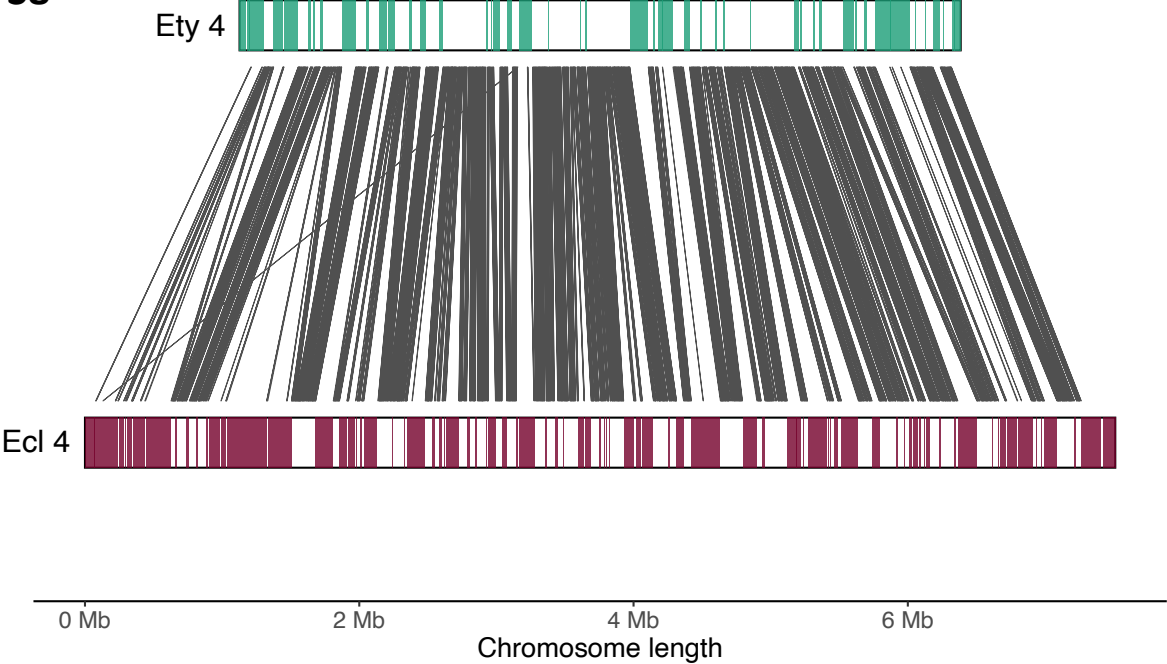


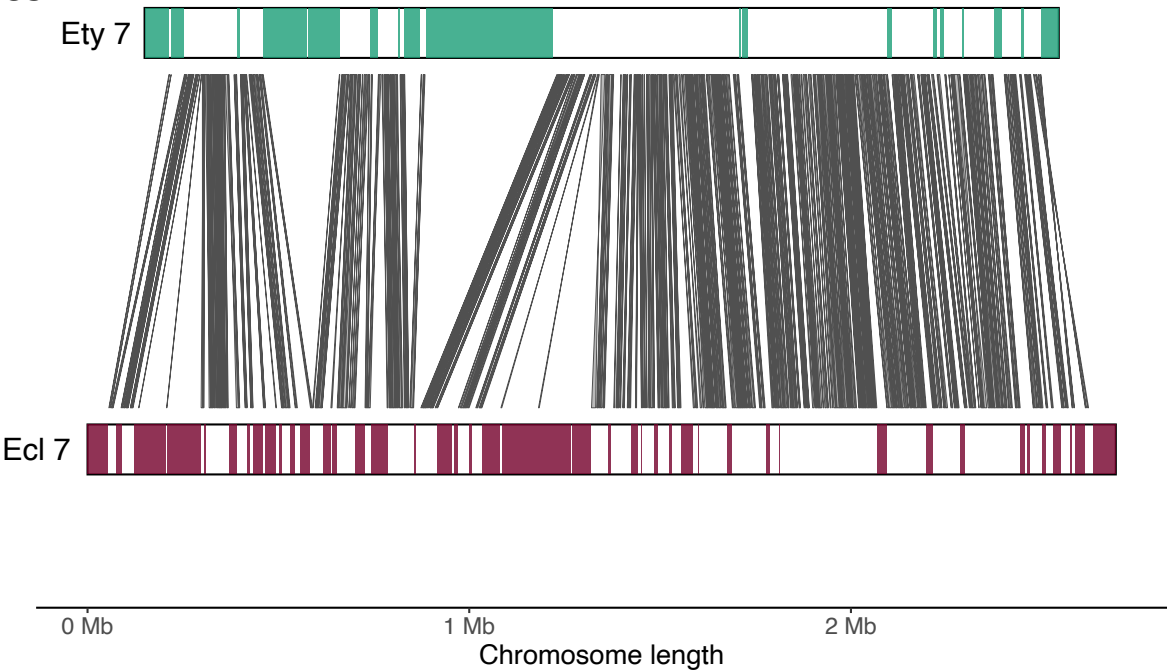
Figure S7: Whole genome alignment between *E. clarkii* and *E. festucae*.

Whole genome dotplot of alignment of *E. clarkii* (Ecl, x-axis) and *E. festucae* (Fl1, y-axis) genomes show rearrangements between chromosomes. Synteny is retained in some parts of the genomes, for example between chromosome 3 in Ecl and the homologous chromosome 3 in Fl1.

S8

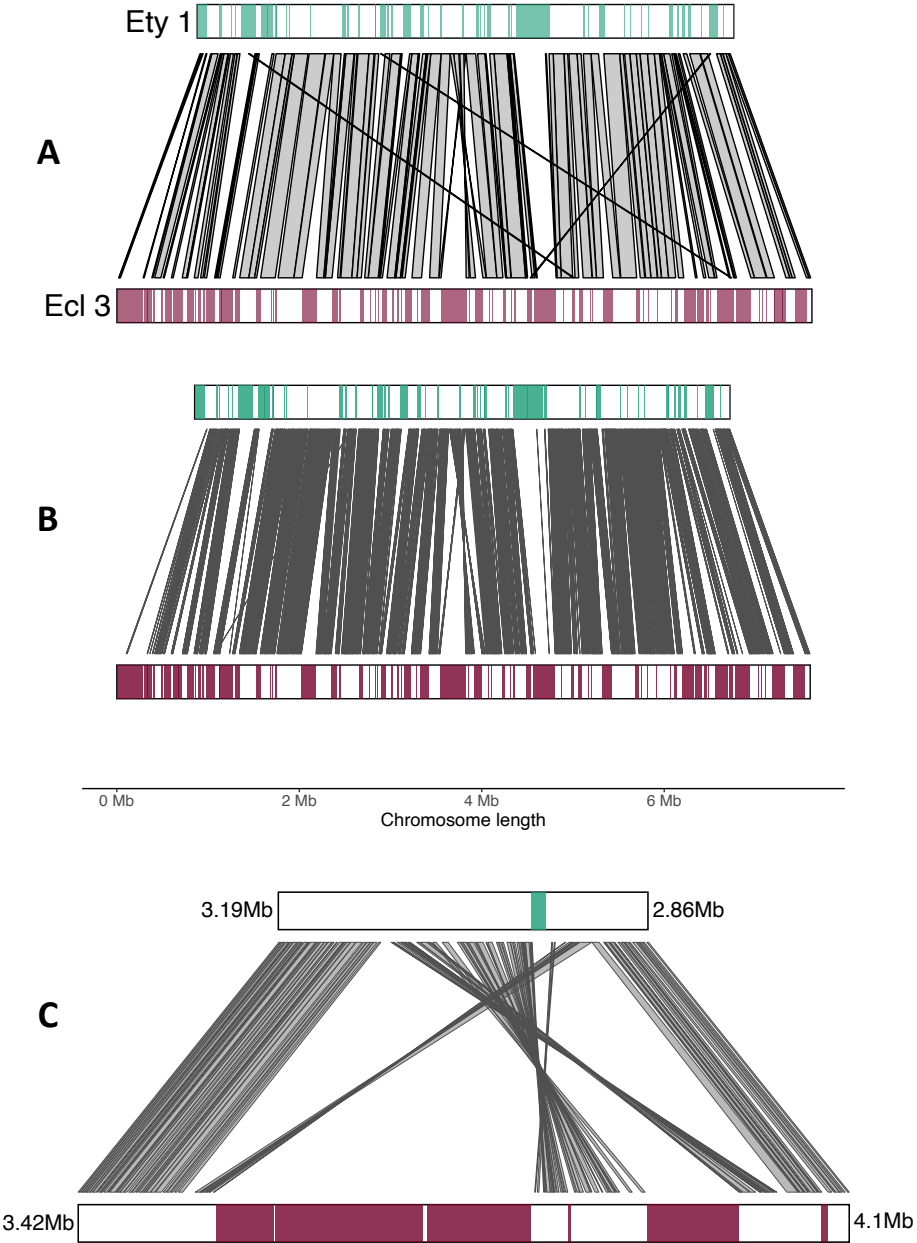


S9



Supplementary Figures S8 and S9: Conservation of gene-level synteny among homologous chromosomes.

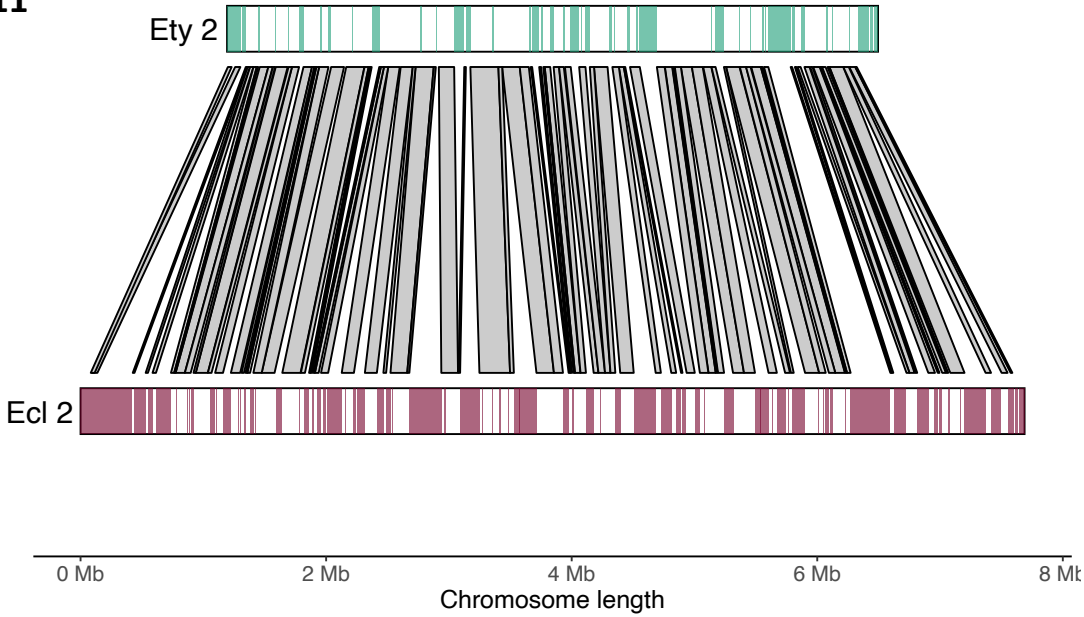
Homologous chromosome pairs of Ety (green) and Ecl (purple) are shown as ideograms with AT-rich regions in color for chromosomes four and seven. Links represent orthologous genes. Fig. S8 shows rearrangement of a single gene out of 1062 in respect to gene order. Fig. S9 shows no rearrangements.



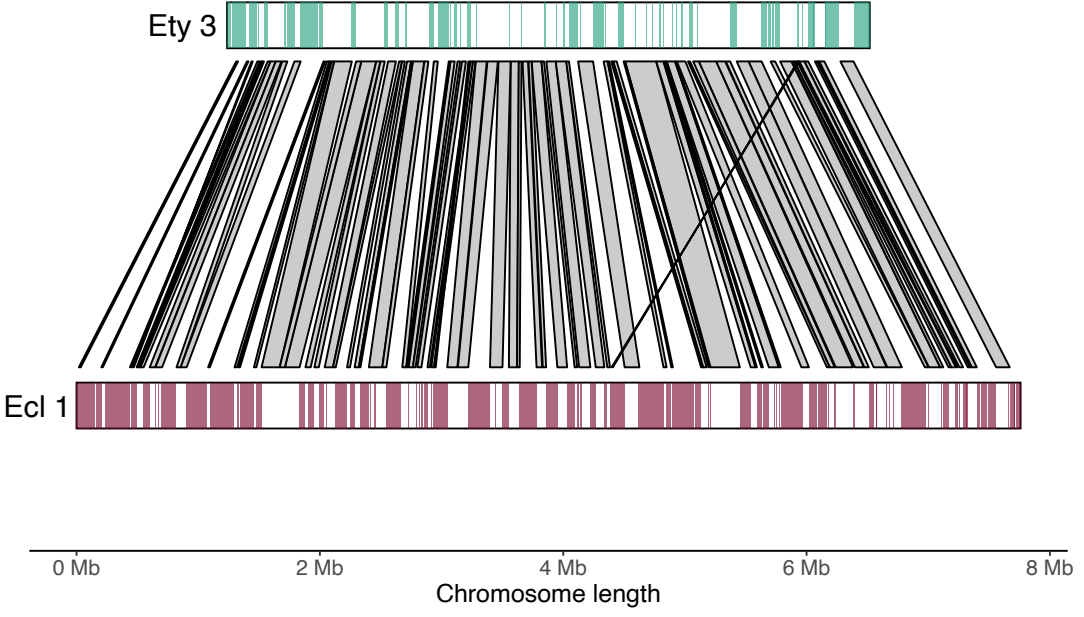
Supplementary Figure S10: Rearrangement between Ecl chr 3 and Ety chr 1.

Homologous chromosome pair of Ety (green) and Ecl (purple) are shown as ideograms with AT-rich regions in color. Links represent alignments **A**) and orthologous genes **B**). The inversions and translocations in the center of the chromosomes comprise 47 orthologous genes. A zoomed-in window is shown in **C**) spanning 676,232 bp on Ecl3 and the homologous region of 324,193 bp on Ety1. Because chromosome 1 in Ety corresponds to the inverted chromosome 3 in Ecl we used the reverse complement of the target sequence, in this case Ety1, for plotting. The corresponding ideograms were modified accordingly to show inverse order of AT-rich regions.

S11



S12



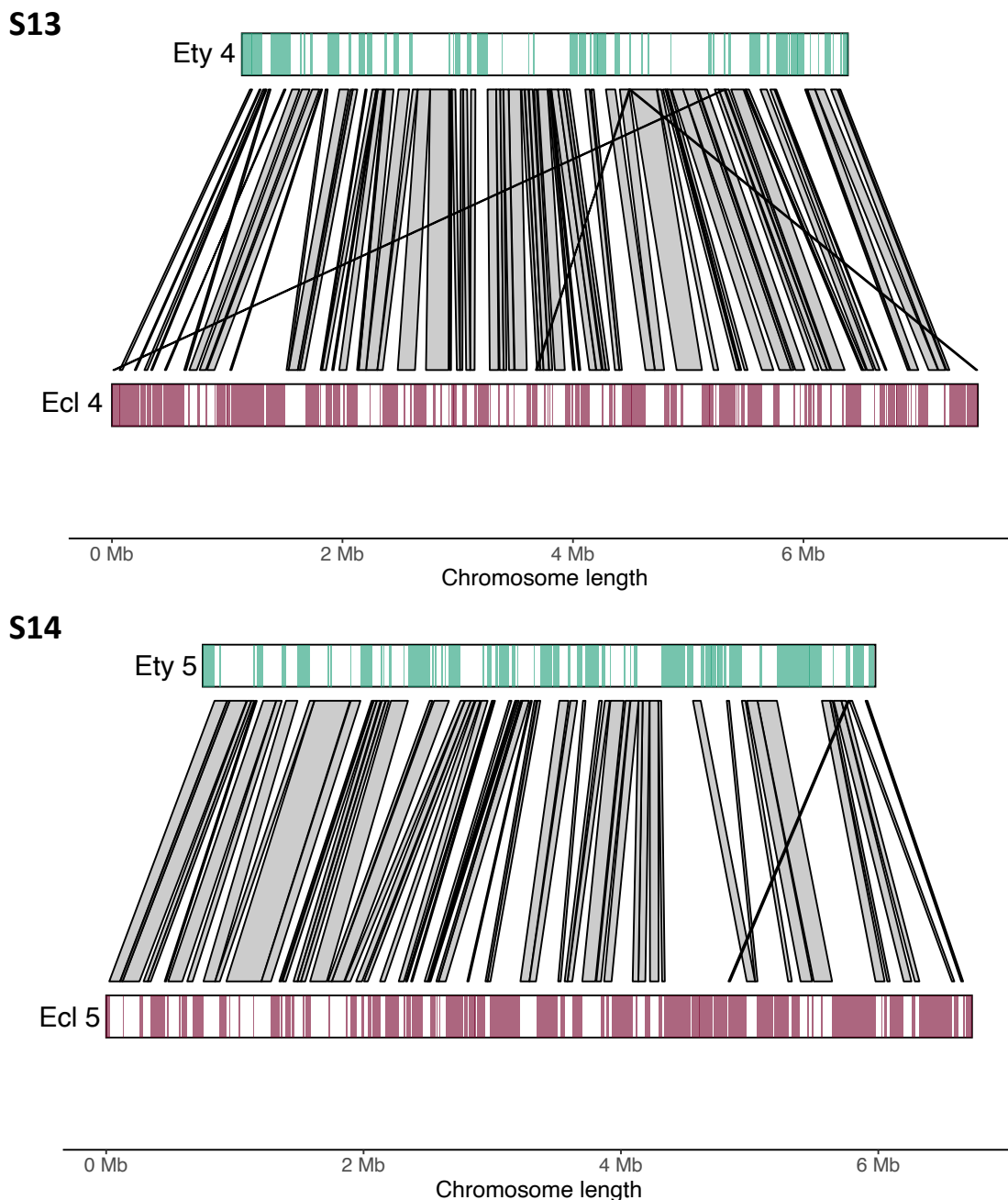


Figure S11-S14: Synteny between homologous chromosomes is conserved in gene-rich regions.

Syntenic regions were identified based on the alignment of homologous chromosomes aligning regions with sequence similarity >90% and a mapping quality >5 (grey polygons). Note that the plots only include alignments >10kb for aesthetics, comparative figures of chromosome 7 with unfiltered alignments can be found in FigS15. Chromosomes of Et4 (green) and Ecl4 (purple) are shown as ideograms with AT-rich regions identified by OcculterCut in color. Syntenic regions are limited to gene-rich regions with the exception of few short alignments. These likely correspond to the alignment of repetitive sequence and putative rearrangements should therefore be interpreted with caution. **S11:** Et2-Ecl2; **S12:** Et3-Ecl1; **S13:** Et4-Ecl4; **S14:** Et5-Ecl5. Because chromosome 3 in Et4 corresponds to the inverted chromosome 1 in Ecl4 we used the reverse complement of the target sequence, in this case Et3, for plotting. The corresponding ideogram was modified accordingly to show inverse order of AT-rich regions.

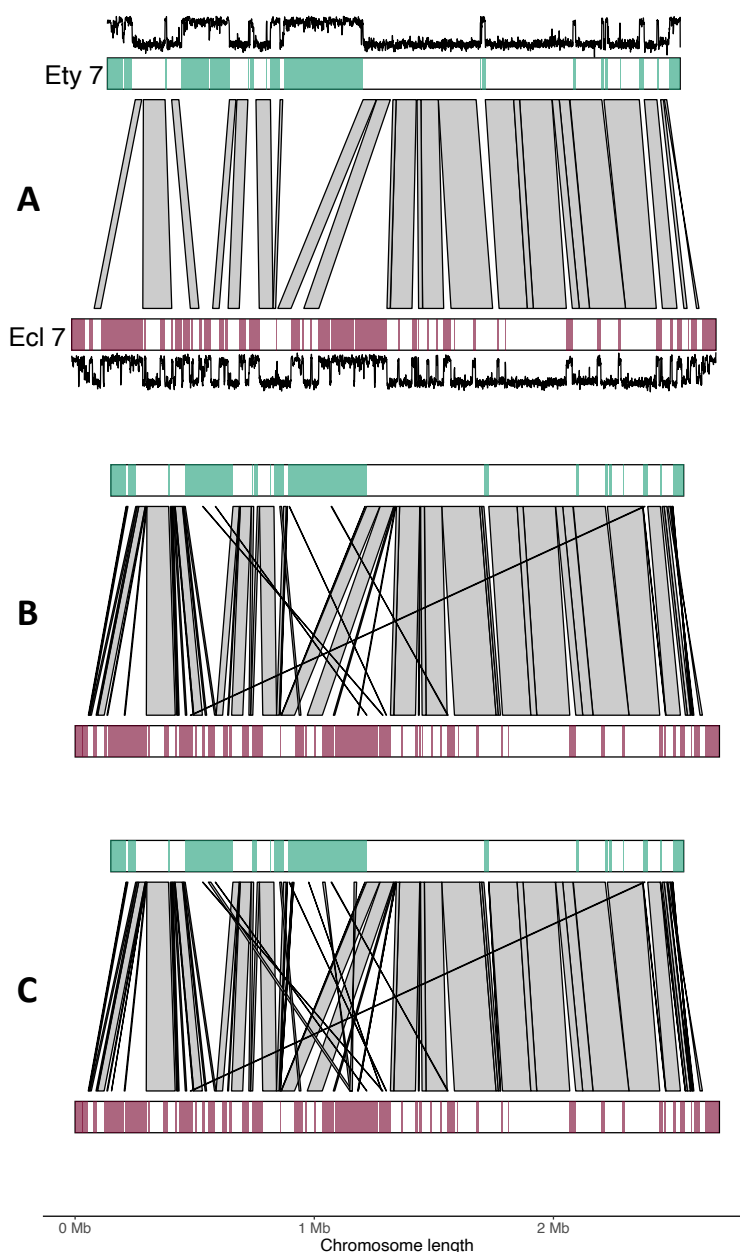


Figure S15: Long alignments represent syntenic regions between homologous chromosomes.

Aligned regions with sequence similarity >90% between homologous chromosomes 7 of Ety (top) and Ecl (bottom) are shown as grey polygons. The green and purple blocks indicate locations of AT-rich regions on the respective chromosomes. **A)** includes alignments $\geq 10\text{kb}$ and mapping quality > 5 , **B)** includes alignments $\geq 1\text{kb}$ and mapping quality > 5 , and **C)** shows unfiltered alignments. Plotting long alignments $> 10\text{kb}$ is representative for visualization of synteny in gene-rich regions. Note the large AT-rich block spanning 330kb in Ety which is not present in Ecl and, vice versa, a similarly large yet fragmented AT-rich region in Ecl which is not present in Ety. Nevertheless, synteny in surrounding gene-rich regions is conserved.

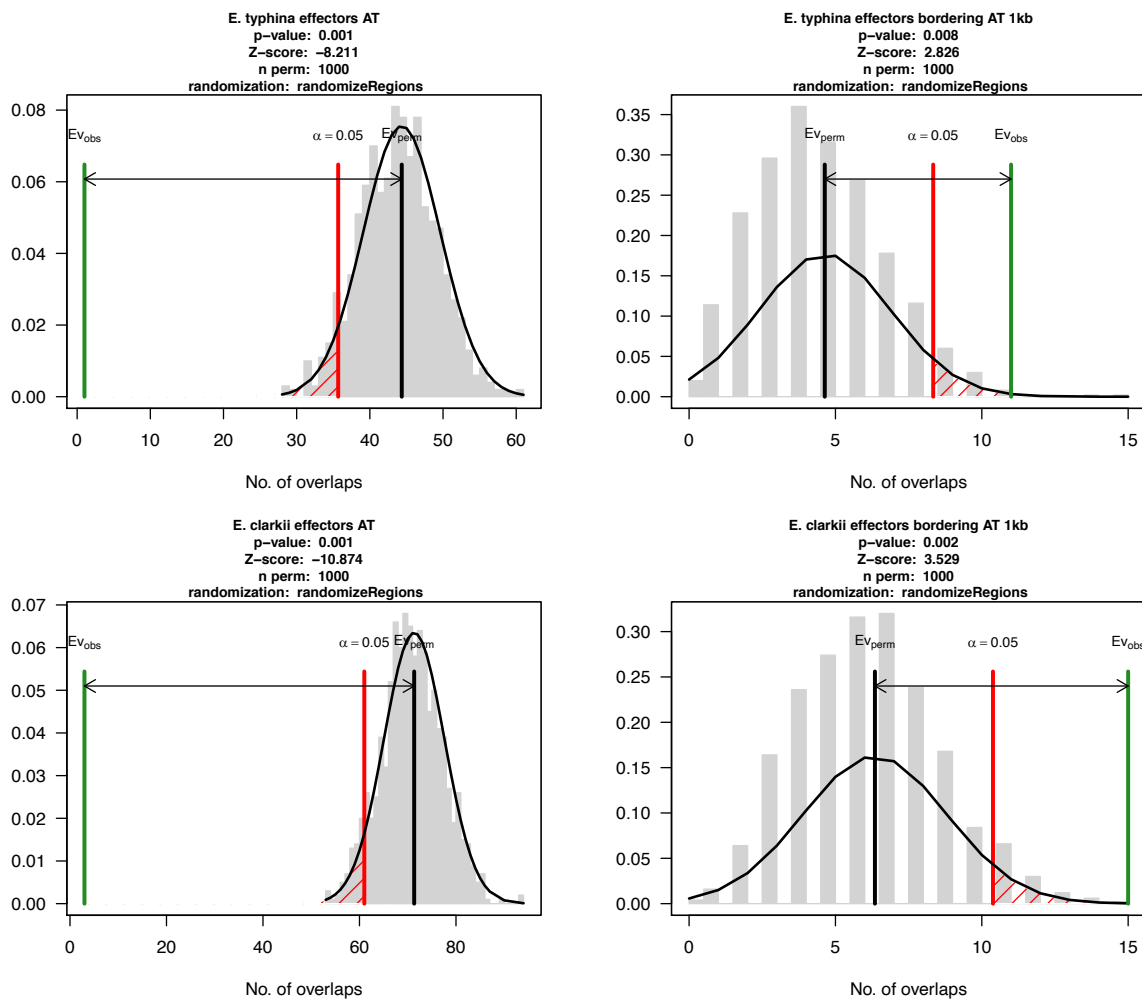


Figure S16: Effectors are associated with AT-bordering regions.

We tested if putative effector genes were more likely to be located in or adjacent to AT-rich regions of *Epichloe* genomes using a permutation tests implemented using the ‘overlapPermTest’ function in the R package regionR with 1000 iterations (Gel et al. 2016). Plots show graphical representation of the permutation test where the grey histogram represents the number of overlaps using a randomized region set with a fitted normal, Ev_{perm} (black line) indicates the mean of the randomized counts and Ev_{obs} (green line) indicates the observed overlap count. Effector genes were significantly less likely to be located in AT-rich regions (left) than expected by chance in both *E. typhina* (top) and *E. clarkii* (bottom). However, effector genes were significantly enriched in the regions directly bordering AT-rich regions (right), 1kb upstream and 1kb downstream. This is relevant as RIP mutations have been shown to leak beyond duplicated regions in other fungi [*Neurospora crassa* (Irelan et al. 1994) and *Leptosphaeria maculans* (Fudal et al. 2009)]. Increased RIP-induced mutation rates may therefore occur in single copy sequences located in close proximity to RIP-affected repeats.

References:

- Fudal, I., S. Ross, H. Brun, A. L. Besnard, M. Ermel, M. L. Kuhn, M. H. Balesdent, and T. Rouxel. 2009. Repeat-Induced Point Mutation (RIP) as an alternative mechanism of evolution toward virulence in leptosphaeria maculans. *Molecular Plant-Microbe Interactions* 22:932-941.
- Gel, B., A. Díez-Villanueva, E. Serra, M. Buschbeck, M. A. Peinado, and R. Malinverni. 2016. regionR: An R/Bioconductor package for the association analysis of genomic regions based on permutation tests. *Bioinformatics* 32:289-291.
- Irelan, J. T., A. T. Hagemann, and E. U. Selker. 1994. High frequency repeat-induced point mutation (RIP) is not associated with efficient recombination in *Neurospora*. *Genetics* 138.

Genetic diversity and population structure of *Epichloe* - Fungal pathogens of plants in natural ecosystems

Artemis D. Treindl^a, Jessica Stapley^b, Adrian Leuchtman^a

^a Plant Ecological Genetics Group, Institute of Integrative Biology, ETH Zurich, 8092 Zurich, Switzerland

^b Plant Pathology Group, Institute of Integrative Biology, ETH Zurich, 8092 Zurich, Switzerland

Manuscript in preparation

ABSTRACT

Understanding the population genetic processes driving the evolution of plant pathogens is of central interest to plant pathologists and evolutionary biologists alike. However, most studies focus on host-pathogen associations in agricultural systems of high genetic and environmental homogeneity and less is known about the genetic structure of pathogen populations infecting wild plants in natural ecosystems. We performed parallel population sampling of two pathogenic *Epichloe* sibling species occurring sympatrically on different native European host grasses in natural and seminatural grasslands: *E. typhina* infecting *Dactylis glomerata* and *E. clarkii* infecting *Holcus lanatus*. We sequenced the genomes of 480 haploid isolates and generated genome-wide SNP datasets to investigate genetic diversity and population structure. Although *E. clarkii* was considered to be specialized on *H. lanatus*, we found that it also occurs on *Holcus mollis*. This finding backs up previous observations from the UK with genetic data and, for the first time, report *H. mollis* as an *E. clarkii* host on European mainland. Clustering analyses showed that in both species geographically separated populations also formed genetically distinct groups which could result from local adaptation generating and maintaining genetic structure, however, separation between populations was less clear in *E. typhina* compared to *E. clarkii*. The patterns of among

population admixture differed between species across the same geographic range: We found higher levels of genetic differentiation between populations and a stronger effect of isolation by distance in *E. clarkii* consistent with lower levels of gene flow compared to *E. typhina*. This pattern may be explained by the different dispersal abilities of the two pathogens linked to their spore morphology while it is probable that the genetic structure of host populations also plays a role. Furthermore, the overall higher genetic diversity within *E. typhina* populations compared to *E. clarkii* indicate highly different effective population sizes and suggest that evolutionary processes such as genetic drift and selection may have differentially affected populations in the two species. Our population genomic analyses provide insights into the factors contributing to population divergence of plant pathogens in natural systems and reveal a strong geographic structure in *Epichloe* which is likely driven by coevolutionary dynamics between host and pathogen at the local scale making this system well suited to investigate the genetic basis of host driven adaptation in the future.

INTRODUCTION

Fungal plant pathogens are fundamental and ubiquitous components of natural biodiversity (Bradley et al. 2008, Burdon and Laine 2019). And yet, fungal plant pathogens are not commonly studied in natural ecosystems but rather in an agricultural setting, because they cause yield losses in important crop plants and may have negative impacts on human or animal health (Fisher et al. 2012). Evolutionary research in the field of plant pathology therefore heavily focuses on mechanisms leading to disease outbreaks in agricultural ecosystems usually associated with the rapid evolution of pathogenicity related traits, and ultimately the evaluation of disease risk or prevention of economic losses (McDonald and Linde 2002). Agricultural systems are useful for the study of antagonistic coevolution because these systems are characterized by strong selection and rapid evolutionary change, driving adaptation and speciation and providing, in a way, short-term evolutionary “experiments” (Stukenbrock and McDonald 2008). A potential drawback, however, is the fact that the nature of agroecosystems defines the starting conditions of all these “experiments” in a similar fashion: pathogens emerge under environmental and genetic uniformity. The impact of fungal plant

pathogens on wild plants has received some attention in the case of invasive diseases (for example *Hymenoscyphus fraxineus* causing ash-dieback (McMullan et al. 2018), or the chestnut blight fungus *Cryphonectria parasitica* (Milgroom et al. 2008), but mechanisms underlying persistence and evolutionary dynamics of host-pathogen associations in native systems remain much more elusive. How do co-evolutionary processes govern these interactions and what defines the evolutionary potential of a pathogen across different evolutionary time scales? What can we learn from nature's experiments and can we assess potential risks posed by future pathogens? Addressing such questions requires more knowledge about the evolutionary biology of plant pathogens in natural ecosystems.

Population genetic structure determines the coevolutionary fate

Population genetic studies which explore genetic structure of populations are central to the study of the evolutionary interactions of fungal pathogens and their host plants. The distribution of genetic variation within populations (genetic diversity) and among populations (genetic differentiation/patterns of gene flow) can provide insights into the evolutionary histories, distributions and disease dynamics of interacting species (Thompson 2005, Chapter 2, Raw Materials for Coevolution I, Barrett et al. 2008). For example, the lack of genetic structure in host populations of cultivated plants favors the contagious spread of few (or a single) fungal pathogen genotypes and has been linked to devastating epidemics in agricultural systems (Stukenbrock and McDonald 2008, Persoons et al. 2017). In contrast, persistent antagonistic coevolution of pathogens with their hosts as it is observed in natural systems requires generation and maintenance of genetic variation in both partners. For example the analysis of natural populations of the flax rust pathogen *Melampsora lini* in Australia showed extensive genetic variation and population differentiation at loci associated with pathogenicity, suggesting that the local influence of evolutionary processes such as selection and genetic drift maintain a high diversity in the fungal pathogen (Barrett et al. 2009). In another well studied model pathosystem, the anther-smut fungus *Microbotryum lynchmidis-dioicae* infecting *Silene latifolia*, patterns of genetic structure across both the pathogens and the hosts European range were consistent with recolonization from ancient glacial refugia (Vercken et al. 2010, Badouin et al. 2017), providing a conclusive example of co-evolutionary processes

shaping congruent population genetic structure between a fungal pathogen and its host plant (Feurtey et al. 2016). It appears that genetic structure of both the pathogen and the host underlies dynamics of disease prevalence at temporal and spatial scales and, for example in the case of powdery mildew *Podosphaera plantaginis* infecting *Plantago lanceolata*, this genetic and environmental heterogeneity may buffer fatal disease outbreaks (Jousimo et al. 2014). However, natural plant pathogen systems in which genetic structure has been investigated still remain scarce compared to their agricultural counterparts. In order to increase our understanding of the evolutionary forces governing plant-pathogen interactions, more detailed empirical studies of genetic diversity in natural populations of fungal pathogens are needed.

The *Epichloe* study system

The sexually reproducing members of the ascomycete genus *Epichloe*, notably the *E. typhina* species complex, are an intriguing system in which to investigate the diversity and population structure of fungal pathogens in natural environments. Fungi of the *E. typhina* complex are highly specialized biotrophic endophytes and pathogens of over 20 different species of pooid grasses forming long-lasting systemic infections (Craven et al. 2001, Leuchtman et al. 2014). Whole plants are usually infected by a single haploid genotype (Leuchtman and Clay 1997), and obligate outcrossing takes place between two genotypes of opposite mating types, each infecting a different host individual (bipolar heterothallic mating system) (White and Bultman 1987, Schardl et al. 2014).

As endophytes they form systemic and symptomless infections in vegetative tissues of the host grasses and then hijack grass flowering tillers to produce gametes and complete their own sexual cycle (Clay and Schardl 2002). The formation of the fungal fruiting structure (stroma) around the grass inflorescence causes choke disease, usually affecting all host inflorescences and therefore effectively sterilizing the entire plant by preventing production of pollen and seed (Chung and Schardl 1997). Exceptions are infections on *Brachypodium sylvaticum*, *Poa nemoralis* and *Puccinellia distans* (Leuchtman et al. 2014). Otherwise, only very rare cases of single flowering tillers which escaped infection and produced seeds have been observed under laboratory settings (personal observation). The pathogen has a detrimental fitness effect on the host plant because systemically infected plants may have no more reproductive output for the rest of their

lifetime. In such sterilizing systems, selection pressures on the host plant to develop resistance and on the pathogen to overcome this resistance are expected to be particularly strong (Ashby and Gupta 2014, Tellier et al. 2014), making the *E. typhina* species complex an intriguing system to study host-pathogen coevolution.

Here we focus on two sister species of the *E. typhina* complex occurring on hosts from two different genera of Poaceae: *E. typhina* infecting *Dactylis glomerata* and *E. clarkii* infecting *Holcus lanatus*. The two sibling taxa are currently assigned the taxonomic rank of subspecies based on their sexual compatibility i.e. their ability to hybridize in experimental crosses (Leuchtman and Schardl 1998, Leuchtman et al. 2014). Indeed, hybrid ascospores between *E. typhina* and *E. clarkii* have also been observed in natural sympatric populations (Bultman et al. 2011), indicating that at least the potential for gene-flow is retained. However, there is no evidence of host grasses infected by hybrid strains in natural populations suggesting the existence of postzygotic barriers. In line with our previous work and supported by evidence of reproductive isolation and genome-wide high levels of genetic differentiation indicating advanced stages of divergence (Schirrmann et al. 2015, 2018, Treindl and Leuchtman 2019), we consider *E. typhina* and *E. clarkii* as closely related yet distinct species.

In previous work, we have generated high-quality genome assemblies of reference strains for both species which provide a critical genomic resource for this study (Chapter I of this thesis). Analyses of the structural organization of *Epichloe* genomes revealed a clear pattern of genome compartmentalization into gene-rich and AT-rich compartments, consistent with the so called “two-speed” model of genome evolution (see also Winter et al. 2018). This particular genome organization shared by some but not all plant-pathogenic fungi is thought to play an important role in driving rapid adaptation in the co-evolutionary arms-race between host and pathogen (Raffaele and Kamoun 2012, Möller and Stukenbrock 2017, Torres et al. 2020). A comparison of the *E. typhina* and *E. clarkii* reference genomes showed a high degree of genomic synteny between the sister species, however, also revealed striking differences in genome size and compartmentalization likely related to an underlying difference in effective population size (Chapter I of this thesis). Besides the differences in genome structure and the obvious distinction based on host specificity, *E. typhina* and *E. clarkii* also differ

in morphological characteristics including the size and disarticulation patterns of their ascospores: *E. typhina* has long, thread-shaped (filiform) ascospores ($176 \pm 34 \times 1.6 \pm 0.2 \mu\text{m}$) that remain unfragmented, whereas *E. clarkii* produces shorter spear-shaped spores ($46 \pm 16 \times 2.3 \pm 0.3 \mu\text{m}$) resulting from fragmentation of filiform spores into four part-spores at maturity (White Jr. 1993, Leuchtman and Clay 1997).

The overarching goal of this study was to understand the population genetic processes affecting fungal pathogens in natural ecosystems. Across the natural distribution range of the host grasses, infections of *E. typhina* on *D. glomerata* are more commonly observed than *E. clarkii* on *H. lanatus* but incidence of both pathogens can be very high in local populations (Bultman et al. 2011). The fact that host grasses harboring their specialized pathogens often co-occur as intermixed populations provides a unprecedented opportunity to conduct a parallel study of two closely related pathogen species specialized on distinct hosts in sympatric populations. We can also determine if the species naturally hybridizes, which is possible because the flowering times of both hosts and thus the reproductive cycle of both pathogens overlap. We performed parallel population sampling of two sympatric pathogen species in natural ecosystems across Europe. We used this population genomic dataset to investigate genetic population structure, that is the genetic diversity and patterns of differentiation within and among populations from different geographical regions. The *E. typhina* - *E. clarkii* study system is exceptional in that it allows us to study genetic structure at the within-species level and a comparison of these patterns between two species (between-species level) across the same geographic range. This provides important insights into how the biology of these species and the interaction with their host plants may have affected their evolutionary trajectories and helps to increase our understanding of the population genetic processes affecting fungal pathogens in natural ecosystems.

MATERIALS AND METHODS

Details about host grasses

The host species *D. glomerata* (orchard grass or cocksfoot), and *Holcus lanatus* (Yorkshire fog) are both long-lived perennial grasses native to Europe (Tutin et al. 1980).

They are wind-pollinated outcrossers, flowering typically May to July and are common and widely distributed in nutrient rich grasslands, where they often co-occur. Cultivars of *D. glomerata* may also be sown as part of commercial forage grass mixtures (e.g. see UFA-Samen, Herzogenbuchsee, Switzerland, and BARENBRUG France, Montévrain), while *H. lanatus* is not commercially sown. Although choke disease caused by *E. typhina* infections has been shown to cause yield reductions in commercial seed production of *D. glomerata* (Large 1954, Pfender and Alderman 2006), to our knowledge there has been no attempt to control pathogen prevalence in natural, semi-natural or pasture grasslands in Europe besides measures of selective breeding in commercial fields of seed producers and recommendations concerning mowing practices (Raynal 1991).

Sample collection and isolation of fungi

Sampling was performed between May and July 2016-2018 at seven locations across Western Europe, where both infected host species (*Dactylis glomerata* and *Holcus lanatus*) occur intermixed or in close proximity, less than 50 meters apart and a large number (>20) individual infected plants were present (large circles in Figure 1, Supplementary Table S1). The majority of sampled populations were located in permanent natural meadows with no apparent evidence of recent high-intensity agricultural use such as regular sowing, fertilizing or grazing by livestock. One population (So) was located in a permanent semi-natural meadow which may be under low-intensity agricultural use. It is noteworthy that the transition between completely natural (no management) and semi-natural (extensive management) grasslands is fluid and even natural meadows often experience anthropogenic influences such as occasional mowing (e.g. in the Aub and Kew populations). However, *Epichloe* infected grasses were not observed in highly managed meadows such as highly fertilized meadows or grass sown as an annual crop under crop rotation for high-intensity silage or hay production although single infected plants were observed around the rims of such fields (personal observation). It appears that agricultural practices that involve regular mowing (i.e. usually before host grasses flower) negatively affect the pathogen prevalence. We consider our populations to be «natural populations» in the sense that they are part of permanent grasslands not under strong human disturbance.

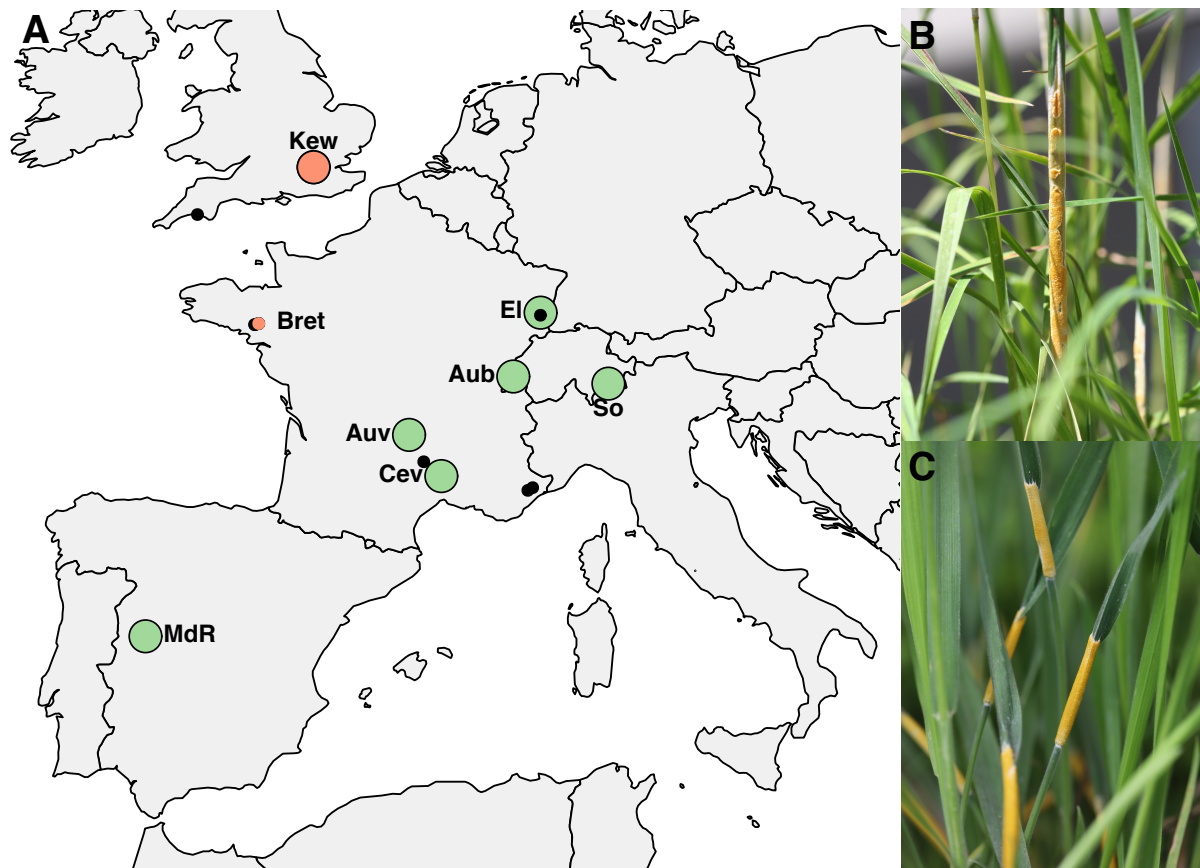


Figure 1: Population sampling of fungal isolates across Europe.

A: Green circles indicate sampling locations of sympatric populations where *E. typhina* is infecting *D. glomerata* and *E. clarkii* is infecting *H. lanatus*. Smaller circles indicate additional samples collected from allopatric populations where only one of the two host species was found to be infected. Black circles indicate locations where *E. typhina* occurs alone (n=2-4, hereafter designated as Other). Red circles indicate locations where *E. clarkii* was found to infect another host grass, *H. mollis*, namely one sympatric location (Kew) and one additional allopatric location (Bret, n=3). The population in Western Switzerland (Aub) was sampled twice thirteen years apart, in 2005 and 2018. B: Ripe fungal fruiting body (stroma) of *E. typhina* infecting *D. glomerata*. C: Stromata of *E. clarkii* infecting *H. lanatus*.

From each host species 21-33 infected plants per site were sampled by picking one choked flowering tiller per plant individual. Each sample represents a different haploid genotype, since an individual plant is assumed to be infected by a single fungal strain resulting from horizontal transmission of meiotic ascospores (Leuchtman and Clay 1997). In Kew Gardens (London, UK) *E. clarkii* was sampled from a previously unknown host species, *Holcus mollis*, occurring in sympatry with *E. typhina* on *D. glomerata*. *Holcus mollis*, unlike *H. lanatus*, can reproduce clonally through the formation of rhizomes in addition to normal sexual reproduction (Ovington and Scurfield 1956). Therefore, we cannot exclude the possibility that some isolates from this population originate from a single fungal genotype systemically propagated via rhizomes within the

population of *H. mollis*, although tillers were collected from plants at a distance of at least five meters. Additional samples were collected from sites, where only one of the two host species was found to be infected (allopatric populations, small circles in Figure 1 and Supplementary Table S2). Here, infected flowering tillers from 1-4 individuals per population and host species were collected. The sympatric population in Western Switzerland (Aub) had been sampled previously thirteen years prior to this study in 2005. We included these for sequencing and hereafter designate the isolates as Aub (sampled in 2018) and AubX (sampled in 2005).

Fungal strains were isolated from freshly collected tillers by splitting open the stromata with a sterile blade to expose the interstitial fungal mycelium between undeveloped inflorescence and leaf blades. Pieces of mycelium were then transferred with sterile tweezers and subsequently grown on supplemented malt extract-agar (SMA) plates containing 1 % malt extract, 1 % glucose, 0.25 % bacto peptone, 0.25 % yeast extract, 1.5 % bacto agar and 0.005% oxytetracycline (Pfizer, New York, NY, U.S.A.). Isolates were carefully checked under the dissecting scope for purity and *Epichloe* identity. Pure cultures were then grown in V8 liquid media on a rotary shaker at 12 rpm for 10-14 days at room temperature (Christensen et al. 1993). Mycelium was vacuum filtered, freeze dried for 24h and then ground using liquid nitrogen. All live axenic cultures were also stored on SMA in tubes covered with mineral oil at 4°C at ETH Zurich.

DNA preparation and sequencing

Genomic DNA was extracted from 10mg freeze dried material using the sbeadex mini plant kit (LGC Genomics, Berlin, Germany), but replacing the lysis buffer with PVP (sbeadex livestock kit) as this buffer had previously shown improved DNA yields. Extraction was automated on a KingFisher system following the manufacturer's protocol (Thermo Fisher Scientific, Waltham, MA, USA). The DNA concentration of extracted samples was quantified using a Spark[®] Multimode Microplate Reader (Tecan Trading AG, Switzerland). Paired-end libraries of 2 x 150 bp fragments with an insert size of 330 bp were prepared with NEBNext[®] Ultra[™] II DNA Library Prep Kits (New England Biolabs, Ipswich, MA, USA), and sequencing was performed on a HiSeq4000 Illumina sequencer (Illumina, San Diego, CA, USA), at 16x coverage on average. We sequenced whole genomes of 240 individuals of *E. typhina* and 240 individuals of *E. clarkii*. All Illumina

sequence data will be made accessible on the NCBI Short Read Archive upon publication.

Mapping and variant calling

Reads were mapped against the high-quality reference genomes of *E. typhina* (Ety_1756, BioProject ID PRJNA533210) and *E. clarkii* (Ecl_1605_22, BioProject ID PRJNA533212) (Chapter I of this thesis). Mapping was performed with BWA (Li and Durbin 2009) and variant calling was done with the Genome Analysis Toolkit (GATK, McKenna et al. 2010), using the GATK Short Variant Discovery best practice workflow (<https://gatk.broadinstitute.org/hc/en-us/articles/360035535932-Germline-short-variant-discovery-SNPs-Indels->). Details and examples of how each command was run are provided on DRYAD (for review purpose these are hosted temporarily on github (https://github.com/adtreindl/Epichloe_genomics)). In brief, the raw reads were trimmed and adaptors removed using trimomatic v0.35 (Bolger et al. 2014). The trimmed reads were mapped to the respective reference genomes using bwa mem v0.7.17 (Li and Durbin 2009). Low quality alignments (qual>10) were removed and duplicates were marked. Mapping statistics were obtained using sambaba v0.6.6 (Tarasov et al. 2015). After mapping, variants were identified using HaplotypeCaller, GatherVCFs and JointGenotyper.

Filtering

Variants were filtered with custom scripts using vcftools (Danecek et al. 2011), bcftools (Li et al. 2009), and R (R Core Team 2013) (for example scripts see github repository (https://github.com/adtreindl/Epichloe_genomics)). We first removed mitochondrial sequences, and only nuclear DNA (Chr1-7) were used in all subsequent analysis. The following hard-filters were applied using vcftools v1.2.8 minor allele count (--mac 3) and maximum mean depth values (--max-meanDP 100). The latter filter removed variants with excessive coverage (mean depth >> mean read depth) across all samples as these are indicative of either repetitive regions or paralogs. We filtered out SNPs that were not well sampled across populations by removing SNPs with >30% missing within a single population. We removed isolates that had >80% missing data and removed indels to retain only SNPs. We then filtered by read quality and read depth following GATK best

practice recommendations (Van der Auwera et al. 2013), and based on our evaluation of the data. We removed sites with low mapping quality (MQ value < 40, the Root Mean Square of the mapping quality of the reads across all samples) and with a high relative variance in read depth (variance in read depth divided by mean read depth > 80 for *E. typhina* and > 30 in *E. clarkii* (Supplementary Figure S1). We visually inspected regions at different SNP densities using Integrative Genomics Viewer (IGV, Robinson et al. 2011) to assess the variant calls and based on these observations chose to excluded sites where ≥ 3 SNPs had been called within 10bp. Lastly, we removed sites that previous filtering steps had rendered monomorphic. Summary statistics including missingness and mean depth of this dataset are shown in Supplementary Figure S2. For some downstream analyses additional filters were applied and are described below and listed in Supplementary Table S3.

Analyses of genetic structure

We analyzed genetic structure of 211 *E. typhina* and 211 *E. clarkii* isolates using several methods: Principal component analysis (PCA), Bayesian clustering analysis (STRUCTURE) and discriminant analyses of principal component (DAPC). PCA, which does not rely on any population genetic model, was implemented in the R package SNPRelate (Zheng et al. 2012). PCAs were done for the complete SNP datasets, LD-pruned datasets (see below) and datasets with no missing data. We repeated analyses after removing the Mdr population from the datasets, as these individuals are so genetically distinct in both species that they obscure the genetic structure of the other populations. We also visualized the amount of sub-structure within each population using PCA (see Supplementary Notes: Population sub-structure). For the STRUCTURE analysis, we applied a minor allele frequency filter of 0.05 using vcftools (`-maf 0.05`) to exclude rare alleles from the analysis as they are likely not representative of a particular population group. We also removed SNPs that are in linkage disequilibrium (LD), pruning variants with a pairwise r^2 greater than 0.3 in 10kb windows using bcftools v1.9. The LD-pruned datasets included 152,883 biallelic SNPs in *E. typhina* and 79,352 SNPs in *E. clarkii*. STRUCTURE v2.3.4 (Pritchard et al. 2000), was run for K=1-12 with 20,000 Monte Carlo Markov Chain (MCMC) iterations following a 10,000 iteration burn-in for each of twelve replicate runs per K. We used an admixture ancestry model with

correlated allele frequencies and no prior information about the demography. The output of the STRUCTURE analysis was processed using STRUCTURE HARVESTER (Earl and vonHoldt 2012). The model with the maximized rate of change in log likelihood values (ΔK) was considered the optimal number of clusters based on methods described by Evanno et al. (2005). We assessed probability of assignment and genetic admixture among populations by plotting the proportion of membership to each of the inferred clusters for every individual, using the optimal model parameters (the best run) for each K.

DAPC was performed using the R package *adegenet* v2.0.0 (Jombart 2008) and the 'find.clusters' function. This method first transforms the data using PCA and then performs a Discriminant Analysis on the retained principal components. For computational reasons we used the LD-pruned dataset and further reduced the number of SNPs to one variant per 2kb using *vcftools* (--thin 2000). This left 12,562 SNPs in *E. typhina* and 11,338 SNPs in *E. clarkii* (LD-pruned and thinned dataset).

Summary statistics of genetic variation

We calculated several population genetic summary statistics using *vcftools* v1.16, using a patch that allows computation with haploid datasets (<https://github.com/vcftools/vcftools/pull/69>). Calculations were performed for the sympatric populations only (removed isolates termed additional samples), leaving 191 individuals in *E. typhina* and 208 individuals in *E. clarkii*. After removing "additional" isolates we again removed any monomorphic variants. Populations sampled in different years from Western Switzerland (Aub and AubX) were analyzed as separate populations although this is the same location. Step-by-step details and example code is available at: https://github.com/adtreindl/Epichloe_genomics/blob/master/summary_stats.md

Nucleotide diversity π , Tajimas D

We quantified within-population genetic diversity by calculating nucleotide diversity in 10kb non-overlapping windows (π , the average number of differences between individuals (Takahata and Nei 1985)), and Tajima's D in 10kb and 40kb non-overlapping windows (Tajima 1989). Genomic regions with reduced levels of nucleotide diversity and Tajima's D may indicate departure from neutrality that can be the result of demographic

processes or positive selection acting on these loci. We calculated the average nucleotide diversity and Tajima's D per population as the mean across all loci.

F_{ST}

To assess relative genetic differentiation between populations we calculated pairwise fixation indexes F_{ST} (Weir and Cockerham 1984), per site between all populations using `vcftools` (haploid switch, see above). We then calculated mean F_{ST} in non-overlapping windows of 5kb (10kb) along each chromosome using R version 4.0.2 (R Core Team 2013), and genome-wide weighted mean F_{ST} between population pairs. To investigate patterns of isolation-by-distance (IBD) in *E. typhina* and *E. clarkii* we tested for a positive correlation between genetic distance (the mean weighted F_{ST}) and geographic distance. We used the R package `lme4` (Bates et al. 2015) to fit a linear mixed-effects model to the data with populations as random effects and distance as a fixed effect. We also calculated divergence as the number of fixed differences between population pairs (SNPs with an F_{ST} of 1).

Analysis of linkage disequilibrium

We removed loci with a minor allele frequency below 0.05 (within a population) because differences in allele frequencies can (upwardly) bias estimation of LD. We calculated the coefficient of correlation (r^2) between pairs of SNPs that were up to 20 kb apart using `plink` (v. 1.90) and calculated mean LD within 100bp windows. We visualized the decay of r^2 over physical distance using a locally weighed polynomial regression (LOESS) model using the function `geom_smooth` in the R package `ggplot2` (Wickham 2016).

Effective population size N_e

To compare effective population sizes within and between species we estimated effective population sizes (N_e) from polymorphism data with the parameter $\theta = 2N_e\mu$ (for more details see Supplementary Notes: Calculation of effective population size). Calculations were based on Watterson's estimators of θ (Watterson 1975), which is based on the number of segregating sites in a population relative to the sequence length, and an assumed mutation rate (μ) of 3.3×10^{-8} per site. This mutation rate is based on estimates in yeast (Lynch et al. 2008) and has been used in other studies (Stukenbrock et al. 2011).

A table with the number of segregating sites per populations can be found in the Supplementary (Table S4). From this we also calculated the proportion of segregating sites within each population relative to the total variation among all populations, as a measure of genetic diversity comparable between species (Supplementary Table S5).

We also analyzed the distribution of genetic variation separately for distinct genomic compartments in the *Epichloe* genomes. Methods and results are presented in the Supplementary (Supplementary Notes: PCA analysis and calculation of summary statistics by compartments). Additionally, we assessed population sex-ratios (Supplementary Notes: Mating types) and interspecific hybridization in two populations (Supplementary Notes: Assessment of hybridization in sympatric populations).

RESULTS

A population genomic dataset of two *Epichloe* species

After removal of isolates with low coverage and replicates we retained genome-wide polymorphic sites for 211 isolates in each of the species (Tables 1 and 2). We obtained 658,021 biallelic SNPs in the *E. typhina* dataset and 400,033 SNPs in the *E. clarkii* dataset after initial filtering (complete dataset, Supplementary Table S3).

Table 1: *E. typhina* isolates analyzed in this study

	Host species	Origin	Collection year	n
Aub	<i>D. glomerata</i>	Switzerland	2018	22
AubX	<i>D. glomerata</i>	Switzerland	2005	16
Auv	<i>D. glomerata</i>	France	2017	27
Cev	<i>D. glomerata</i>	France	2017	31
El	<i>D. glomerata</i>	France	2017	26
Kew	<i>D. glomerata</i>	United Kingdom	2017	14
MdR	<i>D. glomerata</i>	Spain	2018	33
So	<i>D. glomerata</i>	Switzerland	2016	22
Other	<i>D. glomerata</i>	United Kingdom	2017	3
Other	<i>D. glomerata</i>	France	2017	4
Other	<i>D. glomerata</i>	France	2017	3
Other	<i>D. glomerata</i>	France	2017	3
Other	<i>D. glomerata</i>	France	2017	3
Bret	<i>D. glomerata</i>	France	2017	4
		Total		211

Table 2: *E. clarkii* isolates analyzed in this study

	Host species	Origin	Collection year	n
Aub	<i>H. lanatus</i>	Switzerland	2018	27
AubX	<i>H. lanatus</i>	Switzerland	2005	18
Auv	<i>H. lanatus</i>	France	2017	30
Cev	<i>H. lanatus</i>	France	2017	28
El	<i>H. lanatus</i>	France	2017	27
Kew	<i>H. mollis</i>	United Kingdom	2017	19
MdR	<i>H. lanatus</i>	Spain	2018	29
So	<i>H. lanatus</i>	Switzerland	2016	30
Bret	<i>H. mollis</i>	France	2017	3
		Total		211

Genetic structure of populations

Principal component (PC) and Bayesian clustering analyses were performed to investigate spatial genetic patterns among European *E. typhina* and *E. clarkii* populations. Overall both methods found that population structure of both *Epichloe* species reflected the geographic origin of sampled populations, however, patterns of genetic differentiation differed between species and STRUCTURE indicated fewer clusters in *E. typhina* with three populations (Kew, Auv, El) being assigned to the same cluster (see below). PCA grouped genotypes according to their sampling location, with the Spanish population (MdR) forming the most genetically distinct group (Supplementary Figure S3A and S3D). We removed this population in both species and repeated the analysis to explore genetic patterns among the remaining populations (Figure 2 and Figure 3). Clusters representing the sampled *E. typhina* populations could be clearly distinguished in the principal component space with the first axis resolving the West-to-East distribution (explaining 3.93% of the variation), and the second axis resolving the South-to-North distribution (2.58% variation explained, Figure 2B). Spatially intermediate additional samples (Bret and Other) were also genetically intermediate. In *E. clarkii* seven clusters were evident in the PCA, each corresponding to a single geographic population with the exception of the Kew population, which consisted of two genetically distinct groups (Figure 3B). In both species, isolates sampled thirteen years apart from the same Western Switzerland location did not form genetically distinct groups (AubX and Aub sampled 2005 and 2018), suggesting that thirteen years are likely an insufficient time-span to expect major genetic differentiation based on genome-wide variation in *Epichloe* populations. These results did not differ between datasets that were filtered differently as outlined in the methods section (complete, LD pruned, no missing data) (Supplementary Figures S3 and S4). The two populations that were the most spread out in principal component space in overall analyses (*E. typhina* Cev and *E. clarkii* Kew) also showed evidence of sub-structure in PCA performed within populations using only isolates from the same location (see Supplementary Notes and Supplementary Figures S5 and S6).

Consistent with results obtained with PCA, STRUCTURE showed evidence of subdivision according to geography within both *E. typhina* and *E. clarkii*. At all Ks tested and both species (from K=2 to K=12), the isolates from the Spanish population (MdR)

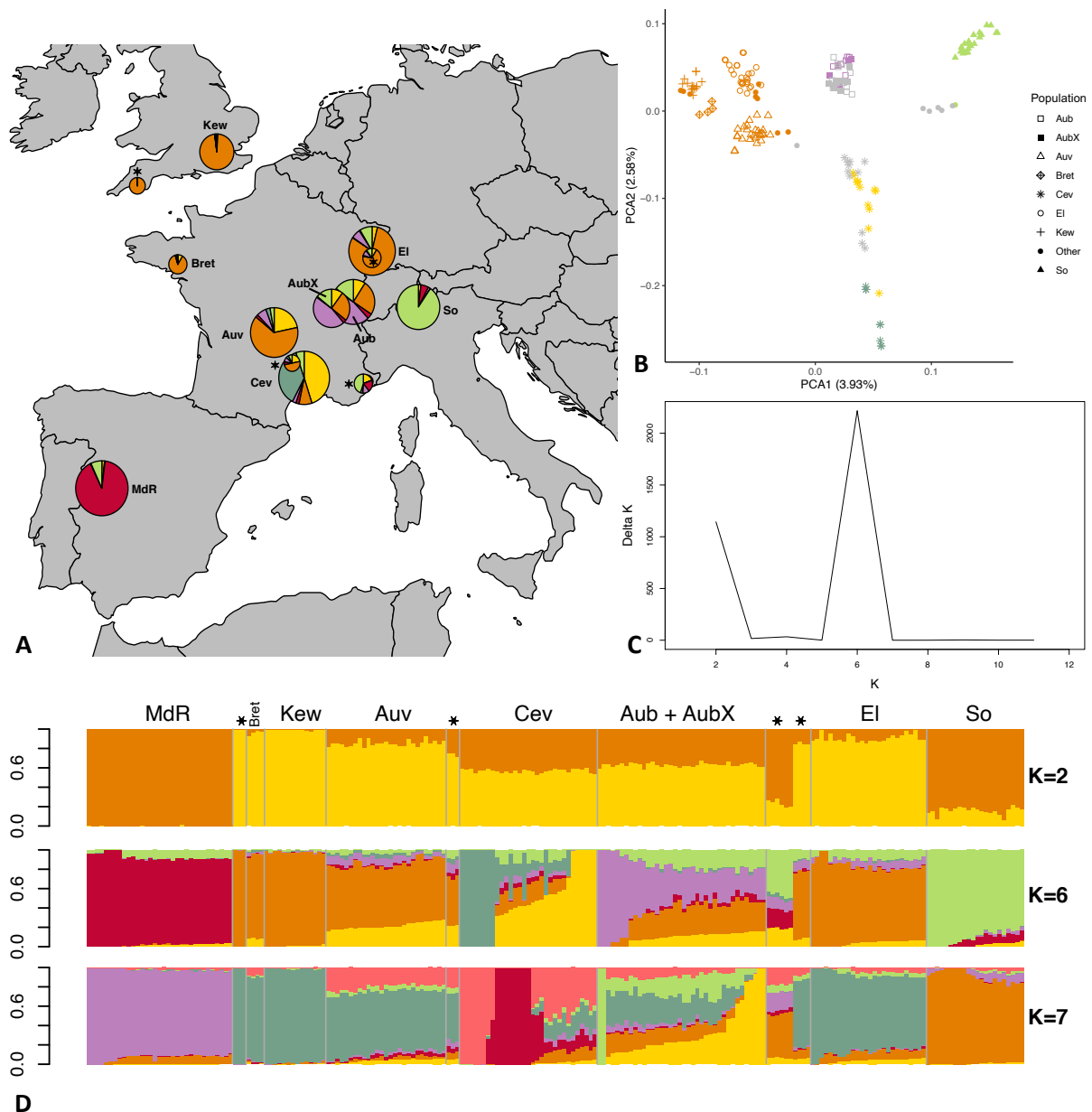


Figure 2: Population structure analyses of 211 *E. typhina* isolates based on genome-wide SNPs.

The LD pruned dataset including 152,883 SNPs was used in the STRUCTURE analyses whereas the complete dataset of 658,021 SNPs was used for the PCA. **A**) Sampling locations of *E. typhina* isolates infecting *D. glomerata*; pie charts represent Bayesian cluster membership proportion for K=6, diameters reflect sample size (range n=3 to n=33). **B**) Principal component analysis of genetic differentiation among isolates excluding the population from Spain (MdR); The percentage of variance explained by the first two principal components is shown in parentheses; symbols indicate sampling locations of populations and colors correspond to the six genetic clusters assigned to each individual by Bayesian clustering for isolates with ≥ 0.5 membership to a single cluster; grey symbols are isolates not assigned to any cluster when using the threshold of 0.5. **C**) Delta K plot of the STRUCTURE Bayesian clustering analyses for Ks of 2–12. **D**) Bar plots of membership proportions (y-axis) as inferred by STRUCTURE for K=2, K=6 and K=7; each isolate displayed as a vertical bar; each color represents a cluster and isolates are ordered according to longitude (West to East); sample sizes are: MdR=33, Bret=4, Kew=14, Auv=27, Cev=31, Aub=22, AubX=16, El=26, So=22, Other(★)=3+3+3+3+4.

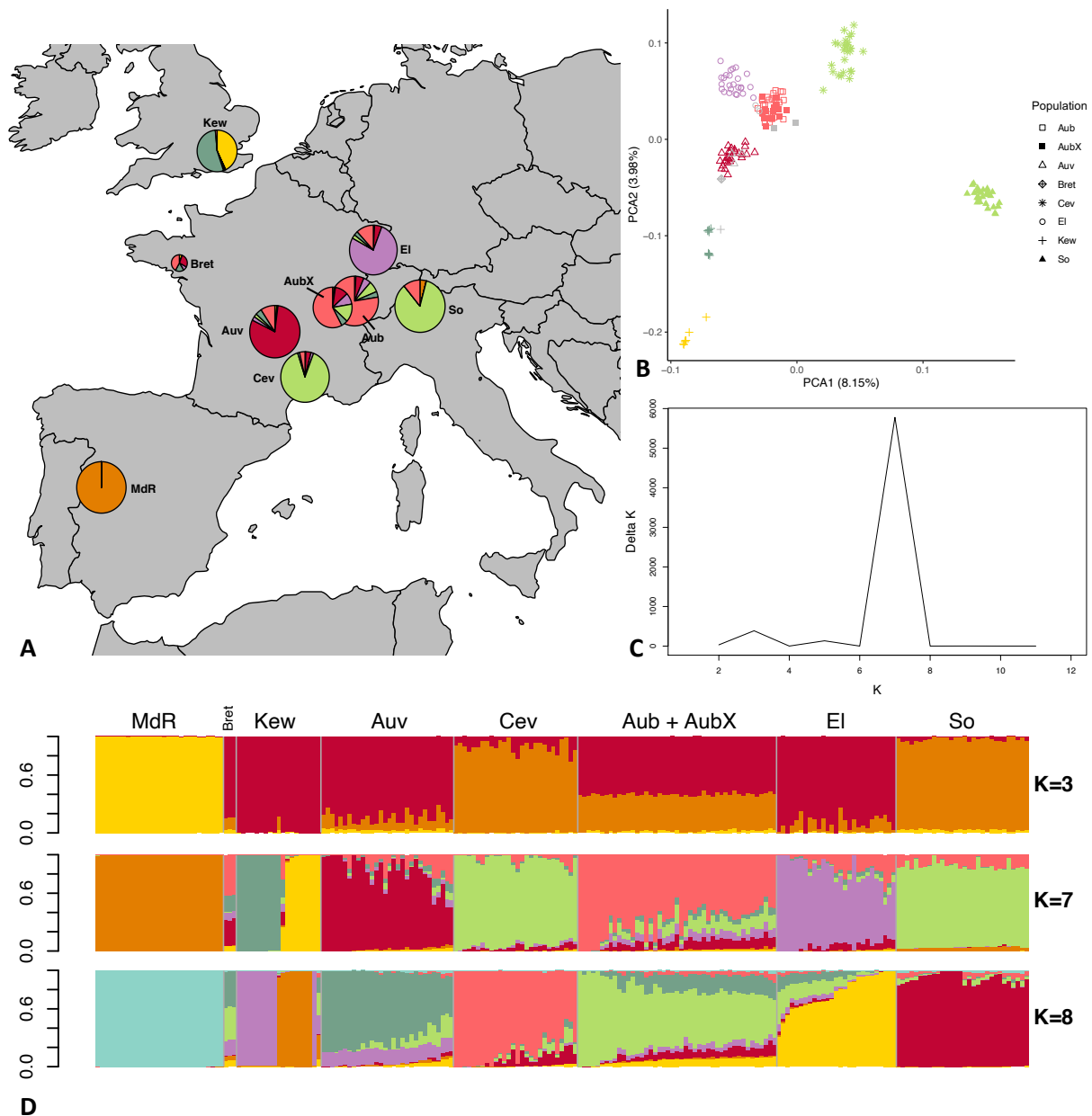


Figure 3: Population structure analyses of 211 *E. clarkii* isolates based on genome-wide SNPs.

The LD pruned dataset including 79,352 SNPs was used in the STRUCTURE analyses whereas the complete dataset of 400,033 SNPs was used for the PCA. **A)** Sampling locations of *E. clarkii* individuals infecting *H. lanatus* and *H. mollis* (Kew & Bret); pie charts represent Bayesian cluster membership proportion for $K=7$, diameters reflect sample size (range $n=3$ to $n=30$). **B)** Principal component analysis of genetic differentiation among isolates excluding the population from Spain (MdR); The percentage of variance explained by the first two principal components is shown in parentheses; symbols indicate sampling locations of populations and colors correspond to the seven genetic clusters assigned to each isolate by Bayesian clustering for isolates with ≥ 0.5 membership to a single cluster; grey symbols are isolates not assigned to any cluster when using the threshold of 0.5. **C)** Delta K plot of the STRUCTURE Bayesian clustering analyses for K s of 2–12. **D)** Bar plots of membership proportions (y-axis) as inferred by STRUCTURE for $K=3$, $K=7$ and $K=8$; each isolate displayed as a vertical bar; each color represents a cluster and isolates are ordered according to longitude (West to East); sample sizes are: MdR=29, Bret=3, Kew=19, Auv=30, Cev=28, Aub=27, AubX=18, EI=27, So=30.

were assigned to a distinct cluster indicating the geographically most remote population is also the most genetically distinct. There was no genetic differentiation between isolates sampled at different time points in the population from Western Switzerland (AubX and Aub sampled 2005 and 2018), in either species. In *E. typhina* the most likely number of clusters was $K=6$ as inferred by the maximized likelihood and ΔK values across twelve replicate runs (Figure 2A,C,D). At $K=6$, two of the six clusters contained all isolates from a single location (red: MdR, green: So). Isolates from most locations appeared to have ancestry from the same orange cluster, to varying degrees. Isolates from more Northern locations (El, Kew, Bret) and three isolates from Plymouth showed the highest ancestry proportion to this cluster. The exception were isolates from Cev, the majority of which had genotypes assigned to two distinct clusters (yellow and darkgreen) but also contained admixed isolates. Additionally, the population from Western Switzerland contained seven isolates assigned to a sixth cluster (purple), three of which were sampled in 2005(AubX) and four in 2018 (Aub) (see Supplementary Figure S7), but the majority of isolates had mixed ancestry. A number of isolates from Aub and Cev as well as six additional isolates collected in the Southeastern tip of France showed a higher proportion of mixed ancestry with membership proportion <0.5 for any of the inferred clusters (grey symbols in PCA). The relationship among genetic clusters and location did not get much clearer at $K=7$.

In *E. clarkii* the most likely number of clusters inferred by STRUCTURE was $K=7$ (Fig 3A,C,D). At $K=7$ all clusters consisted of isolates from a single sampling location except for the populations from Southern France (Cev) and Southern Switzerland (So), which were assigned to the same cluster. At lower and higher K s ($K=5$, $K=6$, $K=8$, Supplementary Figure S8), however, these two populations were assigned to distinct clusters. The population collected at Kew Gardens (Kew) consisted of genotypes from two distinct clusters as well as a single admixed isolate. Although Central European populations contained a higher number of individuals with mixed ancestry, admixture proportions were lower than in *E. typhina*. Overall, ten isolates could not be assigned to a single cluster when using the threshold of 0.5. Of these admixed isolates, all except for three genotypes collected in Western France (Bret) and the admixed isolate from Kew shared the majority of their ancestry with their population of origin.

Results from DAPCs were consistent with STRUCTURE analyses suggesting seven clusters (corresponding to geographic locations) a best-fit in *E. clarkii*. In *E. typhina* DAPC suggested an optimal K of 3 with MdR (Spain) as the most distinct cluster and the other populations separating on the second axis. When removing MdR, K=2 was optimal dividing genotypes into a Northern and a Southern cluster. At higher K's the remaining six populations could be distinguished with corresponding to geographic locations and with admixture among the northern and central populations.

Genetic differentiation between populations and isolation by distance

We calculated relative divergence between populations using F_{ST} (Weir and Cockerham 1984). Weighted genome-wide averages of pairwise F_{ST} were lower between *E. typhina* populations (0.05-0.21, Table 3), supporting higher levels of gene flow particularly among Central European populations for which clustering analysis had inferred mixed ancestry (Kew, Auv, Aub and El). *E. clarkii* populations had higher values of F_{ST} consistent with lower levels of gene flow (0.12-0.49, Table 3). Population pairs with the highest F_{ST} values also had the most fixed differences between them, and overall, there were more fixed differences between *E. clarkii* populations than *E. typhina* (Supplementary table S6). Furthermore, we found a positive correlation between genetic differentiation and geographic distance consistent with a pattern of isolation-by-distance (IBD) in both species (Figure 4). A comparison of the slopes of the two linear models showed that the effect of IBD was stronger in *E. clarkii* compared to *E. typhina* with higher genetic differentiation between population pairs ($Z= 6.964$, $p= 3.314e-12$, 'lm_slopes_compare' from R package EMAtools (Kleiman 2017)).

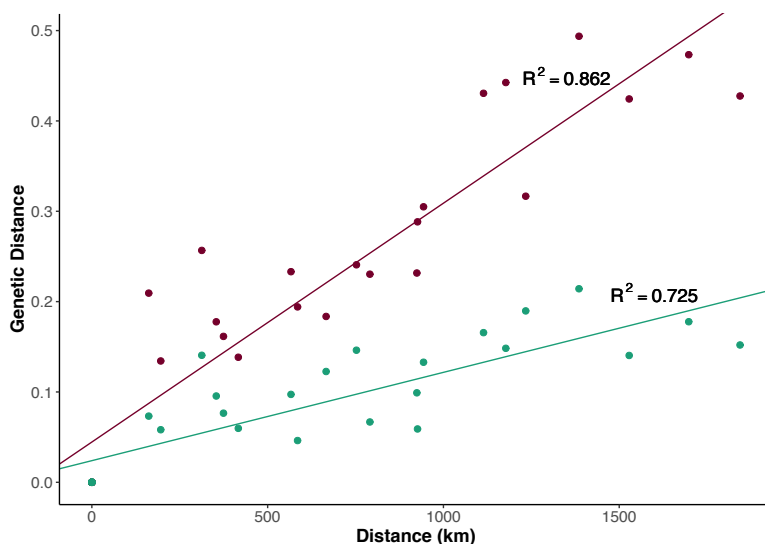
Genetic diversity within populations

We used SNP datasets to compute nucleotide diversity (π) within each population. Mean genome wide nucleotide diversity per site ranged from 0.0024 to 0.0027 in *E. typhina* populations and from 0.0008 to 0.0012 in *E. clarkii* populations (Supplementary Figure S9, Supplementary Table S7). This was consistent with a previous study which found higher diversity in *E. typhina* compared to *E. clarkii* in one sympatric population pair (Schirrmann et al. 2018). Findings suggest an underlying difference in effective population size between the two species, across all sampled populations.

Table 3: Mean genome-wide F_{ST}

	MdR	Kew	Auv	Cev	Aub	AubX	El	So
<i>E. clarkii</i>	MdR		0.21	0.17	0.15	0.14	0.13	0.18
	Kew	0.49		0.07	0.13	0.10	0.10	0.06
	Auv	0.43	0.23		0.07	0.06	0.06	0.05
	Cev	0.44	0.30	0.21		0.08	0.07	0.10
	Aub	0.42	0.23	0.14	0.16		0.01	0.06
	AubX	0.43	0.23	0.13	0.15	0.02		0.06
	El	0.47	0.29	0.19	0.23	0.13	0.12	
	So	0.43	0.32	0.24	0.18	0.18	0.16	0.26
<i>E. typhina</i>	MdR							
	Kew							
	Auv							
	Cev							
	Aub							
	AubX							
	El							
	So							

Pairwise weighted F_{ST} values (Weir and Cockerham 1984) for populations of *E. typhina* (top right) and *E. clarkii* (bottom left), calculated using genome-wide SNPs. Populations are ordered according to longitude (West to East) and higher F_{ST} values are underlaid with darker shades of red.

**Table 4:** Effective population size N_e

	<i>E. typhina</i>	<i>E. clarkii</i>
Aub	43,941	17,606
AubX	41,997	18,010
Auv	42,358	15,284
Cev	40,223	15,989
El	41,235	14,289
Kew	37,034	10,250
MdR	39,535	9,251
So	39,647	18,809

Figure 4: Isolation by distance among *Epichloe* populations.

Relationship between pairwise F_{ST} and geographic distance for *E. typhina* (green) and *E. clarkii* (purple). Correlations for both species were significant (*E. typhina*: $p = 4.19e-08$; *E. clarkii*: $p = 7.187e-13$) with increasing geographic distance having a stronger effect in *E. clarkii* ($p = 3.314e-12$).

We also calculated population-specific Tajima's D , which reflects the distribution of allele frequencies within populations (site frequency spectrum). Tajima's D is > 0 when there is a lack of rare alleles indicative of balancing selection or population contraction, Tajima's $D < 0$ indicates an excess of low frequency alleles as a result of recent positive selection or population expansion, and Tajima's D of 0 suggest neutral evolution. Mean Tajima's D was slightly negative in all populations of *E. typhina*, ranging from -0.369 to -0.061 (Supplementary Figure S10, Supplementary Table S8). In *E. clarkii* Tajima's D showed more variation across populations: Three populations, including the same

population sampled at different time points (Aub and AubX), had slightly negative Tajima's D ranging from -0.354 to -0.015, two populations had slightly positive Tajima's D (Auv = 0.102 and Cev = 0.151) and two populations had Tajima's D around 1 (Kew = 0.903 and MdR = 1.318).

Analysis of linkage disequilibrium

To test the extent of recombination in different populations we calculated linkage disequilibrium as r^2 , the coefficient of correlation between a pair of SNPs. LD decay curves showed a strong decrease below an r^2 of 0.2 in the first 1-2kb in all *E. typhina* populations and most *E. clarkii* populations (Supplementary Figure S11). This indicates high recombination rates resulting in very small linkage blocks and therefore pervasive outcrossing. However, two populations in *E. clarkii* showed different patterns (Supplementary Figure S12): In Kew, LD decayed much more slowly with distance, which is most likely a result of clonal population substructure. The Spanish population MdR had very high levels of LD overall, which could indicate a recent bottleneck. This population also showed the most striking differences between genome compartments: the gene-rich compartment appeared to consist of highly linked sequence while LD was low in the AT-rich compartment (see Supplementary Notes: PCA analysis and calculation of summary statistics by compartments).

Effective population size N_e

We used Watterson's estimators of θ (Watterson 1975) and an approximate mutation rate of 3.3×10^{-8} per site to calculate and compare effective population sizes within and between *Epichloe* species. Effective population sizes in *E. typhina* were between 2-fold and 4-fold larger than in *E. clarkii* populations, which also had more variance in N_e estimates among populations (Table 4: Estimates of effective population size N_e). In *E. typhina* the population from Western Switzerland (Aub) had the largest N_e (43,941) and the British population (Kew) had the smallest (37,034). In *E. clarkii* the population from Southern Switzerland (So) had the largest N_e (18,809) and the Spanish population (MdR) had the smallest (9,251).

DISCUSSION

The amount and distribution of genetic variation within species and populations constitute the raw material for evolutionary processes, both adaptive and neutral. Population genetic studies of fungal plant pathogens have revealed how life-history strategies and the biology of pathogens together with selection mediated by the host strongly affects these patterns and shapes pathogen genetic structure in agricultural ecosystems (McDonald and Linde 2002, Barrett et al. 2008), but we are only beginning to understand how these factors affect natural pathogen populations. We therefore sampled seven sympatric population pairs of two species of *Epichloe* grass pathogens in natural grassland ecosystems across Western Europe. These highly specialized fungi with their host-sterilizing lifestyle present an intriguing biological system to study adaptive processes in natural ecosystems, and the parallel sampling of two closely related sibling species occurring in sympatry gave us the unique opportunity to compare genetic patterns between two species across the same geographic range.

Different levels of gene flow among spatially structured populations

In our analysis of genome-wide variation we found a clear pattern of biogeographic structure in both *E. typhina* and *E. clarkii*. Spatially separated populations were also genetically differentiated and we were able to clearly resolve both the West-to-East and South-to-North distribution of our sampling. In both species genetic distance increased with geographic distance and the most remote population from Spain was genetically highly distinct. Despite the similar genetic structure in *E. typhina* and *E. clarkii* at the European scale, patterns of gene-flow among populations are different. *E. typhina* populations had higher genetic diversity and showed less differentiation between populations indicating greater gene flow particularly among Central European and Northwestern populations. In contrast, we found lower diversity and higher levels of genetic differentiation among *E. clarkii* populations and a stronger effect of isolation by distance. In previous work we compared genome organization and size between the reference genomes of the two species and suggested that genome expansion in *E. clarkii* compared to *E. typhina* may be linked to a reduction in effective population size which may lead to stronger effects of drift compared to *E. typhina* (Chapter I of this thesis).

The lower genetic diversity and lower effective population size within *E. clarkii* populations we report supports this hypothesis. Of the total variation assessed among populations in both species, a similar proportion was polymorphic within most populations indicating that the two species likely have similar mutation rates and that, in spite of a lower overall diversity, *E. clarkii* populations still harbor a considerable amount of standing genetic variation. This can be expected given the obligately outcrossing lifestyle of the pathogen and high recombination rates detected. The exceptions were two populations, Kew and MdR, which were found to have significantly less genetic diversity and high degrees of linkage likely related to population substructure and/or a recent bottleneck which is also reflected in the positive Tajimas'D values for these two populations. Such demographic processes affecting genome-wide variation will need to be taken into consideration in future studies when investigating the impact of selection in these populations.

Determinants of population structure in Epichloe pathogens:

1. Phylogeographic history may shape genetic structure

Genetic variation in specialized obligate pathogens such as *E. typhina* and *E. clarkii* is expected to be strongly influenced by the distribution and genetic structure of their host plants (Wilson et al. 2005, Greischar and Koskella 2007). Previous studies of genetic structure in the anther smut pathogen *Microbotryium lynchnidis-dioicae* and its white campion host (*Silene latifolia*) across a larger sampling range have identified three genetically differentiated clusters spanning greater geographic scales and suggested that clusters reflect the concerted recolonization of the host plant and its fungal pathogen from distinct southern glacial refugia (Vercken et al. 2010, Feurtey et al. 2016, Badouin et al. 2017). The maintenance of genetically differentiated gene pools in these refugia and subsequent northwards recolonization as ice-sheets retreated is thought to have strongly impacted the spatial structure of genetic variation observed across Europe today and indeed, the existence of at least three major ice-free refugia during the last glacial maximum located in Iberia, Italy and the Balkans is a well-established hypothesis supported by phylogeographic studies of a growing number of temperate species including grasses (Hewitt 2000, Provan and Bennett 2008, François et al. 2008). For example, recent studies found genetic structure of natural populations to be consistent

with postglacial colonization history in the *Poaceae* grasses *Festuca pratensis* (Fjellheim et al. 2006), *Festuca rubra* (von Cräutlein et al. 2019) and *Lolium perenne* (McGrath et al. 2007). Genetic diversity in Europe has also been assessed in one of our host grasses, *D. glomerata*, and three genetic clusters could be identified although genetic differentiation was low among natural populations at large geographic scales across Europe (Tuna et al. 2004, Last et al. 2013). The relatively high levels of genetic diversity and genetic admixture reported for the species in these studies are not surprising given that for outcrossing and wind pollinated grass species gene flow among populations is expected to be high (Brown 1978). Furthermore, human mediated transfer likely also plays a role in promoting genetic admixture in grasses which are also used as cultivars in farming (Last et al. 2013, von Cräutlein et al. 2019). Although the phylogeographic origin has likely influenced genetic structure of host populations and this may be reflected in the genetic structure of *E. typhina* and *E. clarkii*, analysis of a larger geographic area ideally spanning the entire distribution range and including Eastern and Southern European populations would be necessary to be able to detect patterns of glacial refugia. The data in this study represents a smaller scale of sampling and the fine-scale population structure of the two pathogens suggests that additional processes play a role in driving genetic differentiation between populations.

2. Distribution and dispersal of pathogens and hosts

Both *D. glomerata* and *H. lanatus* are native and widely distributed plants in grasslands across Europe and share similar long-distance pollen dispersal mediated by wind whereas seeds usually disperse more locally (Düll and Kutzelnigg 2016). Our sampling locations are situated in natural grassland ecosystems where the impact of anthropogenic use is likely minimal. Nevertheless, it is possible that agricultural practices such as sowing for pasture farming or forage production could have differentially affected the genetic make-up of grass populations in the two host species. While *H. lanatus* is not commercially sown due to its low nutritional value as a forage grass and may even be considered a weed, *D. glomerata* cultivars are frequently sown as part of commercial forage grass mixtures. Outcrossing of such cultivar genotypes into natural populations may promote gene flow and have a homogenizing effect on genetic structure in *D. glomerata*, however, we expect admixture to be reduced by the fact that

pastures are usually mown before grasses produce flowers or seeds. We have not assessed the underlying genetic structure of natural host populations in this study, however, higher levels of gene flow and reduced genetic differentiation among *D. glomerata* populations could explain why its pathogen *E. typhina* also shows more admixture and higher diversity. Populations of *H. lanatus* on the other hand are unlikely to experience more gene flow due to anthropogenic practices and populations may therefore remain more differentiated which is reflected in the genetic structure of *E. clarkii*. While both host grass species are similarly widespread across our study area and often co-occur in the same habitat, the abundance of their obligate pathogens differs: *E. clarkii* can be very common locally, such is the case of sympatric populations sampled for this study, but generally has a patchier distribution whereas infections of *E. typhina* on *D. glomerata* are more widely distributed. This pattern is consistent with the genetic data, supporting the smaller population sizes and lower levels of gene flow among *E. clarkii* populations. We hypothesize that the different spore sizes of *E. typhina* and *E. clarkii* differentially impact the dispersal abilities of these pathogens. *E. typhina* produces long, thread-like spores which may entail more efficient long-distance dispersal by wind whereas *E. clarkii* produces shorter and wider part-spores (White Jr. 1993, Leuchtman and Clay 1997). The resulting large number of part-spores may be more efficiently dispersed over shorter distances and increase disease pressure locally.

3. The role of local adaptation

Despite the abundance of host plants and the large production of wind-dispersed meiotic spores in both *E. typhina* and *E. clarkii* which serve contagious spread to new host plants (Leuchtman and Clay 1997), *Epichloe* infections are not everywhere. The strong pattern of genetic structure suggests that effective long-distance dispersal may be limited, for example by the range of susceptible host genotypes. For obligate biotrophic fungi such as *Epichloe* that depend on their living host, the host plant constitutes the distinct ecological niche that the pathogen evolves in. The ongoing coevolutionary arms-race in such host-pathogen systems is thought to be one of the most important drivers of natural selection acting at the population level and is generally assumed to result in local adaptation of pathogen populations to

host genotypes and vice-versa (Greischar and Koskella 2007). This may be reflected in congruence of genetic structure between host and pathogens (e.g. Feurtey et al. 2016; Hartmann et al. 2020), however, the extent to which co-evolutionary processes lead to genetic co-structure and local adaptation strongly depends on the strength of selection exerted by both partners, dispersal abilities related to levels of gene-flow and the role of local environmental factors (Croll and McDonald 2016). As discussed, levels of gene flow for outcrossing and wind pollinated grass species such as *D. glomerata* and *H. lanatus* are expected to be high, nevertheless, the abundance of rare alleles specific to geographically distinct populations supports the assumption that *D. glomerata* populations may be locally adapted (Tuna et al. 2004, Last et al. 2013). For *H. lanatus* genetic diversity has, to our knowledge, not been assessed to date, however, reciprocal transplant experiments provided some evidence for local adaptation of European *H. lanatus* populations (Bischoff et al. 2006). Given that *E. typhina* and *E. clarkii* are highly specialized pathogens that cannot grow or reproduce without establishing a systemic infection in their host plant, it is expected that co-evolutionary processes have shaped the fine-scale genetic structure in this system and that local adaptation may have contributed to the accumulation of genetic differences among *Epichloe* populations. Future studies should aim to further investigate patterns of local adaptation by assessing congruence of genetic structure in these two host-pathogen pairs combined with empirical tests: If pathogen populations are locally adapted they should be more fit on their local host genotypes than on foreign host genotypes (home vs away) and should outperform foreign pathogen genotypes on local hosts (local vs foreign) (Croll and McDonald 2016). This could be assessed experimentally in the *Epichloe* study system by performing reciprocal inoculation experiments with plants from natural populations.

Additional host species confirmed for *E. clarkii*

In our extensive sampling of natural populations of *E. clarkii*, we found two locations in the UK and in Bretagne, Northern France where the pathogen occurred on the host grass *Holcus mollis*. This was surprising given that *E. clarkii* is considered to be highly specialized on *H. lanatus* and, moreover, *H. mollis* is the host of a distinct *Epichloe* species, *E. mollis*. However, a literature review revealed that this finding indeed backs

up previous observations from the UK, where *E. clarkii* was identified on *H. mollis* based on its ascospore morphology (Spooner and Kemp 2005). We now provide additional evidence with genetic data and, for the first time, report *H. mollis* as an *E. clarkii* host on European mainland. It needs to be noted that hybrids between *H. lanatus* and *H. mollis* have been described (*Holcus x hybridus*, Wein 1913), which usually show morphologically intermediate phenotypes but tend to resemble *H. mollis* more closely (Beddows 1961, Carroll and Jones 1962). In our samples, species identification of the putative *H. mollis* host grasses was based on the presence of orthotropic shoots/rhizomes indicating clonal propagation and the presence of longer hairs on the nodes compared to *H. lanatus*. Characters of the inflorescence could not be considered since they were encased in the fungal fruiting structure. Although these morphological characters match well with descriptions of *H. mollis*, we cannot entirely exclude that samples originate from hybrids and this will need to be investigated further. In any case this finding opens up new avenues to investigate the evolution of host specialization and divergence within the *Epichloe* study system.

CONCLUSION

The comparison of genetic structure within and between populations of *E. typhina* and *E. clarkii* indicate distinct evolutionary histories of the two species and this is likely influenced by biological factors such as dispersal ability as well as the underlying genetic structure of host populations which may differentially affect coevolutionary dynamics between the two pathogens and their hosts. It appears that *E. typhina* infecting *D. glomerata* is a more ubiquitous pathogen with higher dispersal abilities among more genetically homogenous host populations and *E. clarkii* infecting *H. lanatus* and *H. mollis* forms more fragmented populations restricted by lower dispersal abilities and adaptation to local host genotypes. We expect that in both systems, sufficiently low levels of gene flow and divergent selection should promote local adaptation to the host, however, *E. clarkii* with its lower effective population size is more likely to experience metapopulation dynamics with frequent local extinction and recolonization events which may lead to loss of genetic diversity within populations and generate genetic differences between populations through bottlenecks and genetic drift. Future work will

investigate signatures of adaptive variation in these two pathogens and aim to disentangle to effects of drift and selection. This study and the dataset we present here represents an essential basis for understanding coevolution and adaptation in the *Epichloe* study system and providing novel insights into the population structure and patterns of gene-flow. It demonstrates how the biology of these species and the interaction with their host plants may influence evolutionary trajectories and helps to increase our understanding of the population genetic processes that generate and maintain diversity of fungal pathogens in natural ecosystems.

ACKNOWLEDGEMENTS

We would like to kindly thank Martyn Ainsworth (Royal Botanical Gardens Kew, UK), Fabrizio Steinebrunner, Jean-Paul Priou, James M. Neenan, Elisabeth Walder and Lilith Treindl for assistance with sample collection, and Claudia Michel and Beatrice Arnold for laboratory assistance. We thank Karsten Rohweder for help with the GPS device and Niklaus Zemp for bioinformatic support. Data presented and analyzed here were generated in collaboration with the Genetic Diversity Centre (GDC), ETH Zurich and the Functional Genomics Center Zurich (FGCZ), UZH.

REFERENCES

- Ashby, B., and S. Gupta. 2014. Parasitic castration promotes coevolutionary cycling but also imposes a cost on sex. *Evolution* 68:2234–2244.
- Van der Auwera, G. A., M. O. Carneiro, C. Hartl, R. Poplin, G. Del Angel, A. Levy-Moonshine, T. Jordan, K. Shakir, D. Roazen, and J. Thibault. 2013. From FastQ data to high-confidence variant calls: the genome analysis toolkit best practices pipeline. *Current Protocols in Bioinformatics* 43:10–11.
- Badouin, H., P. Gladieux, J. Gouzy, S. Siguenza, G. Aguileta, A. Snirc, S. Le Prieur, C. Jeziorski, A. Branca, and T. Giraud. 2017. Widespread selective sweeps throughout the genome of model plant pathogenic fungi and identification of effector candidates. *Molecular Ecology* 26:2041–2062.
- Barrett, L. G., P. H. Thrall, J. J. Burdon, and C. C. Linde. 2008. Life history determines genetic structure and evolutionary potential of host-parasite interactions. *Trends in Ecology and Evolution* 23:678–685.
- Barrett, L. G., P. H. Thrall, P. N. Dodds, M. Van Der Merwe, C. C. Linde, G. J. Lawrence, and J. J. Burdon.

2009. Diversity and evolution of effector loci in natural populations of the plant pathogen *Melampsora lini*. *Molecular Biology and Evolution* 26:2499–2513.
- Bates, D., M. Mächler, B. M. Bolker, and S. C. Walker. 2015. Fitting linear mixed-effects models using lme4. *Journal of Statistical Software* 67:1–48.
- Beddows, A. R. 1961. *Holcus lanatus* L. *Journal of Ecology* 49:421–430.
- Bischoff, A., L. Crémieux, M. Smilauerova, C. S. Lawson, S. R. Mortimer, J. Dolezal, V. Lanta, A. R. Edwards, A. J. Brook, M. Macel, J. Leps, T. Steinger, and H. Müller-Schärer. 2006. Detecting local adaptation in widespread grassland species - The importance of scale and local plant community. *Journal of Ecology* 94:1130–1142.
- Bolger, A. M., M. Lohse, and B. Usadel. 2014. Trimmomatic: A flexible trimmer for Illumina sequence data. *Bioinformatics* 30:2114–2120.
- Bradley, D. J., G. S. Gilbert, and J. B. H. Martiny. 2008. Pathogens promote plant diversity through a compensatory response. *Ecology Letters* 11:461–469.
- Brown, A. H. D. 1978. Isozymes, plant population genetic structure and genetic conservation. *Theoretical and Applied Genetics* 52:145–157.
- Bultman, T. L., A. Leuchtman, T. J. Sullivan, and A. P. Dreyer. 2011. Do *Botanophila* flies provide reproductive isolation between two species of *Epichloë* fungi? A field test. *New Phytologist* 190:206–212.
- Burdon, J. J., and A.-L. Laine. 2019. The diverse and ubiquitous nature of pathogens. Pages 1–28 *Evolutionary dynamics of plant-pathogen interactions*. Cambridge University Press, Cambridge, UK.
- Carroll, C. P., and K. Jones. 1962. Cytotaxonomic studies in *Holcus* - A morphological study of the triploid F1 hybrid between *Holcus lanatus* L. and *H. mollis* L. *New Phytologist* 61:72–84.
- Christensen, M. J., A. Leuchtman, D. D. Rowan, and B. A. Tapper. 1993. Taxonomy of *Acremonium* endophytes of tall fescue (*Festuca arundinacea*), meadow fescue (*F. pratensis*) and perennial ryegrass (*Lolium perenne*). *Mycological Research* 97:1083–1092.
- Chung, K.-R., and C. L. Schardl. 1997. Sexual cycle and horizontal transmission of the grass symbiont, *Epichloë typhina*. *Mycological Research* 101:295–301.
- Clay, K., and C. Schardl. 2002. Evolutionary origins and ecological consequences of endophyte symbiosis with grasses. *The American Naturalist* 160:S99–S127.
- von Cräutlein, M., P. H. Leinonen, H. Korpelainen, M. Helander, H. Väre, and K. Saikkonen. 2019. Postglacial colonization history reflects in the genetic structure of natural populations of *Festuca rubra* in Europe. *Ecology and Evolution* 9:3661–3674.
- Craven, K. D., P. T. W. Hsiau, A. Leuchtman, W. Hollin, and C. L. Schardl. 2001. Multigene phylogeny of *Epichloë* species, fungal symbionts of grasses. *Annals of the Missouri Botanical Garden* 88:14–34.
- Croll, D., and B. A. McDonald. 2016. The genetic basis of local adaptation for pathogenic fungi in agricultural ecosystems. *Molecular Ecology* 26:2027–2040.
- Danecek, P., A. Auton, G. Abecasis, C. A. Albers, E. Banks, M. A. DePristo, R. E. Handsaker, G. Lunter, G. T. Marth, S. T. Sherry, G. McVean, R. Durbin, and 1000 Genomes Project Analysis Group. 2011. The variant call format and VCFtools. *Bioinformatics (Oxford, England)* 27:2156–2158.

- Düll, R., and H. Kutzelnigg. 2016. Taschenlexikon der Pflanzen Deutschlands und angrenzender Länder. Eighth edition. Quelle und Meyer, Wiebelsheim.
- Earl, D. A., and B. M. vonHoldt. 2012. STRUCTURE HARVESTER: a website and program for visualizing STRUCTURE output and implementing the Evanno method. *Conservation Genetics Resources* 4:359–361.
- Evanno, G., S. Regnaut, and J. Goudet. 2005. Detecting the number of clusters of individuals using the software structure: a simulation study. *Molecular Ecology* 14:2611–2620.
- Feurtey, A., P. Gladieux, M. E. Hood, A. Snirc, A. Cornille, L. Rosenthal, and T. Giraud. 2016. Strong phylogeographic co-structure between the anther-smut fungus and its white campion host. *New Phytologist* 212:668–679.
- Fisher, M. C., D. A. Henk, C. J. Briggs, J. S. Brownstein, L. C. Madoff, S. L. McCraw, and S. J. Gurr. 2012. Emerging fungal threats to animal, plant and ecosystem health. *Nature* 484:186.
- Fjellheim, S., O. A. Rognli, K. Fosnes, and C. Brochmann. 2006. Phylogeographical history of the widespread meadow fescue (*Festuca pratensis* Huds.) inferred from chloroplast DNA sequences. *Journal of Biogeography* 33:1470–1478.
- François, O., M. G. B. Blum, M. Jakobsson, and N. A. Rosenberg. 2008. Demographic History of European Populations of *Arabidopsis thaliana*. *PLoS Genetics* 4:e1000075.
- Greischar, M. A., and B. Koskella. 2007. A synthesis of experimental work on parasite local adaptation. *Ecology Letters* 10:418–434.
- Hartmann, F. E., A. Snirc, A. Cornille, C. Godé, P. Touzet, F. Van Rossum, E. Fournier, S. Le Prieur, J. Shykoff, and T. Giraud. 2020. Congruent population genetic structures and divergence histories in anther-smut fungi and their host plants *Silene italica* and the *Silene nutans* species complex. *Molecular Ecology* 29:1154–1172.
- Hewitt, G. 2000. The genetic legacy of the quaternary ice ages. *Nature* 405:907–913.
- Jombart, T. 2008. adegenet: A R package for the multivariate analysis of genetic markers. *Bioinformatics* 24:1403–1405.
- Jousimo, J., A. J. M. Tack, O. Ovaskainen, T. Mononen, H. Susi, C. Tollenaere, and A. L. Laine. 2014. Ecological and evolutionary effects of fragmentation on infectious disease dynamics. *Science* 344:1289–1293.
- Kleiman, E. 2017. EMAtools: Data Management Tools for Real-Time Monitoring/Ecological Momentary Assessment Data. R package version 0.1.3. <https://cran.r-project.org/package=EMAtools>.
- Large, E. C. 1954. Surveys for choke (*Epichloe typhina*) in cocksfoot seed crops, 1951–53. *Plant Pathology* 3:6–11.
- Last, L., F. Widmer, W. Fjellstad, S. Stoyanova, and R. Kölliker. 2013. Genetic diversity of natural orchardgrass (*Dactylis glomerata* L.) populations in three regions in Europe. *BMC Genetics* 14.
- Leuchtmann, A., C. W. Bacon, C. L. Schardl, J. F. White Jr., and M. Tadych. 2014. Nomenclatural realignment of *Neotyphodium* species with genus *Epichloë*. *Mycologia* 106:202–215.
- Leuchtmann, A., and K. Clay. 1997. The population biology of grass endophytes. Pages 185–202 in G. C. Carroll and P. Tudzynski, editors. *The Mycota, Vol. V. Part B, Plant Relationships*. Springer-Verlag,

- Berlin, Heidelberg.
- Leuchtman, A., and C. L. Schardl. 1998. Mating compatibility and phylogenetic relationships among two new species of *Epichloë* and other congeneric European species. *Mycological Research* 102:1169–1182.
- Li, H., and R. Durbin. 2009. Fast and accurate short read alignment with Burrows-Wheeler transform. *Bioinformatics* 25:1754–1760.
- Li, H., B. Handsaker, A. Wysoker, T. Fennell, J. Ruan, N. Homer, G. Marth, G. Abecasis, R. Durbin, and 1000 Genome Project Data Processing Subgroup. 2009. The Sequence Alignment/Map format and SAMtools. *Bioinformatics* 25:2078–2079.
- Lynch, M., W. Sung, K. Morris, N. Coffey, C. R. Landry, E. B. Dopman, W. J. Dickinson, K. Okamoto, S. Kulkarni, D. L. Hartl, and W. K. Thomas. 2008. A genome-wide view of the spectrum of spontaneous mutations in yeast. *Proceedings of the National Academy of Sciences of the United States of America* 105:9272–9277.
- McDonald, B. A., and C. Linde. 2002. Pathogen population genetics, evolutionary potential, and durable resistance. *Annual Review of Phytopathology* 40:349–379.
- McGrath, S., T. R. Hodkinson, and S. Barth. 2007. Extremely high cytoplasmic diversity in natural and breeding populations of *Lolium* (Poaceae). *Heredity* 99:531–544.
- McKenna, A., M. Hanna, E. Banks, A. Sivachenko, K. Cibulskis, A. Kernytsky, K. Garimella, D. Altshuler, S. Gabriel, and M. Daly. 2010. The Genome Analysis Toolkit: A MapReduce framework for analyzing next-generation DNA sequencing data. *Genome Research* 20:1297–1303.
- McMullan, M., M. Rafiqi, G. Kaithakottil, B. J. Clavijo, L. Bilham, E. Orton, L. Percival-Alwyn, B. J. Ward, A. Edwards, and D. G. O. Saunders. 2018. The ash dieback invasion of Europe was founded by two genetically divergent individuals. *Nature Ecology and Evolution* 2:1000–1008.
- Milgroom, M. G., K. Sotirovski, D. Spica, J. E. Davis, M. T. Brewer, M. Milev, and P. Cortesi. 2008. Clonal population structure of the chestnut blight fungus in expanding ranges in southeastern Europe. *Molecular Ecology* 17:4446–4458.
- Möller, M., and E. H. Stukenbrock. 2017. Evolution and genome architecture in fungal plant pathogens. *Nature Reviews Microbiology* 15:756–771.
- Ovington, J. D., and G. Scurfield. 1956. *Holcus mollis* L. *Journal of Ecology* 44:272–280.
- Persoons, A., K. J. Hayden, B. Fabre, P. Frey, S. De Mita, A. Tellier, and F. Halkett. 2017. The escalatory Red Queen: Population extinction and replacement following arms race dynamics in poplar rust. *Molecular Ecology* 26:1902–1918.
- Pfender, W. F., and S. C. Alderman. 2006. Regional development of orchardgrass choke and estimation of seed yield loss. *Plant Disease* 90:240–244.
- Pritchard, J. K., M. Stephens, and P. Donnelly. 2000. Inference of population structure using multilocus genotype data. *Genetics* 155:945–959.
- Provan, J., and K. D. Bennett. 2008. Phylogeographic insights into cryptic glacial refugia. *Trends in Ecology and Evolution* 23:564–571.
- R Core Team. 2013. R: A language and environment for statistical computing. Vienna, Austria.
- Raffaele, S., and S. Kamoun. 2012. Genome evolution in filamentous plant pathogens: Why bigger can be

- better. *Nature Reviews Microbiology* 10:417–430.
- Raynal, G. 1991. Libération des ascospores d'*Epichloë typhina*, agent de la quenouille du dactyle. Conséquences pour l'épidémiologie et la lutte. *Fourrages (Versailles)*:345–358.
- Robinson, J. T., H. Thorvaldsdóttir, W. Winckler, M. Guttman, E. S. Lander, G. Getz, and J. P. Mesirov. 2011. Integrative Genomics Viewer. *Nature Biotechnology* 29:24–26.
- Schardl, C. L., C. A. Young, N. Moore, N. Krom, P.-Y. Dupont, J. Pan, S. Florea, J. S. Webb, J. Jaromczyk, and J. W. Jaromczyk. 2014. Genomes of plant-associated Clavicipitaceae. *Advances in Botanical Research* 70:291–327.
- Schirrmann, M. K., S. Zoller, D. Croll, E. H. Stukenbrock, A. Leuchtmann, and S. Fior. 2018. Genomewide signatures of selection in *Epichloë* reveal candidate genes for host specialization. *Molecular Ecology* 27:3070–3086.
- Schirrmann, M. K., S. Zoller, S. Fior, and A. Leuchtmann. 2015. Genetic evidence for reproductive isolation among sympatric *Epichloë* endophytes as inferred from newly developed microsatellite markers. *Microbial Ecology* 70:51–60.
- Spooner, B. M., and S. L. Kemp. 2005. *Epichloë* in Britain. *Mycologist* 19:82–87.
- Stukenbrock, E. H., T. Bataillon, J. Y. Dutheil, T. T. Hansen, R. Li, M. Zala, B. A. McDonald, J. Wang, and M. H. Schierup. 2011. The making of a new pathogen: Insights from comparative population genomics of the domesticated wheat pathogen *Mycosphaerella graminicola* and its wild sister species. *Genome Research* 21:2157–2166.
- Stukenbrock, E. H., and B. A. McDonald. 2008. The origins of plant pathogens in agro-ecosystems. *Annual Review of Phytopathology* 46:75–100.
- Tajima, F. 1989. Statistical method for testing the neutral mutation hypothesis by DNA polymorphism. *Genetics* 123:585–595.
- Takahata, N., and M. Nei. 1985. Gene genealogy and variance of interpopulational nucleotide differences. *Genetics* 110:325–344.
- Tarasov, A., A. J. Vilella, E. Cuppen, I. J. Nijman, and P. Prins. 2015. Sambamba: Fast processing of NGS alignment formats. *Bioinformatics* 31:2032–2034.
- Tellier, A., S. Moreno-Gámez, and W. Stephan. 2014. Speed of adaptation and genomic footprints of host-parasite coevolution under arms race and trench warfare dynamics. *Evolution* 68:2211–2224.
- Thompson, J. N. 2005. *The geographic mosaic of coevolution*. University of Chicago Press, Chicago.
- Torres, D. E., U. Oggenfuss, D. Croll, and M. F. Seidl. 2020. Genome evolution in fungal plant pathogens: Looking beyond the two-speed genome model. *Fungal Biology Reviews* 34:136–143.
- Treindl, A. D., and A. Leuchtmann. 2019. Assortative mating in sympatric ascomycete fungi revealed by experimental fertilizations. *Fungal Biology* 123:676–686.
- Tuna, M., D. K. Khadka, M. K. Shrestha, K. Arumuganathan, and A. Golan-Goldhirsh. 2004. Characterization of natural orchardgrass (*Dactylis glomerata* L.) populations of the Thrace Region of Turkey based on ploidy and DNA polymorphisms. *Euphytica* 135:39–46.
- Tutin, T. G., V. H. Heywood, N. A. Burges, D. M. Moore, D. H. Valentine, S. M. Walters, and D. A. Webb. 1980. *Flora Europaea* Vol. 5. Page Cambridge University Press. Cambridge University Press,

Cambridge.

- Vercken, E., M. C. Fontaine, P. Gladieux, M. E. Hood, O. Jonot, and T. Giraud. 2010. Glacial refugia in pathogens: European genetic structure of anther smut pathogens on *silene latifolia* and *silene dioica*. *PLOS Pathogens* 6.
- Watterson, G. A. 1975. On the number of segregating sites in genetical models without recombination. *Theoretical Population Biology* 7:256–276.
- Weir, B. S., and C. C. Cockerham. 1984. Estimating F-statistics for the analysis of population structure. *Evolution* 38:1358–1370.
- White, J. F., and T. L. Bultman. 1987. Endophyte-host associations in forage grasses. VIII. Heterothallism in *Epichloë typhina*. *American Journal of Botany* 74:1716–1721.
- White Jr., J. F. 1993. Endophyte-host associations in grasses. XIX. A systematic study of some sympatric species of *Epichloë* in England. *Mycologia* 85:444–455.
- Wickham, H. 2016. *ggplot2: Elegant graphics for data analysis*. Springer.
- Wilson, D. J., D. Falush, and G. McVean. 2005. Germs, genomes and genealogies. *Trends in Ecology and Evolution* 20:39–45.
- Winter, D. J., A. R. D. Ganley, C. A. Young, I. Liachko, C. L. Schardl, P. Y. Dupont, D. Berry, A. Ram, B. Scott, and M. P. Cox. 2018. Repeat elements organise 3D genome structure and mediate transcription in the filamentous fungus *Epichloë festucae*. *PLoS Genetics* 14:1–29.
- Zheng, X., D. Levine, J. Shen, S. M. Gogarten, C. Laurie, and B. S. Weir. 2012. A high-performance computing toolset for relatedness and principal component analysis of SNP data. *Bioinformatics* 28:3326–3328.

SUPPLEMENTARY NOTES

Variant filtration

The use of Rank Sum Tests implemented in GATK (McKenna et al. 2010), is commonly applied in many variant filtration pipelines, including haploid datasets (e.g. Hartmann et al. 2018, Mohd-Assaad et al. 2018). However, RankSum annotations using the GATK VariantFiltration tool are only calculated for heterozygous sites with read support for both the REF and ALT allele. In haploid individuals true variants should only contain reads for either the REF or ALT allele and erroneously “heterozygous” sites with support for both alleles are evidence of sequencing errors or misalignments during mapping. While one of the called alleles may be real, the others are a result of mapping artifacts or sequencing errors and therefore RankSumTest statistics should be positively biased towards the “real” allele. Based on these considerations we have not used and do not recommend using Rank Sum Test statistics (calculated within individuals) for filtering of haploid datasets. For further information see details on ReadPosRankSumTest, MQRankSumTest and BaseQRankSumTest here: <https://gatk.broadinstitute.org/hc/en-us/categories/360002369672-Tool-Index> and notes on hard filtering here: <https://gatk.broadinstitute.org/hc/en-us/articles/360035531112--How-to-Filter-variants-either-with-VQSR-or-by-hard-filtering>.

The filtering step to exclude SNPs with >30% missing data within one or more populations was performed using only the eight sympatric populations and not for the additional allopatric samples which only included few isolates per locality.

Datasets

After filtering we retained 658,021 and 400,033 high-quality biallelic SNPs in the *E. typhina* and the *E. clarkii* dataset respectively. For downstream analyses we applied additional filters to subset this dataset: To calculate population statistics we removed isolates that were sampled from allopatric populations and removed variants from the SNP dataset that were monomorphic among populations. We also generated a dataset containing only SNPs genotyped across all isolates (no missing data). For the STRUCTURE analysis, we applied a minor allele frequency filter of 0.05 and LD-pruned

variants for sites with a pairwise genotypic r^2 greater than 0.3 in 10kb windows. For DAPC, this dataset was additionally thinned to one variant per 2kb.

Replicates

We included technical replicates of five and three individuals respectively. For each replicate, we removed the replicate with fewer mapped reads from the dataset after assessing genotype concordance rates. Average allele concordance between replicates was 98.9% with a standard deviation of 0.1% in Et and 97.7% with a standard deviation of 0.3% in Ec.

Population sub-structure

We investigated genetic sub-structure within sympatric populations of *E. typhina* and *E. clarkii* to identify highly related individuals using PCA as described in the main text. Most populations did not have apparent sub-structure or contain highly related individuals. The exceptions in *E. typhina* were the populations Cev, with 5 individuals that were more distinct from the rest in principal component space, and MdR with 2 clearly separated genotypes. In *E. clarkii* the Kew population showed a high degree of sub-structure with three groups of highly related individuals and a single intermediate genotype. Individuals from each sub-populations corresponded to different areas within the sampling location and the very low level of diversity within sub-populations compared to the whole population suggested that these may indeed be isolates of clonal origin. This was not highly surprising and likely related to the clonal population structure of the host grass *H. mollis*.

PCA analysis and calculation of summary statistics by compartments

Epichloe genomes are compartmentalized into conserved gene-rich regions and AT-rich regions containing a high proportion of repeat elements and only few genes (Chapter I of this thesis). Accurate variant discovery in these polymorphic AT-rich regions using alignment of short-reads is often challenging because reads are more likely to map incorrectly (see Pfeifer 2017). We used stringent filtering criteria to remove erroneous SNP calls and assessed the distribution of variants in our filtered dataset. For this, genome-wide SNPs (complete datasets, see Supplementary Table S3) were split by

compartment to generate two variant files for each species: one containing SNPs called in gene-rich regions and one containing the SNPs called in AT-rich regions. The number of SNPs within each compartment was roughly proportional to the relative size of the compartment: In *E. typhina* the AT-rich compartment made up 31.7% of the genome and contained 23.9% of SNPs whereas in *E. clarkii* the AT-rich compartment made up 48.6% and contained 53.1% of SNPs. We used these SNP datasets to analyze genetic structure among populations using PCA and found that clusters corresponding to sampled populations were generally resolved with variants in either compartments giving us confidence that our SNP datasets reflect true biological variation (Supplementary Figures S13 and S14). However, genetic differentiation between clusters was more apparent in the gene-rich compartment and more variance was explained by these variants in both species. In *E. typhina* clusters were less clearly resolved in the AT-rich compartment which may suggest lower genetic differentiation in these regions of the genome (Supplementary Figure S13 C & E). The overall pattern of population structure was more strongly shaped by strong genetic differentiation in the gene-rich regions and the high number of SNPs there. In *E. clarkii* genetic differentiation between Cev and So was less pronounced in the AT-rich compartment compared to the gene-rich, which could explain why these populations were assigned to the same cluster in the overall STRUCTURE analysis at K=7. We hypothesize that the patterns of genetic differentiation in distinct genome compartments could reflect the differential action of random genetic drift and selection in these genome compartments: Variation in the conserved gene-rich regions may be strongly influenced by local selection driving differentiation between populations while the AT-rich compartment evolves under a stronger influence of drift. We also computed the summary statistics π , Tajima's D, LD r^2 and SNP density for each compartment separately using the same parameters as in the genome-wide analysis. Compiled plots for each population can be found in the Supplementary (Figures S15 and S16). Across all populations, rare alleles were more abundant in the AT-rich compartment (lower Tajima's D than in the gene-rich compartment) and this could reflect the activity of RIP (repeat-induced point mutation) in these regions of the genome (most mutations introduced by RIP are expected to be at low frequency unless they are targeted by selection). Furthermore, SNPs were more strongly linked in the gene-rich compartment. This difference was most striking in *E. clarkii* MdR where the

gene-rich compartment had an extremely high mean r^2 consistent with the idea that this population may have experienced a recent bottleneck. Average SNP density in 10kb windows was generally lower in AT-rich regions and this is likely influenced by the fact that many variants in regions where mapping was very poor were removed during filtering resulting in a relatively large number of “empty” AT-windows that did not contain any more SNPs. Nevertheless, outlier points indicate that we also retained a number of highly polymorphic windows with higher-than-average SNP density.

Calculation of effective population size

To compare effective population sizes within and between species we estimated effective population sizes (N_e) from polymorphism data with the parameter $\theta = 2N_e\mu$. Calculations were based on Watterson's estimators of θ (Watterson 1975), which is based on the number of segregating sites in a population relative to the sequence length, and an assumed mutation rate (μ) of 3.3×10^{-8} per site. For each population we calculated the number of segregating sites in 5kb windows along the genome using the PopGenome R package (Pfeifer et al. 2014). We divided the number of segregating sites by the $(n-1)^{\text{th}}$ harmonic number to calculate Watterson's θ , with n being the number of chromosomes or in case of haploids simply the sample size. Following the formula, we then calculated $N_e = \theta/2\mu$.

Assessment of hybridization in sympatric populations

A previous study by Bultman et al. (2011) found that 9.3 % of fruiting bodies collected from a sympatric population of *E. typhina* and *E. clarkii* contained asci with hybrid spores. We assessed the occurrence of hybridizations in two additional natural populations in Southern and Central France (Cev and Auv) one year after the collection of population samples. Populations were revisited at the end of the flowering season in July 2018 and 1-3 ripe stromata with perithecia of *E. typhina* and *E. clarkii* were collected from 7-13 *D. glomerata* and *H. lanatus* plants. From each stroma three probes were taken and ascospore morphology within perithecia was observed microscopically. When asci contain the distinct spore-morphotypes of both *E. typhina* and *E. clarkii*, this indicates that hybridizations have occurred (see Treindl and Leuchtmann 2019). We detected hybrid asci in 12.5 and 8.3% (mean 10.4%) of probes taken from naturally fertilized

fruiting bodies (stromata) collected from sympatric populations in Southern and Central France (Cev and Auv, Supplementary Table S9-S11). Despite the occurrence of interspecific mating, however, we did not detect any intermediate/ putatively hybrid genotypes in a preliminary analysis using a draft version of the *E. typhina* reference genome and individuals from both species, suggesting lack of gene flow between *E. typhina* and *E. clarkii*. Additionally, we did not find evidence for hybrid or introgressed isolates infecting either *D. glomerata* or *H. lanatus* despite the extensive analyses of over 400 isolates from sympatric populations (where the potential for hybridization exists). These observations suggest that despite incidences of interspecific mating as evidence by the occurrence of hybrid asci in sympatric populations, *E. typhina* and *E. clarkii* are reproductively isolated and that postzygotic reproductive barriers exist that prevent the establishment of hybrid genotypes in the population and maintain the genetic and ecological integrity of the two species.

Mating types

We analyzed the distribution of mating types in sympatric populations as skewed sex ratio within populations reduces effective population size (N_e) and can bias population genomic inferences (Wright 1938). Therefore, we analyzed the distribution of mating types in sympatric populations. Sexual *Epichloe* species have a bipolar heterothallic mating system (White and Bultman 1987). Each haploid individual possesses a mating type locus with one of two alternative idiomorphs designated MTA and MTB or mat-1 and mat-2. The MTA idiomorph includes three genes (mtAA, mtAB, and mtAC) whereas the MTB idiomorph includes only the gene mtBA (Schardl et al. 2012). The reference genomes for both species have the MTA mating type idiomorph (mat-1) so to assess sex ratios in our populations (i.e. the proportions of mating types) we extracted the region including the mating type locus from the alignment files (bam) of every individual and used the function `genomecov -bga` in *bedtools* version 2.28.0, to report read depth in a BedGraph format (Quinlan and Hall 2010). This includes regions with zero coverage and can be visualized in IGV (Integrative Genomics Viewer) (Robinson et al. 2011). Individuals with the same mating type as the reference (mat-1) have normal read coverage whereas individuals with the alternative mating type (mat-2) have a deletion/ no read coverage. Here we included all individuals with sufficient coverage around the

locus, even those with low coverage overall that were removed for subsequent analyses. We assessed the proportions of mating types and found that both mating types were present in all sampled populations. Across all individuals in both *E. typhina* and *E. clarkii* sex ratios were slightly biased towards mating type 1 with 54% in *E. typhina* and 55% in *E. clarkii* (Supplementary Figures S17 and S18). Among populations sex ratios varied from 41% to 68% mat-1 in *E. typhina* and 39% to 71% in *E. clarkii*. This variation is likely to be increased by stochastic processes due to small sample sizes.

References:

- Bultman, T. L., A. Leuchtmann, T. J. Sullivan, and A. P. Dreyer. 2011. Do Botanophila flies provide reproductive isolation between two species of *Epichloë* fungi? A field test. *New Phytologist* 190:206–212.
- Hartmann, F. E., B. A. McDonald, and D. Croll. 2018. Genome-wide evidence for divergent selection between populations of a major agricultural pathogen. *Molecular Ecology* 27:2725–2741.
- McKenna, A., M. Hanna, E. Banks, A. Sivachenko, K. Cibulskis, A. Kernytsky, K. Garimella, D. Altshuler, S. Gabriel, and M. Daly. 2010. The Genome Analysis Toolkit: A MapReduce framework for analyzing next-generation DNA sequencing data. *Genome Research* 20:1297–1303.
- Mohd-Assaad, N., B. A. McDonald, and D. Croll. 2018. Genome-wide detection of genes under positive selection in worldwide populations of the barley scald pathogen. *Genome Biology and Evolution* 10:1315–1332.
- Pfeifer, B., U. Wittelsbürger, S. E. Ramos-Onsins, and M. J. Lercher. 2014. PopGenome: An efficient Swiss army knife for population genomic analyses in R. *Molecular Biology and Evolution* 31:1929–1936.
- Pfeifer, S. P. 2017. From next-generation resequencing reads to a high-quality variant data set. *Heredity* 118:111–124.
- Quinlan, A. R., and I. M. Hall. 2010. BEDTools: A flexible suite of utilities for comparing genomic features. *Bioinformatics* 26:841–842.
- Robinson, J. T., H. Thorvaldsdóttir, W. Winckler, M. Guttman, E. S. Lander, G. Getz, and J. P. Mesirov. 2011. Integrative Genomics Viewer. *Nature Biotechnology* 29:24–26.
- Schardl, C. L., B. Scott, C. A. Young, G. E. Aiken, R. L. McCulley, and J. R. Strickland. 2012. Recommendations for gene nomenclature for *Epichloë* species and related Clavicipitaceae. Pages 84–87 *Epichloae, endophytes of cool season grasses: implications, utilization and biology*. Proceedings of the 7th International Symposium on Fungal Endophytes of Grasses, Lexington, Kentucky, USA, 28 June to 1 July 2010. Samuel Roberts Noble Foundation.
- Treindl, A. D., and A. Leuchtmann. 2019. Assortative mating in sympatric ascomycete fungi revealed by experimental fertilizations. *Fungal Biology* 123:676–686.
- Watterson, G. A. 1975. On the number of segregating sites in genetical models without recombination. *Theoretical Population Biology* 7:256–276.
- White, J. F., and T. L. Bultman. 1987. Endophyte-host associations in forage grasses. VIII. Heterothallism in *Epichloë typhina*. *American Journal of Botany* 74:1716–1721.
- Wright, S. 1938. Size of population and breeding structure in relation to evolution. *Science* 87:430–431.

SUPPLEMENTARY TABLES AND FIGURES

Table S1: Sampling of sympatric populations

Fungal species	Host species	ID	Collection year	Country	Sampling location	Latitude	Longitude	n
<i>E. clarkii</i>	<i>H. lanatus</i>	So	2016	CH	Soglio, GR	46.3429	9.5373	33
<i>E. typhina</i>	<i>D. glomerata</i>	So	2016	CH	Soglio, GR	46.3429	9.5373	27
<i>E. clarkii</i>	<i>H. lanatus</i>	El	2017	F	Abbaye de Marbach, Alsace	48.0266	7.2752	32
<i>E. typhina</i>	<i>D. glomerata</i>	El	2017	F	Abbaye de Marbach, Alsace	48.0275	7.2769	32
<i>E. typhina</i>	<i>D. glomerata</i>	Kew	2017	UK	Kew Gardens, London	51.4783	-0.2990	21
<i>E. clarkii</i>	<i>H. mollis</i>	Kew	2017	UK	Kew Gardens, London	51.4783	-0.2990	30
<i>E. clarkii</i>	<i>H. lanatus</i>	Cev	2017	F	Saint-Paul-la-Coste, Cevennes	44.1483	3.9627	31
<i>E. typhina</i>	<i>D. glomerata</i>	Cev	2017	F	Saint-Paul-la-Coste, Cevennes	44.1483	3.9627	31
<i>E. typhina</i>	<i>D. glomerata</i>	Auv	2017	F	Murat, Auvergne	45.1263	2.8863	30
<i>E. clarkii</i>	<i>H. lanatus</i>	Auv	2017	F	Murat, Auvergne	45.1263	2.8863	32
<i>E. clarkii</i>	<i>H. lanatus</i>	Aub	2018	CH	Aubonne, VD	46.5123	6.3637	32
<i>E. typhina</i>	<i>D. glomerata</i>	Aub	2018	CH	Aubonne, VD	46.5123	6.3637	22
<i>E. typhina</i>	<i>D. glomerata</i>	MdR	2018	E	Montemayor del Rio	40.3469	-5.9091	33
<i>E. clarkii</i>	<i>H. lanatus</i>	MdR	2018	E	Montemayor del Rio	40.3469	-5.9091	29
<i>E. typhina</i>	<i>D. glomerata</i>	AubX	2005	CH	Aubonne, VD	46.5123	6.3637	23
<i>E. clarkii</i>	<i>H. lanatus</i>	AubX	2005	CH	Aubonne, VD	46.5123	6.3637	23

Table S2: Additional samples from allopatric populations

Fungal species	Host species	ID	Collection year	Country	Sampling location	Latitude	Longitude	n
<i>E. typhina</i>	<i>D. glomerata</i>	Other*	2017	UK	Mount Edgecomb, Plymouth	50.3554	-4.1753	3
<i>E. typhina</i>	<i>D. glomerata</i>	Other*	2017	F	Westhalten, Alsace	47.9656	7.2696	4
<i>E. typhina</i>	<i>D. glomerata</i>	Other*	2017	F	Roquestron, Préalpes d'Azur	43.8743	7.0097	3
<i>E. typhina</i>	<i>D. glomerata</i>	Other*	2017	F	Haut Thorenc, Préalpes d'Azur	43.7989	6.8460	3
<i>E. typhina</i>	<i>D. glomerata</i>	Other*	2017	F	Cultures, Mende	44.4890	3.3751	3
<i>E. typhina</i>	<i>D. glomerata</i>	Bret	2017	F	Bretagne	47.7692	-2.1384	4
<i>E. clarkii</i>	<i>H. mollis</i>	Bret	2017	F	Bretagne	47.7687	-2.1300	3

Table S3: SNP datasets used for different analyses

	<i>E. typhina</i> SNPs	<i>E. clarkii</i> SNPs	analyses
complete	658,021	400,033	PCA
populations only	657,966	397,283	population statistics
no missing data	295,446	63,930	PCA
LD-pruned	152,883	79,352	PCA, STRUCTURE
LD-pruned and thinned	12,562	11,338	DAPC
complete gene-rich	500,747	187,711	PCA
complete AT-rich	157,274	212,322	PCA

Table S4: Number of segregating sites per population

	<i>E. typhina</i>	<i>E. clarkii</i>
Aub	362,128	206,283
AubX	317,031	189,525
Auv	368,058	183,846
Cev	361,706	189,053
El	354,895	167,418
Kew	268,886	109,493
MdR	360,954	110,351
So	326,742	226,241

Table S5: Proportion of segregating sites per population [%]

	<i>E. typhina</i>	<i>E. clarkii</i>
Aub	55.0	51.9
AubX	48.2	47.7
Auv	55.9	46.3
Cev	55.0	47.6
El	53.9	42.1
Kew	40.9	27.6
MdR	54.9	27.8
So	49.7	56.9

A similar proportion of sites was polymorphic in most *E. clarkii* populations compared to *E. typhina* populations despite *E. clarkii* having a lower overall diversity. This could indicate similar mutation rates in the two species. Exceptions were Kew and MdR.

Table S6: Fixed differences between populations

	MdR	Kew	Auv	Cev	Aub	AubX	El	So		
<i>E. clarkii</i>	MdR		448	121	9	58	0	177	340	<i>E. typhina</i>
	Kew	10669		22	137	26	197	24	690	
	Auv	4598	685		0	0	83	0	214	
	Cev	5337	1922	83		9	0	126	170	
	Aub	4827	1285	78	0		0	0	11	
	AubX	5165	1493	58	23	0		0	1	
	El	6214	1672	173	44	1	3		57	
	So	5561	3507	762	21	288	263	486		

Number of loci fixed for alternative alleles ($F_{ST}=1$) between population pairs of *E. typhina* (top right) and *E. clarkii* (bottom left). Populations are ordered according to longitude (West to East).

Table S7: Mean nucleotide diversity (π) per site

	<i>E. typhina</i>	<i>E. clarkii</i>
Aub	0.00269	0.00109
AubX	0.00265	0.00108
Auv	0.00249	0.00098
Cev	0.00258	0.00107
El	0.00256	0.00092
Kew	0.00240	0.00081
MdR	0.00249	0.00077
So	0.00241	0.00118

Table S8: Mean Tajimas D calculated over 40kb windows

	<i>E. typhina</i>	<i>E. clarkii</i>
Aub	-0.334	-0.211
AubX	-0.366	-0.354
Auv	-0.369	0.102
Cev	-0.061	0.151
El	-0.307	-0.015
Kew	-0.093	0.903
MdR	-0.183	1.318
So	-0.282	-0.076

Table S9: Hybridization rates in two sympatric populations

	Dg_indv	HI_indv	Et_collected	Ec_collected	total_collected	Et_probes	Ec_probes	total_probes	Et_hyb	Ec_hyb	Et_hyb_prop	Ec_hyb_prop	total_hyb	total_hyb_prop
Cev	13	12	21	27	48	63	81	144	7	11	0.111	0.1358	18	0.1250
Auv	7	9	16	24	40	48	72	120	6	4	0.125	0.0555	10	0.0833

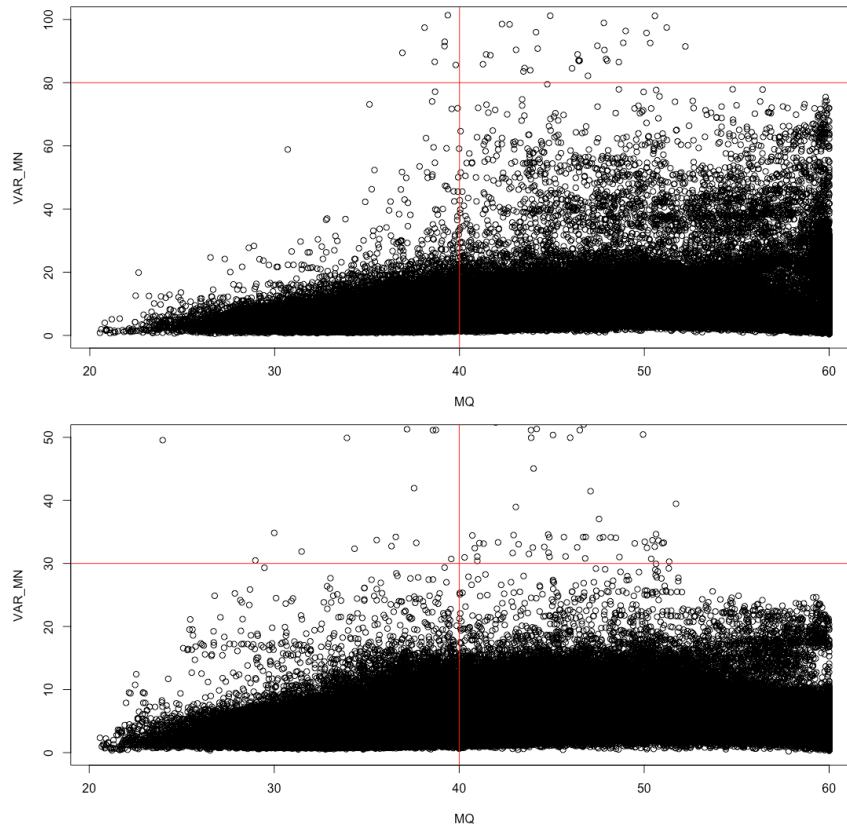
Columns specify the number of sampled individual plants [1,2], the number of stromata collected [3,4,5], the number of perithecia probes taken [6,7,8], the number of probes with hybrid asci [9,10], the proportion of hybrid asci [11,12], the total number of probes with hybrid asci [13] and the total hybrid proportion across both species [14]. Dg = *Dactylis glomerata*; HI = *Holcus lanatus*; Et = *E. typhina*; Ec = *E. clarkii*

Table S10: Individual plants assessed in Cev

plant_indv	stroma_samples	probes	hyb_asci	
Dg1		2	6	0
Dg2		1	3	0
Dg3		2	6	2
Dg4		2	6	1
Dg5		3	9	0
Dg6		1	3	0
Dg7		1	3	0
Dg8		2	6	3
Dg9		1	3	0
Dg10		1	3	1
Dg11		2	6	0
Dg12		2	6	0
Dg13		1	3	0
HI1		2	6	0
HI2		2	6	2
HI3		3	9	3
HI4		3	9	0
HI5		2	6	0
HI6		1	3	0
HI7		3	9	0
HI8		2	6	0
HI9		3	9	4
HI10		2	6	0
HI11		1	3	0
HI12		3	9	2
Total		48	144	18

Table S11: Individual plants assessed in Auv

plant_indv	stroma_samples	probes	hyb_asci	
Dg1		3	9	1
Dg2		3	9	0
Dg3		2	6	0
Dg4		2	6	0
Dg5		3	9	2
Dg6		2	6	3
Dg7		1	3	0
HI1		3	9	0
HI2		3	9	3
HI3		3	9	0
HI4		3	9	0
HI5		2	6	0
HI6		2	6	0
HI7		3	9	1
HI8		2	6	0
HI9		3	9	0
Total		40	120	10



Figures S1: Defining filtering cutoffs.

We retained SNPs with a MQ value > 40 (the Root Mean Square of the mapping quality of the reads across all samples), and with a relative variance in read depth < 80 in *E. typhina* (top) and > 30 in *E. clarkii* (bottom).

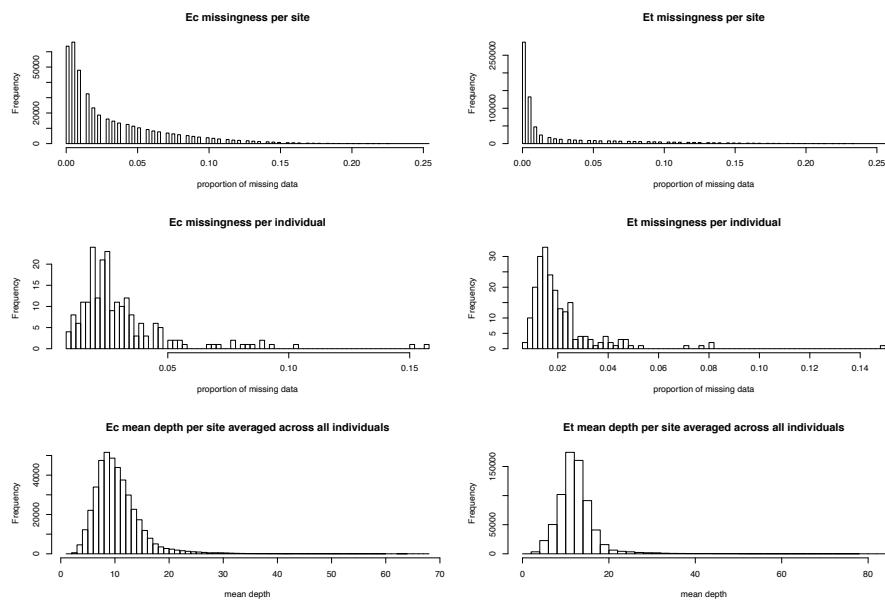


Figure S2: Summary of statistics after filtering.

For the *E. typhina* (right) and *E. clarkii* (left) SNP datasets, missingness per site, missingness per individual and mean depth per site are plotted as histograms. *E. clarkii* has more missing data and lower depth which is probably related to the slightly lower coverage with the larger reference genome and larger proportion of polymorphic regions where accurate mapping is more challenging.

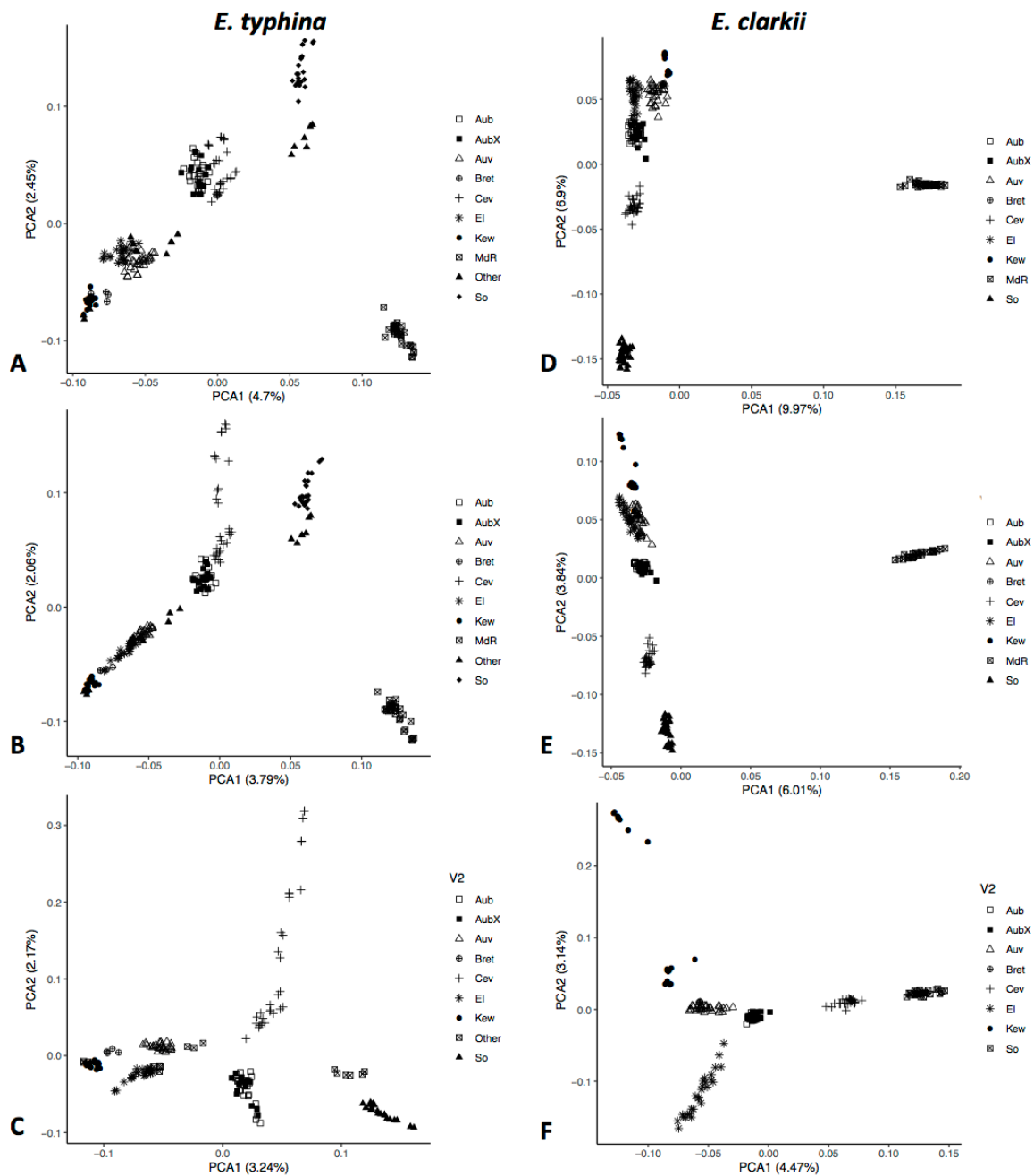


Figure S3: PCAs based on SNP datasets with different filtering criteria.

Percentage of variance explained by the first two principal components is shown in parentheses; symbols indicate sampling locations of populations. (A & D) complete datasets including Spanish populations (MdR); (B & E) LD pruned data with $\text{maf} \geq 0.05$ filter; (C & F) Same as B & E but with MdR removed.

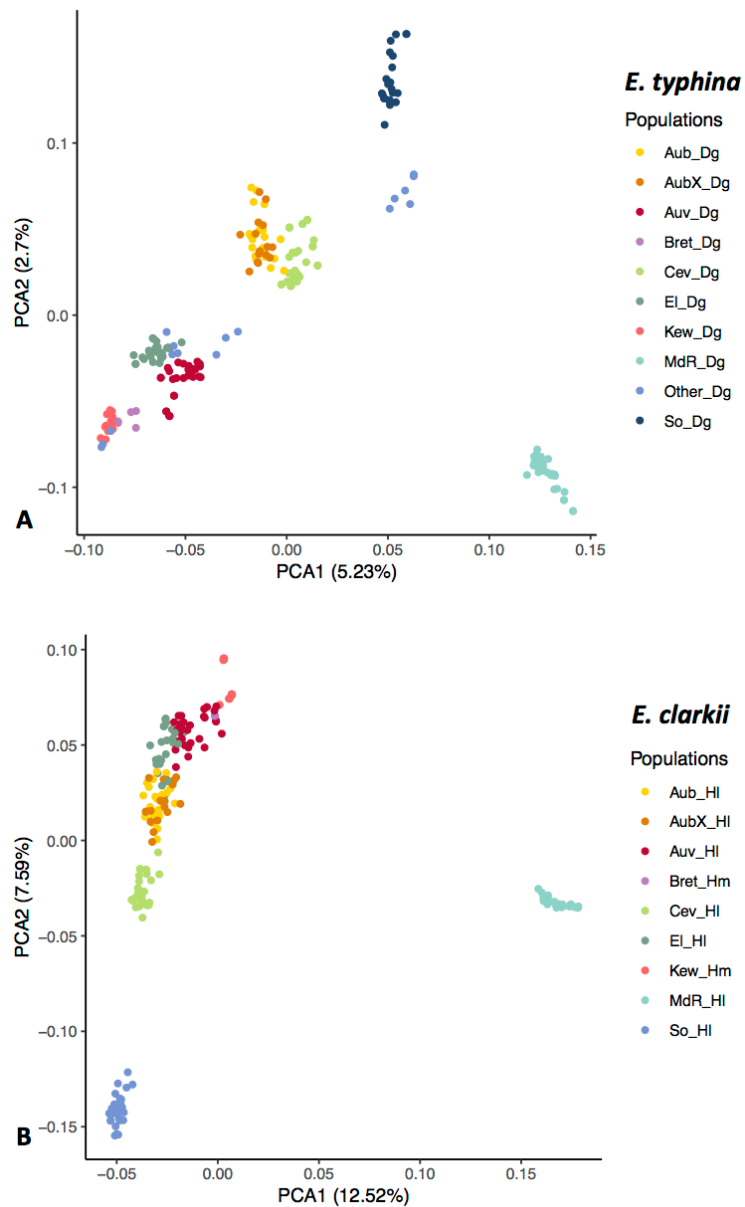
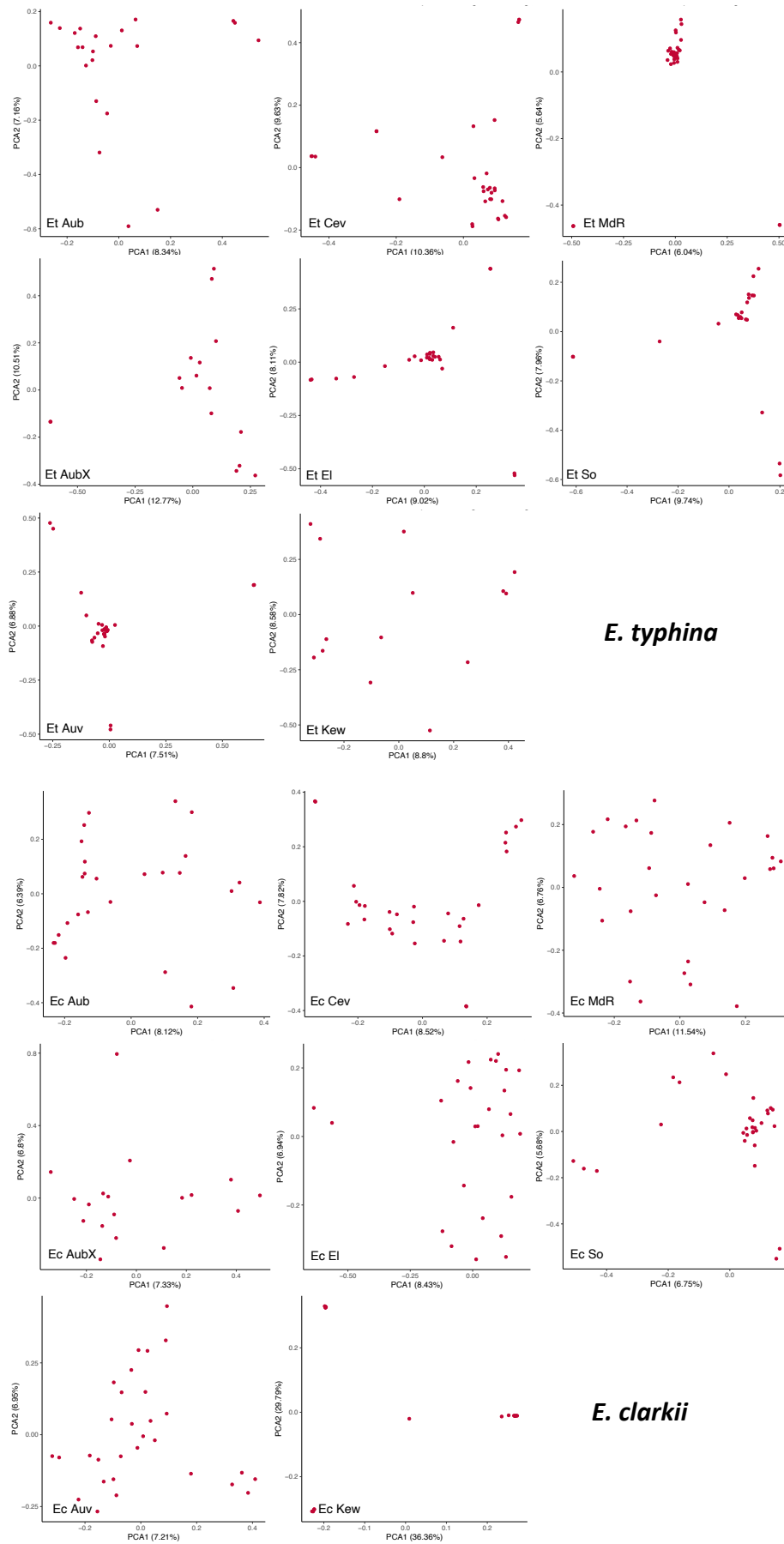
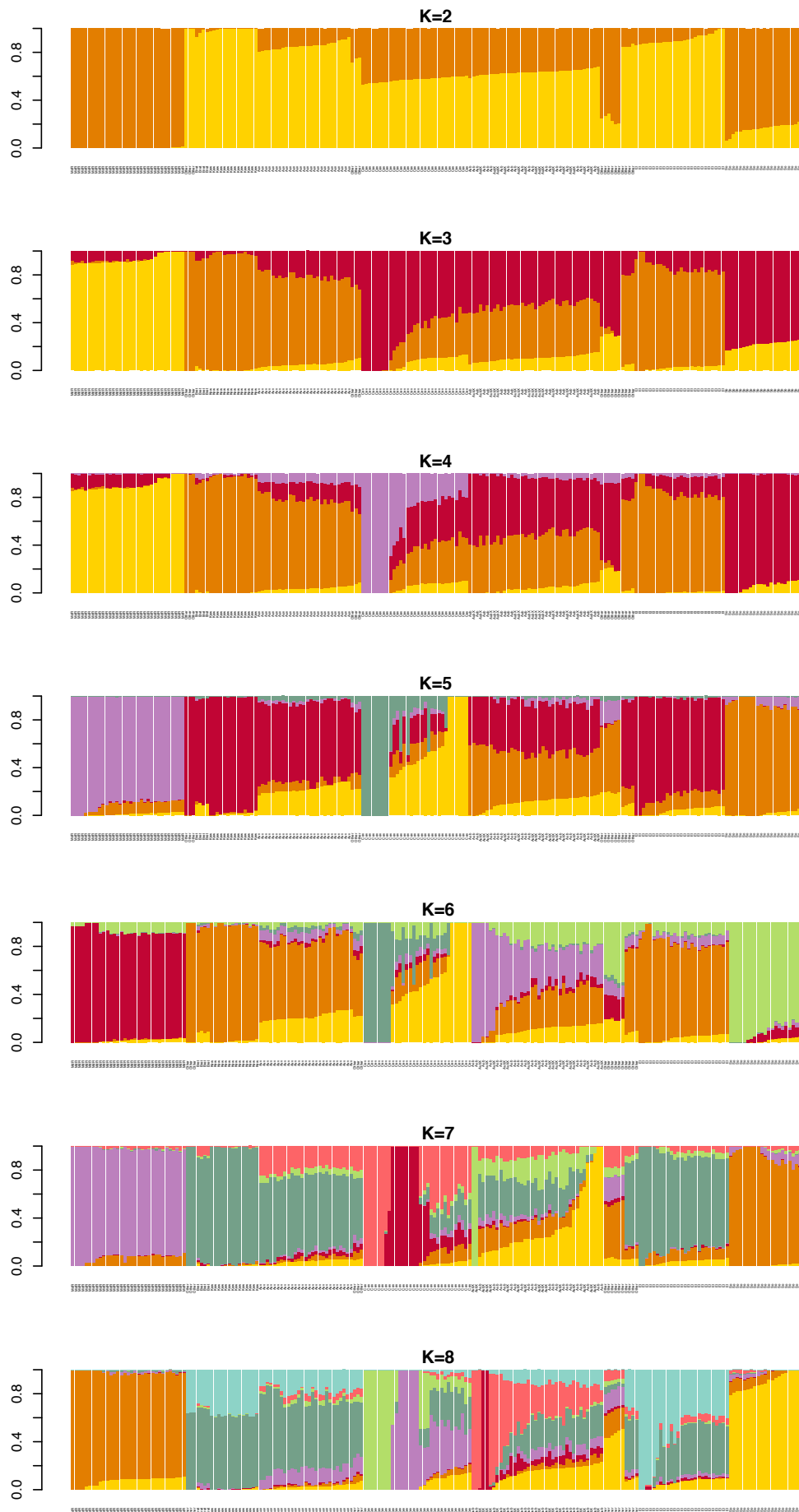


Figure S3: PCAs based on SNP datasets with no missing data.

Percentage of variance explained by the first two principal components is shown in parentheses; colors indicate sampling locations of populations



Figures S5 and S6: PCAs within populations.



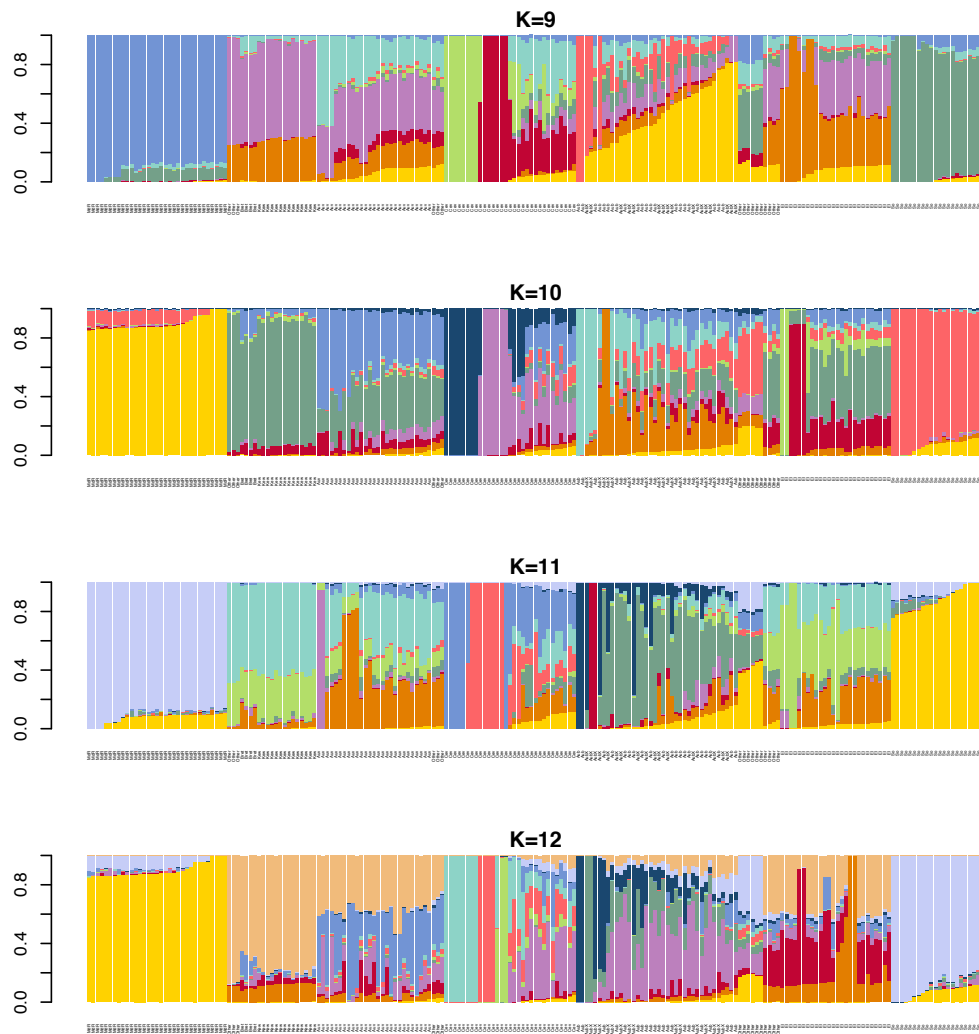


Figure S7: *E. typhina* STRUCTURE results for all Ks.

Bar plots of membership proportions (y-axis) as inferred by STRUCTURE for K=2 – K=12.



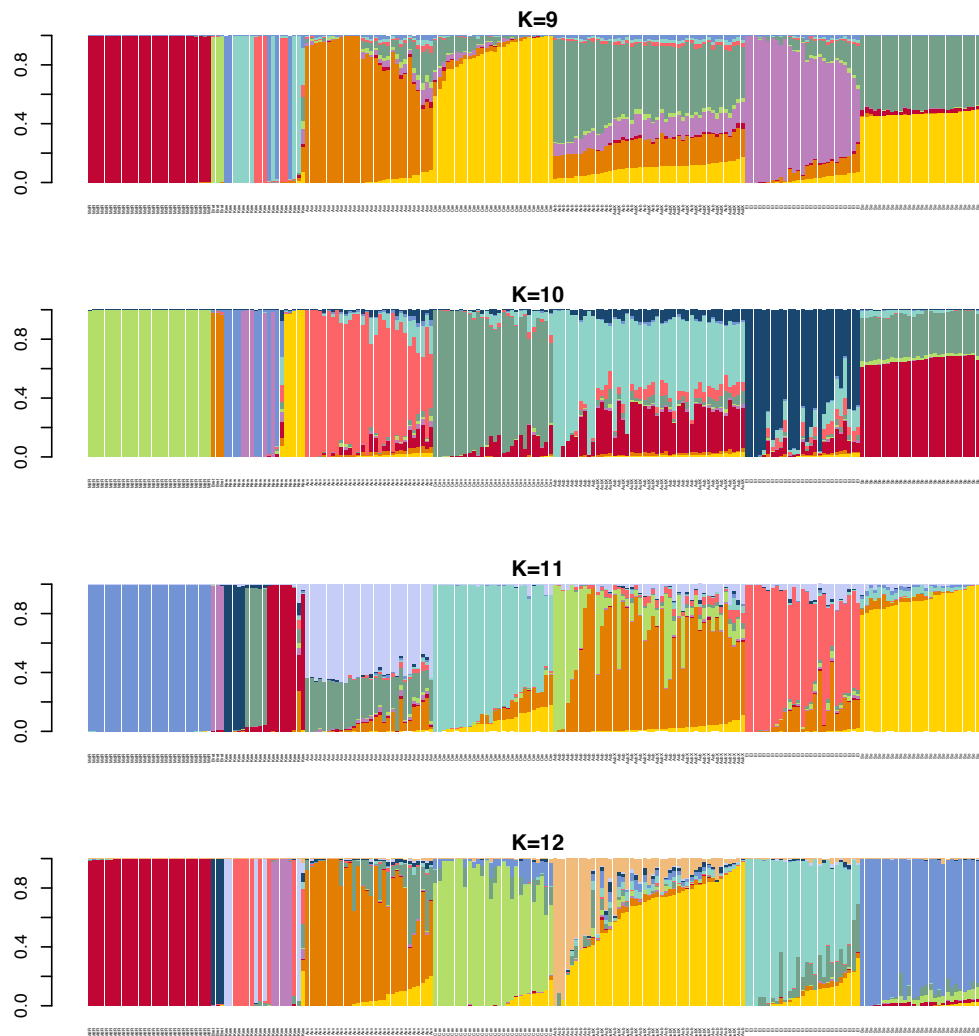


Figure S8: *E. clarkii* STRUCTURE results for all Ks.

Bar plots of membership proportions (y-axis) as inferred by STRUCTURE for K=2 – K=12.

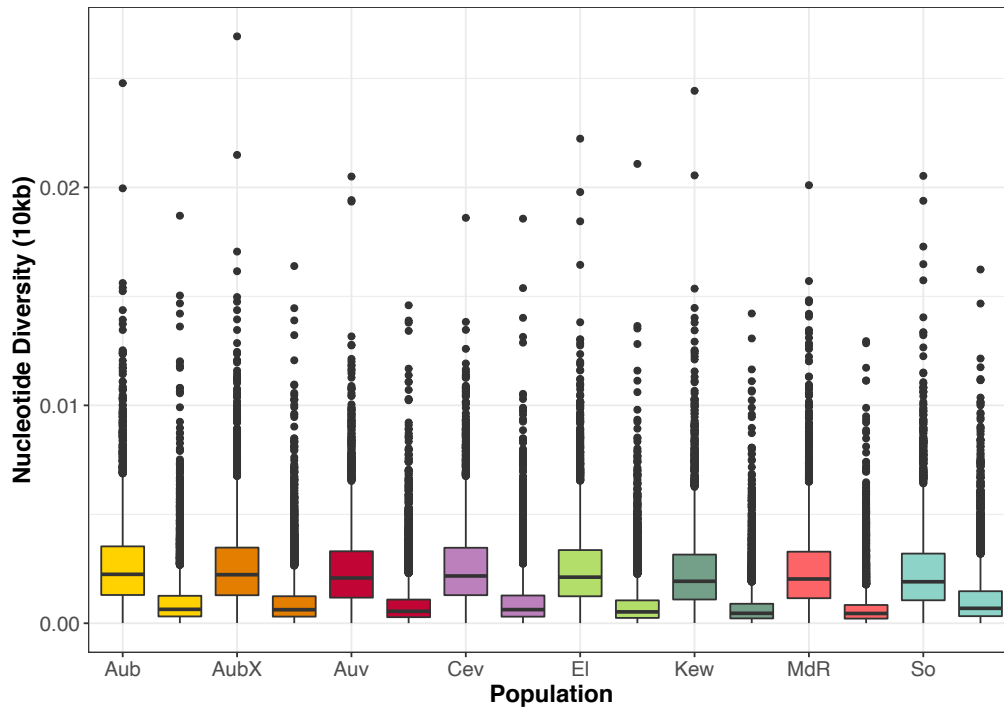


Figure S9: Nucleotide diversity per population.

Nucleotide diversity (π) calculated in 10kb sliding windows along genomes. Colors correspond to sympatric population pairs, with the first boxplot of each color showing *E. typhina* and the second *E. clarkii*. Lower and upper hinges correspond to the first and third quartiles (the 25th and 75th percentiles); whiskers extend to 1.5 times the inter-quartile range and outlier data beyond this value are plotted as individual points.

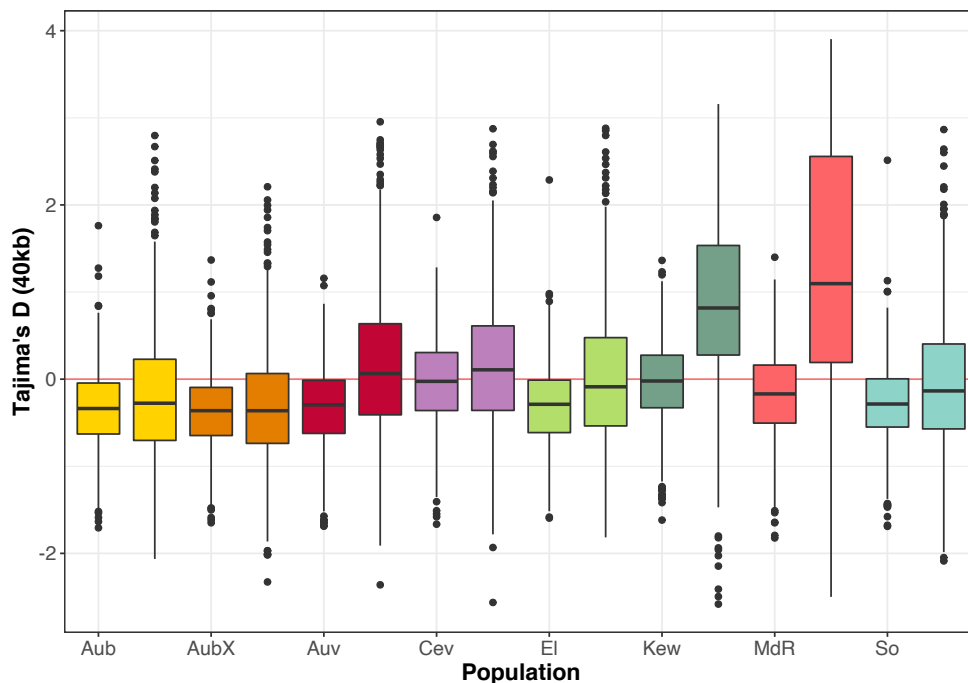


Figure S10: Tajima's D per population.

Tajima's D calculated in 40kb sliding windows along genomes. Colors correspond to sympatric population pairs, with the first boxplot of each color showing *E. typhina* and the second *E. clarkii*. Lower and upper hinges correspond to the first and third quartiles (the 25th and 75th percentiles); whiskers extend to 1.5 times the inter-quartile range and outlier data beyond this value are plotted as individual points.

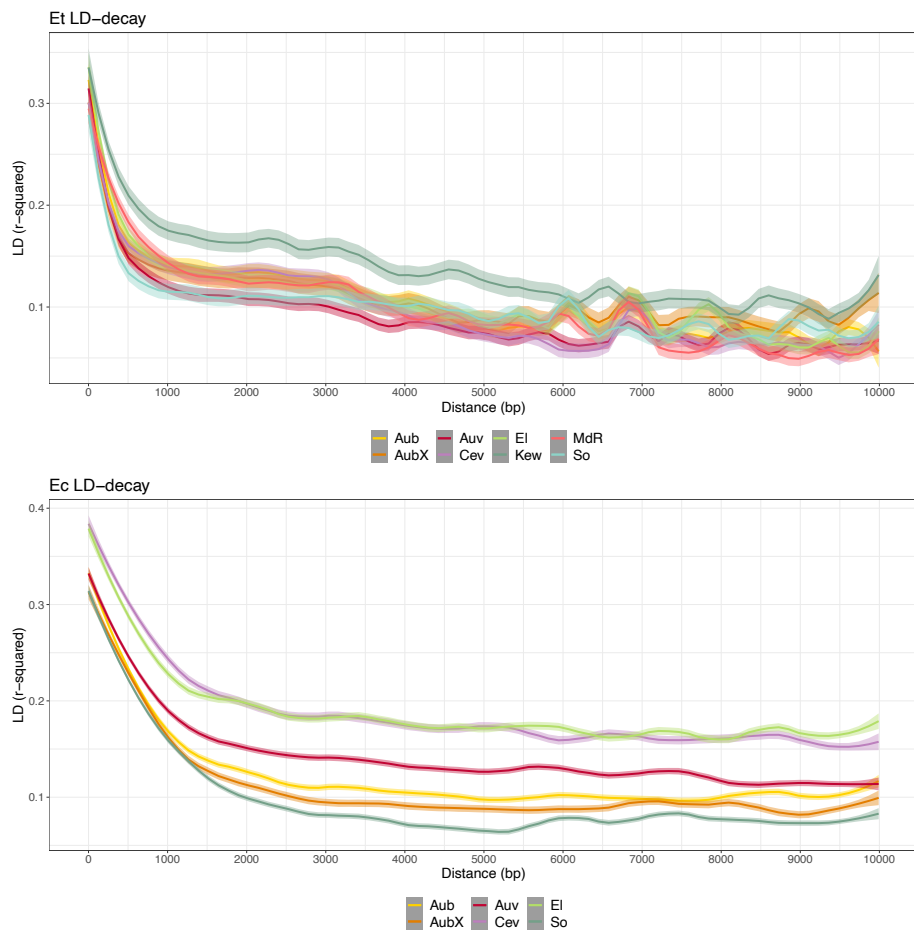


Figure S11: Decay of linkage disequilibrium (r^2) with physical distance in *E. typhina* (top) and *E. clarkii* (bottom) populations.

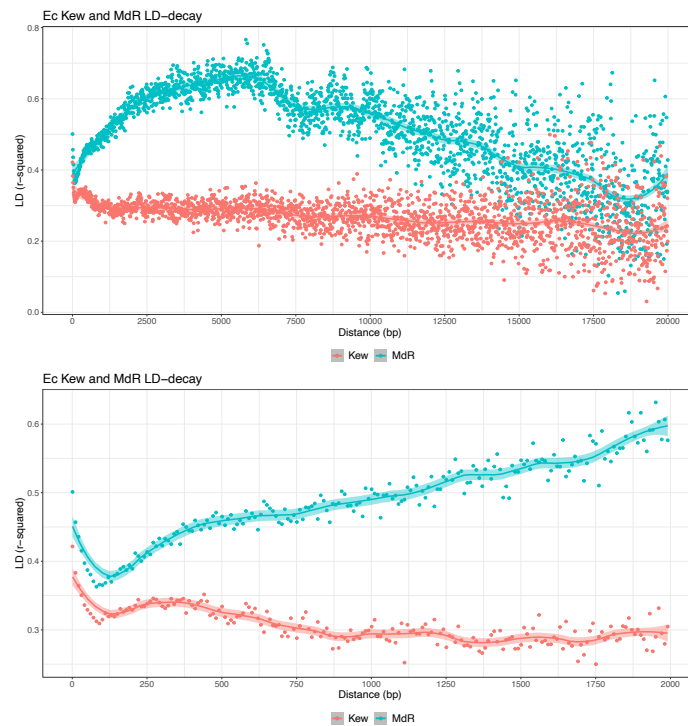


Figure S12: Decay of linkage disequilibrium (r^2) with physical distance in *E. clarkii* Kew and MdR. LD decayed very slowly in Kew (red) and MdR had high LD overall. The bottom plot is zoomed in on the first 2kb showing an initial LD decay within the first 100bp.

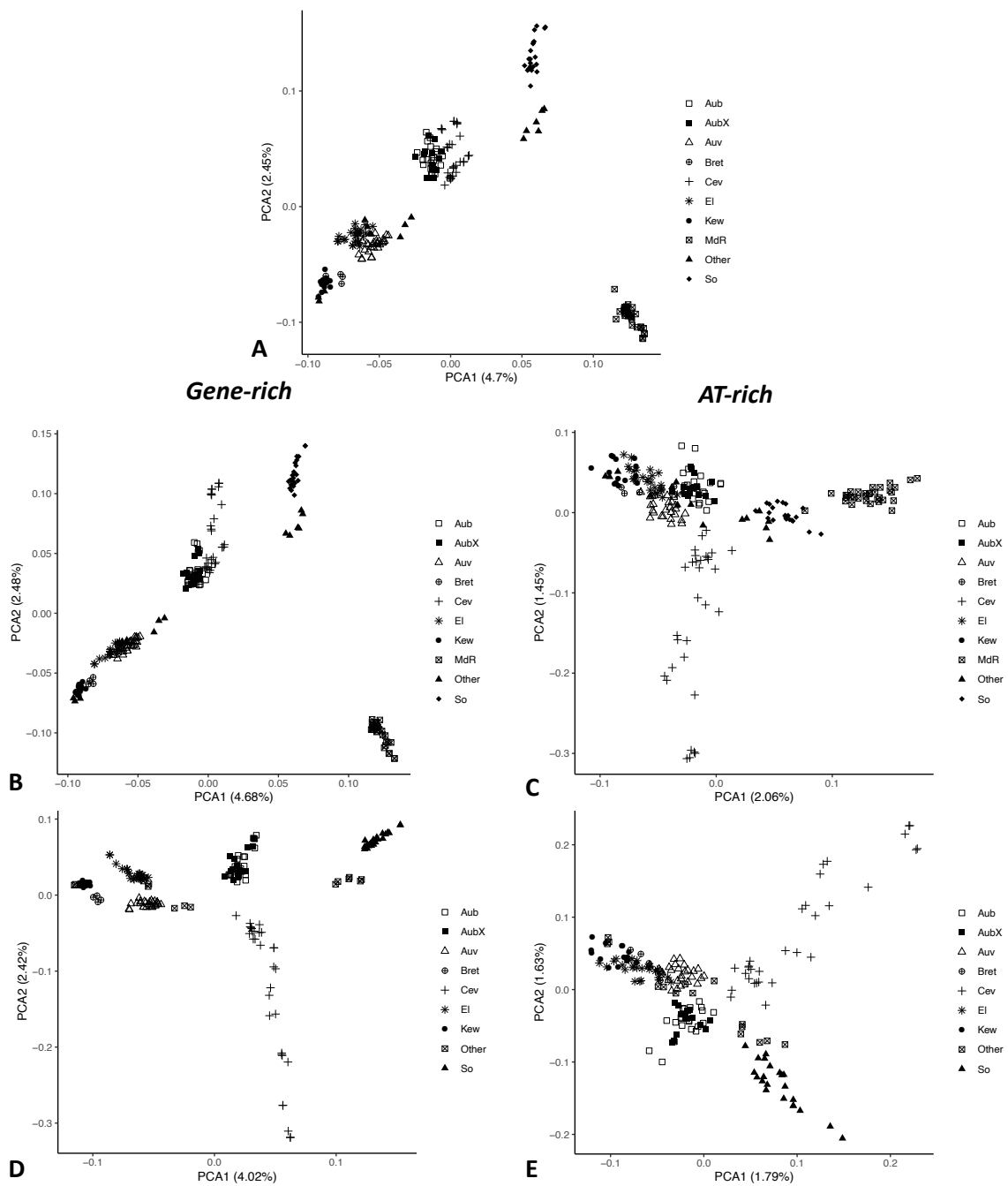


Figure S13: PCAs by genome compartments in *E. typhina*.

Analyses were performed with complete SNP datasets split by compartments. For reference the PCA for complete genome-wide SNPs including MdR is shown (A, same as Figure S3A). B & C: Compartment PCAs including MdR; D & E: Compartment PCAs excluding MdR. Percentage of variance explained by the first two principal components is shown in parentheses; symbols indicate sampling locations of populations.

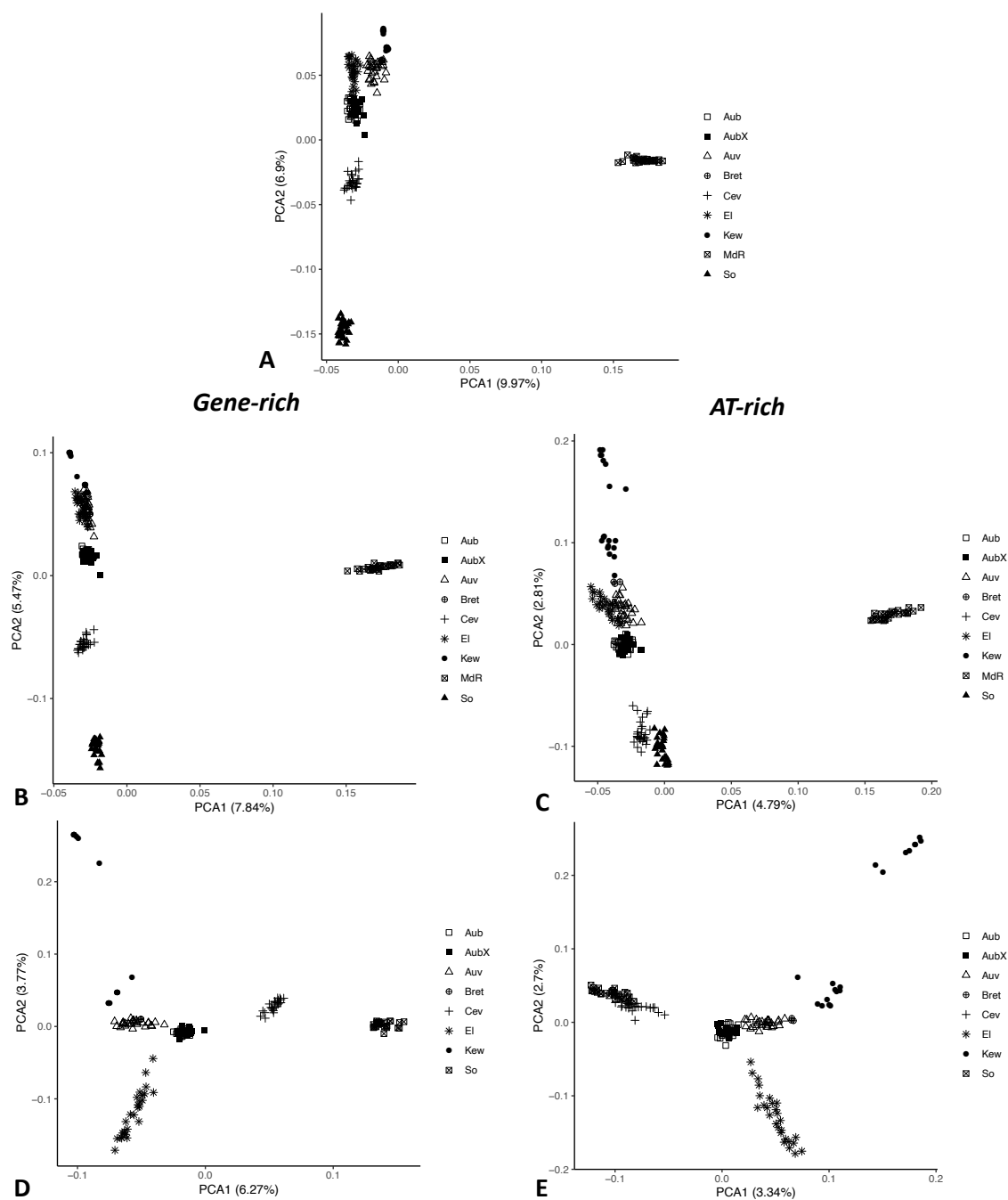
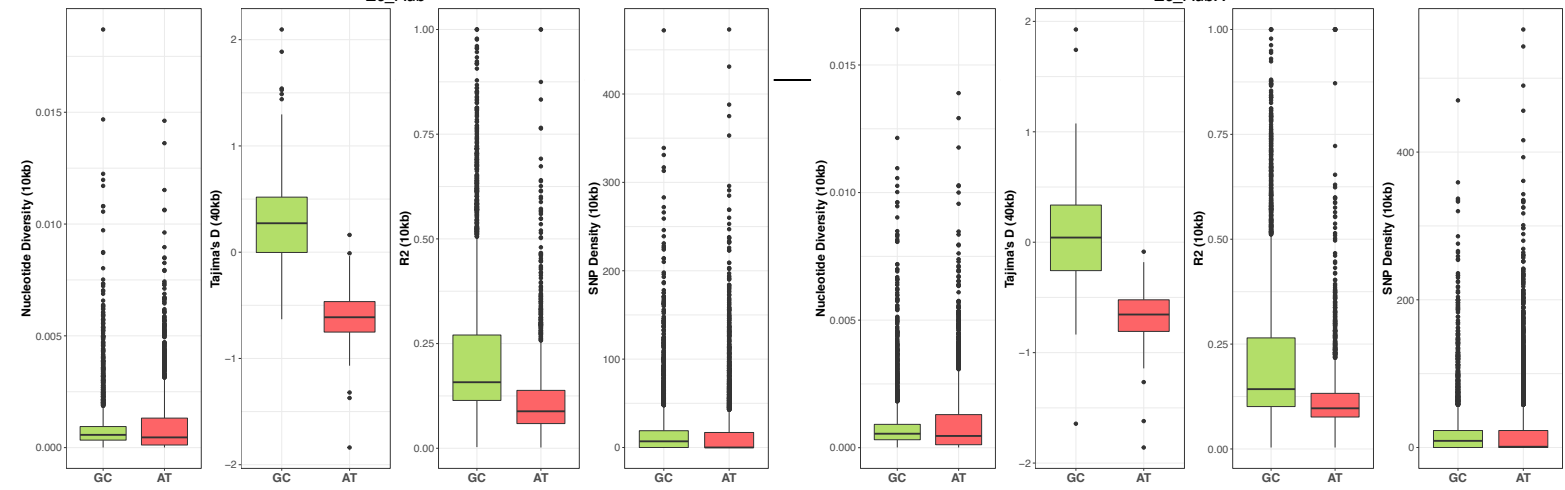
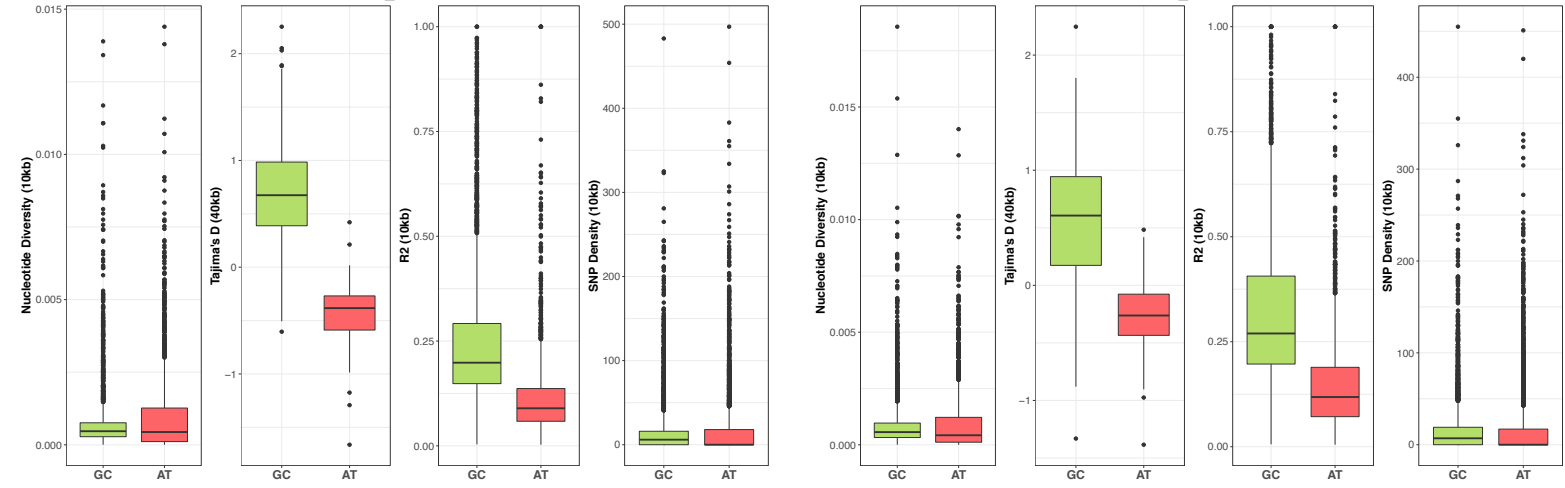
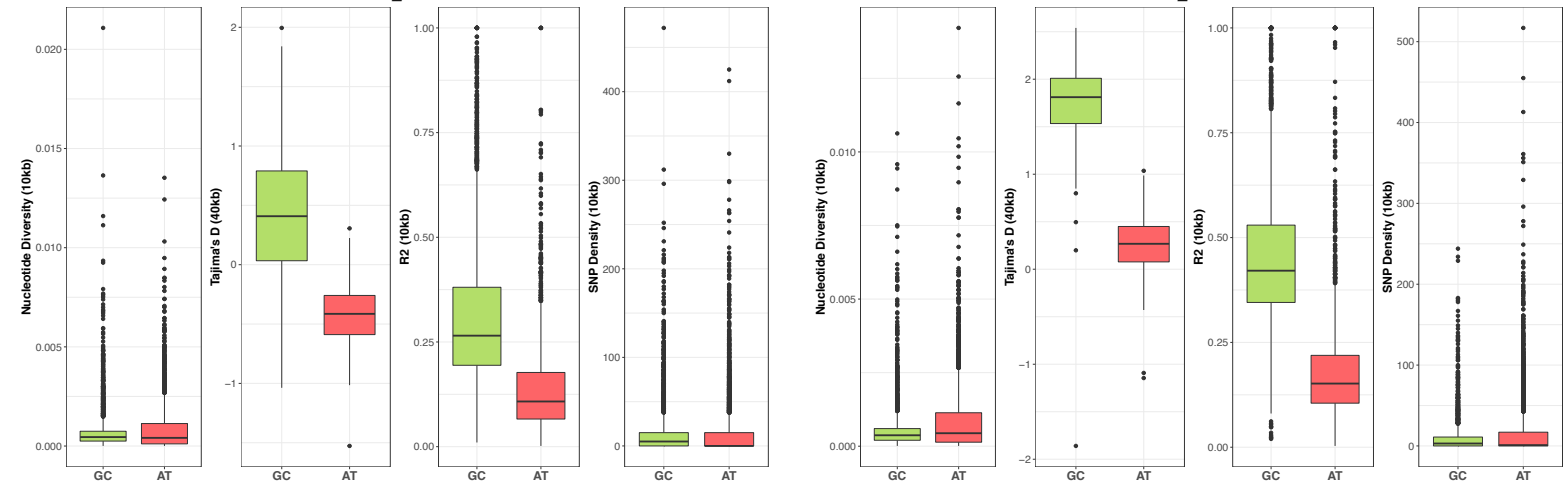
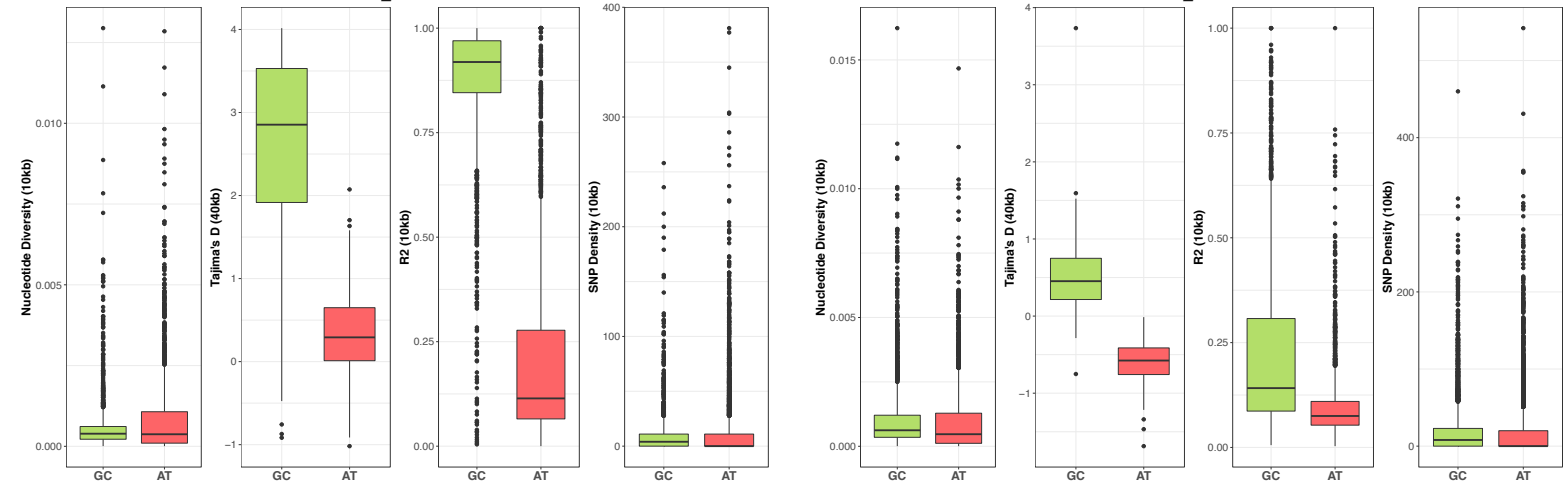


Figure S14: PCAs by genome compartments in *E. clarkii*.

Analyses were performed with complete SNP datasets split by compartments. For reference the PCA for complete genome-wide SNPs including MdR is shown (A, same as Figure S3D). B & C: Compartment PCAs including MdR; D & E: Compartment PCAs excluding MdR. Percentage of variance explained by the first two principal components is shown in parentheses; symbols indicate sampling locations of populations.



*Ec_Aub**Ec_AubX**Ec_Auv**Ec_Cev**Ec_El**Ec_Kew**Ec_MdR**Ec_So*

*Previous two pages:

Figures S15 and S16: Population summary statistics by genome compartments.

In each panel statistics for one population are shown as boxplots, green for the gene-rich compartment (GC) and red for the AT-rich compartment (AT). From left to right: Nucleotide diversity (π) in 10kb windows, Tajima's D in 40kb windows, LD r^2 in 10kb windows, SNP density in 10kb windows. Species and population IDs are indicated in the headers (Et = *E. typhina* and Ec = *E. clarkii*). Lower and upper hinges correspond to the first and third quartiles (the 25th and 75th percentiles); whiskers extend to 1.5 times the inter-quartile range and outlier data beyond this value are plotted as individual points.

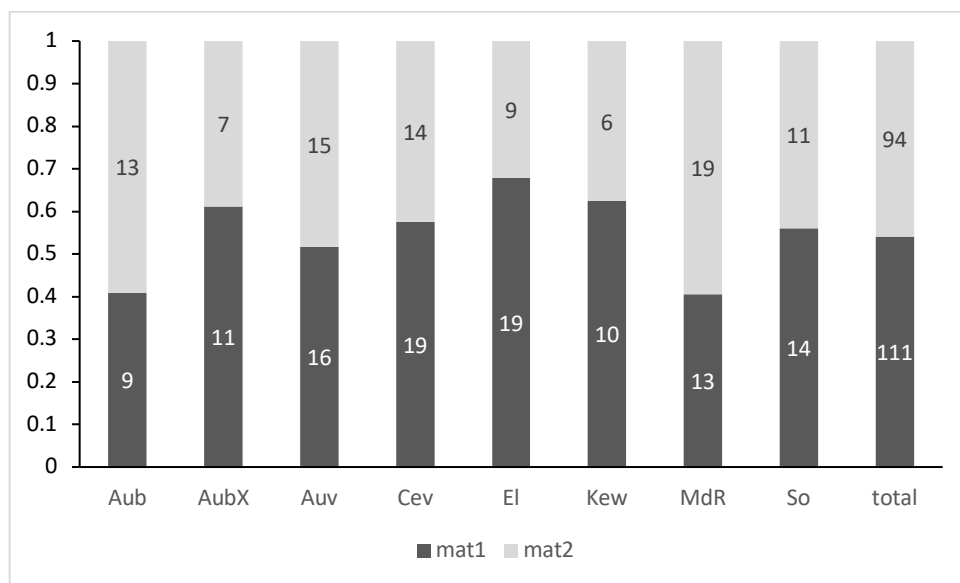


Figure S17: Proportion of mating types in *E. typhina* populations and across all sequenced genotypes. Numbers indicate how many individuals had mat-1 (dark grey) and how many had mat-2 (light grey).

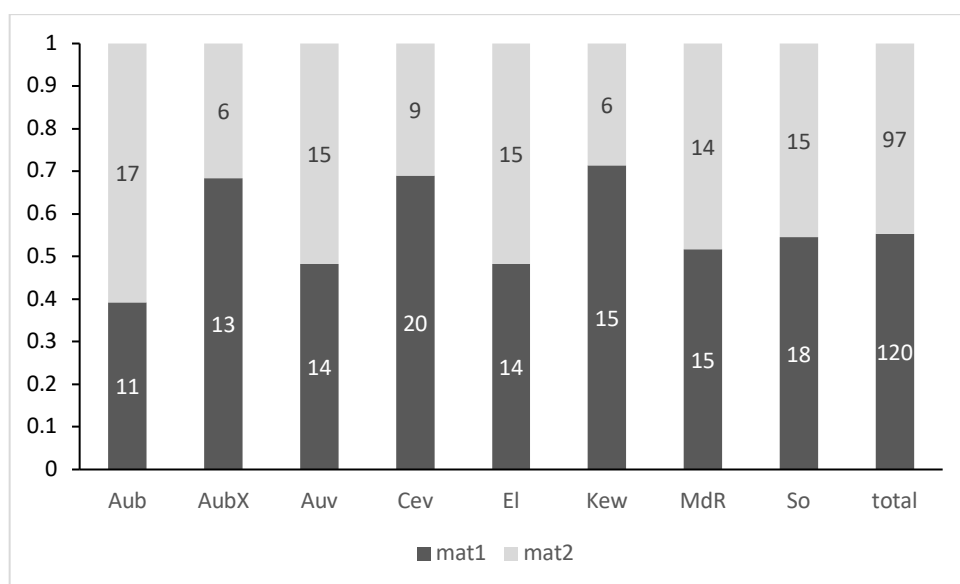


Figure S18: Proportion of mating types in *E. clarkii* populations and across all sequenced genotypes. Numbers indicate how many individuals had mat-1 (dark grey) and how many had mat-2 (light grey).

Genome-wide signatures of selection in natural plant pathogen populations

Artemis D. Treindl^a, Jessica Stapley^b, Adrian Leuchtman^a

^a Plant Ecological Genetics Group, Institute of Integrative Biology, ETH Zurich, 8092 Zurich, Switzerland

^b Plant Pathology Group, Institute of Integrative Biology, ETH Zurich, 8092 Zurich, Switzerland

Manuscript in preparation

ABSTRACT

Genetic variation in host-pathogen systems is generated and maintained by continuous antagonistic selection on traits that provide the host with mechanisms to become resistant to the pathogen and, reciprocally, on traits that aid the pathogen to infect the host and overcome resistance. These ongoing cycles of adaptation and counter-adaptation known as coevolution can drive rapid evolutionary change, and leave distinct molecular footprints of past and ongoing selection in the genome sequences of interacting organisms. Despite the increasing availability of tools needed to identify signatures of selection, the genetic mechanisms underlying coevolutionary interactions and the specific genes involved are still poorly understood, especially in heterogeneous natural environments. The complex spatial structure in these systems necessitates that the genomic signatures of adaptation are assessed across many populations, and ideally multiple ecologically similar species, in order to elucidate the coevolutionary processes at play. In this study we look for evidence of recent selection on pathogen genomes using a tractable eukaryotic study system: the *Epichloe* fungi. This genus includes highly host-specific fungal pathogens that cause sterilization of their host grasses, which generates strong selection pressure and an ideal setting to observe signatures of coevolutionary dynamics in the pathogen genomes. We performed selection scans using genome-wide SNP data from seven natural populations of two co-occurring *Epichloe* sister species specialized on different host grasses. We found widespread evidence of recent (and

ongoing) selective sweeps in both species, however, selective sweeps were more abundant in the species with larger effective population size (N_e). This is consistent with the expectation that the relative effect of selection versus genetic drift should be stronger in large- N_e populations. Sweep regions were not randomly distributed across pathogen genomes and often overlapped with highly polymorphic AT-rich regions supporting a role of these genome compartments in adaptive evolution. Many different loci were under selection in different populations suggesting that heterogeneous selection pressures generate locally adapted populations. Some selected loci were, however, shared across multiple populations suggesting shared selection regimes and a possible role of these loci in adaptation to the host at the species-level. We identified putative candidate genes located in shared sweeps encoding several known effector functions such as small-secreted proteins and cell-wall-degrading enzymes. By investigating the genomic signatures of selection across multiple populations and species this study reveals diverse genetic targets that may be involved in coevolutionary interactions and provides novel insights into the complexity of adaptive processes in a natural fungal plant pathogen system.

INTRODUCTION

Catching evolution in the act: Why host-pathogen systems are interesting evolutionary models to study adaptation

Identifying adaptive traits, elucidating their genetic basis and understanding the forces driving divergence of populations and species lie at the core of evolutionary biology research. Pathogens are a fascinating biological system to study adaptation over short evolutionary time-scales; not only are they ubiquitous in the natural world, pathogens are entangled in tight biotic interactions with their hosts which fosters continuous and rapid coevolutionary dynamics (Paterson et al. 2010, Kurtz et al. 2016). By definition, pathogens have negative effects on their host's fitness and exert strong selection pressure on traits that increase resistance in the host. In turn, pathogens must adapt to overcome host defenses and establish infection as they depend on colonizing their host for survival or to complete their reproductive cycle. This interaction manifests in

ongoing cycles of adaptation and counteradaptation driven by strong antagonistic selection and leading to rapid evolutionary change (Tellier et al. 2014, Papkou et al. 2019). Antagonistic selection between the host and pathogen is expected to result in two main types of coevolutionary dynamics: the *arms-race* model (also called *Escalatory Red Queen*) and the *trench-warfare* model (or *Fluctuating Red Queen*) (Woolhouse et al. 2002, Brockhurst et al. 2014, Ebert and Fields 2020). Arms-race dynamics are characterized by recurrent selective sweeps continuously driving new beneficial variants to fixation within a population. Because variants that are linked to the beneficial allele will also reach high frequency or become fixed (genetic hitchhiking), selection produces a local depletion of genetic variation in the genomic region that surrounds the locus, the signature of a selective sweep. Selective sweeps initially result in strong linkage disequilibrium (LD) across the affected regions and contain, as new mutations arise, an excess of low and high frequency variants (negative Tajima's D (Tajima 1989)). Trench-warfare dynamics on the other hand involve the continuous rise and fall in allele frequencies driven by negative (or antagonistic) frequency dependent selection whereby selection always acts against the most common host or pathogen genotype, resulting in maintenance of polymorphism at the coevolving locus. Because none of the different alleles are expected to become fixed this dynamic is detectable as an excess of intermediate frequency variants (positive Tajima's D) around the genomic region under selection: the genomic signature of balancing selection (Charlesworth 2006).

The arms-race and trench-warfare coevolutionary models represent extreme scenarios of what is in reality a continuum of possible allele frequency dynamics observable at the genomic level (Woolhouse et al. 2002, Brown and Tellier 2011, Tellier et al. 2014, Ebert and Fields 2020). Importantly, while these models describe temporal evolutionary dynamics at a given locus, time is not the only dimension impacting the strength of selection. Selection pressure also varies through space in heterogeneous natural environments creating spatially structured populations where coevolution with local host genotypes can lead to asynchrony in evolutionary dynamics across populations and local adaptation (Woolhouse et al. 2002, Thompson 2005). Theory suggests that how antagonistic selection ultimately plays out in terms of evolutionary dynamics largely depends on population structure and levels of gene flow among host and pathogen populations (Gandon and Michalakis 2002, Gandon and Nuismer 2009).

There is now a need for these theoretical predictions to be “carried to the field” (Ebert 2018), using suitable host-pathogen systems evolving in natural environments (see Burdon and Thrall 2009, Tack et al. 2012, Jousimo et al. 2014), to move beyond experimental evolution studies in laboratories. To further increase our understanding of the processes and targets of antagonistic selection and the nature of antagonistic coevolutionary dynamics acting in spatially and genetically complex biological systems, analyses should ideally include multiple distinct populations. This enables an assessment of the underlying population genetic parameters and thus allows evolutionary processes to be put into context. A particularly powerful opportunity arises when analyses of many populations can be extended to multiple species because differences among populations can be compared providing additional power for the inference of adaptive relevance.

Fungal pathogens as model systems to study antagonistic coevolution

Fungal plant pathogens with their small tractable eukaryotic genomes, microbial lifestyle and large population sizes provide ideal and experimentally manipulable model organisms for investigating the genetic basis of host-driven adaptation (Gladieux et al. 2014). However, depending on the life histories of both the fungus and its host plant, for example the tightness of the host-pathogen interaction (i.e. the degree of specificity or host range), and the level of fitness costs, the mode and speed of coevolutionary dynamics is predicted to vary (Barrett et al. 2008, 2009, Papkou et al. 2019). Among the most well suited biological systems to detect molecular signatures of rapid adaptation are specialized fungi that infect a single host species and bring about a severe reduction in plant fitness, such as pathogens that kill or sterilize their host (Ashby and Gupta 2014). Antagonistic selection is expected to be particularly strong in these systems fostering arms-race dynamics via selective sweeps (Tellier et al. 2014, Persoons et al. 2017). Furthermore, extensive research centered around the rapid surmounting of plant resistance by fungal pathogens in agricultural ecosystems has supplied valuable insights as to which gene functions are important in the antagonistic interaction (McDonald and Stukenbrock 2016). In order to colonize the host, suppress or avoid defense responses, and ultimately receive nutrients from the plant, fungal pathogens produce a specialized set of secreted proteins, so called effectors. Many known effectors are small secreted

cysteine-rich proteins and underlying effector genes are characterized by high expression in planta (reviewed in Lo Presti et al. 2015, van der Does and Rep 2017). Another group of effector genes important for fungal plant pathogenesis are those involved in plant cell wall degradation (Kubicek et al. 2014). These genes encode cell wall degrading enzymes (CWDEs), larger secreted proteins to degrade cellulose, hemicellulose or pectin and have also been shown to be specifically expressed by pathogenic fungi during plant infection. As key players in the pathogenic interaction, effector genes are common targets of selective sweeps and among the most rapidly evolving genes in pathogen genomes (Aguileta et al. 2009, Plissonneau et al. 2017). However, likely due to their rapid evolution, many effector genes lack conserved domains and homologs in closely related species (Franceschetti et al. 2017) making their prediction challenging without the availability of a high quality reference genome and transcriptomic data.

Not only has the analysis of pathogen genomes shed light on the targets of antagonistic selection and led to the identification of genes underlying rapid adaptation, it has also revealed structural characteristics of the genomes that promote said rapid evolution. A number of fungal plant pathogens possess compartmentalized genomes with a conserved gene-rich compartment and a dynamic repeat-rich compartment largely devoid of genes (Croll and McDonald 2012). This bipartite genome organization has been referred to as the “two-speed genome” model because the compartments are thought to experience different evolutionary rates (Raffaele and Kamoun 2012, Dong et al. 2015). Repeat-rich regions are hot spots for duplication, deletion and recombination as a result of transposable element activity rendering these regions highly polymorphic and may be targeted by DNA defense mechanisms such as repeat-induced point mutation (RIP) inducing locally high mutation rates (Stukenbrock and Croll 2014, Möller and Stukenbrock 2017, Seidl and Thomma 2017). Effector genes are often localized in or near these dynamic genome compartments where they may experience faster rates of sequence evolution (e.g. Schirawski et al., 2010; Grandaubert et al., 2014; Dong, Raffaele and Kamoun, 2015; Faino et al., 2016). Together, these observations have highlighted the role of such dynamic compartments in the adaptive evolution of fungal pathogens: They provide a genomic environment rich in genetic novelty that natural selection can act on, thereby driving rapid adaptation (Eschenbrenner et al. 2020).

To date most studies investigating the signatures of selection in fungal pathogen genomes have focused on agricultural systems, and selective sweeps targeting effector genes have been identified as causes of devastating epidemics in crop plants (see McDonald and Stukenbrock 2016). Conversely, knowledge about the impact and targets of selection in natural populations of fungal plant pathogens are still much more elusive, at least in part because fewer high quality reference genomes are available. Additionally, while growing evidence emphasizes the role of dynamic genomic compartments for adaptive evolution in fungal pathogens (Stukenbrock and Croll 2014, Torres et al. 2020) there is a necessity for studies in suitable biological systems to investigate the link between genome organization and adaptation using population genomic approaches. With the large scale study of two closely related species and seven populations each, we aim to fill this gap by investigating the footprints of recent selection in the two-speed genomes of a natural fungal plant-pathogen system, the sexual members of the genus *Epichloe* specialized on wild grasses.

***Epichloe* study system and research objectives**

The pathogenic members of the ascomycete genus *Epichloe* provide an ideal biological system to detect signatures of selection and study adaptive evolution in fungi with compartmentalized genomes. These haploid filamentous ascomycetes infect different species of grasses in natural and semi-natural ecosystems and reproduce sexually through pervasive outcrossing while sterilizing the host plant (White and Bultman 1987, Chung and Schardl 1997, Clay and Schardl 2002). The sterilizing lifestyle and high host-specificity of *Epichloe* is expected to generate strong selection pressure, an ideal precondition to observe signatures of coevolutionary dynamics in the pathogen genomes (Ashby and Gupta 2014, Tellier et al. 2014). Furthermore, rapid adaptation and divergence at both the population and species level may be fueled by their specialized two-speed genome organization with conserved gene-rich and dynamic AT-rich compartments (see Chapter I of this thesis). In a parallel population genomic study based on genome-wide analysis of SNPs, we have investigated seven natural sympatric populations of two *Epichloe* sibling species, *Epichloe typhina* and *Epichloe clarkii*, and show that natural populations harbor substantial genetic variation and high recombination rates (see Chapter II of this thesis). Furthermore, populations showed

clear evidence of genetic structure that may reflect coevolutionary dynamics acting at the population level, driving divergence and generating locally adapted pathogen populations (see Croll and McDonald 2016). A comparison of the two species showed, however, that while they are ecologically very similar, they differ in population demography which may be related to different dispersal abilities; *E. clarkii* populations have lower genetic diversity suggesting smaller N_e and exhibit a stronger effect of isolation by distance indicating lower levels of gene flow among populations, compared to *E. typhina* which has greater genetic diversity, larger N_e and higher rates of geneflow. The relative strength of selection versus stochastic processes such as genetic drift affecting evolutionary change in genomes largely depends on the N_e of a given population or species (see Charlesworth 2009), and we therefore expect that genomic signatures of selection may be more prevalent in *E. typhina* compared to *E. clarkii*.

In this study, we used population genomic datasets of *E. typhina* (n=191) and *E. clarkii* (n=208) sampled from seven locations across Western Europe where the two species occur in sympatry. We used linkage disequilibrium (LD)-based and site frequency spectrum (SFS)-based selection scan methods to detect signatures of selective sweeps and uncover loci under recent selection within and between populations. As a first objective (1) we considered the number and distribution of selected loci to address the following questions: (i) How common are selective sweeps in *E. typhina* and *E. clarkii*? (ii) Are different regions under selection in different populations or are selected regions shared across populations? (iii) What inferences can be drawn from these patterns about coevolutionary dynamics at play across naturally structured populations experiencing geneflow? (iv) Are selective sweeps associated with AT-rich regions and hence, does the dynamic genome compartment contribute to adaptive evolution in *Epichloe*? A second objective (2) focused on the gene content within selected regions to answer the questions: (v) What are the specific functions of genes involved in adaptive evolution? (vi) Is recent selection targeting candidate genes likely involved in host-pathogen interactions such as putative effector genes or genes specifically expressed *in planta*? The population genomic resources and high quality reference genomes of these two *Epichloe* species provide a unique opportunity to identify candidate genes underlying coevolutionary host-pathogen interactions, investigate their genomic distribution, and

to compare the genome-wide patterns of selection between ecologically similar, recently diverged but highly specialized fungal taxa in a natural environment.

MATERIALS AND METHODS

A population genomic dataset of sympatric *Epichloe* sibling species

The systemic endophytes *E. typhina* and *E. clarkii* are two pathogenic sibling species specialized on distinct grass species: *E. typhina* causes choke disease on *Dactylis glomerata* and *E. clarkii* on *Holcus lanatus* and *H. lanatus* respectively. We have previously generated a population genomic dataset for each of the two species comprising whole-genome SNP data from haploid isolates sampled from seven natural sympatric populations in Europe between 2016 and 2018 (Supplementary Tables S1 and S2). The sampling, SNP genotyping and population genetic structure across these populations was described in Chapter II of this thesis. In brief, fungal strains were isolated from fruiting bodies encasing infected flowering tillers of host grasses, genomic DNA was extracted from pure liquid cultures and Illumina paired-end sequencing of 2x150-bp read length, and an insert size of 330 bp was performed yielding on average 6x coverage per isolate. The sampling included isolates (16 *E. typhina* and 18 *E. clarkii*) collected in 2005 from a population in Western Switzerland also sampled in 2018. The 2005 samples showed high genetic congruence with the more recent samples, but we have treated these as a separate population (AubX) in analyses and only report results for the 2018 population unless indicated otherwise. In total we analyzed 191 *E. typhina* isolates (n=14-33) and 208 *E. clarkii* isolates (n=18-30).

Variant detection and data polarization

The mapping, variant calling and filtering procedure is described in detail elsewhere (Chapter II of this thesis). In brief, we used BWA (Li and Durbin 2009) to map reads against the chromosome-level reference genomes of *E. typhina* (Ety_1756, BioProject ID [PRJNA533210](#)) and *E. clarkii* (Ecl_1605_22, BioProject ID [PRJNA533212](#)) and the Genome Analysis Toolkit (GATK, McKenna et al. 2010) for variant calling. After filtering we retained 658,021 biallelic SNPs in the *E. typhina* dataset and 400,033 SNPs in the *E.*

clarkii dataset. To determine ancestral allele state in our focal species we included high coverage sequencing data from the closely related outgroup species *E. festucae* (strain Fl1 ([SRX5056454](#), 2x250 Illumina paired-end, 550 bp insert size, approximately 87x coverage see (Winter et al. 2018)). Only biallelic SNP positions were included, outgroup genotypes were extracted and subsequently used to annotate population vcf files with the ancestral allele information using bcftools (Li et al. 2009), before removing the outgroup individual from the datasets. We were able to assign ancestral alleles for 480,102 SNPs in *E. typhina* and 219,085 SNPs in *E. clarkii* (72% and 55% of the total number of SNPs in *E. typhina* and *E. clarkii* respectively). We annotated and predicted the effect of SNPs using SnpEff 4.3m (Cingolani et al. 2012).

Research objective 1: Evidence of recent selection

Genome scans for signatures of selective sweeps within populations

To detect target regions of recent selection in *E. typhina* and *E. clarkii* populations we used two different selection scan approaches that search for distinct sweep signatures. We performed selection scans separately for each of the seven populations using only SNPs with known ancestral states and we used conservative percentile thresholds (99.9th percentile) of the test statistics distribution to identify selective sweep regions with the strongest signature in each approach. As a first approach, we used extended haplotype homozygosity (EHH), a measure of linkage disequilibrium, which is expected to be increased across haplotypes that contain an allele under selection. These blocks of high EHH are long if the selective sweep was strong and recent, however, the length of haplotype blocks will break down gradually when new mutations and recombination reintroduce diversity. We detected long haplotype blocks by comparing the EHH of the ancestral allele and the derived allele at each SNP locus calculating the integrated haplotype scores (iHS) implemented in the R package REHH v3.0.1 (Gautier et al. 2017). Calculations were performed for each chromosome separately using only SNPs for which the ancestral allele was inferred (polarized data), and with minor allele frequency >5% (min_maf=0.05). The 99.9th percentile of the distribution of absolute iHS values was used as an outlier threshold to identify SNPs under recent selection. Candidate regions were then determined by looking for clusters of outlier SNPs within sliding windows of length 10kb with 5kb overlap for *E. typhina* and 20kb with 10kb overlap for *E. clarkii*

respectively. Different window sizes were used to account for the different SNP-densities in the two species and correspond to windows of 100 SNPs on average in both species. Windows with at least two markers displaying an absolute value of the statistic above the threshold were classified as candidate sweep regions and neighboring windows containing outlier SNPs were merged into one candidate region. We henceforth refer to candidate regions detected by this method as iHS-sweeps.

In a second approach, we detected selective sweeps based on changes in the shape of the site frequency spectrum. As a selected allele reaches high frequency in a population, linked variants nearby also sweep towards fixation (genetic hitchhiking), creating a region of reduced genetic diversity with derived alleles at high frequency. New mutations then gradually appear on this homogenous background but these alleles will initially be at low frequencies creating a surplus of rare alleles in the site frequency spectrum (SFS). This shift constitutes a signature of positive selection, which we analyzed using the software SweeD v3.2.1 (Pavlidis et al. 2013). SweeD implements a composite likelihood ratio (CLR) test based on the SweepFinder algorithm (Nielsen et al. 2005). The CLR test compares the likelihood of the observed polymorphism pattern of a sequence (in our case chromosome) under a null model without a selective sweep to its likelihood under a model of a selective sweep. The null model is based on the global observed SFS along the chromosome and not on a standard neutral equilibrium model, rendering this test more robust to demographic events such as bottlenecks or population expansion (Nielsen et al. 2005). We excluded sites that were monomorphic (invariant) within each population and computed CLR on the unfolded SFS of each chromosome using a grid size of 5000 in *E. typhina* and 3000 in *E. clarkii*. The grid size specifies the number of positions in the sequence where the CLR will be computed and again we chose different window sizes to account for the differences in SNP density between the two species - approximate window sizes of 1kb (*E. typhina*) and 2kb (*E. clarkii*) provide on average 20 SNPs per window. We expect selected haplotypes to be relatively short because recombination rates are high and LD decays quickly in *Epichloe* and therefore chose small window sizes, although computationally more intensive. Larger grid size (= small window size) increases the number of positions evaluated along the sequence and can lower the chances of missing the signature of selection (Alachiotis and Pavlidis 2018). We repeated the analyses using smaller grid sizes with CLR tests

computed every 100th SNP and every 200th SNP on average, and this decreased likelihood scores but did not increase the size of outlier sweep regions indicating that small windows are appropriate to detect sweep signatures. The 99.9th percentile of the distribution of the CLR statistic was used as an outlier threshold to identify SNPs under positive selection. We concatenated adjacent outlier positions into one region whenever they were <5kb apart and extended candidate sweep regions based on genome-wide estimations of LD decay to account for potential blocks of high linkage disequilibrium. In *E. typhina*, LD decayed below 0.2 within 1kb in all populations, so we added 1kb at each end of the region. In *E. clarkii*, we added 1kb to each end in Aub, AubX, Auv and So, 2kb in El and Cev, and 5kb in MdR and Kew based on the slower LD decay in those populations. We henceforth refer to candidate regions detected by this method as CLR-sweeps.

Genome scans for signatures of divergent selection between populations

In addition to within-population EHH analyses we also performed cross-population EHH tests (XP-EHH), again implemented in the R package REHH v3.0.1 (Gautier et al. 2017). This selection scan compares haplotypes across pairs of populations and detects selected alleles that are nearly fixed in one but not both populations. We performed scans for all population pairs, however, for simplicity we present only candidate regions from a subset of three populations, Aub, Auv and El, which had among the smallest geographic distances between them and showed relatively low levels of genetic differentiation based on genome-wide averages of pairwise F_{ST} (*E. typhina*: 0.06, 0.06 and 0.05; *E. clarkii*: 0.14, 0.13 and 0.19 for population pairs Aub-Auv, Aub-El, Auv-El respectively; see Chapter II of this thesis). We considered these population pairs to be among the most likely to experience gene-flow therefore giving us increased power to detect loci under divergent selection. Regions under divergent selection are expected to be protected from gene-flow and show reduced variation compared to the genomic background where gene-flow homogenizes polymorphism across populations. We used the same conditions as outlined above for the within population iHS scans to determine outliers and define candidate regions.

Association between selective sweeps and dynamic AT-rich regions

For each species, we combined iHS- and CLR-sweeps from all populations to compile “species-wide” candidate region sets containing all loci under selection in any population. Shared sweeps that showed a signature of selection in multiple populations were only counted once and overlapping sweep regions from different populations were concatenated to one larger region. We then tested if sweeps and AT-rich regions overlapped more than expected by chance using a permutation tests implemented using the ‘overlapPermTest’ function in the R package regionR with 5000 iterations (Gel et al. 2016). This function creates randomly positioned dummy candidate region sets, which have the same number of regions and the same width as the real candidate region set and compares the observed number of overlaps to random expectations. To ensure that each sweep region was only counted once even if it overlapped more than one AT-rich region, we set the parameter `count.once=TRUE`.

Research objective 2: Functional analyses of genes under selection

Gene ontology enrichment analysis

We performed Gene ontology (GO) enrichment analyses for genes that were located in selective sweep regions detected by the two different methods (iHS and CLR). GO-annotations for conserved protein domain annotations based on the PFAM (Bateman et al. 2002) and PANTHER databases (Thomas et al. 2003) were obtained from the functional annotation done in Chapter I of this thesis. GO enrichment was assessed across all sweep regions (see below for different categories) using the R package ‘topGO’ and implementing a Fisher’s exact test using the function ‘runTest’ (Alexa and Rahnenführer 2009). Here we considered only GO-terms with a minimum of five annotated genes. The analysis was performed separately for iHS sweeps and CLR sweeps. For each selection scan method we tested six different sweep categories based on their (1) location and (2) their uniqueness or sharedness respectively: Categories were (1.1) all sweeps, (1.2) AT-overlap: sweeps partially within AT-rich regions, (1.3) GC-full: sweeps entirely in gene-rich regions, (2.1) Unique: sweeps unique to a single population, (2.2) Shared by 2: sweeps shared by two populations, and (2.3) Shared >2: sweeps shared by three or more populations. We performed the same analysis for divergent sweep regions detected by XP-EHH among three populations (Aub, Auv and El) categorized by (3.1) all

div: all divergent sweeps, (3.2) within/between : divergent sweeps between populations that were also under selection within at least one population based on iHS or CLR, (3.3) Shared by 2: divergent sweeps shared by two out of the three population pairs, and (3.4) Shared by 3: divergent sweeps shared by all three population pairs.

Differential gene expression analysis

RNAseq data that was collected from isolates grown in culture and *in planta* was used to identify differentially expressed genes in these two environments (RNA extraction and sequencing is described in Chapter I of this thesis). For this analysis we mapped the RNAseq reads to their respective indexed reference genomes using STAR v2.5.3a (Dobin et al. 2013) and counted the number of reads for each gene using the R package ‘Rsubread’ (Liao et al. 2019). Read counts from *in planta* technical replicates were merged. Then we used the R package ‘edgeR’ (Robinson et al. 2010) to identify genes that were differentially expressed *in culture* and *in planta*. Because variation in sequencing depth can influence variation in read counts, we normalized for library size. Often the number of reads is used as the library size, but in this experiment the number of mapped reads is more appropriate because the *in planta* samples contained a high proportion of plant RNA and the number of reads therefore does not reflect the sequencing depth at a gene in the fungus. We identified significantly over- and under-expressed genes by estimating dispersion of read counts across all genes and performing a gene-wise exact test for differences in the means between the two different treatment groups (*in planta* and *in culture*). Finally, GO-term enrichment analysis was performed on differentially expressed gene sets as described previously for genes in selective sweeps. Details of the analysis and code will be made available at https://github.com/adtreindl/Epichloe_genomics.

RESULTS

Research objective 1:

(i) Evidence of widespread and recent selection in *Epichloe* populations

To detect loci under recent selection in seven sympatric population pairs of *E. typhina* and *E. clarkii* we scanned genome-wide SNP datasets with inferred ancestral states for signatures of selective sweeps using linkage-based and SFS-based scans. Analyses were performed separately within each population. We used methods that account for the genome-wide patterns of polymorphism to mitigate the confounding effects of demographic events (Nielsen et al. 2005, Vitti et al. 2013), and we applied conservative percentile thresholds (99.9th percentile) of the test statistics distribution to identify sweeps with the strongest signature of selection. While selective sweeps were distributed along all seven chromosomes in both species, we found more selective sweeps among *E. typhina* populations than among *E. clarkii* populations. In *E. typhina*, we detected 228 iHS-sweeps (19-39 per population) and 124 CLR-sweeps (10-22 per population) (Table 1). In *E. clarkii*, we detected 90 iHS-sweeps (9-17 per population) and 87 CLR-sweeps (3-18 per population) (Table 2).

Table 1: Number of selective sweeps identified in *E. typhina* populations

	Aub	Auv	Cev	El	Kew	MdR	So	Total
iHS selection scan	30	39	32	38	19	38	32	228
CLR selection scan	22	21	24	10	13	19	15	124
Method overlap	1	4	3	2	5	2	4	21

Detailed summaries can be found in the Supplementary (Table S3).

Table 2: Number of selective sweeps identified in *E. clarkii* populations

	Aub	Auv	Cev	El	Kew	MdR	So	Total
iHS selection scan	13	12	15	12	9	12	17	90
CLR selection scan	16	6	17	15	3	12	18	87
Method overlap	0	0	1	0	0	0	1	2

Detailed summaries can be found in the Supplementary (Table S4).

On average CLR-sweeps tended to be longer than iHS sweeps and consistently a larger proportion of the genome was covered by CLR-sweeps than iHS-sweeps (*E. typhina* 2.94% vs 1.39% and *E. clarkii* 1.87% vs 0.85% averaged across populations;

Supplementary Tables S3 and S4). We identified concordance of different selection scans as those outlier sweep regions that overlapped between methods within the same population. Overall, sweeps identified by iHS- and CLR-scans showed relatively little overlap (21 sweeps or 5.96% in *E. typhina* and 2 sweeps or 1.12% in *E. clarkii*). An example composite plot of overlapping selective sweep signatures for one chromosome and population pair in *E. clarkii* is shown in Figure 1, and another example for *E. typhina* in Supplementary Figure S1. We generated composite plots for each of the seven chromosomes and pairwise population combinations totaling 294 plots all of which will be made available at a DRYAD data repository (<https://datadryad.org>).

(ii) Different regions of the genome are under selection in different populations but some selected regions are shared across multiple populations

We analyzed the sharedness of selective sweeps among the seven populations in both species to investigate whether the same loci were under selection in different populations. Sweeps were considered shared when genomic positions overlapped between populations and joined into common “sweep regions”. We performed this separately for either selection scan method. (Supplementary Tables S5 and S6; Supplementary Figure S2). The majority of selective sweep regions were unique to a single population (*E. typhina*: 72.4% of iHS-regions and 60% of CLR-regions; *E. clarkii*: 62.9% of iHS-regions and 75.8% of CLR-regions). Among shared selective sweep regions, most were under selection in only two populations (*E. typhina*: 66.7% of iHS-regions and 70.6% of CLR-regions; *E. clarkii*: 82.6% of iHS-regions and 81.3% of CLR-regions). In *E. typhina*, 15 genomic regions showed evidence of a selective sweep in three or more populations based on iHS-scans and 10 regions based on CLR-scans, whereas in *E. clarkii* only four (iHS) and three (CLR) regions were under selection in three or more populations. We investigated which populations shared the most sweeps and tested if significantly more sweeps overlapped by 50% or more between population pairs using the ‘overlapPermTest’ function in the R package regionR with 1000 iterations. Population pairs sharing the most iHS-sweeps in *E. typhina* where Cev-El (7, $Z=7.9697$, $p < 0.001$) and El-MdR(8, $Z= 7.8517$, $p < 0.001$), and population pairs sharing the most CLR-sweeps in *E. typhina* were Auv-El (6, $Z=8.497$, $p < 0.001$), Auv-MdR (6, $Z=7.645$, $p < 0.001$) and Auv-So (6, $p < 0.001$).

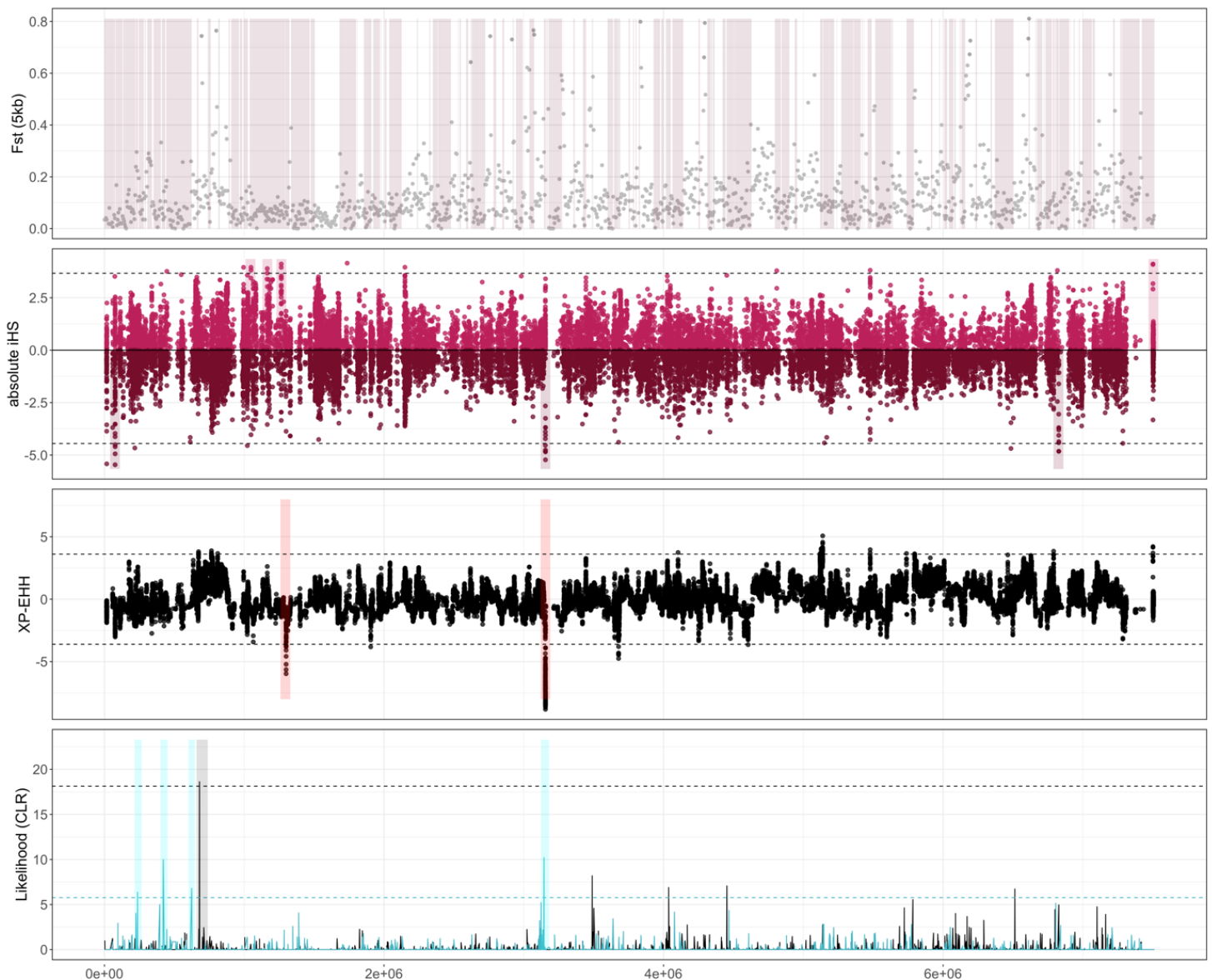


Figure 1: Signatures of selective sweeps on *E. clarkii* chromosome 4.

Chromosome-wide raw data and sweeps identified by different analyses are shown for the population pair Auv–So. The top panel shows pairwise F_{ST} values, averaged across 5kb windows. Shaded rectangles represent the locations of AT-rich regions. The second panel shows the absolute values of the integrated haplotype score (iHS) calculated at each SNP locus for which the ancestral allele state was known. Scores for Auv are shown at the top and scores for So are negatively transformed and showed at the bottom. Horizontal dashed lines indicate the 99.9% percentile threshold which was used as a cutoff to identify outlier SNPs and inferred iHS sweeps are shown as shaded rectangles. The third panel shows the cross-population extended haplotype homozygosity (XP-EHH) scores calculated between the two populations. Dashed lines indicate 99.9% percentile threshold which was used as a cutoff to identify outlier SNPs and inferred divergent sweeps are shown as shaded rectangles. Positive and negative XP-EHH values refer to the direction of selection: positive values indicate selection in Auv, negative values indicate selection in So. In the bottom panel, composite likelihood ratio (CLR) scores are plotted for Auv (black) and So (blue), colored dashed lines indicate respective 99.9% threshold and colored rectangles highlight inferred CLR-sweeps. Note the region at around 3Mb in which a selective sweep was detected by both iHS and CLR-scans. The same region is also under divergent selection between these two populations and overlaps with several relatively short AT-rich regions.

Population pairs sharing the most iHS-sweeps in *E. clarkii* where Cev-So (5, $Z=6.7353$, $p < 0.001$) and Aub-So (6, $Z=9.9234$, $p < 0.001$), and population pairs sharing the most CLR-sweeps in *E. clarkii* were Aub-El (4, $Z=5.857$, $p < 0.003$) and Aub-So (4, $Z=10.684$, $p < 0.001$). In our previous analysis we found that both *E. typhina* and *E. clarkii* showed a pattern of isolation by distance (IBD) with geographically more distant populations also displaying higher levels of pairwise F_{ST} (genome-wide weighted mean, (Weir and Cockerham 1984)) (Chapter II of this thesis). We therefore tested if the number of shared sweeps between populations decreased with increasing genetic or geographic distance. For this we calculated the fraction of overlapping sweeps for each population pair as the mean of the fraction of sweep intervals in one population that overlap with a sweep interval in the second population (that is, (fraction in population A + fraction in population B) / 2) and used the R package lme4 (Bates et al. 2015) to fit a linear mixed-effects model to the data with populations as random effects and distance as a fixed effect. We found no significant correlations between the sharedness of sweeps and distance in either species and with either selection scan method (Supplementary Figures S3 and S4). Although not significant, the proportion of shared CLR-sweeps showed a decreasing trend with distance and this effect was strongest relative to genetic distance in *E. clarkii* ($p = 0.081$), the species that also has a stronger effect of IBD overall.

(iii) Evidence of divergent selection among populations experiencing gene flow

We evaluated candidate regions showing signatures of divergent selection identified by cross-population XP-EHH-scans between three population pairs: Aub, Auv and El. Overall we detected almost twice as many divergent sweeps in *E. typhina* compared to *E. clarkii* (124 and 63 sweeps corresponding to 91 and 45 distinct regions). A similar proportion of divergent sweep regions also showed a signature of selection within at least one of the populations (23% and 22%). The majority of regions were under divergent selection between one population pair (67% and 69%), about a third of regions were under divergent selection between one population compared to both other populations (31% and 31%), and in *E. typhina*, two regions were under divergent selection between all three population pairs.

(iv) Selective sweeps are more common in AT-rich regions

To investigate if AT-rich regions, the dynamic genome compartments in *Epichloe* genomes, potentially play a role in adaptive evolution we analyzed the distribution of regions with a signature of recent selection and tested if these regions clustered in particular genomic compartments (Figure 2). Permutation tests showed that there was significantly more overlap than expected by chance between AT-rich regions and sweeps identified by iHS-scans: 80.6% of regions overlapped in *E. typhina* (125 out of 155, $p = 0.0002$) and 94.6% of regions overlapped in *E. clarkii* (53 out of 56, $p = 0.0002$) and the same was true for CLR-scans in both species: 86.8% of regions overlapped in *E. typhina* (66 out of 76, $p = 0.0002$) and 92.2% of regions overlapped in *E. clarkii* (59 out of 64, $p = 0.0004$).

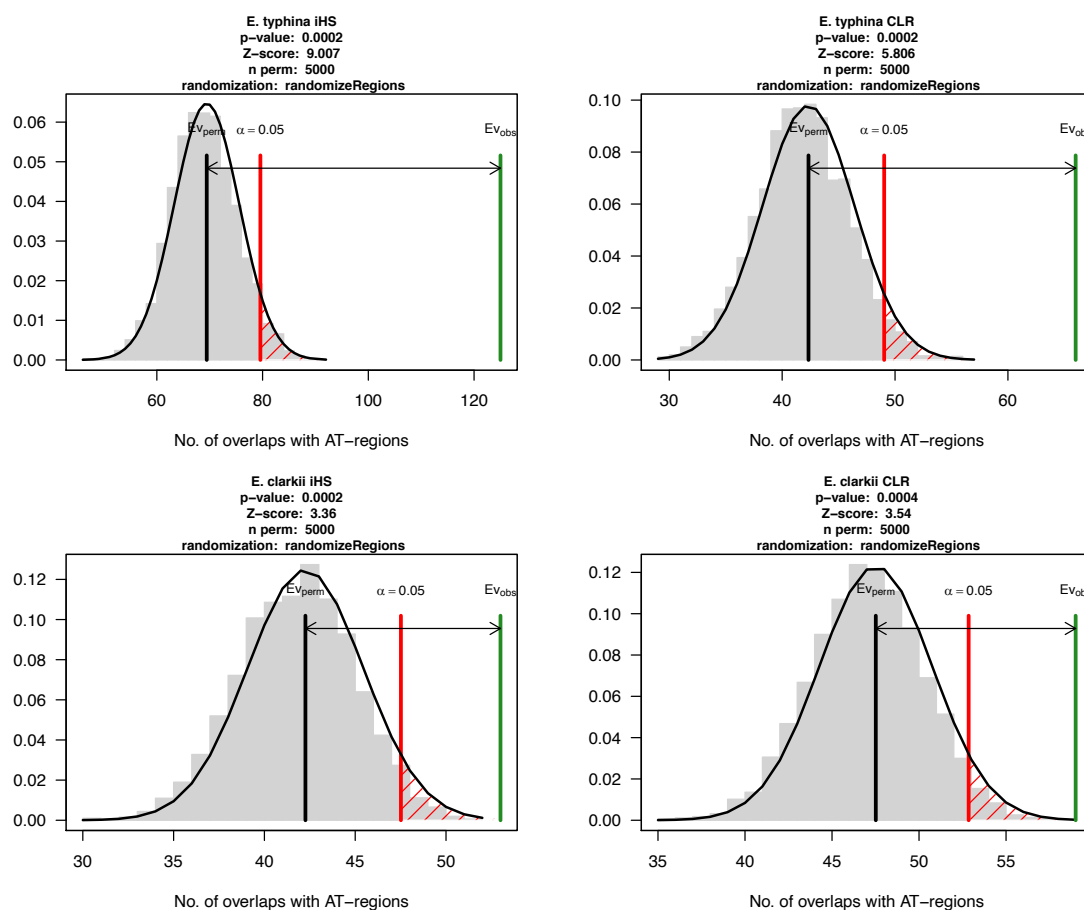


Figure 2: Association between regions under selection and AT-rich regions.

Plots show graphical representation of the permutation test performed where the grey histogram represents the number of overlaps using a randomized region set with a fitted normal, $E_{v_{perm}}$ (black line) indicates the mean of the randomized counts and $E_{v_{obs}}$ (green line) indicates the observed overlap count. In both *E. typhina* (top) and *E. clarkii* (bottom) and for sweeps detected by either selection scan-method (iHS left, CLR right), regions under selection were more likely to overlap with AT-rich regions than expected by chance.

Research objective 2:**(v) Gene content analyses of regions under selection**

We analyzed the gene content of different sweep regions (Table 3). Genome-wide sweep region sets contain all loci under selection within each of the two species resulting from the concatenation of all sweeps across the seven populations. Since some proportion of sweeps were shared among several populations, the number of sweep regions is lower than the total number of sweeps detected, e.g. we detected 228 iHS sweeps across seven populations in *E. typhina*, 110 of them were shared between at least two populations resulting in a total of 163 sweep regions. Many sweep regions were located fully in AT-rich regions (Supplementary Figure S5), however these regions contained only two predicted genes in either species. Although none of them had a functional annotation, a blast search revealed homology to genes or gene clusters known to be involved in alkaloid biosynthesis in *Epichloe* (Schardl et al. 2013b): Ety_004235 showed homology to lolF (86% query cover, 78% identity), Ety_007055 showed homology to the partial EAS gene cluster (93% query cover, 71% identity); Ecl_003076 showed homology to the PER gene cluster (89% query cover, 89% identity) and Ecl_004371 showed homology to lolF (89% query cover, 87% identity).

Table 3: Number and gene content of selective sweep regions in *E. typhina* and *E. clarkii* genomes

		<i>E. typhina</i>		<i>E. clarkii</i>		
		sweep regions	genes	sweep regions	genes	
<i>method</i>	iHS	163	297	62	162	
	CLR	85	388	66	632	
	overlap	15	89	2	3	
<i>compartment</i>	GC	iHS	31	172	3	47
		CLR	14	78	8	46
	overlap	iHS	66	124	36	114
		CLR	48	323	42	605
	AT	iHS	69	1	25	2
		CLR	27	1	19	0

To identify functional groups of genes that showed evidence of recent selection, we tested for overrepresented GO terms in selective sweeps compared to the genome-wide background for genes with an annotated PFAM and or PANTHER domain. Here we report the result for the region sets which contained the most genes and for which we therefore had more power to detect significant enrichments. The most significantly enriched GO-terms ($p < 0.01$) for all sweeps independent of their genomic location (category 1.1, see Methods section) and sweeps overlapping with AT-rich regions (category 1.2) are reported in (Table 4) with the respective selection scan method indicated. A complete table including significant GO-terms up to $p = 0.05$ and for the remaining categories 1.3, 2.1-2.3 and 3.1-3.4 will be available in an online data repository (Supplementary Materials on DRYAD).

In *E. typhina* protein functions related to oxidation-reduction processes (GO:0055114) and amide biosynthetic processes (GO:0043604) were enriched across all CLR-sweeps, and both GO-terms were also enriched in iHS-sweeps, but not significant at our conservative p-value cutoff of $p < 0.01$ (GO:0055114: $p = 0.037$ and GO:0043604: $p = 0.021$). Proteins related to oxidation-reduction processes and intracellular signal transduction (GO:0035556) were also enriched in sweeps detected by both methods. In selective sweeps overlapping with AT-rich regions, protein functions involved in cellular amide metabolic processes (GO:0043603) were enriched in iHS- and CLR-sweeps, but below our conservative cutoff (GO:0043603: $p = 0.012$). The strongest evidence of enrichment was found for biological functions related to protein phosphorylation (GO:0006468) across all CLR-sweeps and the same GO-term was enriched in AT-overlapping iHS- and CLR-sweeps ($p = 0.002$ and $p = 0.00012$ respectively). When reducing the necessary minimum number of annotated genes per GO-term to two rather than five an additional category showed a significant enrichment in selective sweep regions: fatty acid transmembrane transport (GO:1902001) with three genome-wide annotations, all of which were located in iHS sweeps located in GC-rich regions.

In *E. clarkii* the same GO-term for protein phosphorylation showed the strongest evidence of enrichment across all CLR-sweeps and for CLR-sweeps overlapping with AT-rich regions. Proteins related to mycotoxin biosynthetic processes (GO:0043386) and establishment of localization in cell (GO:0051649) were also enriched across all sweeps and AT-overlapping sweeps detected by CLR-based scans and iHS-based scans

respectively. Among all iHS-sweeps functional domains involved in cellular modified amino acid biosynthetic processes (GO:0042398) were enriched. When removing the filter for >5 annotated genes per GO-term two additional categories showed a significant enrichment in selective sweep regions: sterol biosynthesis (GO:0016126) and alkaloid metabolism (GO:0009820).

Among the 59 genes in *E. typhina* selective sweep regions with significant GO-term annotations, 20 encoded protein kinases (GO:0006468), 7 encoded ribosomal proteins (GO:0043603 and GO:0043604) and 4 encoded Cytochrome P450 monooxygenases (GO:0055114). Among the 32 genes in *E. clarkii* selective sweep regions with significant GO-term annotations, 21 encoded protein kinases (GO:0006468) and 5 encoded proteins putatively involved in the production of toxic cyclic peptides (ustYa-like proteins, GO:0043386).

Table 4: Biological functions overrepresented among selective sweeps in *E. typhina* and *E. clarkii* (Fisher's exact test: $p < 0.01$)

	GO ID	GO term	Significant category	Genome-wide annotations	Significant annotations	Enrichment p-value
<i>E. typhina</i>	GO:0006468	protein phosphorylation	all sweeps CLR	148	18	6.5E-05
	GO:0035556	intracellular signal transduction	all sweeps iHS-CLR-overlap	30	2	5.4E-03
	GO:0043604	amide biosynthetic process	all sweeps CLR	127	8	8.0E-03
	GO:0055114	oxidation-reduction process	all sweeps iHS-CLR-overlap ^a	349	18	1.6E-05
	GO:0006468	protein phosphorylation	AT-overlap CLR and iHS	148	15/7	1.2E-04/2.0E-03
	GO:0043603	cellular amide metabolic process	AT-overlap iHS	143	6	4.3E-03
<i>E. clarkii</i>	GO:0006468	protein phosphorylation	all sweeps CLR	147	21	2.5E-04
	GO:0043386	mycotoxin biosynthetic process	all sweeps CLR	15	5	1.7E-03
	GO:0042398	cellular modified amino acid biosynthetic process	all sweeps iHS	12	2	7.2E-03
	GO:0051649	establishment of localization in cell	all sweeps iHS	81	4	9.9E-03
	GO:0006468	protein phosphorylation	AT-overlap CLR	147	20	3.4E-04
	GO:0043386	mycotoxin biosynthetic process	AT-overlap CLR	15	5	1.4E-03
	GO:0051649	establishment of localization in cell	AT-overlap iHS	81	4	1.6E-03

Significant category indicates the sweep subsets on which enrichment tests were performed; Genome-wide annotations indicates the total number of genes annotated to the respective GO-term and Significant annotations indicates how many of those genes were located in selective sweeps.

^aalso significant in CLR sweeps: 25 annotations ($p=9.4E-03$) and iHS-sweeps: 17 annotations ($p=3.7E-02$)

Differential gene expression analysis

We compared expression levels of genes between *Epichloe* reference strains grown on culture plates (*in culture*) and in their respective host plants (*in planta*). Overall, we found 1,037 and 1,598 genes with higher expression levels *in planta* corresponding to

12.6% and 22.1% of genes in *E. typhina* and *E. clarkii* respectively. A total of 1,440 (17.5%) and 1,458 (20.1%) of genes had lower expression levels *in planta*. Enriched GO terms in differentially expressed genes between *in planta* and *in culture* ($p < 0.01$) are listed in Supplementary Table S7. We found similar GO terms among differentially expressed genes in both *E. typhina* and *E. clarkii*: genes involved in mycotoxin biosynthetic processes and oxidation-reduction processes were upregulated *in planta* relative to *in culture*. Genes related to transmembrane transport were enriched among both up- and down-regulated genes, and specifically genes involved in copper ion transmembrane transport had lower expression levels *in planta* relative to *in culture*. Additionally, in *E. typhina*, genes related to pathogenesis and sterol metabolic processes were down-regulated *in planta*. Other enriched GO-terms ($p \leq 0.05$) among genes up-regulated *in planta* in *E. typhina* included several functions related to the regulation of gene expression and transcription (online Supplementary Materials, DRYAD). In *E. clarkii* down-regulated genes were enriched for functions related to gluconeogenesis, fungal cell wall organization and chitin biosynthesis while other significant GO-terms ($p \leq 0.05$) among genes upregulated *in planta* in *E. clarkii* included functions related to autophagy and chromosome segregation (online Supplementary Materials, DRYAD).

(vi) Effector genes under selection and other candidate genes putatively involved in host-pathogen interactions

Among all genes located in selective sweep regions we investigated two categories of genes more closely with the aim to pin down genes which may be involved in antagonistic coevolution with the host plant: genes that were annotated as putative effectors and genes located in regions where we detected signatures of selective sweeps with both iHS- and CLR-based selection scans (method overlap).

Analysis of putative effector genes

Regions under selection in the genomes of *E. typhina* and *E. clarkii* were not more likely to contain putative effector genes than expected at random (Supplementary Figure S6). In total 19 out of 138 (*E. typhina*) and 23 out of 144 effectors were located in selective sweep regions identified by either iHS or CLR. Out of these, 9 and 10 genes in each species had functional annotations. One putative effector gene (Ec1_000006) with

functional homology to ribotoxins, fungal extracellular RNases, was located in a CLR-sweep region shared by all *E. clarkii* populations except MdR and the orthologous locus in *E. typhina* (Ety_004220) also showed evidence of selection in one population (So). Interestingly, Ecl_000006 was significantly upregulated *in planta* (1.2 log₂-fold-change (logFC)) whereas the ortholog Ety_004220 was significantly downregulated *in planta* (5.0 logFC). We investigated the haplotype structure at these loci in both species (Figures 3 and 4, Supplementary Figure S7). In *E. clarkii* four populations in which the CLR-scan identified a sweep were fixed for one haplotype with some low frequency variants present consistent with arms-race dynamics and MdR was fixed for another haplotype. An alternative haplotype with comparatively reduced variation was present in one population at intermediate frequency (So) and also in another population at low frequency (Cev). At the homologous locus in *E. typhina*, our analysis revealed the presence of two alternative haplotypes at intermediate frequencies within populations consistent with trench-warfare dynamics.

Among the effector genes under selection in *E. typhina* one encoded a protein with a CyanoVirin-N Homology (CVNH) domain (Ety_004204), one with a LysM domain (Ety_004208), a Copper/zinc superoxide dismutase (SODC, Ety_007724), and a pectinesterase CWDE (Ety_004245). The latter gene was located in the subtelomeric region of chromosome four in a CLR-sweep region detected in the El population and overlapping CLR-sweeps in four additional populations (Aub, Auv, Kew and MdR), and was also highly upregulated *in planta* (4.7 logFC). In *E. clarkii* selected effector genes also encoded a protein with a CVNH domain (Ecl_005839), one with a LysM domain (Ecl_000304) and two putative CWDEs, a cutinase (Ecl_005058) and a polysaccharide deacetylase (Ecl_008800). One putative effector gene (Ecl_001468) was located in a CLR-sweep region on chromosome two shared by four populations (Aub, Auv, El and Kew) and was highly upregulated *in planta* (9.1 logFC). The gene had no annotated biological functions but showed homology to a meiotically up-regulated gene family in yeast (*Schizosaccharomyces pombe*, PFAM domain: PF15474). Plotting the cross-population haplotype structure at this locus revealed the presence of two distinct haplotypes at different frequencies within populations (Supplementary Figures S8 and S9). Populations for which the CLR-scan inferred a recent selective sweep were fixed for

one haplotype consistent with arms-race dynamics, however, one population was fixed for the alternative haplotype and two populations contained both haplotypes at intermediate frequencies.

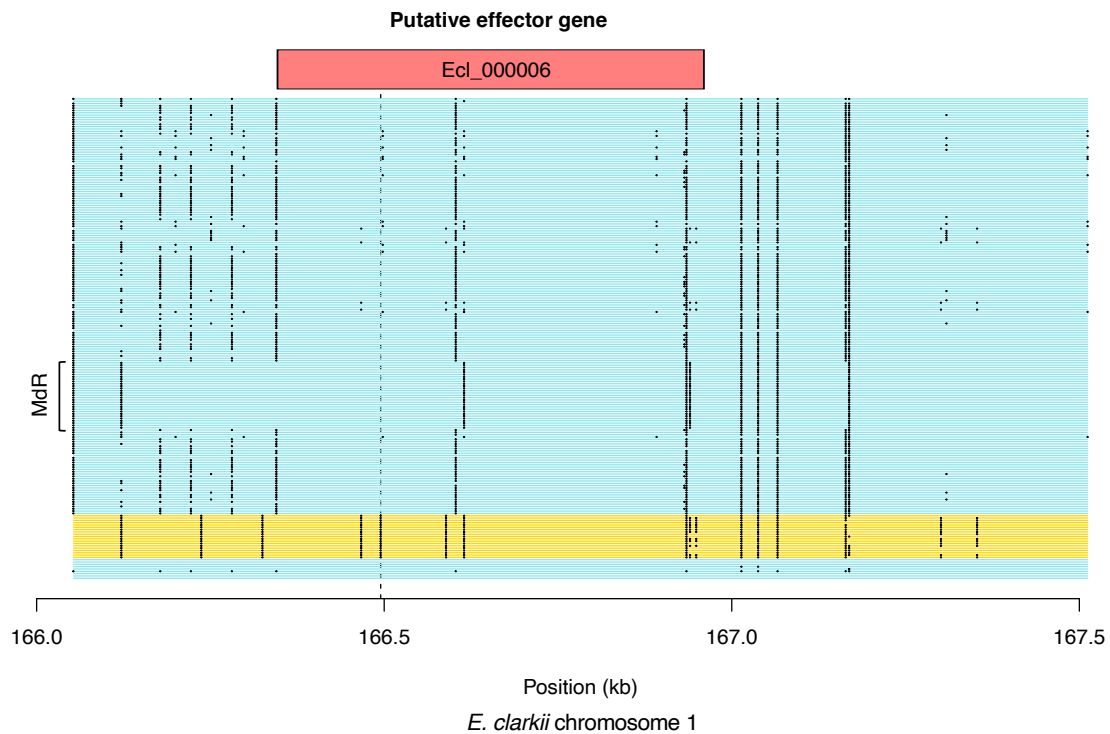


Figure 3: Cross-population haplotype structure for candidate effector locus in *E. clarkii* (chr1).

The haplotype structure of a predicted effector gene (Ecl_000006) located in a CLR-sweep region shared by six populations (all except MdR) is shown, which was differentially expressed *in planta*. The genomic location lies in a relatively short GC-rich region (4,982bp) sandwiched between two AT-rich regions at the start of chromosome one and contains two other predicted genes besides the effector, one of them encoding a protein kinase domain (Ecl_000004). Haplotypes for all isolates from all populations are shown (n=190) and were colored relative to the first nonsynonymous SNP position located in the genic region of Ecl_000005 (166,495; blue=A, yellow=G). Blue lines at the bottom of the plot indicate individuals that have not been genotyped at this locus (9). In total three different haplotypes can be distinguished at the locus: The most common blue haplotype is present in all populations except MdR where isolates are fixed for a different haplotype (also blue) as indicated in the plot. The yellow haplotype is present at intermediate frequency in So and at low frequency in Cev (see Supplementary Figure S7i showing haplotypes at this locus for each population separately). The homologous locus was also located within a selective sweep in *E. typhina* (see Figure 4).

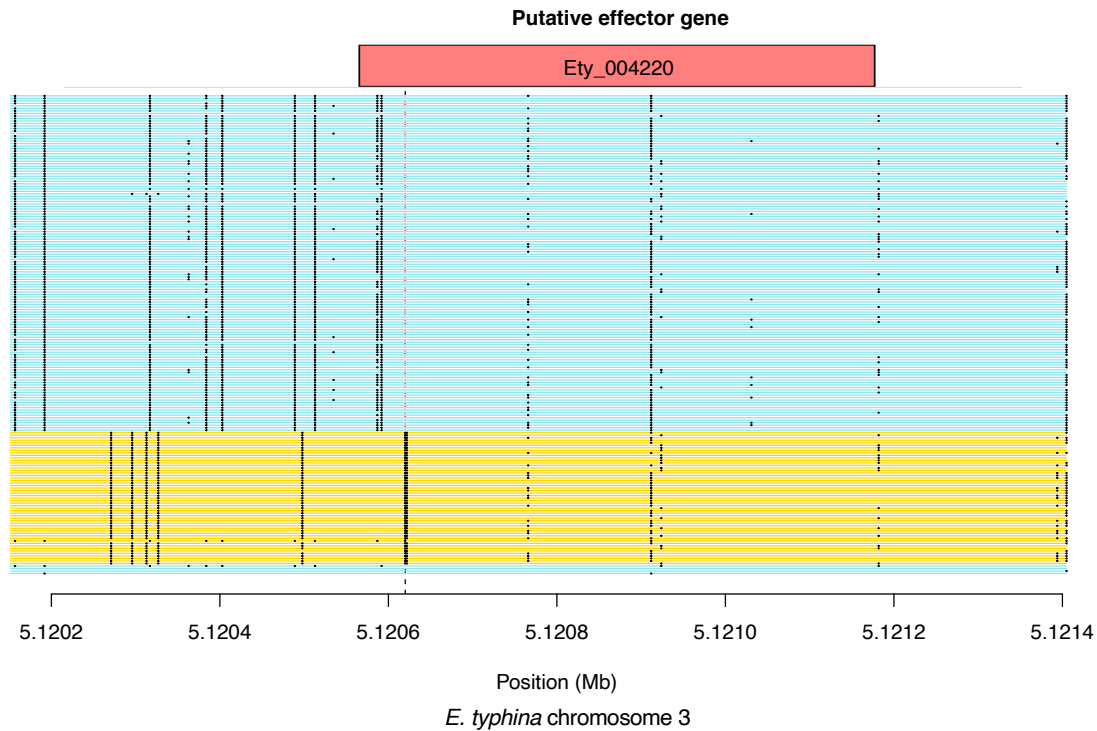


Figure 4: Cross-population haplotype structure for candidate effector locus in *E. typhina* (chr3).

The haplotype structure of a predicted effector gene (Ety_004220) located in a CLR-sweep region in one population (So) which was also under selection in *E. clarkii* is shown. The gene was significantly downregulated *in-planta*. The genomic location lies in a longer GC-rich region (31.5kb) at the end of chromosome three (which is the homologous inversion of *E. clarkii* chromosome one). Haplotypes for all isolates from all populations are shown (n=175) and were colored relative to the second nonsynonymous SNP position located in the genic region of Ety_004220 (5,120,620; blue=A, yellow=G). Blue lines at the bottom of the plot indicate individuals that have not been genotyped at this locus (4). Two different haplotypes can be distinguished at the locus both of which are present at different frequencies in all seven populations (see Supplementary Figure S7ii showing haplotypes at this locus for each population separately). The population for which the CLR-scan inferred a sweep (So) is almost fixed for the blue haplotype with some low frequency variants present and one isolate with the yellow haplotype.

Genes located in selective sweeps identified by different methods

We found only 21 cases where both methods inferred a selective sweep within *E. typhina* populations, corresponding to 15 distinct genomic regions as some of the overlapping sweeps were also shared between populations (four sweeps were shared by two populations, one sweep was shared by three populations). Ten of these 15 regions contained 89 predicted proteins and among them 70 genes with functional annotations. Genes encoded protein kinases (4), major facilitator superfamily transporters (2), and different functional domains related to oxidation-reduction processes (13) such as

Cytochrome P450 monooxygenases (3). None of them were annotated as putative effectors but one showed homology to a LysM effector domain (Ety_004233). We investigated this latter locus more closely as it showed evidence of selective sweeps in all populations inferred by either or both methods (CLR-sweeps in Aub, Auv, El, Kew, MdR overlapping with iHS-sweeps in Cev, Kew, MdR, So), and some of these sweeps were fully located in an AT-rich region. The AT-rich region contained a gene, the aforementioned Ety_004235, showing homology to the alkaloid biosynthesis gene lolF. The haplotype structure at this locus showed that most isolates share a similar haplotype despite a relatively high level of polymorphism which is likely a result of the genomic localization, however one population (So) contained a different haplotype at intermediate frequencies with one particular SNP encoding a splice site variant (Supplementary Figure S10).

Within *E. clarkii* populations, we detected two regions with sweep signatures inferred by both methods, one of which contained 3 genes while the other contained no annotated genes. Of the three genes located in the same sweep region on chromosome four, two encoded pectate lyase superfamily proteins (PFAM:PF12708) and the third had no functional annotation, however, a blast search revealed similarity to a LysM domain-containing protein in *Metarhizium acridum* (84% query cover, 68% identity). None of the genes were significantly over- or under-expressed *in planta*, however, these genes were located just upstream of an 11kb AT-rich region. Signatures of selection at this locus were detected with both methods in So and three additional populations shared the iHS sweep: Aub, Cev and El. We investigated the cross-population haplotype structure at this locus (Supplementary Figures S11 and S12). Populations for which the iHS-scan inferred a sweep contained a shared haplotype at intermediate frequencies and three other haplotypes were present at different frequencies in different populations. As a comparison we also investigated the haplotype structure at the homologous region in *E. typhina*, which did not contain any sweep outliers. It also contained the two pectate lyase superfamily proteins (Ety_004682 and Ety_004683) but lacked a putative effector gene annotation (Supplementary Figure S13 Figure X: Cross-population haplotype structure for homologous locus not under selection on chromosome four in *E. typhina*). Here, all populations are fixed for the same haplotype except MdR which contained a different haplotype at intermediate frequency.

DISCUSSION

We identified genome-wide signatures of selective sweeps in seven field populations sampled from sympatric locations in Europe using two different selection scan methods. Selective sweeps were widespread across the genomes of *Epichloe* and were more abundant in *E. typhina* compared to its sibling species *E. clarkii*. The lower number of selective sweeps in *E. clarkii* populations is in agreement with our expectations based on the lower rates of gene flow, the lower genetic diversities and smaller effective population sizes described previously (Chapter II of this thesis). Allele frequencies in populations with a low N_e are governed by stronger relative effects of stochastic processes such as genetic drift while the efficacy of selection is reduced compared to populations with a large N_e (Charlesworth 2009). Consequently, alleles are more likely to become fixed or be lost from small populations due to chance resulting in reduced genetic diversity and reduced levels of potentially adaptive genetic variation on which selection can act. In addition, the statistical power to detect selective sweeps from sequence data is lower in populations or species that are less genetically diverse overall. Selective sweeps were also less common in the lower-diversity and smaller- N_e anther smut fungus *Microbotryum silenes-dioicae* compared to its sister species *M. lychnidis-dioicae* (Badouin et al. 2017). Similarly, genome-scans identified fewer sweeps in less genetically diverse field populations of the wheat pathogen *Zymoseptoria tritici* (Hartmann et al. 2018).

Selective sweep regions were narrow as expected based on the high recombination rates and rapid decay of LD in these sexually reproducing fungi. On average, inferred sweeps were narrower (haplotypes shorter) in *E. typhina* consistent with the faster decay in LD compared to *E. clarkii* (Chapter II of this thesis). This is obviously influenced by the different window sizes: we used larger windows in *E. clarkii* than *E. typhina* accounting for different densities of polymorphism. Nevertheless, sweeps detected by iHS-scans contained, on average, a similar number of SNPs, a similar number of outliers and similar absolute values of iHS statistics among outliers across populations of both species and CLR-sweeps had similar likelihood scores giving us confidence that the analyses was performed at similar levels of accuracy and a comparison of results between species is reasonable. We detected the fewest and widest selective sweeps using CLR-scans in the Kew population in *E. clarkii*, which had the smallest sample size ($n=19$),

harbored significantly less genetic diversity and high degrees of linkage likely related to population substructure (see Chapter II of this thesis). As selection is expected to be less efficient in small N_e populations and larger blocks of LD additionally increase the likelihood that beneficial mutations appear in linkage with slightly deleterious mutations, thus reducing the true number of sweeps. However, given the strong LD in this population and confounding demographic effects, the power to detect selective sweeps is inherently limited and the length of selected haplotypes inferred by genome scans may be inflated (Koropoulis et al. 2020). Similarly, the MdR population in *E. clarkii* also harbored low genetic diversity and large blocks of LD in this case likely resulting from a recent bottleneck (Chapter II of this thesis). The development and implementation of sophisticated methods that disentangle such demographic effects on genetic loci from true adaptive signals is an ongoing challenge in genomic studies and, despite using methods that are relatively robust to demographic variations, some selective sweeps may have been missed or represent false positives (Weigand and Leese 2018, Horscroft et al. 2019).

Little overlap between selective sweeps detected by different statistics

The concordance between selective sweeps detected by different genome scan methods in our study was relatively low. The overlap between methods may be an underestimate because we used a stringent significance cutoff (99.9th percentile), but low overlap is consistent with other studies and thus not surprising (Pavlidis and Alachiotis 2017, Hartmann et al. 2018). We expect little overlap as the methods rely on the detection of different signatures of selection and have increased power at different time points relative to the onset of selection: SFS-based scans such as CLR perform better at detecting selected variants close to fixation, while LD-based scans such as iHS are best suited to identify ongoing sweeps with alleles at intermediate frequencies (Vitti et al. 2013, Weigand and Leese 2018). By combining two complementary methods we were able to provide a more integrated evaluation of the impact of selection in *Epichloe* genomes.

Signatures of recent selection are commonly associated with AT-rich regions

Sweep regions were not randomly distributed across *Epichloe* genomes. In fact, a large proportion (82.7% in *E. typhina* and 93.3% in *E. clarkii*) of sweeps detected by either method partially or fully overlapped with AT-rich regions. This finding emphasizes the role that repeat-rich genome compartments play in rapid and ongoing adaptation in *Epichloe* by demonstrating a link between the structural organization and signatures of selection in fungal plant pathogen genomes. “Two-speed” genomes or more precisely their dynamic genome compartments in the form of whole accessory chromosomes [e.g. in *Fusarium oxysporum* (Ma et al. 2010) and *Z. tritici* (Plissonneau et al. 2018)], or scattered mosaic-like sequence blocks [e.g. in *Leptosphaeria maculans* (Van de Wouw et al. 2010, Rouxel et al. 2011, Grandaubert et al. 2014) and *Verticillium dahlia* (Faino et al. 2016)] compose an environment of increased genetic and structural variation, providing fuel for natural selection. However, the signatures of selection within and around these particular genomic regions have rarely been studied. Instead, it is usually the preferential association of effector genes with dynamic regions that constitutes the evidence for their putative adaptive role. The results presented here, however, provide new support for the adaptive potential of dynamic and repeat-rich genomic regions in two-speed genomes that is purely based on patterns of DNA-sequence variation at the population genomic level. Analyzing these patterns using an alignment-based approach and resolving nucleotide variation not only in the conserved genome compartment but also in the dynamic one was made possible by the chromosome-level reference genomes which provided an accurate resolution of the patchwork genome structure of *Epichloe* (Chapter I of his thesis). And still, characterizing genetic variation in highly polymorphic regions presents particular challenges (Eschenbrenner et al. 2020). This is largely because there are limitations to the accuracy of variant discovery based on the alignment of short reads in highly polymorphic repeat-rich regions, because sequences may be incorrectly mapped to the reference genome leading to ambiguities in the variant calling (Treangen and Salzberg 2012, Pfeifer 2017). Furthermore, which regions can be resolved strongly depends on the reference sequence and structural variation such as presence absence polymorphisms, for example when a particular AT-rich region is present in our reference genome but missing in some mapped individuals or vice versa, can lead to erroneous inferences of SNPs. In order to correct for these issues and achieve an accurate

evaluation of genetic variation we developed a thorough variant filtering procedure. Firstly, the fact that our dataset was haploid made ambiguous loci detectable as falsely “heterozygous” sites. Secondly, we employed stringent filtering criteria that limited our down-stream analyses to high confidence SNPs with similar genotyping rates across all populations thereby excluding genomic regions with large scale presence/absence polymorphism (see Chapter II of this thesis for details). Finally, for the genome-scans performed in this study we included only conserved SNPs that had high mapping quality of the outgroup species (See Methods section: Variant detection and inference of ancestral alleles). In Chapter II of this thesis we demonstrate that the genetic structure among populations based on SNPs called in the conserved gene-rich regions is also resolved in the patterns of nucleotide variation when including only the SNPs called in the dynamic AT-rich regions giving us confidence that our analyses are based on true and biologically relevant genetic variation (see Chapter II Supplementary Notes: PCA analysis and calculation of summary statistics by compartments).

Evidence of heterogenous selection pressure among populations

Selective sweeps in *E. typhina* and *E. clarkii* were predominantly population-specific suggesting that selection pressures vary in space and different loci are adaptive in different populations. The extensive evidence of divergent selection identified between populations connected by gene-flow further confirmed this observation. The results of this study are in line with the expectation that in heterogenous environments such as natural ecosystems different traits should be favored by selection in different populations (Thompson 2005), while acknowledging that differences in population genetic parameters and demographic history (and therefore the power to detect selective sweeps, for example in the Kew and MdR population in *E. clarkii*) may also have an influence. Spatially variable selection pressures could result from differences in the genetic makeup of host populations and arms-race coevolution with local host genotypes, however, the pattern may also indicate adaptation to different abiotic environmental conditions independent of the host plant such as fluctuations in humidity or temperature. Both theoretical work and empirical studies substantiate this assumptions: Environmental heterogeneity including both biotic and abiotic factors, reduced levels of gene flow and spatially variable patterns of selection drive

coevolutionary dynamics at the local scale and generate locally adapted populations (Laine 2005, Gandon and Nuismer 2009, Croll and McDonald 2016). As a result, distinct loci show signatures of selection in different populations across diverse fungal systems ranging from agricultural pathogens to wild mutualistic mycorrhiza (Ellison et al. 2011, Branco et al. 2017, Hartmann et al. 2018, Mohd-Assaad et al. 2018).

However, not all selective sweeps were unique to one *Epichloe* population. Indeed, some loci were under selection in multiple populations across large geographic distances suggesting shared selection regimes and a possible role of the underlying loci in coevolutionary dynamics at a larger scale. The combined evidence for both unique and shared regions under selection across populations might be due to fact that by sampling multiple local populations at large geographic distances we may catch them at different points along coevolutionary cycles. Ultimately this underscores the multidimensional nature of the coevolutionary process. Thompson (2005) called it “the geographic mosaic of coevolution” aptly represented by two interacting metapopulations, here the fungal pathogen and its host plant, occupying a heterogeneous landscape. Antagonistic selection pressures are asynchronized over space and time as they act locally within populations but are affected by among-population processes depending on the underlying species distribution, their dispersal rates (migration) and levels of gene flow (Gandon and Nuismer 2009, Smith et al. 2011, Ebert and Fields 2020). For example, a beneficial new mutation enabling the infection of a previously resistant host genotype may first appear in one local pathogen population, sweep to high frequency and simultaneously, given sufficient levels of gene flow, spread to other populations where the same host genotype is present and also increase in frequency generating a shared signature of a selective sweep across multiple populations. Alternatively, unique sweeps may be observed when a new mutation only provides an adaptive advantage locally or fails to spread to other populations. Importantly, as coevolution acts at the level of phenotypes and their causative genes and not species as a whole, and sexual recombination can decouple these loci from their genomic background, different regions of the genome may experience different dynamics at the same time (Ebert 2018). We investigated the genomic signatures of selective sweeps under the assumption that strong selection pressures in the *Epichloe* system should foster arms-race coevolutionary dynamics and found widespread evidence of selective sweeps in the pathogens genomes

at the population level. However, upon investigating the distribution of adaptive haplotypes across populations we also found evidence of maintained polymorphism (see e.g. Supplementary Figure S12) with the same alleles being present in populations at different frequencies. This pattern appears consistent with trench-warfare coevolutionary dynamics and frequency dependent selection affecting the same loci across populations. Selection pressures in a local population are expected to vary through time, an aspect we cannot address with our dataset, but selection pressures also vary in space among populations as outlined above. We hypothesize that the variation in allele frequencies observed among populations may result from the asynchrony of selection pressures and be promoted by isolation-by-distance (Gandon and Nuismer 2009, Tack et al. 2012). Future analyses of our datasets could explore this hypothesis by testing if candidate loci identified here indeed show signatures of balancing selection across the whole species. The study of multiple local pathogen populations as presented here provides a single glimpse of one half of a complex coevolutionary interaction (not including the host) and yet, manages to resolve distinct evolutionary trajectories within that snapshot.

Genes under recent selection include candidates involved in host adaptation

Selective sweeps contained genes encoding a broad range of different biological functions, but one notable group were those involved in protein phosphorylation, which was enriched in both species and comprised genes encoding protein kinase domains. Protein kinases catalyze the phosphorylation of proteins thereby acting as regulators of protein activity and gene expression. In fungal pathogens they govern a variety of cellular pathways related to plant infection including processes involved in cell surface remodeling to avoid recognition by the host, differentiation of infection-related structures and invasive growth (Turrà et al. 2014). A protein kinase (*sakA*) in *Epichloe festucae* has also been shown to be important for the regulation of hyphal growth *in planta* and maintenance of a mutualistic interaction with the host plant *Lolium perenne*, while a deletion of this gene lead to pathogenic interactions (Eaton et al. 2010). It appears that protein kinases take over conserved functions across diverse fungal systems (Turrà et al. 2014), and our finding that the genes encoding them are often targets of

recent selection suggests that they constitute an important component of the *Epichloe* armory.

The analysis of loci under recent selection revealed a number of other interesting genes related to the biosynthesis of secondary metabolites. Genes involved in the metabolism of mycotoxins including alkaloids and sterols were enriched in selective sweeps in *E. clarkii* but not *E. typhina*, and mycotoxin genes were also highly expressed *in planta* in both species (Supplementary Table S7). Selected regions also contained additional putative alkaloid genes, however, these were not included in the enrichment analysis as they lacked a functional annotation, likely because they were often highly mutated and located in polymorphic AT-rich regions of the genomes. An extensive body of research has highlighted the structural diversity of alkaloid genes across the *Epichloe* genus and specifically emphasized the importance of alkaloids produced predominantly by non-pathogenic *Epichloe* in providing benefits to the host plant as herbivore deterrents (Schardl et al. 2013a, Young et al. 2015). Alkaloids are therefore considered the “currencies maintaining mutualistic symbiosis” (Schardl et al. 2013b). A detailed analysis of these secondary metabolite gene clusters at the population level would have gone beyond the scope of this particular study, nevertheless, our results suggest that these genes may not only play a role in mutualistic interactions between asexual *Epichloe* and their host grasses but also in sexual pathogenic *Epichloe* and provide opportunity for further avenues of research.

We identified a set of putative effector genes located in selective sweep regions, some of which were also differentially expressed and many of which encoded biological functions that may be relevant at some stage of host infection. Secreted CWDEs included a cutinase which may be important for penetration of the plant cuticle (Chen et al. 2013), as well as pectate lyases and a pectinesterase putatively involved in the degradation of pectin in the plant cell wall (Yang et al. 2018). Small, cysteine-rich effector proteins with CVNH domains have been shown to be important for infection and growth in the stem rot pathogen *Sclerotinia sclerotiorum* (Lyu et al. 2015, Seifbarghi et al. 2017), and a gene encoding another CVN-effector (Ety_004758/ Ecl_005079) was also among the candidate genes involved in host specialization in our *Epichloe* system showing high levels of differentiation between *E. typhina* and *E. clarkii* (Schirrmann et al. 2018). LysM effectors are involved in the binding of chitin oligosaccharides,

degradation products of fungal cell wall that are released during plant invasion (e.g. in *Verticillium dahliae*) and are recognized by host chitin receptors (Gao et al. 2019). The secretion of LysM effectors functions as an effective strategy to prevent chitin-triggered host immunity. We also found an enrichment of genes in regions under recent selection that are presumably not related to host-pathogen interactions such as genes involved in intracellular signal transduction, cellular localization or amide metabolism. Genes related to oxidation-reduction processes were enriched in selective sweeps in *E. typhina* but not *E. clarkii*, though they were enriched among upregulated genes in both species. This was not surprising given that redox reactions play a major role in biological functions of all organisms.

Genes showing evidence of selection are assumed to be of functional relevance, but what evolutionary researchers bedazzled by coherent biological narratives sometimes forget is that not every genome-scan outlier is adaptive. Especially when our *a priori* hypotheses consider large sets of genes to be potentially involved in adaptation based on diverse biological functions, it is very likely to retrieve a subset of those candidate genes from genome scans for which a plausible interpretation is then easily constructed (Pavlidis et al. 2012). After all, the majority of genes in an organism's genome have important biological functions. Given the heterotrophic lifestyle and absorptive nutrition of fungi, it seems only logical that many genes in fungal genomes should encode secreted enzymes or transport proteins. Ultimately, biological interpretations should be treated with care and the contribution of putative variants or genes to an adaptive trait should not only rely on genomic evidence but ideally be validated using functional experiments.

Searching for candidate genes involved in host-pathogen interactions without any *a priori* expectation is challenging. In many systems, phenotypic data provides the initial insight and can help to identify putative loci. However, this approach can limit genomic analyses to few large effect loci under strong selection and the conclusion may be drawn that only these loci are involved in coevolutionary interactions. But in reality, pathogenicity is a complex, ever-changing phenotype - a perpetual war between pathogen and host fought simultaneously on many fronts. Our analyses corroborate this by showing that selective sweeps in the genomes of *Epichloe* sibling species are widespread and that many different genes appear to be under recent selection in

different populations including putative effector genes such as small-secreted proteins and cell-wall-degrading enzymes. We further provide evidence that dynamic genome compartments are important for adaptative evolution in *Epichloe* genomes. Thereby this study provides a unique view on the complexity of host-pathogen interactions by shedding light on the genetic basis of host-adaptation in natural populations of fungal plant pathogens and on the processes driving antagonistic coevolution.

ACKNOWLEDGEMENTS

We would like to kindly thank Claudia Michel and Beatrice Arnold for laboratory assistance and Niklaus Zemp for bioinformatic support. Data presented and analyzed here were generated in collaboration with the Genetic Diversity Centre (GDC), ETH Zurich and the Functional Genomics Center Zurich (FGCZ), UZH.

REFERENCES

- Aguileta, G., G. Refrégier, R. Yockteng, E. Fournier, and T. Giraud. 2009. Rapidly evolving genes in pathogens: Methods for detecting positive selection and examples among fungi, bacteria, viruses and protists. *Infection, Genetics and Evolution* 9:656–670.
- Alachiotis, N., and P. Pavlidis. 2018. RAiSD detects positive selection based on multiple signatures of a selective sweep and SNP vectors. *Communications Biology* 1.
- Alexa, A., and J. Rahnenführer. 2009. Gene set enrichment analysis with topGO. *Bioconductor Improv* 27:1–26.
- Ashby, B., and S. Gupta. 2014. Parasitic castration promotes coevolutionary cycling but also imposes a cost on sex. *Evolution* 68:2234–2244.
- Badouin, H., P. Gladieux, J. Gouzy, S. Siguenza, G. Aguileta, A. Snirc, S. Le Prieur, C. Jeziorski, A. Branca, and T. Giraud. 2017. Widespread selective sweeps throughout the genome of model plant pathogenic fungi and identification of effector candidates. *Molecular Ecology* 26:2041–2062.
- Barrett, L. G., J. M. Kniskern, N. Bodenhausen, W. Zhang, and J. Bergelson. 2009. Continua of specificity and virulence in plant host-pathogen interactions: Causes and consequences. *The New Phytologist* 183:513–529.
- Barrett, L. G., P. H. Thrall, J. J. Burdon, and C. C. Linde. 2008. Life history determines genetic structure and evolutionary potential of host-parasite interactions. *Trends in Ecology and Evolution* 23:678–685.

- Bateman, A., E. Birney, L. Cerruti, R. Durbin, L. Ewinger, S. R. Eddy, S. Griffiths-Jones, K. L. Howe, M. Marshall, and E. L. L. Sonnhammer. 2002. The Pfam protein families database. *Nucleic Acids Research* 30:276–280.
- Bates, D., M. Mächler, B. M. Bolker, and S. C. Walker. 2015. Fitting linear mixed-effects models using lme4. *Journal of Statistical Software* 67:1–48.
- Branco, S., K. Bi, H. L. Liao, P. Gladieux, H. Badouin, C. E. Ellison, N. H. Nguyen, R. Vilgalys, K. G. Peay, J. W. Taylor, and T. D. Bruns. 2017. Continental-level population differentiation and environmental adaptation in the mushroom *Suillus brevipes*. *Molecular Ecology* 26:2063–2076.
- Brockhurst, M. A., T. Chapman, K. C. King, J. E. Mank, S. Paterson, and G. D. D. Hurst. 2014. Running with the Red Queen: The role of biotic conflicts in evolution. *Proceedings of the Royal Society B: Biological Sciences* 281:20141382.
- Brown, J. K. M., and A. Tellier. 2011. Plant-parasite coevolution: Bridging the gap between genetics and ecology. *Annual Review of Phytopathology* 49:345–367.
- Burdon, J. J., and P. H. Thrall. 2009. Coevolution of plants and their pathogens in natural habitats. *Science* 324:755–756.
- Charlesworth, B. 2009. Fundamental concepts in genetics: Effective population size and patterns of molecular evolution and variation. *Nature Reviews Genetics* 10:195–205.
- Charlesworth, D. 2006. Balancing selection and its effects on sequences in nearby genome regions. *PLOS Genetics* 2:379–384.
- Chen, S., L. Su, J. Chen, and J. Wu. 2013. Cutinase: Characteristics, preparation, and application. *Biotechnology Advances* 31:1754–1767.
- Chung, K.-R., and C. L. Schardl. 1997. Sexual cycle and horizontal transmission of the grass symbiont, *Epichloë typhina*. *Mycological Research* 101:295–301.
- Cingolani, P., A. Platts, L. L. Wang, M. Coon, T. Nguyen, L. Wang, S. J. Land, X. Lu, and D. M. Ruden. 2012. A program for annotating and predicting the effects of single nucleotide polymorphisms, SnpEff. *Fly* 6:80–92.
- Clay, K., and C. Schardl. 2002. Evolutionary origins and ecological consequences of endophyte symbiosis with grasses. *The American Naturalist* 160:S99–S127.
- Croll, D., and B. A. McDonald. 2012. The accessory genome as a cradle for adaptive evolution in pathogens. *PLOS Pathogens* 8:e1002608.
- Croll, D., and B. A. McDonald. 2016. The genetic basis of local adaptation for pathogenic fungi in agricultural ecosystems. *Molecular Ecology* 26:2027–2040.
- Dobin, A., C. A. Davis, F. Schlesinger, J. Drenkow, C. Zaleski, S. Jha, P. Batut, M. Chaisson, and T. R. Gingeras. 2013. STAR: ultrafast universal RNA-seq aligner. *Bioinformatics* 29:15–21.
- van der Does, H. C., and M. Rep. 2017. Adaptation to the host environment by plant-pathogenic fungi. *Annual Review of Phytopathology* 55:427–450.
- Dong, S., S. Raffaele, and S. Kamoun. 2015. The two-speed genomes of filamentous pathogens: Waltz with plants. *Current Opinion in Genetics and Development* 35:57–65.
- Eaton, C. J., M. P. Cox, B. Ambrose, M. Becker, U. Hesse, C. L. Schardl, and B. Scott. 2010. Disruption of

- signaling in a fungal-grass symbiosis leads to pathogenesis. *Plant Physiology* 153:1780–1794.
- Ebert, D. 2018. Open questions: What are the genes underlying antagonistic coevolution? *BMC Biology* 16:7–9.
- Ebert, D., and P. D. Fields. 2020. Host–parasite co-evolution and its genomic signature. *Nature Reviews Genetics* 21:754–768.
- Ellison, C. E., C. Hall, D. Kowbel, J. Welch, R. B. Brem, N. L. Glass, and J. W. Taylor. 2011. Population genomics and local adaptation in wild isolates of a model microbial eukaryote. *Proceedings of the National Academy of Sciences of the United States of America* 108:2831–2836.
- Eschenbrenner, C. J., A. Feurtey, and E. H. Stukenbrock. 2020. Population genomics of fungal plant pathogens and the analyses of rapidly evolving genome compartments. Pages 337–355 in J. Y. Duthiel, editor. *Statistical Population Genomics, Methods in Molecular Biology*. 2090th edition. Humana, New York, NY.
- Faino, L., M. F. Seidl, X. Shi-Kunne, M. Pauper, G. C. M. Van Den Berg, A. H. J. Wittenberg, and B. P. H. J. Thomma. 2016. Transposons passively and actively contribute to evolution of the two-speed genome of a fungal pathogen. *Genome Research* 26:1091–1100.
- Franceschetti, M., A. Maqbool, M. J. Jiménez-Dalmaroni, H. G. Pennington, S. Kamoun, and M. J. Banfield. 2017. Effectors of filamentous plant pathogens: Commonalities amid diversity. *Microbiology and Molecular Biology Reviews* 81:e00066-16.
- Gandon, S., and Y. Michalakis. 2002. Local adaptation, evolutionary potential and host-parasite coevolution: Interactions between migration, mutation, population size and generation time. *Journal of Evolutionary Biology* 15:451–462.
- Gandon, S., and S. L. Nuismer. 2009. Interactions between genetic drift, gene flow, and selection mosaics drive parasite local adaptation. *American Naturalist* 173:212–224.
- Gao, F., B. Sen Zhang, J. H. Zhao, J. F. Huang, P. S. Jia, S. Wang, J. Zhang, J. M. Zhou, and H. S. Guo. 2019. Deacetylation of chitin oligomers increases virulence in soil-borne fungal pathogens. *Nature Plants* 5:1167–1176.
- Gautier, M., A. Klassmann, and R. Vitalis. 2017. rehh 2.0: A reimplement of the R package rehh to detect positive selection from haplotype structure. *Molecular Ecology Resources* 17:78–90.
- Gel, B., A. Díez-Villanueva, E. Serra, M. Buschbeck, M. A. Peinado, and R. Malinverni. 2016. regioneR: An R/Bioconductor package for the association analysis of genomic regions based on permutation tests. *Bioinformatics* 32:289–291.
- Gladieux, P., J. Ropars, H. Badouin, A. Branca, G. Aguilera, D. M. De Vienne, R. C. Rodríguez de la Vega, S. Branco, and T. Giraud. 2014. Fungal evolutionary genomics provides insight into the mechanisms of adaptive divergence in eukaryotes. *Molecular Ecology* 23:753–773.
- Grandaubert, J., R. G. T. Lowe, J. L. Soyer, C. L. Schoch, A. P. Van De Wouw, I. Fudal, B. Robbertse, N. Lapalu, M. G. Links, B. Ollivier, J. Linglin, V. Barbe, S. Mangenot, C. Cruaud, H. Borhan, B. J. Howlett, M. H. Balesdent, and T. Rouxel. 2014. Transposable element-assisted evolution and adaptation to host plant within the *Leptosphaeria maculans*-*Leptosphaeria biglobosa* species complex of fungal pathogens. *BMC Genomics* 15:1–27.

- Hartmann, F. E., B. A. McDonald, and D. Croll. 2018. Genome-wide evidence for divergent selection between populations of a major agricultural pathogen. *Molecular Ecology* 27:2725–2741.
- Horscroft, C., S. Ennis, R. J. Pengelly, T. J. Sluckin, and A. Collins. 2019. Sequencing era methods for identifying signatures of selection in the genome. *Briefings in Bioinformatics* 20:1997–2008.
- Jousimo, J., A. J. M. Tack, O. Ovaskainen, T. Mononen, H. Susi, C. Tollenaere, and A. L. Laine. 2014. Ecological and evolutionary effects of fragmentation on infectious disease dynamics. *Science* 344:1289–1293.
- Koropoulis, A., N. Alachiotis, and P. Pavlidis. 2020. Detecting positive selection in populations using genetic data. Pages 87–123 in J. Y. Dutheil, editor. *Statistical Population Genomics, Methods in Molecular Biology*. Humana Press Inc. New York.
- Kubicek, C. P., T. L. Starr, and N. L. Glass. 2014. Plant cell wall-degrading enzymes and their secretion in plant-pathogenic fungi. *Annual Review of Phytopathology* 52:427–451.
- Kurtz, J., H. Schulenburg, and T. B. H. Reusch. 2016. Host–parasite coevolution—rapid reciprocal adaptation and its genetic basis. *Zoology* 119:241–243.
- Laine, A. L. 2005. Spatial scale of local adaptation in a plant-pathogen metapopulation. *Journal of Evolutionary Biology* 18:930–938.
- Li, H., and R. Durbin. 2009. Fast and accurate short read alignment with Burrows-Wheeler transform. *Bioinformatics* 25:1754–1760.
- Li, H., B. Handsaker, A. Wysoker, T. Fennell, J. Ruan, N. Homer, G. Marth, G. Abecasis, R. Durbin, and 1000 Genome Project Data Processing Subgroup. 2009. The Sequence Alignment/Map format and SAMtools. *Bioinformatics* 25:2078–2079.
- Liao, Y., G. K. Smyth, and W. Shi. 2019. The R package Rsubread is easier, faster, cheaper and better for alignment and quantification of RNA sequencing reads. *Nucleic Acids Research* 47:e47–e47.
- Lyu, X., C. Shen, Y. Fu, J. Xie, D. Jiang, G. Li, and J. Cheng. 2015. Comparative genomic and transcriptional analyses of the carbohydrate-active enzymes and secretomes of phytopathogenic fungi reveal their significant roles during infection and development. *Scientific Reports* 5:15565.
- Ma, L.-J., H. C. Van Der Does, K. A. Borkovich, J. J. Coleman, M.-J. Daboussi, A. Di Pietro, M. Dufresne, M. Freitag, M. Grabherr, and B. Henrissat. 2010. Comparative analysis reveals mobile pathogenicity chromosomes in *Fusarium*. *Nature* 464:367–373.
- McDonald, B. A., and E. H. Stukenbrock. 2016. Rapid emergence of pathogens in agro-ecosystems: Global threats to agricultural sustainability and food security. *Philosophical Transactions of the Royal Society B: Biological Sciences* 371:20160026.
- McKenna, A., M. Hanna, E. Banks, A. Sivachenko, K. Cibulskis, A. Kernytsky, K. Garimella, D. Altshuler, S. Gabriel, and M. Daly. 2010. The Genome Analysis Toolkit: A MapReduce framework for analyzing next-generation DNA sequencing data. *Genome Research* 20:1297–1303.
- Mohd-Assaad, N., B. A. McDonald, and D. Croll. 2018. Genome-wide detection of genes under positive selection in worldwide populations of the barley scald pathogen. *Genome Biology and Evolution* 10:1315–1332.
- Möller, M., and E. H. Stukenbrock. 2017. Evolution and genome architecture in fungal plant pathogens.

- Nature Reviews Microbiology 15:756–771.
- Nielsen, R., S. Williamson, Y. Kim, M. J. Hubisz, A. G. Clark, and C. Bustamante. 2005. Genomic scans for selective sweeps using SNP data. *Genome Research* 15:1566–1575.
- Papkou, A., T. Guzella, W. Yang, S. Koepper, B. Pees, R. Schalkowski, M. C. Barg, P. C. Rosenstiel, H. Teotónio, and H. Schulenburg. 2019. The genomic basis of red queen dynamics during rapid reciprocal host–pathogen coevolution. *Proceedings of the National Academy of Sciences of the United States of America* 116:923–928.
- Paterson, S., T. Vogwill, A. Buckling, R. Benmayor, A. J. Spiers, N. R. Thomson, M. Quail, F. Smith, D. Walker, B. Libberton, A. Fenton, N. Hall, and M. A. Brockhurst. 2010. Antagonistic coevolution accelerates molecular evolution. *Nature* 464:275–278.
- Pavlidis, P., and N. Alachiotis. 2017. A survey of methods and tools to detect recent and strong positive selection. *Journal of Biological Research-Thessaloniki* 24:7.
- Pavlidis, P., J. D. Jensen, W. Stephan, and A. Stamatakis. 2012. A critical assessment of storytelling: Gene ontology categories and the importance of validating genomic scans. *Molecular Biology and Evolution* 29:3237–3248.
- Pavlidis, P., D. Živković, A. Stamatakis, and N. Alachiotis. 2013. SweeD: Likelihood-based detection of selective sweeps in thousands of genomes. *Molecular Biology and Evolution* 30:2224–2234.
- Persoons, A., K. J. Hayden, B. Fabre, P. Frey, S. De Mita, A. Tellier, and F. Halkett. 2017. The escalatory Red Queen: Population extinction and replacement following arms race dynamics in poplar rust. *Molecular Ecology* 26:1902–1918.
- Pfeifer, S. P. 2017. From next-generation resequencing reads to a high-quality variant data set. *Heredity* 118:111–124.
- Plissonneau, C., J. Benevenuto, N. Mohd-Assaad, S. Fouché, F. E. Hartmann, and D. Croll. 2017. Using population and comparative genomics to understand the genetic basis of effector-driven fungal pathogen evolution. *Frontiers in Plant Science* 8:1–15.
- Plissonneau, C., F. E. Hartmann, and D. Croll. 2018. Pangenome analyses of the wheat pathogen *Zymoseptoria tritici* reveal the structural basis of a highly plastic eukaryotic genome. *BMC Biology* 16:5.
- Lo Presti, L., D. Lanver, G. Schweizer, S. Tanaka, L. Liang, M. Tollot, A. Zuccaro, S. Reissmann, and R. Kahmann. 2015. Fungal effectors and plant susceptibility. *Annual Review of Plant Biology* 66:513–545.
- Raffaele, S., and S. Kamoun. 2012. Genome evolution in filamentous plant pathogens: Why bigger can be better. *Nature Reviews Microbiology* 10:417–430.
- Robinson, M. D., D. J. McCarthy, and G. K. Smyth. 2010. edgeR: A Bioconductor package for differential expression analysis of digital gene expression data. *Bioinformatics* 26:139–140.
- Rouxel, T., J. Grandaubert, J. K. Hane, C. Hoede, A. P. van de Wouw, A. Couloux, V. Dominguez, V. Anthouard, P. Bally, S. Bourras, A. J. Cozijnsen, L. M. Ciuffetti, A. Degrave, A. Dilmaghani, L. Duret, I. Fudal, S. B. Goodwin, L. Gout, N. Glaser, J. Linglin, G. H. J. Kema, N. Lapalu, C. B. Lawrence, K. May, M. Meyer, B. Ollivier, J. Poulain, C. L. Schoch, A. Simon, J. W. Spatafora, A. Stachowiak, B. G.

- Turgeon, B. M. Tyler, D. Vincent, J. Weissenbach, J. Amselem, H. Quesneville, R. P. Oliver, P. Wincker, M.-H. Balesdent, and B. J. Howlett. 2011. Effector diversification within compartments of the *Leptosphaeria maculans* genome affected by Repeat-Induced Point mutations. *Nature Communications* 2:202.
- Schardl, C. L., S. Florea, J. Pan, P. Nagabhyru, S. Bec, and P. J. Calie. 2013a. The epichloae: alkaloid diversity and roles in symbiosis with grasses. *Current Opinion in Plant Biology* 16:480–488.
- Schardl, C. L., C. A. Young, J. Pan, S. Florea, J. E. Takach, D. G. Panaccione, M. L. Farman, J. S. Webb, J. Jaromczyk, N. D. Charlton, P. Nagabhyru, L. Chen, C. Shi, and A. Leuchtmann. 2013b. Currencies of mutualisms: Sources of alkaloid genes in vertically transmitted epichloae. *Toxins* 5:1064–1088.
- Schirawski, J., G. Mannhaupt, K. Münch, T. Brefort, K. Schipper, G. Doehlemann, M. Di Stasio, N. Rössel, A. Mendoza-Mendoza, D. Pester, O. Müller, B. Winterberg, E. Meyer, H. Ghareeb, T. Wollenberg, M. Münsterkötter, P. Wong, M. Walter, E. Stukenbrock, U. Güldener, and R. Kahmann. 2010. Pathogenicity determinants in smut fungi revealed by genome comparison. *Science* 330:1546–1548.
- Schirrmann, M. K., S. Zoller, D. Croll, E. H. Stukenbrock, A. Leuchtmann, and S. Fior. 2018. Genomewide signatures of selection in *Epichloë* reveal candidate genes for host specialization. *Molecular Ecology* 27:3070–3086.
- Seidl, M. F., and B. P. H. J. Thomma. 2017. Transposable elements direct the coevolution between plants and microbes. *Trends in Genetics* 33:842–851.
- Seifbarghi, S., M. H. Borhan, Y. Wei, C. Coutu, S. J. Robinson, and D. D. Hegedus. 2017. Changes in the *Sclerotinia sclerotiorum* transcriptome during infection of *Brassica napus*. *BMC Genomics* 18:1–37.
- Smith, D. L., L. Ericson, and J. J. Burdon. 2011. Co-evolutionary hot and cold spots of selective pressure move in space and time. *Journal of Ecology* 99:634–641.
- Stukenbrock, E. H., and D. Croll. 2014. The evolving fungal genome. *Fungal Biology Reviews* 28:1–12.
- Tack, A. J. M., P. H. Thrall, L. G. Barrett, J. J. Burdon, and A. L. Laine. 2012. Variation in infectivity and aggressiveness in space and time in wild host-pathogen systems: Causes and consequences. *Journal of Evolutionary Biology* 25:1918–1936.
- Tajima, F. 1989. Statistical method for testing the neutral mutation hypothesis by DNA polymorphism. *Genetics* 123:585–595.
- Tellier, A., S. Moreno-Gámez, and W. Stephan. 2014. Speed of adaptation and genomic footprints of host-parasite coevolution under arms race and trench warfare dynamics. *Evolution* 68:2211–2224.
- Thomas, P. D., M. J. Campbell, A. Kejariwal, H. Mi, B. Karlak, R. Daverman, K. Diemer, A. Muruganujan, and A. Narechania. 2003. PANTHER: A library of protein families and subfamilies indexed by function. *Genome Research* 13:2129–2141.
- Thompson, J. N. 2005. *The geographic mosaic of coevolution*. University of Chicago Press, Chicago.
- Torres, D. E., U. Oggenfuss, D. Croll, and M. F. Seidl. 2020. Genome evolution in fungal plant pathogens: Looking beyond the two-speed genome model. *Fungal Biology Reviews* 34:136–143.
- Treangen, T. J., and S. L. Salzberg. 2012. Repetitive DNA and next-generation sequencing: Computational challenges and solutions. *Nature Reviews Genetics* 13:36–46.
- Turrà, D., D. Segorbe, and A. Di Pietro. 2014. Protein kinases in plant-pathogenic fungi: conserved

- regulators of infection. *Annual Review of Phytopathology* 52:267–288.
- Vitti, J. J., S. R. Grossman, and P. C. Sabeti. 2013. Detecting natural selection in genomic data. *Annual Review of Genetics* 47:97–120.
- Weigand, H., and F. Leese. 2018. Detecting signatures of positive selection in non-model species using genomic data. *Zoological Journal of the Linnean Society* 184:528–583.
- Weir, B. S., and C. C. Cockerham. 1984. Estimating F-statistics for the analysis of population structure. *Evolution* 38:1358–1370.
- White, J. F., and T. L. Bultman. 1987. Endophyte-host associations in forage grasses. VIII. Heterothallism in *Epichloë typhina*. *American Journal of Botany* 74:1716–1721.
- Winter, D. J., A. R. D. Ganley, C. A. Young, I. Liachko, C. L. Schardl, P. Y. Dupont, D. Berry, A. Ram, B. Scott, and M. P. Cox. 2018. Repeat elements organise 3D genome structure and mediate transcription in the filamentous fungus *Epichloë festucae*. *PLoS Genetics* 14:1–29.
- Woolhouse, M. E. J., J. P. Webster, E. Domingo, B. Charlesworth, and B. R. Levin. 2002. Biological and biomedical implications of the co-evolution of pathogens and their hosts. *Nature Genetics* 32:569–577.
- Van de Wouw, A. P., A. J. Cozijnsen, J. K. Hane, P. C. Brunner, B. A. McDonald, R. P. Oliver, and B. J. Howlett. 2010. Evolution of linked avirulence effectors in *Leptosphaeria maculans* is affected by genomic environment and exposure to resistance genes in host plants. *PLOS Pathogens* 6:e1001180.
- Yang, Y., Y. Zhang, B. Li, X. Yang, Y. Dong, and D. Qiu. 2018. A *Verticillium dahliae* pectate lyase induces plant immune responses and contributes to virulence. *Frontiers in Plant Science* 9:1–15.
- Young, C. A., C. L. Schardl, D. G. Panaccione, S. Florea, J. E. Takach, N. D. Charlton, N. Moore, J. S. Webb, and J. Jaromczyk. 2015. Genetics, genomics and evolution of ergot alkaloid diversity. *Toxins* 7:1273–1302.

SUPPLEMENTARY TABLES AND FIGURES

Table S1: *E. typhina* populations analyzed in this study

	<i>Host species</i>	<i>Origin</i>	Collection year	n
Aub	<i>D. glomerata</i>	Aubonne, Switzerland	2018	22
AubX	<i>D. glomerata</i>	Aubonne, Switzerland	2005	16
Auv	<i>D. glomerata</i>	Auvergne, France	2017	27
Cev	<i>D. glomerata</i>	Cevennes, France	2017	31
El	<i>D. glomerata</i>	Alsace, France	2017	26
Kew	<i>D. glomerata</i>	London, United Kingdom	2017	14
MdR	<i>D. glomerata</i>	Montemayor del Rio, Spain	2018	33
So	<i>D. glomerata</i>	Soglio, Switzerland	2016	22
Total n				191

Table S2: *E. clarkii* populations analyzed in this study

	<i>Host species</i>	<i>Origin</i>	Collection year	n
Aub	<i>H. lanatus</i>	Aubonne, Switzerland	2018	27
AubX	<i>H. lanatus</i>	Aubonne, Switzerland	2005	18
Auv	<i>H. lanatus</i>	Auvergne, France	2017	30
Cev	<i>H. lanatus</i>	Cevennes, France	2017	28
El	<i>H. lanatus</i>	Alsace, France	2017	27
Kew	<i>H. mollis</i>	London, United Kingdom	2017	19
MdR	<i>H. lanatus</i>	Montemayor del Rio, Spain	2018	29
So	<i>H. lanatus</i>	Soglio, Switzerland	2016	30
Total n				208

Table S3: Summary statistics of selective sweeps identified in *E. typhina* populations

<i>E. typhina</i>	Aub	Auv	Cev	El	Kew	MdR	So	all
iHS selection scan								
number of sweep regions	30	39	32	38	19	38	32	228
mean length of sweeps ^a	14834	14488	15157	14475	15790	14869	14689	14900
median length of sweeps ^a	15001	15001	15001	15001	15001	15001	15001	15001
mean no. SNPs per sweep	89.0	94.4	109.6	92.6	81.0	93.4	80.4	91.5
mean no. extremal SNPs	3.2	2.8	3.5	2.8	2.9	2.9	2.6	2.9
mean iHS extremal SNPs	3.83	3.93	3.99	3.58	4.10	4.38	3.89	4.0
max length ^a	30001	15001	25001	20001	25001	25001	25001	-
total length ^a	445030	565039	485032	550038	300019	565038	470032	-
% of total genome	1.32%	1.67%	1.43%	1.63%	0.89%	1.67%	1.39%	1.43%
overlaps with AT-rich regions	26	33	27	29	14	33	28	190
% overlap AT-rich regions	86.67%	84.62%	84.38%	76.32%	73.68%	86.84%	87.50%	83.33%
CLR selection scan								
number of sweep regions	22	21	24	10	13	19	15	124
mean length of sweeps ^a	28782	26749	22133	56845	53637	30835	66308	40755
median length of sweeps ^a	18922	21017	18864	33334	34153	28796	35699	27255
mean CLR outlier windows	15.47	15.05	10.65	41.80	26.38	16.03	27.92	21.90
mean CLR alpha ^b	2.4E-03	2.5E-03	2.5E-03	8.2E-04	8.9E-04	1.2E-03	6.9E-04	1.6E-03
max length ^a	103393	81644	68089	208817	213647	75354	233416	-
total length ^a	633193	561726	531203	568447	697275	585867	994621	-
% of total genome	1.87%	1.66%	1.57%	1.68%	2.06%	1.73%	2.94%	1.93%
overlaps with AT-rich regions	19	18	18	10	12	17	14	108
% overlap AT-rich regions	86.36%	85.71%	75.00%	100%	92.31%	89.47%	93.33%	87.10%
Method overlap	1	4	3	2	5	2	4	21

^a refers to length in basepairs

^b α -value indicates estimated strength of selection as a function of the selection coefficient, the recombination rate and the effective population size

% overlap AT-rich regions refers to the proportional number of sweeps overlapping AT-rich regions, not the proportion in basepairs

Table S4: Summary statistics of selective sweeps identified in *E. clarkii* populations

<i>E. clarkii</i>	Aub	Auv	Cev	El	Kew	MdR	So	all
iHS selection scan								
number of sweep regions	13	12	15	12	9	12	17	90
mean length of sweeps ^a	30001	30001	31334	29168	31112	30001	30589	30315
median length of sweeps ^a	30001	30001	30001	30001	30001	30001	30001	30001
mean no. SNPs per sweep	81.5	68.5	59.0	91.8	102.6	83.2	80.9	81.0
mean no. extremal SNPs	2.8	3.3	3.5	2.6	2.9	2.6	4.2	3.1
mean iHS extremal SNPs	3.95	3.89	4.15	3.83	7.01	6.14	5.06	4.9
max length ^a	30001	30001	50001	30001	50001	40001	60001	-
total length ^a	390013	360012	470015	350012	280009	360012	520017	-
% of total genome	0.85%	0.79%	1.03%	0.77%	0.61%	0.79%	1.14%	0.85%
overlaps with AT-rich regions	11	12	15	11	9	11	17	86
% overlap AT-rich regions	84.62%	100%	100%	91.67%	100%	91.67%	100%	95.56%
CLR selection scan								
number of sweep regions	16	6	17	15	3	12	18	87
mean length of sweeps ^a	28376	114654	31849	45902	613902	111575	24136	138628
median length of sweeps ^a	22354	72471	20672	22970	611432	60136	14205	117748
mean CLR outlier windows	14.65	27.06	11.19	11.70	63.88	19.70	11.54	22.82
mean CLR alpha ^b	1.9E-03	2.5E-04	1.4E-03	1.2E-03	3.3E-05	7.1E-04	2.0E-03	1.1E-03
max length ^a	122921	342360	76551	237223	788713	432394	77266	-
total length ^a	454011	687924	541433	688524	1841705	1338904	434439	-
% of total genome	1.00%	1.51%	1.19%	1.51%	4.04%	2.93%	0.95%	1.87%
overlaps with AT-rich regions	16	5	15	10	3	12	17	78
% overlap AT-rich regions	100%	83.33%	88.24%	66.67%	100%	100%	94.44%	89.66%
Method overlap	0	0	1	0	0	0	1	2

^a refers to length in basepairs

^b α -value indicates estimated strength of selection as a function of the selection coefficient, the recombination rate and the effective population size

Table S5: Sharedness of selective sweeps among *E. typhina* populations

<i>E. typhina</i> sharedness	iHS-sweeps	iHS-regions	CLR-sweeps	CLR-regions
unique to 1	118	118	51	51
shared by 2	62	30	40	24
shared by 3	30	10	16	6
shared by 4	12	3	12	3
shared by 5	5	1	5	1
shared by 6	0	0	0	0
shared by 7	0	0	0	0
Total	228	163	124	85

Sweeps were categorized based on how many populations showed an overlapping sweep signature using the same selection scan method (columns iHS-sweeps and CLR-sweeps); The columns iHS-regions and CLR-regions indicate how many distinct genomic regions show signature of selection across populations within the species.

Table S6: Sharedness of selective sweeps among *E. clarkii* populations

<i>E. clarkii</i> sharedness	iHS-sweeps	iHS-regions	CLR-sweeps	CLR-regions
unique to 1	39	39	50	50
shared by 2	36	19	24	13
shared by 3	3	1	3	1
shared by 4	12	3	4	1
shared by 5	0	0	0	0
shared by 6	0	0	6	1
shared by 7	0	0	0	0
Total	90	62	87	66

Sweeps were categorized based on how many populations showed an overlapping sweep signature using the same selection scan method (columns iHS-sweeps and CLR-sweeps); The columns iHS-regions and CLR-regions indicate how many distinct genomic regions show signature of selection across populations within the species.

Table S7: Biological functions overrepresented among genes up- or down-regulated in-planta versus in-culture for *E. typhina* and *E. clarkii* (Fisher's exact test: $p < 0.01$)

	GO ID	GO term	Genome-wide annotations	Significant annotations	Enrichment p-value	
<i>E. typhina in planta</i>	up	GO:0043386	mycotoxin biosynthetic process	15	6	6.9E-06
		GO:0055085	transmembrane transport	310	37	7.6E-05
		GO:0055114	oxidation-reduction process	349	20	1.9E-03
	down	GO:0055085	transmembrane transport	310	37	7.6E-05
		GO:0009405	pathogenesis	7	4	3.7E-04
		GO:0035434	copper ion transmembrane transport	4	3	9.5E-05
	GO:0016125	sterol metabolic process	3	2	2.5E-03	
<i>E. clarkii in planta</i>	up	GO:0043386	mycotoxin biosynthetic process	15	5	1.5E-03
		GO:0055114	oxidation-reduction process	362	35	2.5E-03
		GO:0055085	transmembrane transport	310	31	2.7E-03
		GO:0046174	polyol catabolic process	4	3	8.6E-04
		GO:0030004	cellular monovalent inorganic cation homeostasis	2	2	2.3E-03
	down	GO:0055085	transmembrane transport	310	24	1.1E-03
		GO:0035434	copper ion transmembrane transport	4	3	6.0E-05
		GO:0006094	gluconeogenesis	4	2	3.7E-03
		GO:0031505	fungal-type cell wall organization	4	2	3.7E-03
		GO:0006031	chitin biosynthetic process	3	2	3.8E-03
	GO:0071423	malate transmembrane transport	3	2	1.9E-03	
	GO:0007623	circadian rhythm	2	2	1.3E-03	
	GO:0006094	gluconeogenesis	3	2	3.8E-03	

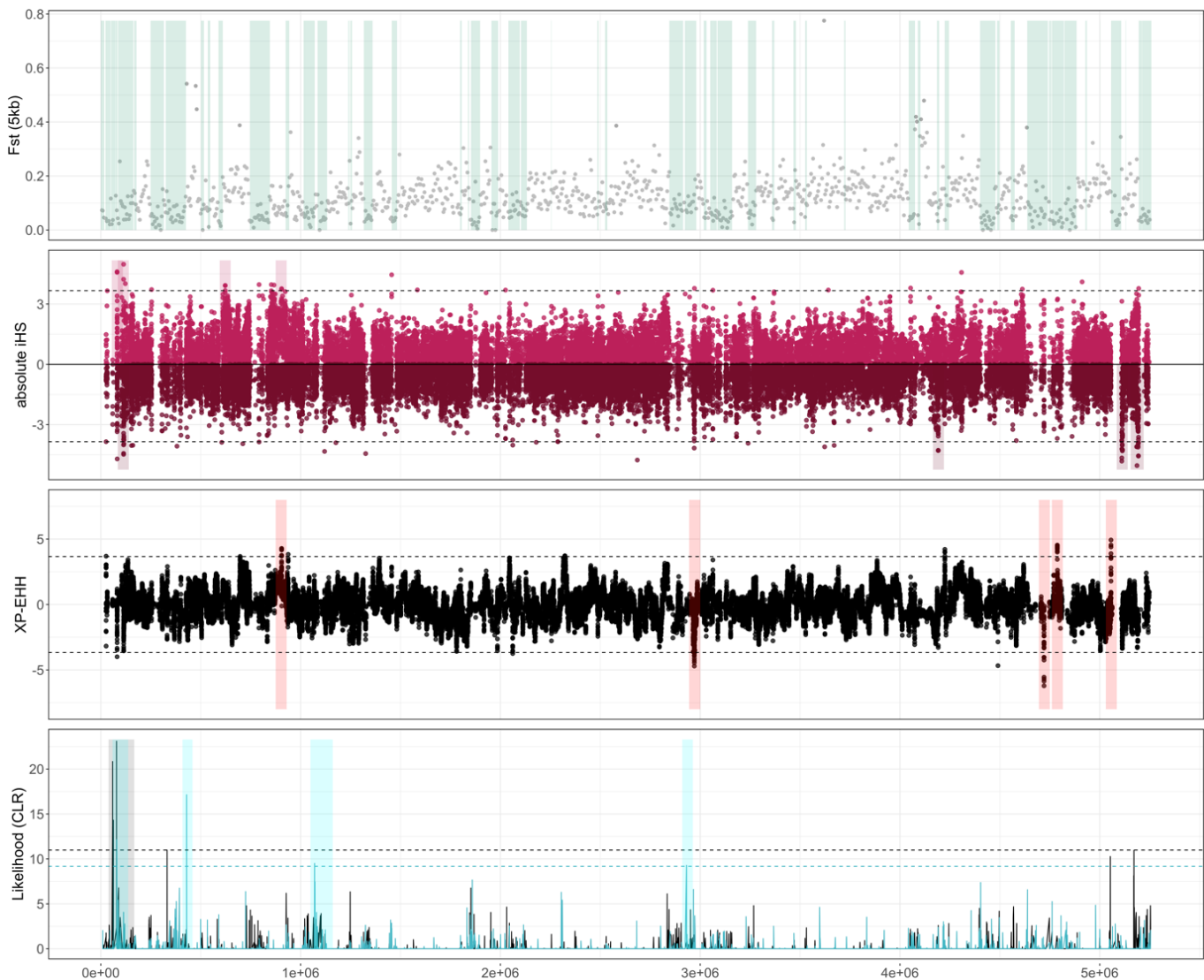


Figure S1: Signatures of selective sweeps on *E. typhina* chromosome 4

Chromosome-wide raw data and sweeps identified by different analyses are shown for the population pair Kew–MdR. In the top panel, pairwise F_{ST} values are plotted, averaged across 5kb windows. Shaded rectangles represent the locations of AT-rich regions. The second panel shows the absolute values of the integrated haplotype score (iHS) calculated at each SNP locus for which the ancestral allele state was known. Scores for Kew are shown at the top and scores for So are negatively transformed and showed at the bottom. Horizontal dashed lines indicate the 99.9% percentile threshold which was used as a cutoff to identify outlier SNPs and inferred iHS sweeps are shown as shaded rectangles. The third panel shows the cross-population extended haplotype homozygosity (XP-EHH) scores calculated between the two populations. Dashed lines indicate 99.9% percentile threshold which was used as a cutoff to identify outlier SNPs and inferred divergent sweeps are shown as shaded rectangles. Positive and negative XP-EHH values refer to the direction of selection: positive values indicate selection in Kew, negative values indicate selection in So. In the bottom panel, composite likelihood ratio (CLR) scores are plotted for Kew (black) and So (blue), colored dashed lines indicate respective 99.9% threshold and colored rectangles highlight inferred CLR-sweeps. Note the sub-telomeric region at the beginning of the chromosome in which a selective sweep was detected in both populations and by both iHS and CLR-scans. The sweep region overlaps with an AT-rich region.

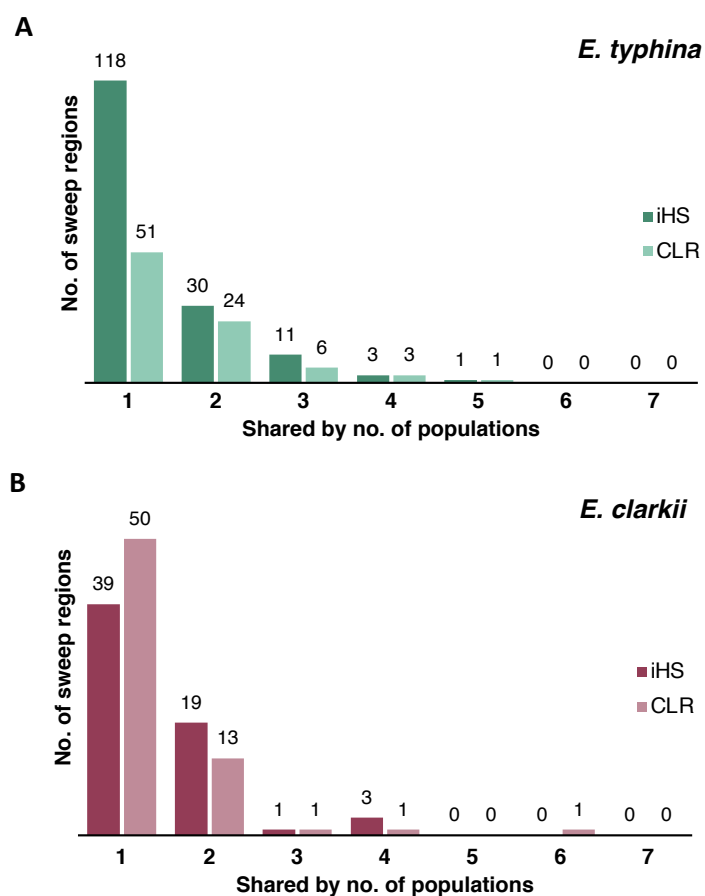


Figure S2: Sharedness of selective sweep regions among *Epichloe* populations.

Most selective sweeps were unique to a single population or shared by two populations. Some regions showed evidence of selection across multiple populations. Sweeps were considered shared when genomic positions overlapped between populations and joined into one sweep region. We performed this separately for either selection scan method. For each sweep region, we counted the number of populations that shared a common signature of a selective sweep.

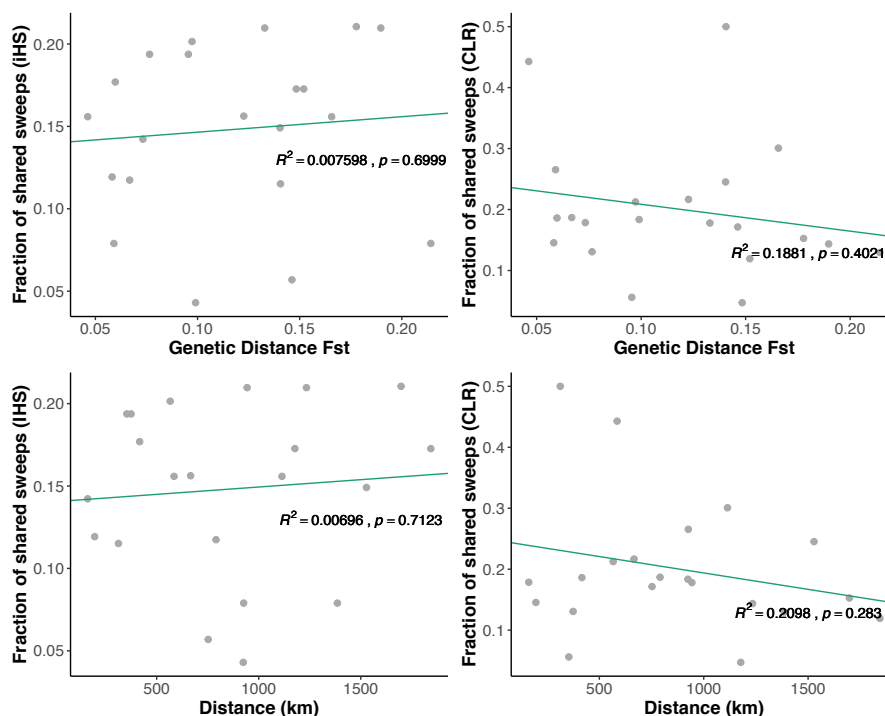


Figure S3: Proportion of shared sweeps relative to genetic and geographic distance among *E. typhina* populations.

We found no significant correlations between the sharedness of sweeps and distance with either selection scan method, however the proportion of shared CLR-sweeps showed a decreasing trend with both increasing genetic and geographic distance.

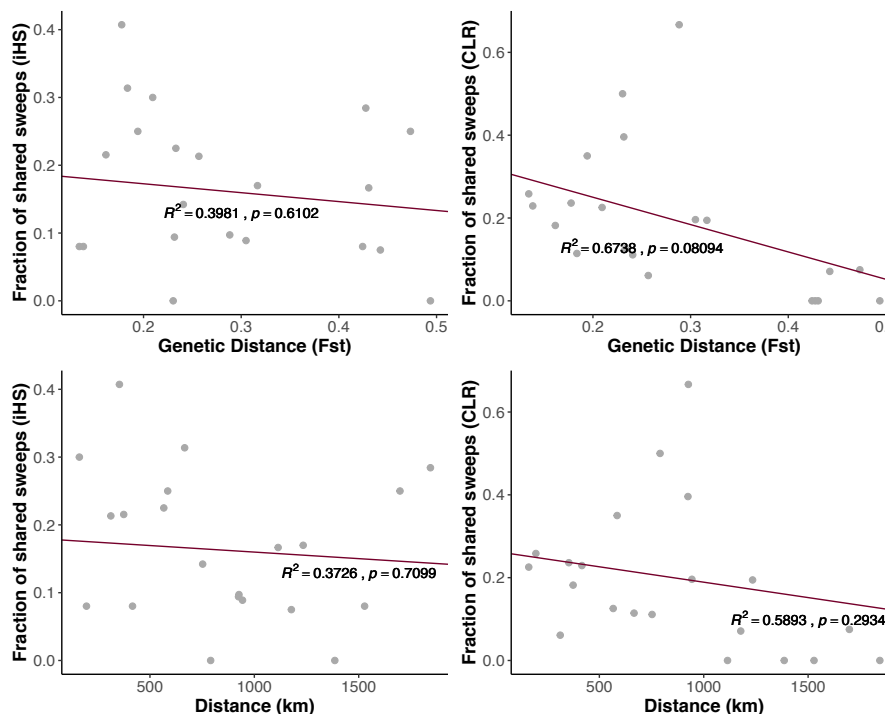


Figure S4: Proportion of shared sweeps relative to genetic and geographic distance among *E. clarkii* populations.

We found no significant correlations between the sharedness of sweeps and distance with either selection scan method, however the proportion of shared CLR-sweeps showed a decreasing trend with both increasing genetic and geographic distance.

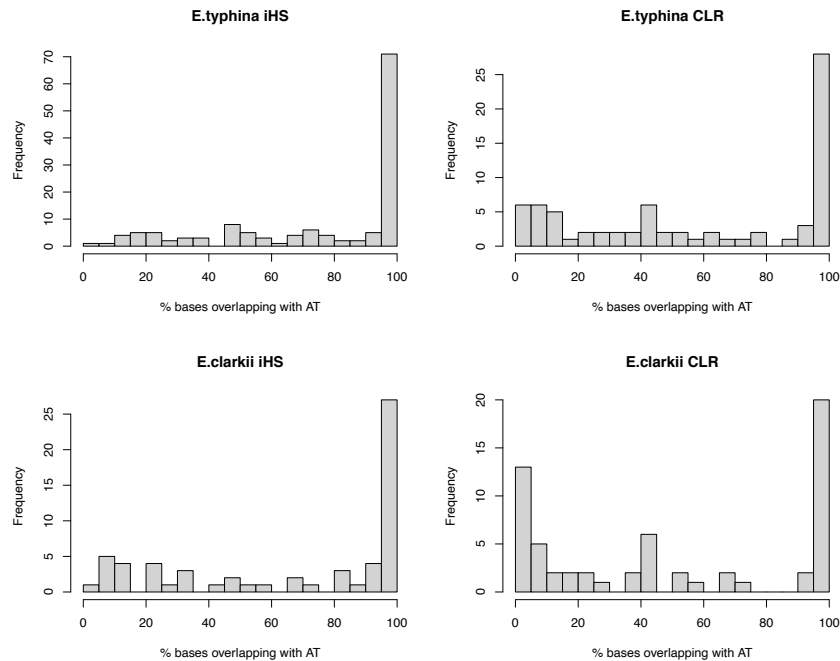


Figure S5: Proportional overlap of selective sweep regions with AT-rich regions.

Overlap was calculated as the percentage of base pairs located in AT-rich regions in *Epichloe* genomes. Many sweep regions were located fully in AT-rich regions.

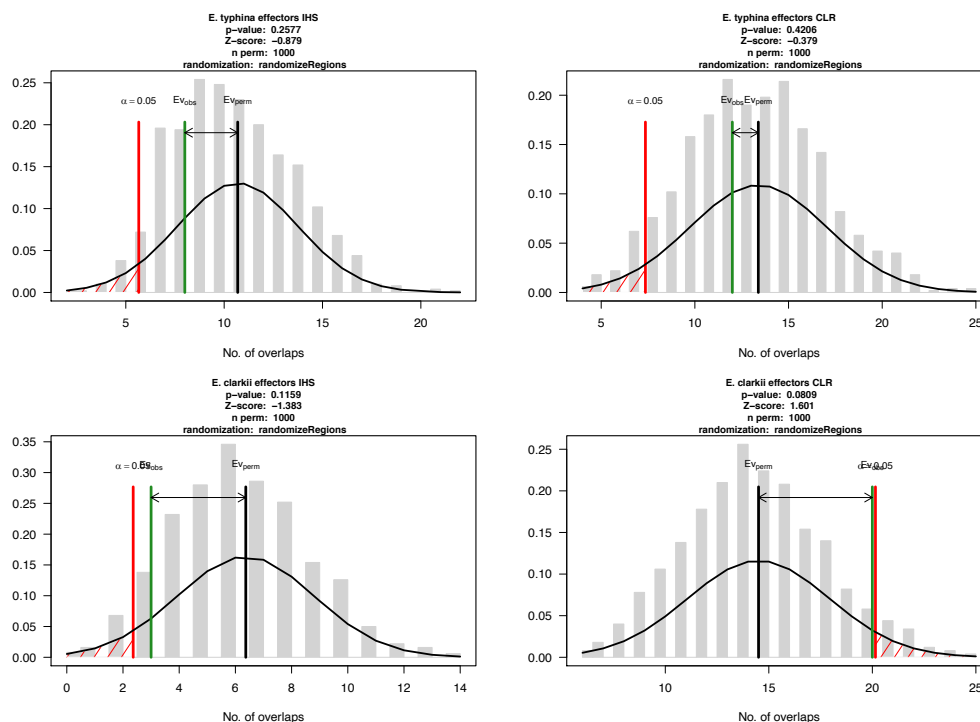
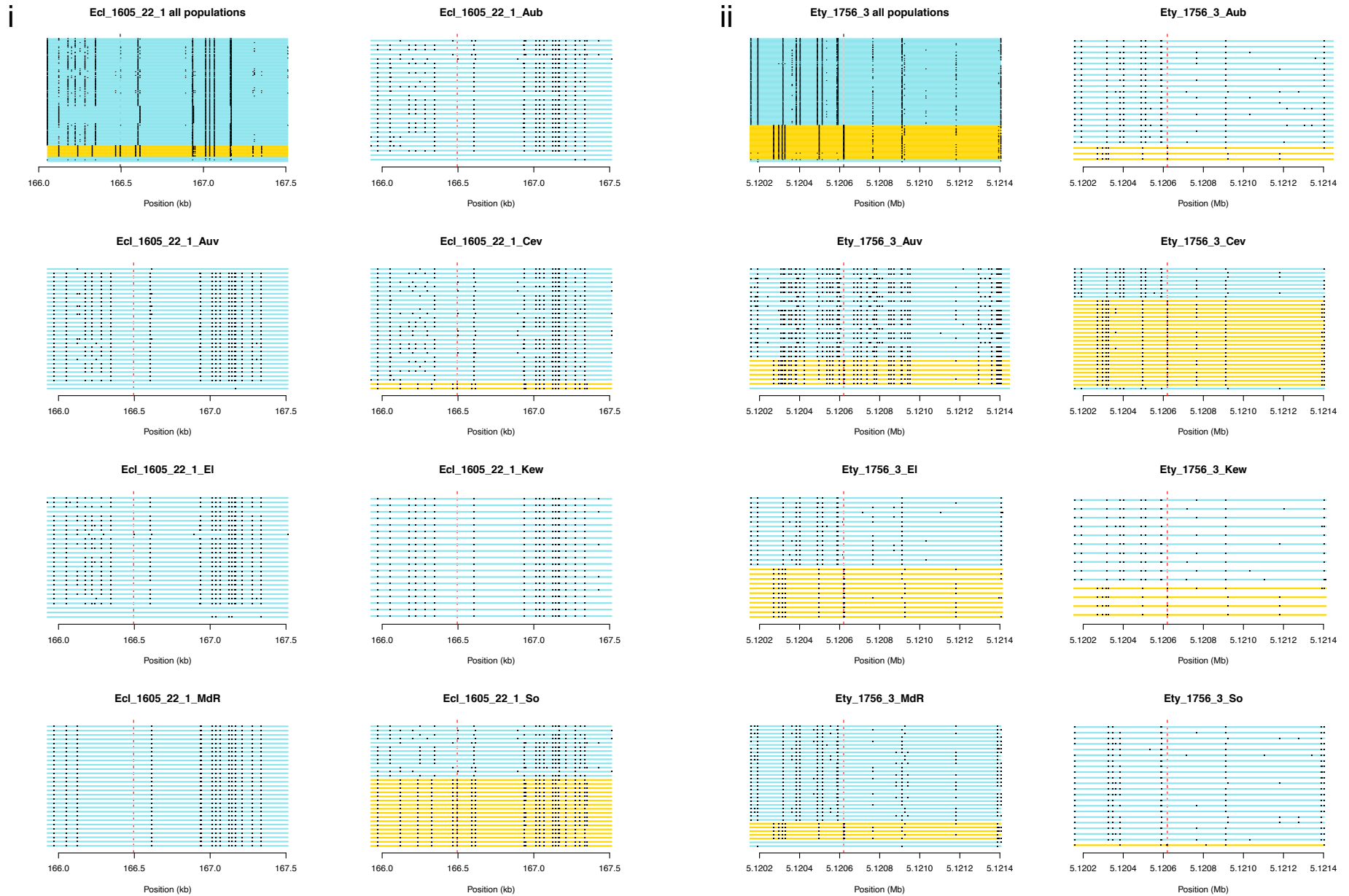


Figure S6: Association between effectors and selective sweeps.

Plots show graphical representation of the permutation test performed where the grey histogram represents the number of overlaps using a randomized region set with a fitted normal, $E_{v_{perm}}$ (black line) indicates the mean of the randomized counts and $E_{v_{obs}}$ (green line) indicates the observed overlap count. In both *E. typhina* (top) and *E. clarkii* (bottom) and for sweeps detected by either selection scan-method (iHS left, CLR right), putative effector genes were not more likely to be located in selective sweep regions than expected by chance



*Relates to previous page: **Figure S7: Within-population haplotype structure for homologous candidate effector locus on chromosome one in *E. clarkii* (i) and chromosome three in *E. typhina* (ii).** Haplotypes for all isolates are shown in the top left and in the remaining seven plots for each population separately. Haplotypes were colored relative to the highlighted SNP position (see Figures 3 and 4 in main text for details about candidate SNP loci). Blue lines at the bottom of the plots indicate individuals that have not been genotyped at this locus. **(i)** The most common blue haplotype is present in all populations for which a CLR-sweep was inferred, all except MdR which is fixed for a different haplotype. Five populations (Aub, Auv, El, Kew and MdR) are fixed for one haplotype. The yellow haplotype is present at intermediate frequency in So with less variation compared to the alternative (blue) haplotype and at low frequency in Cev; **(ii)** Two different haplotypes can be distinguished at the locus both of which are present at different frequencies in all seven populations. So, the population for which the CLR-scan inferred a sweep is almost fixed for the blue haplotype with some low frequency variants present and one isolate with the yellow haplotype.

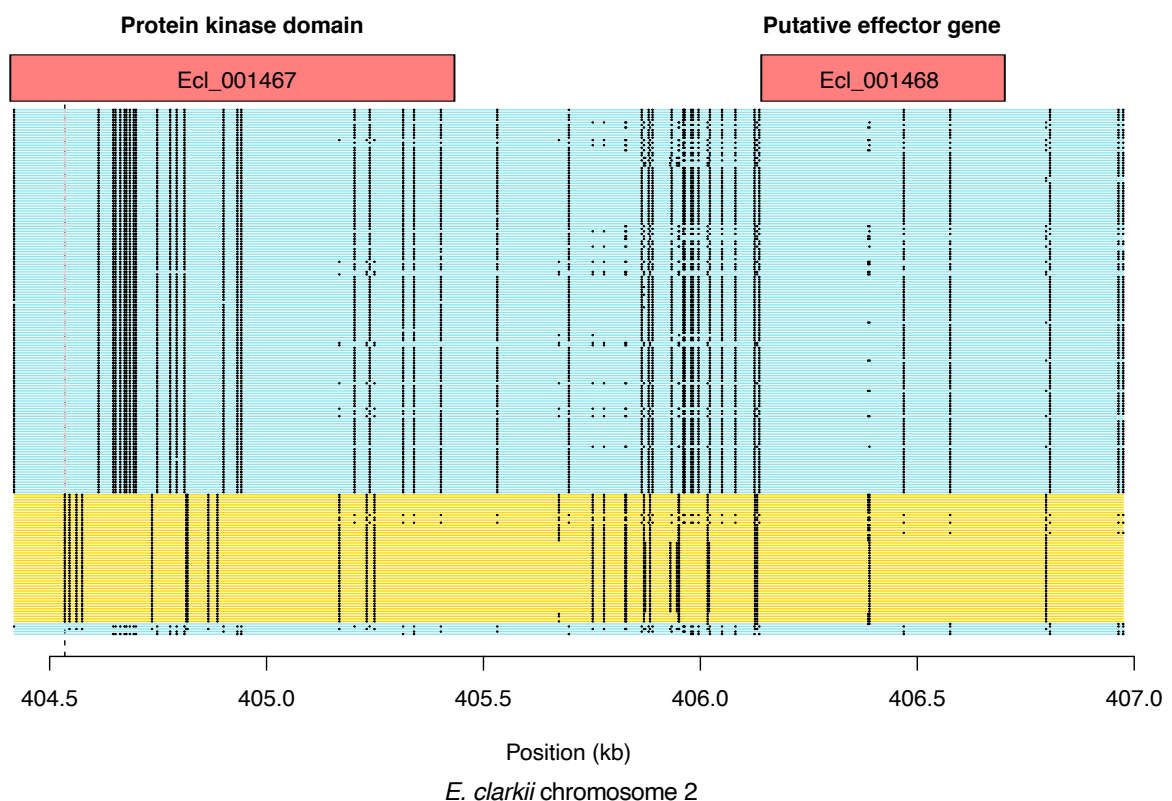
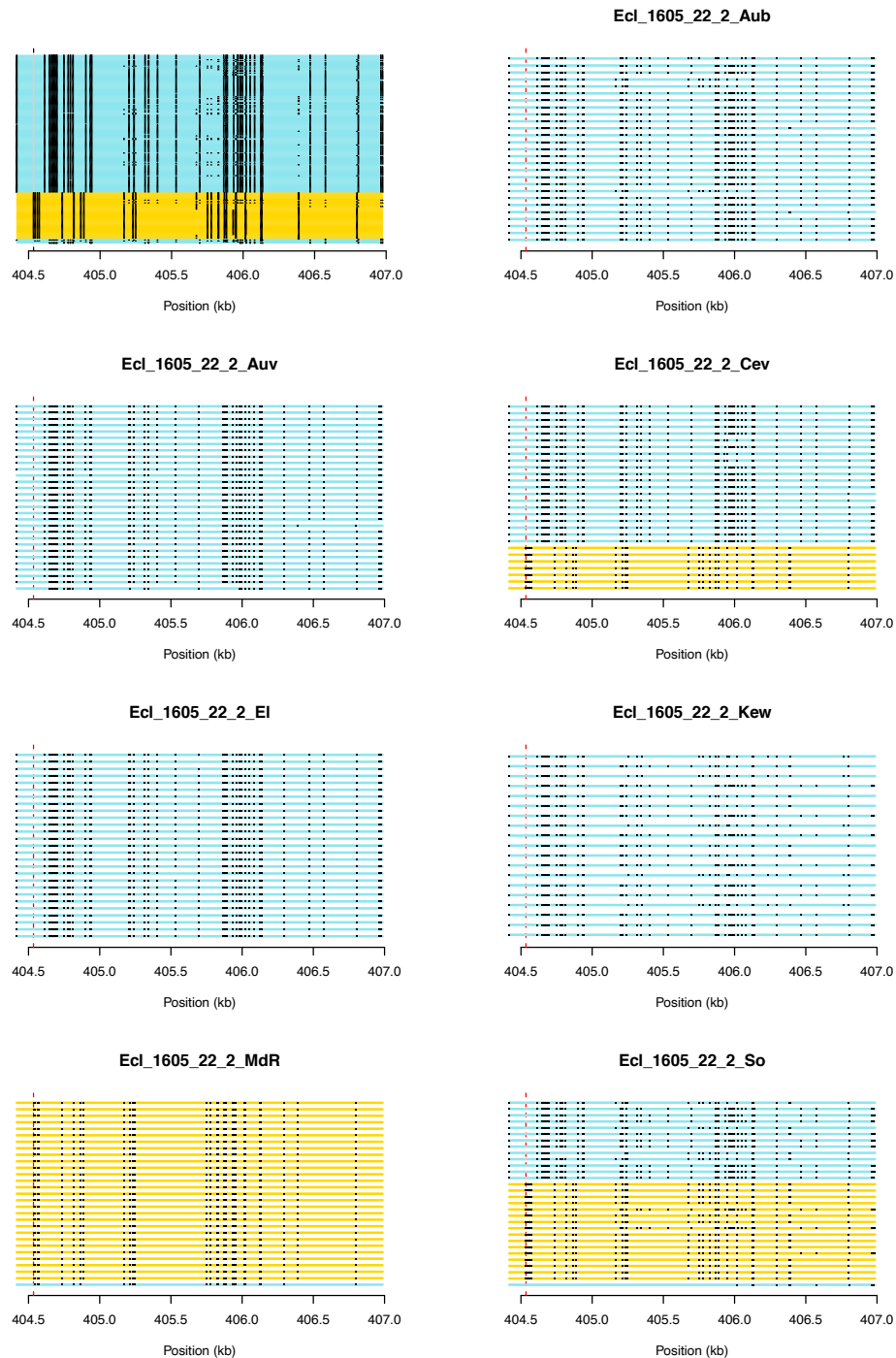


Figure S8: Cross-population haplotype structure for candidate locus in *E. clarkii* (chr2).

We investigated the haplotype structure at a CLR-sweep region shared by four populations (Aub, Auv, El and Kew) and containing two annotated genes, 700bp apart: one encoding a protein-kinase like domain (Ecl_001467) and the other an effector gene (Ecl_001468) which was highly upregulated in planta (9.1 logFC). Haplotypes for all isolates are shown (n=190) and were colored relative to the first nonsynonymous SNP position located in the genic region of Ecl_001467 (404,546; blue=A, yellow=G). Blue lines at the bottom of the plot indicate individuals that have not been genotyped at this locus (5). Populations for which our analysis inferred a recent selective sweep are fixed for the blue haplotype (see Supplementary Figure S9 showing haplotypes at this locus for each population separately). The yellow haplotype is fixed in one population (MdR) and present at intermediate frequencies in the remaining two populations (Cev and So). Three populations (Aub, Kew and So) contain three, five and six recombinant isolates respectively.



S9: Within-population haplotype structure for candidate locus on chromosome two in *E. clarkii*.

Haplotypes for all isolates are shown in the top left and in the remaining seven plots for each population separately. Haplotypes were colored relative to the highlighted SNP position (see Figure S8 for details about candidate SNP loci). Blue lines at the bottom of the plots indicate individuals that have not been genotyped at this locus. Populations for which our analysis inferred a recent selective sweep (Aub, Auv, EI, Kew) are fixed for the blue haplotype. The yellow haplotype is fixed in one population (MdR) and present at intermediate frequencies in the remaining two populations (Cev and So). Three populations (Aub, Kew and So) contain three, five and six recombinant isolates respectively.

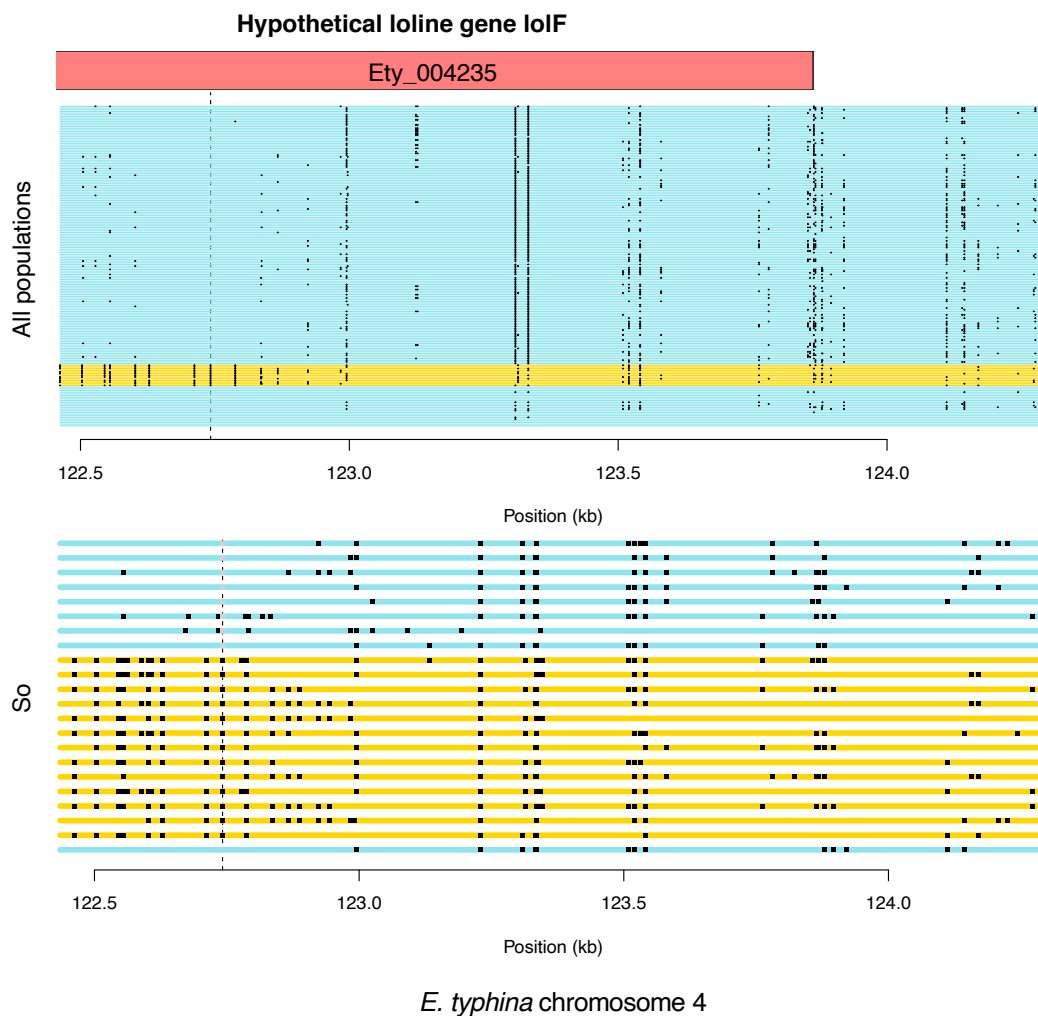


Figure S10: Cross-population haplotype structure for candidate locus in *E. typhina* (chr4).

The haplotype structure of the hypothetical alkaloid gene loIF (Ety_004235) located in a shared sweep within an AT-rich region is shown. Haplotypes for all isolates from all populations are shown ($n=175$) and were colored relative to a splice site mutation (high impact variant) located in the genic region of Ety_004235) (122,742; blue=T, yellow=C). Blue lines at the bottom of the plot indicate individuals that have not been genotyped at this locus (22). The alternative haplotype (yellow) is present in only one population (So, bottom plot) while all other populations contain the blue haplotype with relatively high levels of polymorphism as expected based on the localization in the dynamic genomic compartment.

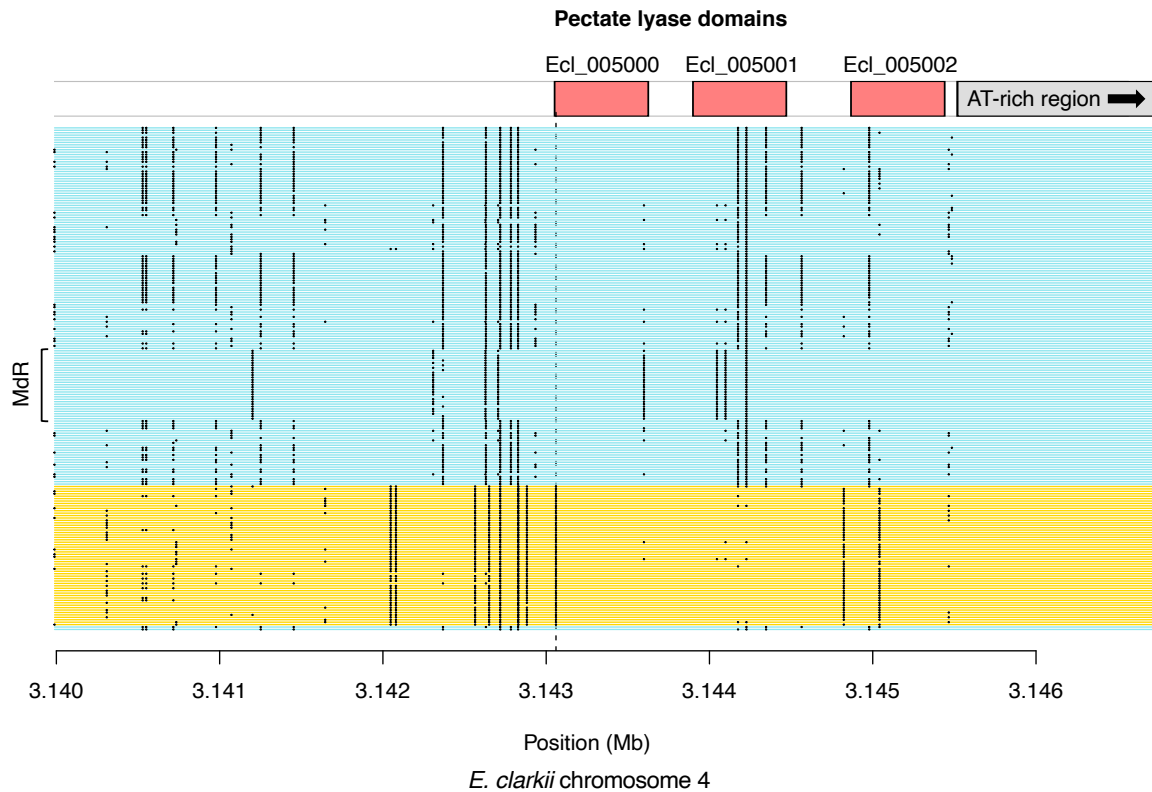
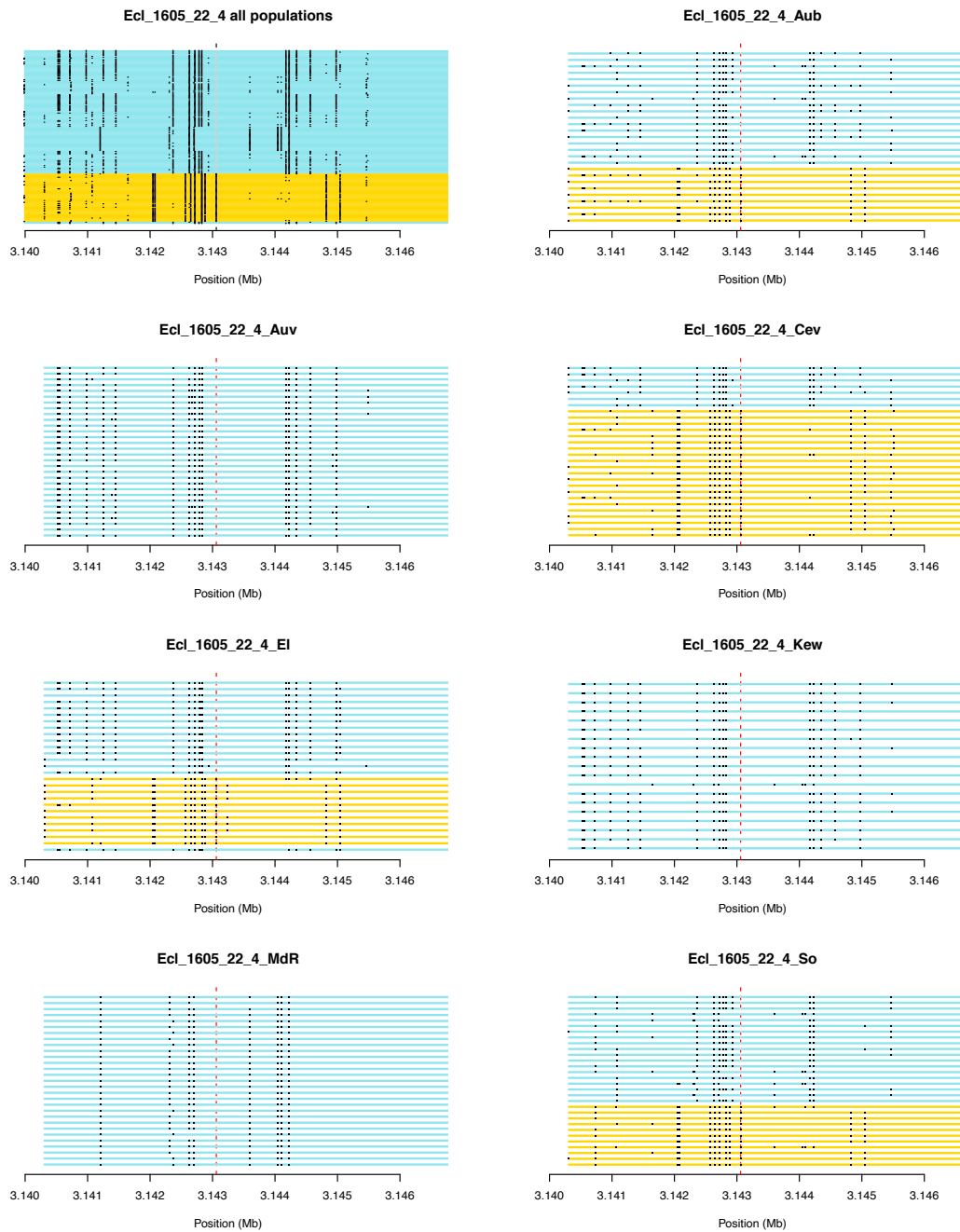


Figure 11: Cross-population haplotype structure for candidate locus in *E. clarkii* (chr4).

We investigated the haplotype structure at a sweep region detected with both methods in So and an iHS-sweep shared by four populations (Aub, Cev, El and So). The locus corresponds to the overlapping sweep signatures highlighted in Figure 1 and contained three genes: two encoded pectate lyase domains (Ecl_005000 and Ecl_005001) and the third had no functional annotation, but showed homology to putative effector with a LysM domain (Ecl_005002). None of the genes were significantly over- or under-expressed *in planta*. Haplotypes for all isolates are shown (n=190) and were colored relative to the SNP position with the most significant iHS score (3,143,060; blue=C, yellow=T). Blue lines at the bottom of the plot indicate individuals that have not been genotyped at this locus (2). In total four different haplotypes can be distinguished at the locus. Populations for which our analysis inferred a recent selective sweep contain the yellow haplotype at intermediate frequencies (see Supplementary Figure S12 showing haplotypes at this locus for each population separately). Mdr is fixed for a different haplotype as indicated in the plot.



S12: Within-population haplotype structure for candidate locus on chromosome four in *E. clarkii*.

Haplotypes for all isolates are shown in the top left and in the remaining seven plots for each population separately. Haplotypes were colored relative to the highlighted SNP position (see Figure S11 for details about candidate SNP loci). Blue lines at the bottom of the plots indicate individuals that have not been genotyped at this locus. Multiple haplotypes are present at different frequencies in different populations. Populations for which our analysis inferred a recent selective sweep contain the yellow haplotype at intermediate frequencies (Aub, Cev, EI, So). Isolates colored in blue contain three additional haplotypes: the most common one is fixed in Auv, almost fixed in Kew and at intermediate frequencies in the populations showing a sweep signature; another one is only present in MdR where it is fixed, and a third haplotype is present in Aub (3), Kew (1) and So (3).

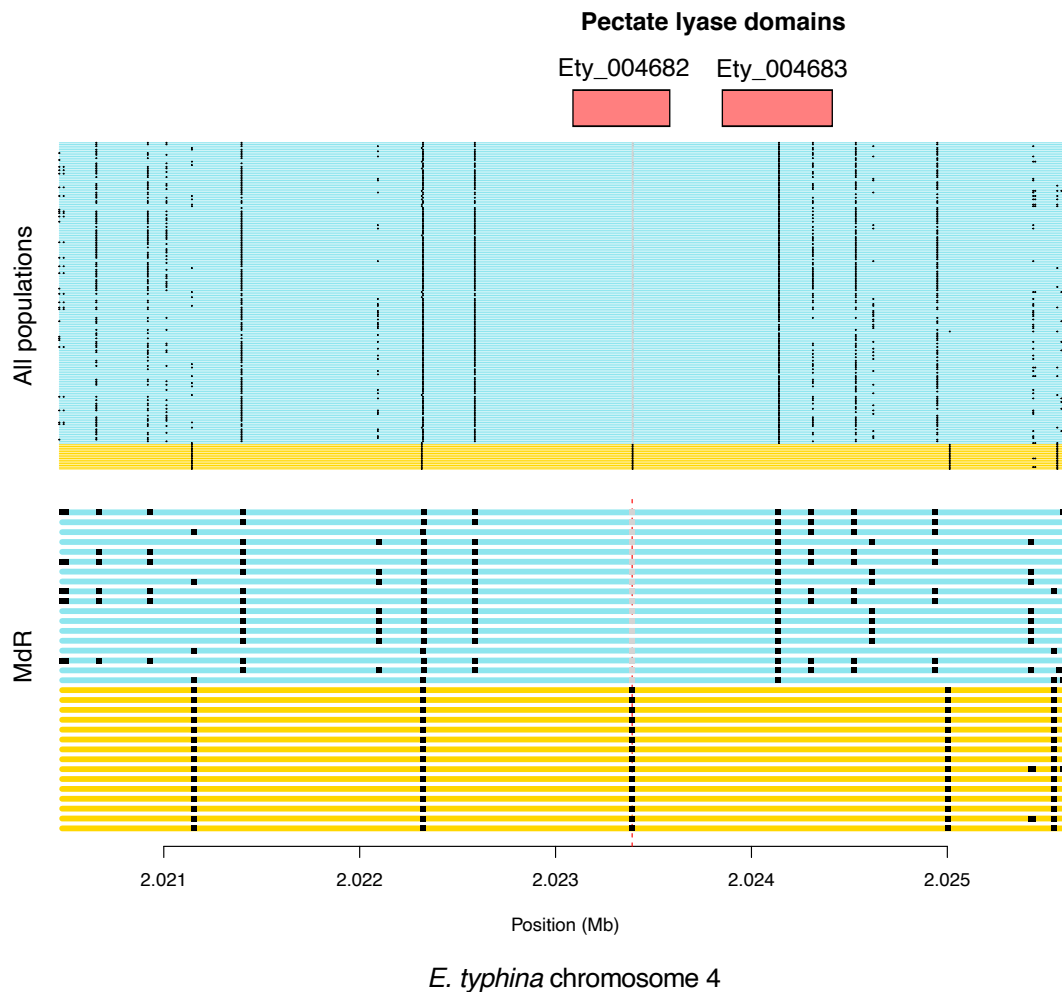


Figure S13: Cross-population haplotype structure for locus not under selection in *E. typhina* (chr4).

The haplotype structure at a locus not under selection in *E. typhina* is shown, which is homologous to a candidate region under selection in *E. clarkii* (Supplementary Figures S11 and S12). The locus contained the two orthologous pectate lyase superfamily proteins (Ety_004682 and Ety_004683) but lacked a putative effector gene annotation present in *E. clarkii*. Haplotypes for all isolates from all populations are shown (n=175) and were colored relative to a splice site mutation (high impact variant) located in the genic region of Ety_004682 (2,023,391; blue=A, yellow=C). All populations are fixed for the same haplotype (blue) except MdR (bottom plot) which contained a different haplotype (yellow) at intermediate frequency with less variation compared to the blue haplotype.

Resolving the phylogeny of the *Epichloe typhina* species complex using genome-wide SNPs

Artemis D. Treindl^a, Jessica Stapley^b, Adrian Leuchtman^a

^a Plant Ecological Genetics Group, Institute of Integrative Biology, ETH Zurich, 8092 Zurich, Switzerland

^b Plant Pathology Group, Institute of Integrative Biology, ETH Zurich, 8092 Zurich, Switzerland

ABSTRACT

The *Epichloe typhina* complex comprises several sexual species and subspecies associated with a wide range of Eurasian grass hosts. Previous analyses using single marker genes were not able to fully resolve the phylogenetic relationships among these taxa. Phylogenetic analysis was performed on genome-wide SNP datasets of isolates collected from 14 different host grass species across Europe and additional genomic data from public data bases. The new phylogeny resolved seven well-supported clades that were associated with different hosts or host groups. Some of them corresponded to existing taxa while others may represent new cryptic species, and we provide evidence for occasional host jumps within some clades. In addition, several isolates from *E. typhina* hosts grouped with the *E. festucae* clade outside the *E. typhina* complex suggesting this taxon may have a more generalist host-spectrum than previously assumed.

BACKGROUND

Phylogenetics forms the basis of modern species concepts and aims to make inferences about the evolutionary relationships among groups of organisms. The analysis of genomic data makes it possible to tackle taxonomic problems by helping to delimit species boundaries in difficult groups or species complexes where morphological data is not informative. This is particularly useful in fungi, which are challenging to classify because of limited characters available for phenotype based taxonomy and the frequent occurrence of cryptic species (Matute and Sepúlveda 2019). The *E. typhina* species complex is a good example where genome sequence data could help resolve species' relationships, as previous studies could not always delimit monophyletic groups, for example in the case of the target species of this thesis, *E. typhina* and *E. clarkii*.

The fungus causing choke disease on *Poaceae* grasses was first described in 1798 as *Sphaeria typhina* (Persoons 1798) and later became the type species of the new genus *Epichloe* (Tulasne and Tulasne 1865). For almost two centuries all choke disease found on numerous host grasses was referred to as one species, *Epichloe typhina* (Pers.) Brockm. Taxonomic research in the genus began to flourish when it was discovered that symptomless *Epichloe* infections of pasture grasses caused toxicosis in grazing livestock (Bacon et al. 1977, Latch and Christensen 1982), which led to the taxonomic description of the responsible fungal endophytes as separate asexual species (Morgan-Jones and Gams 1982, Latch et al. 1984). Several new choke forming species were subsequently described from grasses in Europe and North America showing similarity with *E. typhina* (White Jr. 1993, 1994, Leuchtman et al. 1994, Leuchtman and Schardl 1998, Schardl and Leuchtman 1999). These sexual species were distinguished based on their distinct morphology, association with specific host grass species and supporting evidence from allozyme loci and DNA sequences of marker genes, as well as reproductive incompatibility with other known species. The latter constituted an important criterium for species recognition following a biological species concept (Mayr 1970) and was evaluated by experimental mating tests. Crosses that produced viable ascospores after transfer of spermatia between heterothallic stromata were considered sexually compatible, those that failed to produce ascospores were considered sexually incompatible. In the latest taxonomic reassessment of the genus eleven sexual *Epichloe*

species were recognized, whereas three species were relegated to subspecies as they did not conform to the incompatibility criteria (Leuchtman et al. 2014).

In the *E. typhina* species complex taxonomic delimitation has proven particularly challenging (Leuchtman and Schardl 1998; Craven et al. 2001). This complex comprises a group of sexual *Epichloe* taxa associated with a wide range of Eurasian grass species and as of today, contains two described biological species and several subspecies that are specialized on distinct host grasses but remain interfertile in mating tests (Leuchtman et al. 2014). However, phylogenetic analyses based on several single copy genes revealed instances of paraphyly among biologically defined taxa contradicting a phylogenetic species concept (Taylor et al. 2000, Craven et al. 2001). Furthermore, isolates of *E. typhina* from several different host species were shown to form distinct subclades within the complex suggesting cryptic speciation driven by host specialization (Schardl et al. 2007).

Another example of discrepancy between biological and phylogenetic species concepts are the focal taxa of this thesis *E. typhina* infecting *Dactylis glomerata* and *E. clarkii* infecting *Holcus lanatus* and *H. mollis* which are sexually compatible and therefore currently assigned the taxonomic rank of subspecies within the species complex despite their morphological differences. However, the analyses presented in this thesis and previous studies (Schirrmann et al. 2015, 2018) indicate that these taxa represent distinct evolutionary lineages at advanced stages of divergence and do not form actual interbreeding populations in nature. It is obvious that the available phylogenetic analyses using few marker genes fail to fully resolve the taxonomic relationships among these closely related taxa. Here we present a new phylogeny based on genome-wide SNP data from 144 isolates collected from 14 different host grass species that span the whole species complex. Our analysis should provide a more robust basis to clarify the unresolved inter- and intraspecies relationships within the *E. typhina* complex.

MATERIALS AND METHODS

Sample collection, fungal isolation, DNA extraction and sequencing

Epichloe infected grasses were collected from locations across Western Europe between 1991 and 2020 (Figure 1 and Table 1, see Supplementary Table S1 for more detail). One flowering tiller showing evidence of choke (i.e. with inflorescences enveloped by the stromata) from 1-3 individual plants per location and host species were sampled. Host plants were considered individuals when grass clumps were at least 1 m apart and one infected tiller per plant was collected, each sample representing one haploid genotype since infection of shoot apical meristems and the whole tiller (and usually the whole plant) is established by a single haploid ascospore. *Epichloe* species occurring on *Holcus* spp. (*E. clarkii*) and on *Dactylis glomerata* (*E. typhina*) were part of a more detailed population genomics study and sampled more extensively. For the present phylogenetic analysis, we selected the three individuals with the highest number of Illumina reads from each host species and site. Fungal strains were isolated from stromata by removing pieces of the interstitial fungal mycelium between undeveloped inflorescence and leaf blades after splitting open the stromata with a sterile blade. In a few cases endophytes were isolated from surface disinfected stems of infected tillers. Isolates were grown on supplemented malt extract-agar (SMA) containing 1 % malt extract, 1 % glucose, 0.25 % bacto peptone, 0.25 % yeast extract, 1.5 % bacto agar and 0.005% oxytetracycline (Pfizer, New York, NY, U.S.A.). Pure cultures were subsequently grown in V8 liquid media on a rotary shaker at 12 rpm for 10-14 days at room temperature. Mycelium was vacuum filtered, freeze dried for 24h and then ground using liquid nitrogen. Genomic DNA was extracted from ground mycelium and libraries prepared as described in Chapter II. A total of 107 isolates were sequenced on an Illumina HiSeq 4000 (Illumina, San Diego, CA, USA) at 16x coverage on average, while additional 40 isolates were prepared separately using the same library kit and sequenced on an Illumina NovaSeq 6000 at approximately 40x coverage. One additional isolate (strain NFe76) was included from another host species of the *E. typhina* complex native to North America (*Bromus laevipes*), which has been collected by Carolyn Young, Noble Foundation, Oklahoma, USA and sequenced by Murray Cox, Massey University, New Zealand.

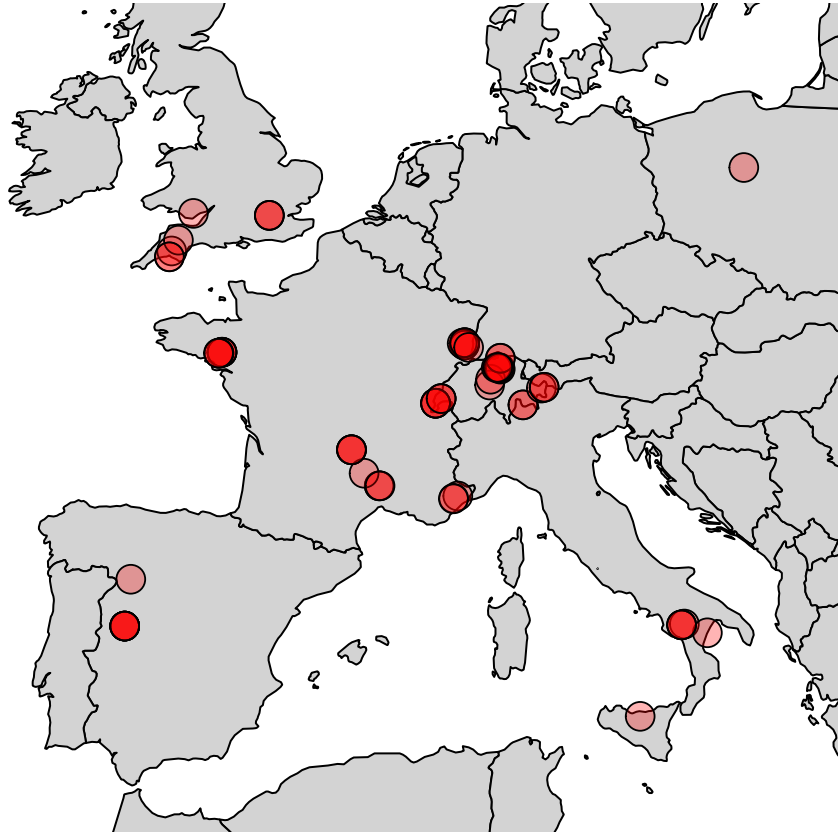


Figure 1: Collection sites of the 144 *Epichloe* isolates infecting different host grass species in Europe. At many sites several infected host species occurring in sympatry were sampled. This is indicated by different shades of red with darker shades corresponding to more host species (n=1-6).

Table 1: Host species, number of sampling sites and number of fungal isolates from the *E. typhina* species complex used for phylogenetic analysis.

Host species	No. of locations	No. of isolates
<i>Anthoxanthum odoratum</i>	2	6
<i>Brachypodium phoenicoides</i>	1	1
<i>Brachypodium pinnatum</i>	7	9
<i>Brachypodium sylvaticum</i>	8	12
<i>Dactylis glomerata</i>	14	34
<i>Holcus lanatus</i>	8	25
<i>Holcus mollis</i>	2	6
<i>Lolium perenne</i>	2	4
<i>Phleum pratense</i>	2	5
<i>Poa nemoralis</i>	6	10
<i>Poa pratensis</i>	4	8
<i>Poa sylvicola</i>	1	1
<i>Poa trivialis</i>	8	22
<i>Puccinellia distans</i>	1	1
Total		144

Genome sequences of the reference strain of *E. festucae* (Fl_Fll, SRR5428140), an additional *E. festucae* isolate (Fp_OG, SRR8238189) and one of *E. bromicola* (Ed_OG, SRR5420176) were downloaded from NCBI Sequence Read Archive (SRA) (<https://www.ncbi.nlm.nih.gov/sra>, see Supplementary Materials for details).

Mapping and variant calling

The paired sequence reads were trimmed to remove adaptors and poor sequence, then mapped to the *E. festuca* reference genome (EfFl1_v0.2.fna, SRR8238189) using BWA MEM (Li and Durbin 2009). Only alignments with a mapping quality greater than 10 were retained and three isolates were omitted as they had too few reads. Variants were called from mapped reads using Genome Analysis Toolkit (GATK) Short Variant Discovery best practice workflow (<https://gatk.broadinstitute.org/hc/en-us/articles/360035535932-Germline-short-variant-discovery-SNPs-Indels->) (McKenna et al. 2010). In brief, haplotype caller and jointcaller were used to identify variants. Variants were retained for the final analysis if they met these criteria: within nuclear DNA only (mitochondrial sequence removed), biallelic SNPs only (no indels, number of alleles ≤ 2), no missing data, quality depth > 2 , root mean square mapping quality > 40 , strand bias odd ratio (SOR) < 3 , depth variance/mean depth < 50 , and SNPs in approximate linkage equilibrium (remove sites with $r^2 > 0.3$). A total of 48,997 SNPs in 148 isolates were retained for TREE building. The haploid vcf file was converted to diploid, the SNPs were concatenated and a phylip file was created using custom perl script (vcf_to_phylip.py, <https://github.com/mscharmann/tools>).

Building SNP phylogenetic tree

Phylogenetic trees were constructed using a maximum likelihood approach implemented in RaxML (Randomized Accelerated Maximum Likelihood, RaxML v8.2.4 raxmlHPC-PTHREADS-AVX; Stamatakis 2014). We used the ASC_GTRGAMMA substitution model as recommended for concatenated SNP data where all sites are variable (Lewis 2001, Leaché et al. 2015). The *E. bromicola* isolate (Ed_OG, SRR5420176) from *Elymus dahuricus* was specified as the outgroup, and as a result the tree is rooted at the branch leading to this group. RaxML was used to construct 100 trees and repeated

10 times, each time using a different starting tree. We checked that each replicate converged on the same topology and selected the best-tree with the highest likelihood (Figure 2). Branch support values were estimated by bootstrapping with 100 bootstrap replicate data sets. The bootstrap values represent the frequency that a group occurred across the replicate data sets and is a measure of support for that particular grouping in our data (Rokas 2011).

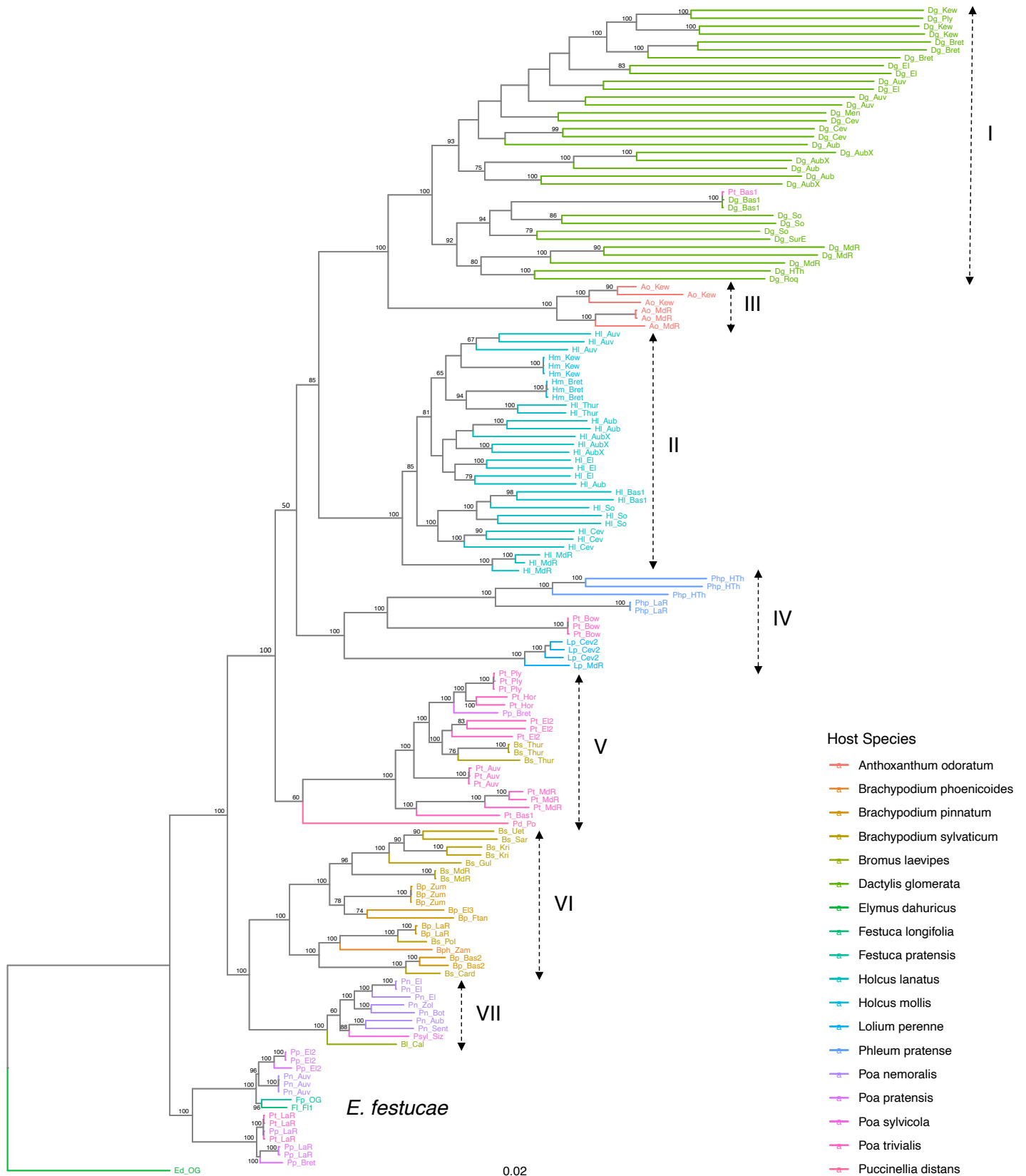


Figure 2: Phylogram of the *E. typhina* species complex based on genome-wide SNP data.

The tree was inferred by maximum likelihood methods as implemented in RaxML v8.2.4 (cit. Stamatakis 2014). Branch support values were estimated by 100 bootstrap replicates. Dashed lines denote different clades within the *E. typhina* complex and are designated by Roman numerals. A number of isolates grouped outside the *E. typhina* complex with *E. festucae* reference strain FI_FI1. The tree is rooted with *E. bromicola* isolated from *Elymus dahuricus* (Ed_OG).

RESULTS AND DISCUSSION

A total of 144 isolates were sampled from 14 different host grass species in Western Europe, all of which are known hosts of members of the *E. typhina* species complex. One isolate from an additional host (*Bromus laevipes*) came from North America (see Table S1 for details). In addition, sequence data for three isolates obtained from NCBI Sequence Read Archive (SRA) from outside the *E. typhina* complex (*E. festucae* and *E. bromicola*) were included. We genotyped these 148 individuals and built a phylogenetic tree using 48,997 unlinked SNPs. The most likely tree based on our data is presented in Figure 2. Our analysis resolved a total of seven clearly separate clades within the *E. typhina* complex of which all except one (clade IV) had $\geq 85\%$ bootstrap support. Some clades were associated with particular host genera or single species while others included isolates from multiple genera of different *Poaceae* tribes. Clade I included all isolates from *Dactylis glomerata* and a single isolate from *Poa trivialis* (see below), and resolved the genetic structure among populations with isolates from the same sampling locations grouping together. This clade corresponds to a taxon currently recognized as *E. typhina* subsp. *typhina* (Pers.) Brockm. and designated throughout this thesis as *E. typhina*. Clade II included all isolates from *Holcus lanatus* and *Holcus mollis* among which population structure was also resolved. This clade corresponds to a taxon described previously and throughout this thesis as *E. clarkii*, however, had been relegated to subspecies as *E. typhina* subsp. *clarkii* (J.F. White) Leuchtm. & Schardl (Leuchtmann et al. 2014). Clade III included isolates from a single host species *Anthoxanthum odoratum* sampled from two locations, one in Montemayor del Rio, Spain and one in the Kew Gardens, UK. Clade IV only had 50 % bootstrap support but consisted of several subclades that had high (100 %) support and included isolates from

Phleum pratense, *Poa trivialis* and *Lolium perenne*. Clade V included most isolates from *Poa trivialis* from six different sampling locations, one isolate from *Poa pratensis* and isolates from *Brachypodium sylvaticum* sampled from one location in Switzerland (Kleinandelfingen, Thur). A single isolate from *Puccinellia distans* from Poland also grouped with this clade. Clade VI comprised isolates from three *Brachypodium* host species including *B. sylvaticum* from which the biological species *E. sylvatica* Leuchtm. & Schardl has been described (Leuchtmann and Schardl 1998). Clade VII included isolates from *Poa nemoralis* and *Poa sylvicola* as well as the North American isolate from *Bromus laevipes* and corresponds to *E. poae* (Tadych et al. 2012), a species currently recognized as *E. typhina* subsp. *poae* (Tadych, K.V. Ambrose, F.C. Belanger & J.F. White) Tadych (Leuchtmann et al. 2014).

In the most recent phylogeny based on portions of beta-tubulin (*tubB*) gene sequences, four subclades within the *E. typhina* species complex were resolved (Leuchtmann et al. 2014). These corresponded to two recognized biological species (*E. typhina sensu stricto* and *E. sylvatica*) and two subspecies (ssp. *poae* and ssp. *pollinensis*). A third subspecies (ssp. *clarkii*) infecting *Holcus lanatus* was nested within the *E. typhina* main clade and could not be separated from *E. typhina* isolates from other hosts including *D. glomerata*. In our new phylogeny based on genome-wide SNP data, seven well-supported subclades of the *E. typhina* complex were evident. Notably, all isolates from the two *Holcus* hosts (*H. lanatus* and *H. mollis*) grouped in a distinct clade separate from *E. typhina* isolates infecting *D. glomerata*, which indicates that populations on these hosts remain genetically distinct even in sympatric settings despite potential compatibility between them. Furthermore, isolates from all *Brachypodium* hosts formed a monophyletic clade that included the distinct biological species *E. sylvatica* and isolates from *B. pinnatum* and *B. phoenicoides* that are sexually compatible with *E. typhina* but incompatible with *E. sylvatica*. This suggests that it may not be justified to distinguish different phylogenetic species infecting these different *Brachypodium* hosts. Previous phylogenies have resolved two subgroups among isolates infecting *Brachypodium* hosts (one being recognized as *E. sylvatica* ssp. *pollinensis*), but these subgroups could not be confirmed in your phylogeny. Finally, three additional clades (III, IV and V) could be distinguished in our analyses that have not been recognized in the previous *tubB* phylogeny by Leuchtmann et al. (2014) and that likely represent novel cryptic species. Each of these

clades included isolates from one or several different hosts grasses suggesting that these, often sympatric, populations are genetically isolated on particular hosts. These findings confirm and clarify results from a previous study analyzing *tubB* sequences of *E. typhina* isolates from populations of several hosts across Switzerland detecting no genetic exchange between populations associated with different hosts, e.g. *Poa trivialis* (Schardl et al. 2007).

Overall, isolates grouped according to their host species or species group and not their sampling location with few exceptions: Several isolates from *B. sylvaticum* grouped with isolates from *Poa* hosts (clade V) and one isolate from *P. trivialis* grouped with isolates from *D. glomerata* (clade I) collected from the same sampling site in Basilicata, Southern Italy. On the other hand, a second *P. trivialis* isolate from this location grouped with clade V in line with the majority of *P. trivialis* isolates analyzed here. There is no reason to attribute this to a labelling or pipetting error as samples were stored and processed separately. This finding therefore suggests the possibility of local host switches between host grasses occurring in sympatry even for a highly host-specific species such as *E. typhina* on *D. glomerata*. Some isolates from *P. trivialis* also grouped with clade IV suggesting that *P. trivialis* may be somewhat of a generalist host susceptible to multiple *Epichloe* species. Surprisingly, some isolates from *Poa* hosts (*P. nemoralis*, *P. pratensis* and *P. trivialis*) grouped outside the *E. typhina* complex with *E. festucae*. This suggests that *E. festucae* has a broader host spectrum than previously assumed and shares some common host species with members of the *E. typhina* complex. We did not sample hosts from the same locations typically associated with *E. festucae* (e.g. *Festuca rubra*), and therefore could not assess if host switches were more likely to occur in sympatry.

A striking feature of the phylogeny are the long terminal branches observed in some clades, particularly clades I and II. The longer branches may reflect greater genetic diversity in these clades. Greater diversity could result from the fact that we have sampled a larger number of isolates in these groups, i.e. *D. glomerata* (n=34), *Holcus* hosts (n=31), and are therefore more likely to identify sites that are only polymorphic within these clades while monomorphic in other clades with fewer isolates. This could be tested by down-sampling to include a similar number of isolates in each clade. However, increased branch lengths may also reflect true differences in nucleotide diversity (π), e. g. diversity in *E. typhina* populations from *D. glomerata* was larger

compared to *E. clarkii* populations from *Holcus* spp. (Chapter II of this thesis). Furthermore, gene flow (between populations) and recombination (within populations) is expected to play a role in the generation of diversity within clades of these sexually reproducing fungi and affect the topology of the phylogenetic tree (see Sakoparnig et al. 2021). The resolution of population structure within clades is likely a result of the fact that different lineages, that is the genetic ancestors of different strains, have recombined at different rates. Recombination events are much more common among individuals from the same location and this is reflected in many more possible changes of the within-population phylogenetic tree along the alignment and therefore longer terminal branches. Following this rationale, the differences in terminal branch lengths among clades may also be related to the underlying size of the recombining population (effective population size N_e) from which isolates were sampled (e.g. see Neafsey et al. 2008). A pervasively outcrossing population with a large N_e would thus harbor higher diversity and have longer branches than a clonal population with a small N_e . We did not specifically test this hypothesis across the whole phylogeny, however, branch lengths within *E. typhina* from *D. glomerata* (clade I) and *E. clarkii* from *Holcus* spp. (clade II) correspond well to the relative differences in N_e inferred from population genomic datasets (see Chapter II of this thesis): *E. clarkii* had lower N_e than *E. typhina* and the populations Kew and MdR in *E. clarkii* had especially low N_e .

The analysis of whole *Epichloe* genomes therefore not only proves useful to assess the overall phylogenetic structure within this species complex, it also offers potential information about evolutionary processes acting within lineages. Indeed, genome-wide SNP datasets have been successfully used in many different systems to address diverse phylogenetic questions ranging from recent divergence at the population level to more ancient diversification (e.g. Emerson et al. 2010, Wagner et al. 2013). The new phylogeny presented here resolves genetically differentiation among closely related cryptic species within the *E. typhina* complex and thus provides a basis for future taxonomic reconsiderations.

ACKNOWLEDGMENTS

We would like to kindly thank Martyn Ainsworth (Royal Botanical Gardens Kew, UK), Marlena Lembicz (Adam Mickiewicz University, Poland), Jean-Paul Priou, James M. Neenan, Elisabeth Walder and Lilith Treindl for assistance with sample collection, Claudia Michel and Beatrice Arnold for laboratory assistance, and Murray P. Cox (Massey University, New Zealand) for providing sequencing reads of fungal isolate NFe76. Data presented and analyzed here were generated in collaboration with the Genetic Diversity Centre (GDC), ETH Zurich and the Functional Genomics Center Zurich (FGCZ), UZH.

REFERENCES

- Bacon, C. W., J. K. Porter, J. D. Robbins, and E. S. Luttrell. 1977. *Epichloe typhina* from toxic tall fescue grasses. *Applied and Environmental Microbiology* 34:576–581.
- Craven, K. D., P. T. W. Hsiau, A. Leuchtman, W. Hollin, and C. L. Schardl. 2001. Multigene phylogeny of *Epichloe* species, fungal symbionts of grasses. *Annals of the Missouri Botanical Garden* 88:14–34.
- Emerson, K. J., C. R. Merz, J. M. Catchen, P. A. Hohenlohe, W. A. Cresko, W. E. Bradshaw, and C. M. Holzapfel. 2010. Resolving postglacial phylogeography using high-throughput sequencing. *Proceedings of the National Academy of Sciences* 107:16196–16200.
- Latch, G. C. M., and M. J. Christensen. 1982. Ryegrass endophyte, incidence, and control Ryegrass endophyte, incidence, and control 443. *New Zealand Journal of Agricultural Research* 25:443–448.
- Latch, G. C. M., M. J. Christensen, and G. J. Samuels. 1984. Five endophytes of *Lolium* and *Festuca* in New Zealand. *Mycotaxon* 20:535–550.
- Leaché, A. D., B. L. Banbury, J. Felsenstein, A. N. M. De Oca, and A. Stamatakis. 2015. Short tree, long tree, right tree, wrong tree: New acquisition bias corrections for inferring SNP phylogenies. *Systematic Biology* 64:1032–1047.
- Leuchtman, A., C. W. Bacon, C. L. Schardl, J. F. White Jr., and M. Tadych. 2014. Nomenclatural realignment of *Neotyphodium* species with genus *Epichloë*. *Mycologia* 106:202–215.
- Leuchtman, A., and C. L. Schardl. 1998. Mating compatibility and phylogenetic relationships among two new species of *Epichloë* and other congeneric European species. *Mycological Research* 102:1169–1182.
- Leuchtman, A., C. L. Schardl, and M. R. Siegel. 1994. Sexual compatibility and taxonomy of a new species

- of *Epichloë* symbiotic with fine fescue grasses. *Mycologia* 86:802–812.
- Lewis, P. O. 2001. A likelihood approach to estimating phylogeny from discrete morphological character data. *Systematic Biology* 50:913–925.
- Li, H., and R. Durbin. 2009. Fast and accurate short read alignment with Burrows-Wheeler transform. *Bioinformatics* 25:1754–1760.
- Matute, D. R., and V. E. Sepúlveda. 2019. Fungal species boundaries in the genomics era. *Fungal Genetics and Biology* 131:103249.
- Mayr, E. 1970. *Populations, species, and evolution: an abridgment of animal species and evolution*. Harvard University Press, Cambridge, Massachusetts.
- McKenna, A., M. Hanna, E. Banks, A. Sivachenko, K. Cibulskis, A. Kernytsky, K. Garimella, D. Altshuler, S. Gabriel, and M. Daly. 2010. The Genome Analysis Toolkit: A MapReduce framework for analyzing next-generation DNA sequencing data. *Genome Research* 20:1297–1303.
- Morgan-Jones, G., and W. Gams. 1982. Notes on Hyphomycetes. XLI. An endophyte of *Festuca arundinacea* and the anamorph of *Epichloë typhina*, new taxa in one of two new sections of *Acremonium*. *Mycotaxon* 15:311–318.
- Neafsey, D. E., S. F. Schaffner, S. K. Volkman, D. Park, P. Montgomery, D. A. Milner, A. Lukens, D. Rosen, R. Daniels, N. Houde, J. F. Cortese, E. Tyndall, C. Gates, N. Stange-Thomann, O. Sarr, D. Ndiaye, O. Ndir, S. Mboup, M. U. Ferreira, S. do L. Moraes, A. P. Dash, C. E. Chitnis, R. C. Wiegand, D. L. Hartl, B. W. Birren, E. S. Lander, P. C. Sabeti, and D. F. Wirth. 2008. Genome-wide SNP genotyping highlights the role of natural selection in *Plasmodium falciparum* population divergence. *Genome Biology* 9:R171.
- Persoons, C. H. 1798. *Icones et descriptiones Fungorum minus cognitorum*. Bibliopolii Breitkopf-Haerteliani impenis, Leipzig.
- Rokas, A. 2011. Phylogenetic analysis of protein sequence data using the Randomized Axelerated Maximum likelihood (RAXML) program. *Current Protocols in Molecular Biology* 96:19.11.1-19.11.14.
- Sakoparnig, T., C. Field, and E. van Nimwegen. 2021. Whole genome phylogenies reflect the distributions of recombination rates for many bacterial species. *eLife* 10:1–61.
- Schardl, C. L., and A. Leuchtman. 1999. Three new species of *Epichloë* symbiotic with North American grasses. *Mycologia* 91:95–107.
- Schardl, C. L., A. Leuchtman, and B. A. McDonald. 2007. Relationships of *Epichloë typhina* isolates from different host grasses. Pages 451–455 in A. J. Popay and E. R. Thom, editors. *Proceedings of the 6th International Symposium on Fungal Endophytes of Grasses* New Zealand Grassland Association (Inc.), Dunedin, New Zealand.
- Schirrmann, M. K., S. Zoller, D. Croll, E. H. Stukenbrock, A. Leuchtman, and S. Fior. 2018. Genomewide signatures of selection in *Epichloë* reveal candidate genes for host specialization. *Molecular Ecology* 27:3070–3086.
- Schirrmann, M. K., S. Zoller, S. Fior, and A. Leuchtman. 2015. Genetic evidence for reproductive isolation among sympatric *Epichloë* endophytes as inferred from newly developed microsatellite markers.

Microbial Ecology 70:51–60.

- Tadych, M., K. V. Ambrose, M. S. Bergen, F. C. Belanger, and J. F. White. 2012. Taxonomic placement of *Epichloe poae* sp. nov. and horizontal dissemination to seedlings via conidia. *Fungal Diversity* 54:117–131.
- Taylor, J. W., D. J. Jacobson, S. Kroken, T. Kasuga, D. M. Geiser, D. S. Hibbett, and M. C. Fisher. 2000. Phylogenetic species recognition and species concepts in fungi. *Fungal Genetics and Biology* 31:21–32.
- Tulasne, L. R., and C. Tulasne. 1865. *Selecta Fungorum Carpologia. Nectriei-Phacidiei-Pezizei*. Imperial, Paris.
- Wagner, C. E., I. Keller, S. Wittwer, O. M. Selz, S. Mwaiko, L. Greuter, A. Sivasundar, and O. Seehausen. 2013. Genome-wide RAD sequence data provide unprecedented resolution of species boundaries and relationships in the Lake Victoria cichlid adaptive radiation. *Molecular Ecology* 22:787–798.
- White Jr., J. F. 1993. Endophyte-host associations in grasses. XIX. A systematic study of some sympatric species of *Epichloë* in England. *Mycologia* 85:444–455.
- White Jr., J. F. 1994. Endophyte-host associations in grasses. XX. Structural and reproductive studies of *Epichloë amarillans* sp. nov. and comparisons to *E. typhina*. *Mycologia* 86:571.

SUPPLEMENTARY MATERIALS

Supplementary Information for samples from online resources

Outgroup sample:

SRR5420176 *Epichloe bromicola*, China: Nalati, Xinjiang, Host: *Elymus dahuricus*
<https://www.ncbi.nlm.nih.gov/biosample/SAMN06352147>

Additional samples:

SRR5428140 *Epichloe festucae* AL9436, collected in Switzerland: Changins, Host:
Festuca pratensis

<https://www.ncbi.nlm.nih.gov/biosample/SAMN06698315>

SRR8238189 *Epichloe festucae* F11, New Zealand: laboratory culture originally
collected in the UK, Host: *Festuca longifolia*

<https://www.ncbi.nlm.nih.gov/biosample/SAMN08391367>

Supplementary Table

Table S2: Host species, site names, year of sampling and coordinates of sampling locations of *Epichloe* fungal isolates included in phylogenetic analysis. Number indicates how many individual host plants were sampled.

Host species	Site name	Year	Latitude	Longitude	Number
<i>Anthoxanthum odoratum</i>	Kew Gardens	2017	51.47825	-0.29903	3
<i>Anthoxanthum odoratum</i>	Montemayor del Rio	2018	40.34686	-5.9091	3
<i>Brachypodium phoenicoides</i>	Zamora	2000	41.61621	-5.67289	1
<i>Brachypodium pinnatum</i>	Basilicata	2019	40.40248	15.79637	2
<i>Brachypodium pinnatum</i>	Elsass	2017	47.9003	7.4291	1
<i>Brachypodium pinnatum</i>	Ftan	2020	46.79731	10.249	1
<i>Brachypodium pinnatum</i>	La Rippe	2016	46.38698	6.14711	2
<i>Brachypodium pinnatum</i>	Zumikon	1994	47.34275	8.62986	3
<i>Brachypodium sylvaticum</i>	Cardiff	1998	51.53239	-3.24455	1
<i>Brachypodium sylvaticum</i>	Guldenen	2007	47.31193	8.65647	1
<i>Brachypodium sylvaticum</i>	Kleinandelfingen	2020	47.60585	8.65983	3
<i>Brachypodium sylvaticum</i>	Kriens	2020	47.03802	8.28069	2
<i>Brachypodium sylvaticum</i>	Montemayor del Rio	2018	40.34686	-5.9091	2
<i>Brachypodium sylvaticum</i>	Policoro	2019	40.16496	16.68808	1
<i>Brachypodium sylvaticum</i>	Sarnen	2020	46.8804	8.24373	1
<i>Brachypodium sylvaticum</i>	Uetliberg	1991	47.35173	8.50014	1
<i>Bromus laevipes</i> *	California	NA	NA	NA	1
<i>Dactylis glomerata</i>	Aubonne	2005/2018	46.51232	6.36365	3
<i>Dactylis glomerata</i>	Auvergne	2017	45.1263	2.88626	3
<i>Dactylis glomerata</i>	Basilicata	2019	40.38004	15.68896	2
<i>Dactylis glomerata</i>	Bretagne	2017	47.76993	-2.13512	3
<i>Dactylis glomerata</i>	Cevennes	2017	44.14829	3.96273	3
<i>Dactylis glomerata</i>	Elsass	2017	48.02659	7.27517	3
<i>Dactylis glomerata</i>	Haut Thorenc	2017	43.79888	6.84598	3
<i>Dactylis glomerata</i>	Kew Gardens	2017	51.47825	-0.29903	3
<i>Dactylis glomerata</i>	Mende	2017	44.48904	3.3751	1
<i>Dactylis glomerata</i>	Montemayor del Rio	2018	40.34686	-5.9091	3
<i>Dactylis glomerata</i>	Mount Edgecomb, Plymouth	2017	50.35542	-4.17531	1
<i>Dactylis glomerata</i>	Roquestron	2017	43.87427	7.00969	1
<i>Dactylis glomerata</i>	Soglio	2016	46.34289	9.53726	3
<i>Dactylis glomerata</i>	Sur En	2020	46.81953	10.36487	1
<i>Holcus lanatus</i>	Aubonne	2005/2018	46.51232	6.36365	3
<i>Holcus lanatus</i>	Auvergne	2017	45.1263	2.88626	3
<i>Holcus lanatus</i>	Basilicata	2019	40.38004	15.68896	2
<i>Holcus lanatus</i>	Cevennes	2017	44.14829	3.96273	3
<i>Holcus lanatus</i>	Elsass	2017	48.02659	7.27517	3
<i>Holcus lanatus</i>	Kleinandelfingen	2020	47.60585	8.65983	2
<i>Holcus lanatus</i>	Montemayor del Rio	2018	40.34686	-5.9091	3
<i>Holcus lanatus</i>	Soglio	2016	46.34289	9.53726	3

<i>Holcus mollis</i>	Bretagne	2017	47.76867	-2.13004	3
<i>Holcus mollis</i>	Kew Gardens	2017	51.47825	-0.29903	3
<i>Lolium perenne</i>	Cevennes	2017	44.14778	3.99339	3
<i>Lolium perenne</i>	Montemayor del Rio	2018	40.34686	-5.9091	1
<i>Phleum pratense</i>	La Rippe	2018	46.38698	6.14711	3
<i>Poa nemoralis</i>	Aubonne	2018	46.51232	6.36365	1
<i>Poa nemoralis</i>	Auvergne	2017	45.1263	2.88626	3
<i>Poa nemoralis</i>	Elsass	2017	48.02659	7.27517	3
<i>Poa nemoralis</i>	Sent	2020	46.81931	10.35399	1
<i>Poa nemoralis</i>	Zollikerberg	1999	47.33803	8.60505	1
<i>Poa nemoralis</i>	Bot. Garden Zurich	2005	47.35879	8.55926	1
<i>Poa pratensis</i>	Bretagne	2017	47.76583	-2.13854	2
<i>Poa pratensis</i>	Elsass	2017	48.02659	7.27517	3
<i>Poa pratensis</i>	La Rippe	2018	46.38698	6.14711	3
<i>Poa sylvicola</i>	Castelbuono	2015	37.90021	14.08284	1
<i>Poa trivialis</i>	Auvergne	2017	45.1263	2.88626	3
<i>Poa trivialis</i>	Basilicata	2019	40.38004	15.68896	2
<i>Poa trivialis</i>	Bow, Devon	2017	50.80663	-3.82524	3
<i>Poa trivialis</i>	Elsass	2017	48.02659	7.27517	3
<i>Poa trivialis</i>	Horrabridge, Devon	2017	50.51391	-4.08839	2
<i>Poa trivialis</i>	La Rippe	2018	46.38698	6.14711	3
<i>Poa trivialis</i>	Montemayor del Rio	2018	40.34686	-5.9091	3
<i>Poa trivialis</i>	Mount Edgecomb, Plymouth	2017	50.35542	-4.17531	3
<i>Puccinellia distans</i>	Giebna, Poland	2005	52.7758	18.10075	1

* Additional isolate collected by Carolyn Young

Synthesis and Outlook

The discipline of evolutionary biology sprouted from grounds laid by Darwin and Wallace and has since dedicated itself to the search for evidence of evolution by identifying adaptive traits, elucidating their genetic basis and understanding the forces leading to divergence of populations and species. What commenced by observing phenotypes of organisms has in recent years been given fresh impetus with the ability to study evolution at the level of genes and even whole genomes: methodological and technical progress in the field of evolutionary genomics goes beyond counting wrinkly peas (Mendel 1866) and has made past evolutionary forces comprehensible, ongoing evolutionary processes measurable and future evolutionary impact predictable. These advances have led to a paradigm shift in the way we approach evolutionary research: from testing hypotheses, that is identifying genetic loci underlying a trait that has been previously classified as adaptive based on phenotypic data, to generating hypotheses about biological meaning and adaptive processes solely from the patterns of variation within genomic data. The possibility to sequence and analyze whole genomes across a plethora of different life forms has also taught us that adaptive variation can occur at different scales ranging from single base pair substitutions (SNPs) and short indels to presence-absence variants encompassing multiple genes and large-scale genomic rearrangements. While it is certainly the combined efforts across a multitude of organisms that advances our understanding of evolutionary processes as a whole, nowhere are adaptive genomic signatures thought to be so prevalent as in the genomes of pathogens that are subject to strong continual selection to coevolve with their hosts (Woolhouse et al. 2002, Paterson et al. 2010, Brockhurst et al. 2014). Pathogens therefore constitute ideal biological models to study evolution at the molecular level and among them, fungal plant pathogen systems have emerged to become some of the most enlightening.

They have particularly high evolutionary rates not only due to their pathogenic lifestyle but also experience the highest recombination rates among eukaryotes (Stapley et al. 2017). In many of them the haploid phase dominates the life cycle and rates of

adaptation have been shown to be faster in haploids compared to diploids as selection is more efficient (Otto and Gerstein 2008, Gerstein et al. 2011, Marad et al. 2018). This is because any beneficial mutations can be directly targeted by selection in a haploid, while recessive beneficial mutations are masked in their heterozygote state in a diploid and are therefore more likely to be lost (Haldane's sieve, Haldane 1927). Similarly, deleterious mutations cannot "hide" as recessives and will be eliminated more efficiently by purifying selection. Another aspect of fungal plant pathogens that boosts their evolutionary potential is the fact that their genomes tend to harbor highly dynamic compartments enriched in repeat elements (Möller and Stukenbrock 2017, Frantzeskakis et al. 2019), also called the «accessory» genome (Bertazzoni et al. 2018). Compartmentalized genomes are by no means unique to pathogenic fungi, indeed it seems that "getting organized" is a shared trend across many branches of the tree of life (Torres et al. 2020). What they all tend to have in common is that repeat elements seem to play an important role in the emergence of dynamic compartments and simultaneously as drivers of rapid sequence evolution in these regions of the genome (Seidl and Thomma 2017). They may be hot spots for mutations and recombination but also duplications, deletions large-scale genomic rearrangements as a direct result of transposable element activity as well as the genomic defense mechanisms counteracting their proliferation. Even in the absence of meiotic recombination genetic variation on which selection can then act may thereby be generated in and around dynamic genome compartments, an important consideration as the life cycle of fungal plant pathogens often entails phases of both sexual and asexual reproduction (Seidl and Thomma 2014).

Taken together, the study of fungal pathogens across a diverse range of ecosystems helps to shed light on fundamental mechanisms of eukaryotic genome evolution (Möller and Stukenbrock 2017, Torres et al. 2020) and adaptive divergence both within and between species (Gladieux et al. 2014). Analyzing their genomes reveals how coevolutionary dynamics and strong selection can lead to rapid adaptation and the emergence of major epidemics in homogenous managed ecosystems (e.g. Persoons et al. 2017, 2021), or maintain polymorphism across space and time in heterogenous natural environments (Thrall et al. 2012, Tack et al. 2014).

Rooted within this diverse field of novel methods and research questions ranging from «What are the sources of genomic variation?» to «How do pathogens coevolve with their hosts?» lie the objectives of the work presented in this thesis. While becoming very specialized may be a very successful strategy for a fungal pathogen, I believe it to be the pitfall in the multidimensional research field of plant pathology (a term I tend to avoid as it lacks the word “fungus” and might imply I would prefer to study plants). Diverse systems and diverse questions often require the need for flexibility in thinking and, although I am left with the feeling that I merely managed to scratch the surface, I would like to think that the most enticing feature of this study is not the application of new technology and methods but the fascinating *Epichloe* study system itself. Below are a selection of reasons and outlooks why.

EPICHLOE ENDOPHYTES HAVE DYNAMIC REGIONS ON STABLE CHROMOSOMES

Epichloe genomes have a remarkable structure and are highly compartmentalized whereby transposable elements organize chromosomes into a patchwork of distinct conserved and dynamic sequence regions. The size of the dynamic compartment shows striking variation among closely related species and exploratory analyses not presented here suggested that genome sizes also vary significantly among populations of the same species – likely as a result of different repeat content. While the synteny is highly conserved between the sibling species *E. typhina* and *E. clarkii*, and it seems conceivable that they should be able to produce viable hybrid offspring, major rearrangements are prevalent with more distantly related species beyond the *E. typhina* species complex. Nevertheless chromosome number appears to be conserved across the whole genus! As part of an ongoing collaboration focusing on comparative genomics across the whole *Epichloe* genus including asexual and sexual species, we now have twelve fully assembled genomes available all of which have exactly seven chromosomes (M. Cox and D. Winter, personal communication). The *Epichloe* endophytes represents an excellent group of organisms to study dynamic genome compartment among closely related species that exhibit different modes of reproduction and therefore disentangle how transposable element activity, genome defense mechanisms such as RIP and recombination affect genome evolution.

EPICHLOE SPAN THE WHOLE SYMBIOTIC CONTINUUM OF PLANT-FUNGAL INTERACTIONS

For the pathogenic *Epichloe* species that stand in the focus of this thesis sexual outcrossing is in all probability the pervasive form of reproduction. The high levels of variation observed within and among populations results, at least in part, from sexual recombination generating the raw material for evolutionary processes such as the continuous coevolution with the host. The impact that *Epichloe* pathogens have on their hosts are severe and, as highlighted sufficiently throughout this thesis, this provides an ideal prerequisite for the study of selective signatures at the genomic levels. However, the sexual *Epichloe* pathogens have so far received relatively little attention as model fungal pathogens for evolutionary research, probably in part because they infect wild grasses in natural ecosystems that are of no major concern to human agricultural practices. Nevertheless, a large body of research surrounds the *Epichloe* system and an extensive arsenal of molecular techniques has been developed; not because the pathogenic “bad guys” threaten our food supply but because the mutualistic “nice guys” harm not their host plant but the herbivores feeding on it and have therefore been commercialized as biological control agents in pasture farming (Caradus and Johnson 2020). The ability to cultivate, experimentally manipulate and genetically modify these fungi efficiently is a further benefit of this system and provides nearly boundless opportunities on which future research focusing on any *Epichloe*-grass association along the symbiotic continuum can build.

PLASTIC GENOMES WITHIN AND BETWEEN SPECIES

Both comparative genomics and population genomic studies such as this thesis that aim to investigate genetic variation in highly polymorphic sequence regions usually rely on curated genome assemblies and high quality alignments. These alignment-to-reference based approaches efficiently identify SNPs and also structural variation such as short presence absence polymorphisms (indels) or larger deletions given that they are present in the reference sequence but not in the sequences of the aligned individual (e.g. see the determination of mating types in Chapter II). However, the ability to detect other types

of genomic variation are limited, for example, large structural variants that are not found in the reference genome, or highly diverged from it, will be overlooked by most analyses (Eschenbrenner et al. 2020). In fact, regardless of its quality and completeness, a single reference genome cannot comprise all genetic and genomic variation of a given species (Parfrey et al. 2008). Novel approaches combine multiple genome sequences within species to build so called pan-genomes that account for this intraspecific variability (Gordon et al. 2017, McCarthy and Fitzpatrick 2019, Brockhurst et al. 2019). The small eukaryotic genomes of fungal plant pathogens are ideally suited to such approaches and allow for genomic variation to be studied at an unprecedented level elucidating the sources of evolutionary novelty and the role of dynamic genomes for adaptive evolution (Plissonneau et al. 2018, Badet and Croll 2020). Pangenome level analysis within the *Epichloe* study system using the extensive whole genome sequencing data generated in this study offers great potential to more accurately capture different levels of genomic variation and explore its evolutionary impact specifically in the highly dynamic regions of *Epichloe* genomes.

HOST JUMPS, DIVERGENCE AND SPECIATION

The population genomic analyses presented in this thesis provides insights into the adaptive processes and coevolutionary dynamics governing the interaction between *Epichloe* pathogens and their host grasses at the population level, the microevolutionary scale. Meanwhile, comparative genomic analyses have thus far barely been touched upon. The availability of whole genome sequences from a large number of closely related species within the *E. typhina* species complex also make this system suitable to study adaptation at the macroevolutionary scale and identify lineage-specific innovations that may have been integral for speciation. A particularly intriguing constellation is apparent within *E. clarkii* which occurs on the two distinct but related host grass species *H. lanatus* and *H. mollis* which appears to be the result of a relatively recent host jump and may represent a case of incipient speciation. In an exploratory analyses the results of which are not presented here, I have performed a genome-wide association study (GWAS) to identify putative genes underlying this host switch and detected genomic regions that show significant divergence between isolates specialized on different hosts. However, the underlying population structure potentially has a very strong impact as *H.*

mollis isolates have so far only been sampled from two locations. Future field work should focus on a more extensive sampling of *E. clarkii* from *H. mollis* and ideally finding locations where both host species are infected. The analyses of such sympatric populations in combination with experimental assessment of reciprocal infection success could provide key insights concerning host specialization and divergence in *Epichloe*.

A COMPLEX PHENOTYPE IN A COMPLEX ENVIRONMENT

While I have found exploring the various levels of genetic variation and the mechanisms generating this variation to be intellectually satisfying, it is ultimately only the first step in understanding the basis of adaptive evolution and underlying coevolutionary dynamics in the *Epichloe* system. In fact, to confidently talk about «coevolution» an evident next move would be to include data from the host plants and investigate their underlying genetic structure. My study species *E. typhina* and *E. clarkii* are diverse, highly specialized pathogens and their adaptation to and in a heterogenous natural environment involves many different genes as evident from the widespread signatures of recent selection in their genomes. Independent of the biological system under investigation, pathogenicity is among the most complex phenotypes in the natural world – which to many biologists seems to be the trigger of curiosity to study it. Pathogenicity is the consequence of a fine-tuned interaction between pathogen and the host it infects. This interaction is governed by continually varying selection pressures for the pathogen (and vice-versa): The host environment, independent of the biological system, is a complex one in which the immune system must be outwitted, adequate nutrition ensured, and competition avoided - and everything fluctuates in response to external biotic and abiotic conditions. In more general terms: adaptive evolution is a continuous process occurring in response to a heterogenous and ever fluctuating natural world. For most biological systems, observations, measurements and patterns become much more enlightening when assessed repeatedly over longer periods of time and I consider it a pity that with today's academic pulse, the opportunities to perform long-term field studies appear to have become rare. Nevertheless, sophisticated (co)evolutionary models, in the face of the current world affairs more popular than ever, are now able to

account for this heterogeneity making theoretical predictions and the formulation of testable hypothesis possible.

The ideal biological system that meets all criteria, fulfills all model assumptions and provides answers to all our questions does not exist - in fact progress in our understanding of the evolution of life in general or fungal plant pathogens in particular requires diverse disciplines working on diverse systems to answer diverse question – but I sincerely hope that this thesis can be considered proof that the *Epichloe* study system deserves a spot among the fungal model systems in evolutionary research.

FINAL NOTES ON A MARVELOUS KINGDOM

It has probably become obvious to attentive readers (or generally people that have gotten to know me) that while this document might qualify me as an evolutionary biologist on paper, I still remain a mycologist at heart. Fungi never cease to evoke my curiosity and passion. Fungi are enigmatic, ancient and everywhere, they often live and grow as networks of filaments embodying the essence of species' entanglements in the natural world. Their hyphal bodies stretch into realms unseen and tales untold, both wonderful and grim. Diving deep into the study of these organisms lets me discover more mysteries and inspires more questions than I could ever wish to solve. I suppose I have myself become entangled with this marvelous kingdom.

REFERENCES

- Badet, T., and D. Croll. 2020. The rise and fall of genes: Origins and functions of plant pathogen pangenomes. *Current Opinion in Plant Biology* 56:65–73.
- Bertazzoni, S., A. H. Williams, D. A. Jones, R. A. Syme, K.-C. Tan, and J. K. Hane. 2018. Accessories make the outfit: Accessory chromosomes and other dispensable DNA regions in plant-pathogenic fungi. *Molecular Plant-Microbe Interactions* 31:779–788.
- Brockhurst, M. A., T. Chapman, K. C. King, J. E. Mank, S. Paterson, and G. D. D. Hurst. 2014. Running with the Red Queen: The role of biotic conflicts in evolution. *Proceedings of the Royal Society B: Biological Sciences* 281:20141382.
- Brockhurst, M. A., E. Harrison, J. P. J. Hall, T. Richards, A. McNally, and C. MacLean. 2019. The ecology and evolution of pangenomes. *Current Biology* 29:R1094–R1103.
- Caradus, J. R., and L. J. Johnson. 2020. Epichloë fungal endophytes - From a biological curiosity in wild grasses to an essential component of resilient high performing ryegrass and fescue pastures. *Journal of Fungi* 6:1–44.
- Eschenbrenner, C. J., A. Feurtey, and E. H. Stukenbrock. 2020. Population genomics of fungal plant pathogens and the analyses of rapidly evolving genome compartments. Pages 337–355 in J. Y. Dutheil, editor. *Statistical Population Genomics, Methods in Molecular Biology*. 2090th edition. Humana, New York, NY.
- Frantzeskakis, L., S. Kusch, and R. Panstruga. 2019. The need for speed: Compartmentalized genome evolution in filamentous phytopathogens. *Molecular Plant Pathology* 20:3–7.
- Gerstein, A. C., L. A. Cleathero, M. A. Mandegar, and S. P. Otto. 2011. Haploids adapt faster than diploids across a range of environments. *Journal of Evolutionary Biology* 24:531–540.
- Gladieux, P., J. Ropars, H. Badouin, A. Branca, G. Aguileta, D. M. De Vienne, R. C. Rodríguez de la Vega, S. Branco, and T. Giraud. 2014. Fungal evolutionary genomics provides insight into the mechanisms of adaptive divergence in eukaryotes. *Molecular Ecology* 23:753–773.
- Gordon, S. P., B. Contreras-Moreira, D. P. Woods, D. L. Des Marais, D. Burgess, S. Shu, C. Stritt, A. C. Roulin, W. Schackwitz, L. Tyler, J. Martin, A. Lipzen, N. Dochy, J. Phillips, K. Barry, K. Geuten, H. Budak, T. E. Juenger, R. Amasino, A. L. Caicedo, D. Goodstein, P. Davidson, L. A. J. Mur, M. Figueroa, M. Freeling, P. Catalan, and J. P. Vogel. 2017. Extensive gene content variation in the *Brachypodium distachyon* pan-genome correlates with population structure. *Nature Communications* 8:2184.
- Haldane, J. B. S. 1927. A mathematical theory of natural and artificial selection, part V: Selection and mutation. *Mathematical Proceedings of the Cambridge Philosophical Society* 23:838–844.
- Marad, D. A., S. W. Buskirk, and G. I. Lang. 2018. Altered access to beneficial mutations slows adaptation and biases fixed mutations in diploids. *Nature Ecology and Evolution* 2:882–889.
- McCarthy, C. G. P., and D. A. Fitzpatrick. 2019. Pan-genome analyses of model fungal species. *Microbial Genomics* 5.
- Mendel, G. 1866. Versuche über Pflanzenhybriden. *Verhandlungen des naturwissenschaftlichen Vereins Brünn*:3–47.

- Möller, M., and E. H. Stukenbrock. 2017. Evolution and genome architecture in fungal plant pathogens. *Nature Reviews Microbiology* 15:756–771.
- Otto, S. P., and A. C. Gerstein. 2008. The evolution of haploidy and diploidy. *Current Biology* 18:1121–1124.
- Parfrey, L. W., D. J. G. Lahr, and L. A. Katz. 2008. The dynamic nature of eukaryotic genomes. *Molecular Biology and Evolution* 25:787–794.
- Paterson, S., T. Vogwill, A. Buckling, R. Benmayor, A. J. Spiers, N. R. Thomson, M. Quail, F. Smith, D. Walker, B. Libberton, A. Fenton, N. Hall, and M. A. Brockhurst. 2010. Antagonistic coevolution accelerates molecular evolution. *Nature* 464:275–278.
- Persoons, A., K. J. Hayden, B. Fabre, P. Frey, S. De Mita, A. Tellier, and F. Halkett. 2017. The escalatory Red Queen: Population extinction and replacement following arms race dynamics in poplar rust. *Molecular Ecology* 26:1902–1918.
- Persoons, A., A. Maupetit, C. Louet, A. Andrieux, A. Lipzen, K. W. Barry, H. Na, C. Adam, I. V Grigoriev, V. Segura, S. Duplessis, P. Frey, F. Halkett, and S. De Mita. 2021. Genomic signatures of a major adaptive event in the pathogenic fungus *Melampsora larici-populina*. *bioRxiv:2021.04.09.439223*.
- Plissonneau, C., F. E. Hartmann, and D. Croll. 2018. Pangenome analyses of the wheat pathogen *Zymoseptoria tritici* reveal the structural basis of a highly plastic eukaryotic genome. *BMC Biology* 16:5.
- Seidl, M. F., and B. P. H. J. Thomma. 2014. Sex or no sex: Evolutionary adaptation occurs regardless. *BioEssays* 36:335–345.
- Seidl, M. F., and B. P. H. J. Thomma. 2017. Transposable elements direct the coevolution between plants and microbes. *Trends in Genetics* 33:842–851.
- Stapley, J., P. G. D. Feulner, S. E. Johnston, A. W. Santure, and C. M. Smadja. 2017. Variation in recombination frequency and distribution across eukaryotes: patterns and processes. *Philosophical Transactions of the Royal Society B: Biological Sciences* 372:20160455.
- Tack, A. J. M., F. Horns, and A. L. Laine. 2014. The impact of spatial scale and habitat configuration on patterns of trait variation and local adaptation in a wild plant parasite. *Evolution* 68:176–189.
- Thrall, P. H., A. L. Laine, M. Ravensdale, A. Nemri, P. N. Dodds, L. G. Barrett, and J. J. Burdon. 2012. Rapid genetic change underpins antagonistic coevolution in a natural host-pathogen metapopulation. *Ecology Letters* 15:425–435.
- Torres, D. E., U. Oggenfuss, D. Croll, and M. F. Seidl. 2020. Genome evolution in fungal plant pathogens: Looking beyond the two-speed genome model. *Fungal Biology Reviews* 34:136–143.
- Woolhouse, M. E. J., J. P. Webster, E. Domingo, B. Charlesworth, and B. R. Levin. 2002. Biological and biomedical implications of the co-evolution of pathogens and their hosts. *Nature Genetics* 32:569–577.

Acknowledgments

First and foremost, I would like to express my gratitude to **Adrian Leuchtmann**, my main supervisor throughout the last four years and 56 days and the main person responsible, besides myself, for getting me into doing a PhD. It was your contagious enthusiasm for the wonderful world of fungi, especially "our" *Epichloe*, and your confidence in my abilities that got me hooked and brought me to this point. You are definitely the only person I know (myself excluded) who enjoys receiving a bouquet of diseased grass tillers for his birthday. I think it speaks for your skills as a mentor and advisor that the further this work has advanced, the more frequent it has become that I succeed to tell you something about *Epichloe* that you didn't already know (or come up with wild hypotheses that shake the things we thought we knew). You gave me the freedom to shape this PhD project into what I wanted it to be and always found the right words of encouragement when I needed them the most. Thank you for the knowledge, the trust and the opportunity to dive into this fascinating world.

To **Jessica Stapley**, I am deeply grateful for your mentorship and valuable guidance, especially in the last year of "writing-up" any "by-no-means-doing-any-more-analyses", and for your essential contributions to the work presented in this thesis. I may have been well equipped with mycological knowledge by the time we started working together but you have shaped my way of thinking and nurtured me into becoming an *evolutionary* mycologist. Most importantly, besides all the data-analyses and data-organization skills, you have taught me that being critical of one's work doesn't have to stand in conflict with being proud of it. I thank you for being my role model and friend, as the exemplary scientist you are (although you know more about recombining lizards than fungi, but nobody's perfect).

I further thank all the people who have contributed to this work or have supported the process of me attaining this degree. Specifically, I thank the members of my PhD committee, **Bruce McDonald** and **Daniel Croll**, for providing their support and for reading, scrutinizing and carefully evaluating this thesis. I am also very grateful to our

collaborators in New Zealand, **Murray Cox** and **David Winter**, who's skills and experience with assembling *Epichloe* genomes have been vital to build the foundation of this scientific work (the two reference genomes). I also thank you for your helpful comments and input as co-authors – working with you has been a true pleasure and I look forward to our continued collaboration in pursuing many other exciting research questions in the future.

Scientific work can flourish when planted in fertile grounds and given a suitable environment. For providing just that – a nurturing research environment - my thanks go to the people who have composed it: The **Plant Ecological Genetics group**.

In particular, I thank **Alex** for his perpetual effort and commitment as a group leader, for his support throughout these last few months of my PhD, for his kind serenity and heartfelt laughs. I thank **Franziska** for her tireless help with administrative issues and who has always had an open ear for both serious and silly business. You have made me feel very at home in this research group. I thank my friends and former office mates **Monica** and **Kevin**, my work-family, for making all the time spent at our desks bearable – no – enjoyable! Thank you for the hours committed to both sensible and senseless discussions and especially, to Monica, for the “mushroom afternoons” and to Kevin for his contribution of utter ignorance. I thank **Martin** for having significantly increased the proportion of fun in the last years of working together; not necessarily by being a sounding board for scientific questions (which you always tried to be even when completely overworked yourself) but simply by being my favorite person to tease. Maybe when the third baby is underway, and all your hair is gone you will understand that sometimes less is more (also applies to haploids versus diploids). I thank **Alessia** who's office door was always open, who passionately commits to any conversation no matter the content and who was always happy to leave her computer and join for a beer. Thanks also goes to my brothers and sisters in arms aka the other PEG PhD students: **Morgane**, **Ursina**, **Simon**, **Hirzi**, **Aksel**, **André** and **Fidel**. Sometimes sharing one's burden makes it a lot easier to carry. I thank **Karsten** for being a role model of anger management when it came to IT-related pitfalls, **Bea** for her essential supply of gossip and cigarettes and **Claudia** for tireless and reliable help with lab-work. I'm also very grateful to the **Genetic Diversity Center** staff, specifically **Aria**, **Silvia** and **Niklaus** for their scientific

support but most importantly for always welcoming be for a friendly chat or internal news exchange.

After a comprehensive evaluation of the PhD-process I conclude that a valuable prerequisite if not an essential consequence is the establishment of a thick skin. This includes dealing with doubt, frustration, stress and disappointment – challenges normal life itself tends to throw at us one way or another but somehow doing a PhD seems to put oneself into an especially susceptible condition. I do not consider myself to having been equipped with said pre-adaptation (the thick skin), and it has been a process of learning, cushioned by the many people in my life I am lucky to call my friends and family.

I am eternally grateful to **Sonja, Helen, Catalina, Valo, Julian, Mark, Paco** and **Adrian** for always having my back no matter the distance, for bearing with me and putting up with my ongoing struggles, for the countless big and small attentions and for the welcome distractions in the form of food, drink and more or less inspirational conversations. My unconditional appreciation and love go to my mother **Elisabeth**, the source of my connection to the natural world, who has given me the roots to grow and the wings to fly. To my father **Alois**, the source of my ever-critical mind, for his everlasting support and who has provided me with a wonderful environment to write this thesis in times of need. To my sister **Lilith**, for thirty years my partner in crime and my companion through thick and thin, through fields, forests, undergrowth, mud, and rivers, through happiness and despair. I could not have done this without you. Finally, my love and gratitude go to **Tobias**, who draws me out of my shell, who makes me laugh when I want to cry and for whom it is never a bad timing to help me out with literally anything and who has learned to extract concrete instructions from my more or less subliminal remarks (bringing me *another* tea or coffee, running to the shop *again* for milk or chocolate, feeding Horst the sourdough starter, or playing *another* song). For the never ceasing energy and passion you put in things – it's been infective. And for all your love.

CURRICULUM VITAE

ARTEMIS DIANA TREINDL

born on the 22nd of February 1991 in Zurich

Steinwiesstrasse 4

CH-8032 Zurich

Tel.: +41 78 785 74 83

Email: artemis.treindl@env.ethz.ch



EDUCATION

04/2017 - present

PhD candidate at the Institute of Integrative Biology
Department of Environmental Systems Science, ETH Zürich

09/2015 – 01/2017

MSc Biology, Major in Ecology and Evolution
ETH Zurich

02/2015 – 09/2015

MSc Biology, Major in Systematics and Evolution
University of Zurich

09/2010 – 12/2014

BSc Biology
University of Zurich

RESEARCH PROJECTS

PhD thesis: »Into the wild – Adaptation genomics of *Epichloe* grass pathogens in natural ecosystems«

Supervisor: Prof. Dr. Adrian Leuchtman, ETH Zurich

Master thesis: »Assortative mating - Reproductive isolation in sexual *Epichloë* endophytes«

Supervisor: Prof. Dr. Adrian Leuchtman, ETH Zurich

Other research projects:

- »Hybrid infectivity and hybrid vs. parent competition in *Epichloë* endophytes« (*Prof. Dr. Adrian Leuchtman, ETH Zurich*)
- »The distribution of rust fungi in plant communities of Switzerland« (*Dr. Reinhard Berndt, ETH Zurich*)
- »Identification and nematotoxicity-assay of secondary metabolites and proteins from mushroom extracts« (*Niels van der Velden, ETH Zurich*)
- »Vector Entomology: The effects of pyriproxyfen on the emergence of *Aedes japonicus* adults« (*Dr. Holly Tuten, UZH*)
- »Experimental Evolution in *E. coli*: Spontaneous evolution of acid stress tolerance« (*Dr. Kathleen Sprouffs, UZH*)

SCIENTIFIC CONTRIBUTIONS

Publications:

- Chromosome-level genomes provide insights into genome evolution, organization and size in *Epichloe* fungi (2021 – in review)
Artemis D. Treindl, Jessica Stapley, David J. Winter, Murray P. Cox, Adrian Leuchtmann
In review at Genomics, Elsevier
- A king amongst dwarfs: *Boletus edulis* forms ectomycorrhiza with dwarf willow in the Swiss Alps (2019)
Artemis D. Treindl and Adrian Leuchtmann
Alpine Botany, vol. 129: no. 2, pp. 185-189, Springer
[DOI: 10.1007/s00035-019-00218-2](https://doi.org/10.1007/s00035-019-00218-2)
- Assortative mating in sympatric ascomycete fungi revealed by experimental fertilizations (2019)
Artemis D. Treindl and Adrian Leuchtmann
Fungal Biology, vol. 123: no. 9, pp. 676-686, Elsevier
[DOI: 10.1016/j.funbio.2019.06.005](https://doi.org/10.1016/j.funbio.2019.06.005)

Conference talks:

- Studying plant-fungal interactions in natural ecosystems: Adaptation genomics in *Epichloë* endophytes
GDC Symposium 2020, ETH Zürich Zentrum, September 10, 2020
- Adaptive evolution and genome structure in *Epichloë* endophytes
Zurich Mycology Symposium 2020, Agroscope Wädenswil, January 24, 2020
- Positive assortative mating and reproductive isolation in sexual *Epichloë*
10th International Symposium on Fungal Endophytes of Grasses (ISFEG 10), Salamanca, Spain, June 18-21, 2018

Further activities:

- Collaboration on the introductory chapter to the new edition of the book »2000 Pilze« by Ruedi Winkler, Gaby Keller and Margrith Montalta (Hauptverlag, 2022): »Steps to molecular biological identification and reclassification of a species using the example of *Amanita citrina* var. *intermedia*« as well as collaboration on the encoding of fungal characteristics for the development of an associated mushroom-identification-App.
- PhD representative in the Institute Council (Institute for Integrative Biology, IBZ) and participation in the PhD committee to promote interaction among the research groups within the IBZ

TEACHING EXPERIENCE

- | | |
|-------------|---|
| 2015 – 2020 | Teaching assistant to Prof. Dr. Adrian Leuchtmann, Dr. Reinhard Berndt and Dr. Beatrice Senn-Irlet:
»Biodiversity and Ecological Significance of Fungi«
Course 551-0333-00L <i>ETH Zurich</i> |
| 2016 – 2020 | Teaching assistant to Prof. Dr. Adrian Leuchtmann:
»Field Course in Mycology«
Course 551-0216-00L <i>ETH Zurich</i> |

FURTHER SKILLS AND QUALIFICATIONS

- Federally certified mushroom inspector (VAPKO)
- Microscopy techniques: Light microscopy, SEM, fluorescence microscopy
- Laboratory work: isolation and cultivation of fungi as well as work under sterile conditions; DNA/RNA extractions with various methods; next-generation as well as third-generation sequencing of DNA/RNA (Illumina, PacBio, Nanopore)
- Genotyping and analysis of genomic data: SNPs, marker genes, microsatellites
- Programming: Data analysis and visualization using R as well as Unix tools (bash); use of high performance computing clusters via LSF task arrays

Languages:

- German: C2
- English: C2
- French: B1-B2