Received: 12 December 2021 | Revised: 14 March 2022 | Accepted: 16 March 2022

DOI: 10.1111/mec.16435

ORIGINAL ARTICLE

MOLECULAR ECOLOGY WILEY

The population genetics of adaptation through copy number variation in a fungal plant pathogen

Luzia Stalder[1] | Ursula Oggenfuss[1] | Norfarhan Mohd-Assaad[2,3] | Daniel Croll[1]

[1]Laboratory of Evolutionary Genetics, Institute of Biology, University of Neuchâtel, Neuchâtel, Switzerland

[2]Plant Pathology, Institute of Integrative Biology, ETH Zurich, Zurich, Switzerland

[3]Department of Applied Physics, Faculty of Science and Technology, Universiti Kebangsaan Malaysia, Bangi, Selangor, Malaysia

Correspondence
Daniel Croll, Laboratory of Evolutionary Genetics, Institute of Biology, University of Neuchâtel, Neuchâtel, Switzerland.
Email: daniel.croll@unine.ch

Funding information
D.C. received support from the Swiss National Science Foundation (grant 177052) in the framework of the COST Action HUPLANTcontrol (CA16110). N.M.A. was supported by the Ministry of Higher Education Malaysia (MOHE) and Universiti Kebangsaan Malaysia (UKM) under the SLAI scheme. N.M.A. has also received additional funding from UKM (grant GGPM-2019-042)

Abstract

Microbial pathogens can adapt rapidly to changing environments such as the application of pesticides or host resistance. Copy number variations (CNVs) are a major source of adaptive genetic variation for recent adaptation. Here, we analyse how a major fungal pathogen of barley, *Rhynchosporium commune*, has adapted to the host environment and fungicide applications. We screen the genomes of 125 isolates sampled across a worldwide set of populations and identify a total of 7,879 gene duplications and 116 gene deletions. Most gene duplications result from segmental chromosomal duplications. Although CNVs are generally under negative selection, we find that genes affected by CNVs are enriched in functions related to host exploitation (i.e., effectors and cell-wall-degrading enzymes). We perform genome-wide association studies (GWAS) and identify a large segmental duplication of *CYP51A* that has contributed to the emergence of azole resistance and a duplication encompassing an effector gene affecting virulence. We show that the adaptive CNVs were probably created by recently active transposable element families. Moreover, we find that specific transposable element families are important drivers of recent gene CNV. Finally, we use a genome-wide single nucleotide polymorphism data set to replicate the GWAS and contrast it with the CNV-focused analysis. Together, our findings show how extensive segmental duplications create the raw material for recent adaptation in global populations of a fungal pathogen.

KEYWORDS
adaptation, copy number variation, fungi, plant pathogen

1 | INTRODUCTION

The incidence of emerging pathogens has been increasing and presents a major threat to worldwide food security today (Brown et al., 2012; Fisher et al., 2012, 2018; Fones et al., 2020). Among pests, fungi cause the most crop devastation, accounting for ~30% of perennial yield losses worldwide (Fisher et al., 2018). Major challenges come from the ability of fungal pathogens to rapidly adapt to environmental changes such as the application of fungicides (Fisher et al., 2018). Point mutations and structural variants, such as copy number variants (CNVs), inversions or translocations provide genetic raw material for selection to act upon. In particular, gene duplication or deletion variants show greater phenotypic effects than point mutations and present a major source of phenotypic variation

driving species adaptation (Mérot et al., 2020; Peter et al., 2018; Wellenreuther et al., 2019). In the case of pathogens, the loss of genes that are recognized by the immune system of the host can be highly beneficial (Hartmann & Croll, 2017; Morris et al., 2012; Olson, 1999). In contrast, duplications of genes targeted by fungicides have been shown to confer resistance in several species (Coste et al., 2007; Leroux & Walker, 2013; Ma et al., 2006; Morschhäuser, 2016; Sánchez-Torres, 2021; Sionov et al., 2010; Steenwyk & Rokas, 2018; Zhang et al., 2019). Yet, despite their high adaptive potential, systematic population-wide analyses of adaptive CNVs have been scarce (Mérot et al., 2020).

The high adaptive potential of CNVs comes from their high genomic abundance, as well as from their manyfold phenotypic effects (Tang & Amon, 2013). In adaptive evolution experiments with nutrient-deprivation schemes, genes that facilitate the uptake and/or metabolism of the limiting factor are upregulated often through gene duplication (Tang & Amon, 2013). For example, glucose limitation in *Saccharomyces cerevisiae* cultures resulted in gene duplications encoding glucose transporters, while sulphate limitation resulted in a gene duplication encoding a high-affinity sulphate transporter (Mishra & Whetstine, 2016). In line with these observations, CNVs of genes targeted by fungicides have been repeatedly associated with resistance (Coste et al., 2007; Leroux & Walker, 2013; Ma et al., 2006; Morschhäuser, 2016; Sánchez-Torres, 2021; Sionov et al., 2010; Steenwyk & Rokas, 2018; Zhang et al., 2019). Azoles are the most widely used fungicides both in agricultural practices and in clinical settings to treat fungal infections (Azevedo et al., 2015; Cools & Fraaije, 2013). The dominant mechanisms to gain azole resistance are nonsynonymous mutations as well as CNVs of the gene *CYP51* (*ERG11* in yeast), a member of the cytochrome P450 family and the direct molecular target of azoles. Human pathogenic fungi including members of the genera *Candida* and *Cryptococcus* have repeatedly acquired azole resistance (Coste et al., 2007; Morschhäuser, 2016; Sionov et al., 2010; Todd & Selmecki, 2020). During infection, *Candida* species show loss of heterozygosity, allowing for more efficient selection as quasihaploids (Selmecki et al., 2006; Todd & Selmecki, 2020). Similarly, many crop pathogens have rapidly gained azole resistance through mutations or duplications of *CYP51* (Lucas et al., 2015). A particularly interesting case is the barley pathogen *Rhynchosporium commune*. In addition to adaptive mutations in *CYP51*, the *R. commune* genome encodes three *CYP51* paralogues including *CYP51A*, *CYP51B* and *CYP51A-p* (a loss of function copy of *CYP51A*) (Brunner et al., 2016; Hawkins et al., 2014). The *CYP51A* presence/absence polymorphism in *R. commune* is strongly correlated with azole resistance on a global scale (Brunner et al., 2016). Populations from the UK had very low *CYP51A* frequencies between 1890 and 1985 which then increased very rapidly, presumably as a response to increased azole treatments (Hawkins et al., 2014). CNVs in *R. commune* have also impacted genes contributing to virulence, including the genes coding for the necrosis-inducing proteins *NIP1*, *NIP2* and *NIP3* (Penselin et al., 2016). *NIP1* is recognized by the barley resistance gene *Rrs1*, impeding infection in a gene-for-gene model (Kirsten et al., 2012; Rohe et al., 1995; Schürch et al., 2004). Consequently, the *NIP1* gene has been undergoing several gene duplication events with some pre-dating speciation (Mohd-Assaad et al., 2019). In addition, adaptive deletion of the *NIP1* gene allows *R. commune* to infect barley cultivars carrying the cognate resistance factor *Rrs1* (Mohd-Assaad et al., 2019).

In eukaryotic microbial pathogens, adaptive CNVs are generated during both mitosis and meiosis (Hastings et al., 2009). During meiosis, nonhomologous recombination between specific chromosomal regions can cause the duplication and deletion of sequences among progeny (Hastings et al., 2009). In the absence of sexual reproduction, nonallelic homologous recombination during mitosis or aneuploidy are critical to generate genetic variability. Importantly, generation of CNVs is often driven by transposable elements (TEs), ubiquitous repetitive DNA sequences that can move from one location in the genome to another (Bourque et al., 2018; Feschotte, 2008; Wells & Feschotte, 2020). TEs can either passively facilitate structural variation by increasing the likelihood of nonhomologous recombination, or actively insert into new regions of the genome and create CNVs in populations (Bourque et al., 2018; Feschotte, 2008; Wells & Feschotte, 2020). The recent advent of genome analyses has allowed the mechanisms giving rise to structural variation to be to established (Mérot et al., 2020). However, how specific genomic features in microbial pathogens influence the creation and how population structure influences the trajectory of adaptive CNVs remain largely unknown (Mérot et al., 2020).

Here we investigated how the barley pathogen *R. commune* has adapted to fungicide applications, host resistance and different temperature regimes at a global scale. Using genome-wide association mapping, we identified CNV loci probably contributing to variation in life-history traits, including large segmental duplications of the fungicide resistance locus encoding *CYP51A*. We show that CNVs associated with phenotypic trait variation are close to specific TE categories. We contrasted the findings on CNV associations with mapping outcomes based on genome-wide single nucleotide polymorphisms (SNPs). We found that CNVs are enriched for gene functions related to host colonization and environmental adaptation. Together, our findings highlight the population genetic context and genomic environment of CNV-based adaptation in populations.

2 | MATERIAL AND METHODS

2.1 | Field collection of *R. commune*

We analysed 125 isolates of *Rhynchosporium commune* collected from naturally infected barley fields. Populations from nine different countries and four different continents were obtained: New Zealand (NZ), Australia (AU), Ethiopia (ET), Switzerland (CH), Spain (SP), Norway (NO), Finland (FI), Iceland (IS) and USA (US). We used all publicly available whole-genome sequencing data covering 12–14 isolates per population (Mohd-Assaad et al., 2016; Tryggvi S. Stefansson et al., 2013; Tryggvi S. Stefansson et al., 2014). The

sequencing data are available at the NCBI Sequence Read Archive under the BioProject accession no. PRJNA327656.

2.2 | Fungicide sensitivity assay

Fungicide sensitivity was assessed as previously described (Mohd-Assaad et al., 2016). Isolates were taken from −80°C storage by plating on lima bean agar plates (LBA; 60 g L^{-1} lima bean, 12 g L^{-1} bacteriological agar, 50 mg L^{-1} kanamycin). After 10 days at 18°C in the dark, spores were harvested by flooding the plate with sterile water, aided by scraping with a microscope slide and filtering through two layers of sterile cheesecloth. The quantification of spore densities was done using KOVATM Glasstic Slides that were adjusted to a concentration of 10^5 spores ml^{-1}. Cyproconazole (Syngenta) was serially diluted to achieve final concentrations of 375, 37.5, 3.75, 0.375, 0.0375, 0.00375 and 0.000375 mg L^{-1}. Synthetic Nutrient Broth (150 µl, SNB; 0.2% [w/v] KH_2PO_4, 0.2% [w/v] KNO_3, 0.1% [w/v] $MgSO_4.7H_2O$, 0.1% [w/v] KCl and 0.2% [w/v] sucrose) together with cyproconazole was prepared in each well and 50 µl of the spore suspension or sterile water as a negative control was added. Each isolate was tested in four technical replicates. Parafilm-covered plates were incubated in the dark at 18°C with 80% humidity for 7 days. Then, 20 µl of filter-sterilized 550 µM resazurin (RZ; Sigma-Aldrich) dissolved in phosphate-buffered saline solution (PBS, pH 7.4) was added to each well, followed by 24 hr of incubation. RZ measures the metabolic activity of fungal growth, and hence the reduction in RZ was used as a measure of fungal hyphal growth in the microtitre plate assay (Vega et al., 2012). The reduction in RZ was measured using a SpectraMax_i3 Multi-Mode Microplate Reader Platform with the SpectraMax-MiniMax Imaging Cytometer (Molecular Devices) at excitation and emission wavelengths of 565 and 590 nm, respectively. The dose–response curve was fitted for each assay individually using R with the four-parameter log-logistic model implemented in the DRC package (Ritz et al., 2015).

2.3 | Virulence assay

Virulence was assessed as previously described (T. S. Stefansson et al., 2014). Spore solutions were prepared from −80°C storage as described above. The quantification of spore density was done using a haemocytometer and concentrations were adjusted to 10^6 spores ml^{-1}. Spore solution volume was adjusted to 5 ml and 5 µl Tween 20 was added. Virulence was measured on the spring barley cultivar Beatrix (Viskosa 9 Pasadena, Saaten Union, breeders' Reference NS01/2449). Pots with soil and four seeds each were placed in a single greenhouse chamber, where plants were grown at 18°C during the day and 15°C during the night. Day length was set to 14 hr and humidity was set at 60%. Watering occurred from the bottom. After emergence of the second leaf (~12 days), each plant was inoculated until run-off in a sterile chamber. A spore suspension of 5 ml was used per isolate and each isolate was replicated three times. Plants were dried for 15–20 min and then arranged randomly in mobile humidity chambers that were placed in the growth chamber. Relative humidity was kept at 100% for 48 hr with the temperature and day length conditions as described above. The plants were then randomly arranged in a single greenhouse chamber and the relative humidity was set to 60%. Scoring for symptoms was done for 1 week beginning 9 days after inoculation. Fifteen days after inoculation, the second leaf on all plants was cut at the leaf base and mounted onto a light-blue paper sheet (size A4). Using a Canon EOS 60D camera and a 50-mm lens, digital images were taking in a darkroom. On each side of the mounted leaves, a fixed light source (Philips PF 319 E/44, 150 W) was placed at a distance of 30 cm. The percentage of leaf area covered by lesions was measured for each leaf using the image analysis software APS ASSESS (Lamari, 2002). The total number of isolates included in the analysis was 114 (with 11 isolates lacking data). Analyses were conducted using the manual panel and thresholds were set by hue in APS ASSESS. Virulence was logit-transformed prior to analysis due to zero inflation.

2.4 | Growth rates and colony melanization at different temperatures

Growth rates at 12, 18 and 22°C were assessed as previously described (Tryggvi S. Stefansson et al., 2013). Temperature sensitivity was calculated as the standard deviation of the growth rate at 12, 18 and 22°C. Hence, a lower standard deviation indicates higher growth resilience to temperature variation. Melanization was measured at 12, 18 and 22°C in the same setting as described above. The digital images of fungal colonies were analysed for melanization using automated image analysis based on IMAGEJ (Schneider et al., 2012). The method was identical to the one described for *Zymoseptoria tritici*, with the only difference that the greyscale (GS) used to measure melanization was inverted using the formula 255 − GS so that 255 equals black and 0 equals white (Lendenmann et al., 2015). Hence, higher values indicate higher levels of melanization.

2.5 | SNP and CNV calling

Single nucleotide polymorphisms calling was done as previously described (Mohd-Assaad, McDonald, & Croll, 2016). The HaplotypeCaller and GenotypeGVCF tools of the GATK 3.3–0 suite were used (McKenna et al., 2010). SNPs with a phred-scaled quality score <500 were removed and SNPs were retained if they satisfied the following conditions: −2 < BaseQualityRankSumTest <2, −2 < ReadPosRankSumTest <2, RMSMappingQuality <30, −2 < MappingQualityRankSumTest <2, QualByDepth >20 and FisherStrand <10. The SNP data set was filtered to retain only SNPs with a genotyping rate >90% and a minor allele frequency (MAF) >5%.

To call CNVs, the software CNVNATOR was used (Abyzov et al., 2011). Statistical analysis of short read coverage along the genome

was performed to identify CNVs in all 125 sequenced isolates compared to the UK7 reference genome (Penselin et al., 2016). The reference genome is assembled into 163 scaffolds with a scaffold N50 of 800.5 kb (Penselin et al., 2016). For each isolate, CNV events were assessed in 100-bp bins as recommended. CNVs were retained if the CNV genotype p-value < .05 and $q_0 < 0.5$. Additionally, CNVs were filtered for the normalized average read depth. Deletions with a normalized average read depth of <0.4 and duplications with normalized average read depth of >1.6 were retained.

Allele frequency of SNPs was calculated using PLINK version 1.07 (Purcell et al., 2007). For CNVs, allele frequency was calculated on a per-gene basis; that is, for each gene, the number of isolates affected by CNV was divided by the total of number of isolates.

To characterize the genomic environment of CNV loci, we split scaffolds into nonoverlapping 10-kb windows using *splitter* in the EMBOSS version 6.6.0.0 package (Rice et al., 2000). We used *intersect* in BEDTOOLS version 2.27.1 to calculate the percentage of basepair overlaps either with annotated genes or TEs (Quinlan & Hall, 2010). We used CIRCOS version 0.69.9 to visualize gene and TE density, as well as TE superfamilies for all scaffolds >0.5 Mbp (Krzywinski et al., 2009).

2.6 | RNA-seq-assisted gene model prediction, functional annotation and differential expression analyses

The transcription profiles of all genes were assessed using RNA-sequencing (RNA-seq) data generated for the reference isolate UK7. RNA-seq experiments were conducted for four conditions, two *in vitro* and two *in plantae*. *In vitro* culture conditions included growth on Luria-Bertani (LB) and Potato Dextrose Broth (PDB) media. For the *in planta* experiments, plants of the barley cultivar Beatrix (Viskosa 9 Pasadena, Saaten Union, breeders' Reference NS01/2449) were infected, and leaves were collected at 9 and 13 days post-infection (dpi). All experiments were conducted in triplicate. Total RNA was extracted using TRIzol (Invitrogen) following the manufacturer's recommendations. RNA integrity and quantity were assessed on a Bioanalyzer 2100 (Agilent) and a Qubit fluorometer (Life Technologies). Libraries were prepared using the TruSeq stranded mRNA sample prep kit (Illumina). Total RNA was ribosome-depleted by using polyA selection and reverse-transcribed into double-stranded cDNA. Raw reads are available at the NCBI Sequence Read Archive under BioProject accession no. PRJNA804666.

The software TOPHAT version 2.0.14 was used to align short reads to the reference genome (Stefansson, Willi, et al., 2014). To identify high-quality gene models, the reference genome was newly annotated. Intron splice site hints were generated using bam2hints, included in the AUGUSTUS version 3.2.1 software (Stanke et al., 2006). Due to the very high RNA sequencing depth available, intron splice hints were filtered for a minimum coverage of 20 reads to avoid an impact of spurious splice signals on gene prediction. To produce *ab initio* gene models, the BRAKER version 1.0 pipeline combining GeneMark-ET *ab initio* gene model predictions and AUGUSTUS version 3.2.1 (Hoff et al., 2016). GeneMark-ET was trained using the RNA-seq-based splice information as hints. AUGUSTUS was automatically trained using *ab initio* gene models that were fully supported by splice information. Finally, AUGUSTUS was used to predict gene models using both RNA-seq splice information and coding sequence hints based on exonerate protein alignments as extrinsic evidence. The gene models are available at https://doi.org/10.5281/zenodo.5730007. All newly annotated gene models for the UK7 reference genome were functionally characterized using INTERPROSCAN 5 version 79.0 (Jones et al., 2014). All proteins predicted to be secreted using SIGNALP version 4.1 were also screened for carbohydrate-active enzymes (CAZymes) using the carbohydrate-active enzyme annotation (dbCAN) release 5.0 (Petersen et al., 2011; Yin et al., 2012). Only protein hits that were predicted by DIAMOND, HMMER and HOTPEP were retained. Orthologues to virulence- and resistance-associated genes were identified using BLASTP against the pathogen–host interaction database PHI-base (Urban et al., 2020). Proteins were extracted with the search terms "effector" or "resist." Mapped reads overlapping gene models were counted using HTSEQ-COUNT version 0.11.1, setting the matching mode to "union" and filtering out reads with an alignment quality below 10 (Anders et al., 2015). Reads per kilobase of transcript per million mapped reads (RPKM) was calculated using R with the package edgeR version 3.30.0 (Robinson et al., 2009).

2.7 | Identification of conserved and orphan CNV genes

To identify genes affected by CNV, the percentage of overlap between CNVs and genes were calculated using the BEDTOOLS version 2.27.1 "intersect" command (Quinlan & Hall, 2010). A gene was conservatively considered as affected by a CNV if the CNV was overlapping >80% of the gene length. CNV overlap was defined as the sum of the overlapping deletion or duplication events. Genes affected by both deletions and duplications were excluded. To classify CNV genes as conserved or orphan CNV genes, we inferred orthologues of all *R. commune* genes using ORTHOFINDER version 2.1.2 with default parameters (Emms & Kelly, 2019). We used 68 Leotimycetes genomes including four *Rhynchosporium* genomes and four Sordariomycetes genomes for our analysis. We defined conserved *R. commune* genes as genes with orthologues in all *Rhynchosporium* species, >55 Leotimycetes and four Sordariomycetes. We defined *R. commune* orphan genes as genes with no orthologues in other species of the genus *Rhynchosporium*, Leotimycetes or Sordariomycetes.

2.8 | TE consensus identification, classification and annotation

To obtain a consensus sequence for each TE family, REPEATMODELER version open-4.0.7 (http://www.repeatmasker.org/RepeatModeler/) was run on the *R. commune* UK7 reference genome. The classification was based on the GIRI Repbase (version 2018) using

REPEATMASKER version open-4.0.7. (Bao et al., 2015; Smit et al., 2015). We used WICKERSOFT to finalize the classification of TE consensus sequences (Breen et al., 2010) including the following steps. (i) Screening for copies of known consensus sequences from other fungal species with BLASTN filtering for sequence identity >80% and sequence length >80%. (Altschul et al., 1997). (ii) Flanks of 10,000 bp were added and visually inspected for sequence similarity and terminal repeats with dot plots. Subsequent multiple sequence alignments were performed with 10–15 sequences using CLUSTALW (Thompson et al., 1994). (iii) Alignment boundaries were visually inspected and trimmed if necessary. (iv) Consensus sequences were classified according to the presence and type of terminal repeats, as well as homology of the encoded proteins based on BLASTX against the NCBI protein database. Consensus sequences were named according to the three-letter classification system (Wicker et al., 2007).

The reference genome was annotated with the curated consensus sequences using REPEATMASKER version open-4.0.7 with a cut-off value of 250 (Smit et al., 2015). Simple repeats, low-complexity regions and annotated elements shorter than 100 bp were filtered out and adjacent identical TEs overlapping by more than 100 bp were merged as belonging to the same TE family. Different TE families overlapping by more than 100 bp were considered as nested insertions and were renamed accordingly. Identical elements separated by less than 200 bp are indicative of interrupted elements and were grouped into a single element. TEs overlapping genes were recovered using the BEDTOOLS version 2.27.1 suite and the "overlap" function (Quinlan & Hall, 2010). In total, we classified 10,112 individual TE sequences grouped into 68 TE families. The complete TE annotation is available at https://doi.org/10.5281/zenodo.6010023. We used calcDivergenceFromAlign in REPEATMASKER to calculate the Kimura distances of each TE copy from the alignment.

2.9 | Population genetic analyses

Genetic structure among isolates was estimated using unsupervised model-based Bayesian clustering implemented in STRUCTURE version 2.34 as previously described (Mohd-Assaad et al., 2018; Pritchard et al., 2008). For principal component analysis (PCA) the "prcomp" function of the base R package stats was used. For estimatesof linkage disequilibria decay, the previously described filtered SNP data set was used. All calculations were made with VCFTOOLS version 0.1.12a (Danecek et al., 2011). Scans for selection were performed as described previously (Mohd-Assaad et al., 2018). We calculated integrated haplotype scores (iHS) and cross-population extended haplotype homozygosity (XP-EHH) using the R package rehh version 2.02 (Gautier et al., 2017). The analysis was performed for each genetic cluster separately. iHS estimates the decay of the EHH between an ancestral and derived allele at each SNP position (Sabeti et al., 2002; Voight et al., 2006). An extension to within-population EHH analyses is XP-EHH, which compares haplotypes in a pair of populations and is effective at detecting nearly fixed selective sweeps (Sabeti et al., 2002). Analyses of EHH are particularly suited to detect incomplete selective sweeps (Voight et al., 2006). We retrieved evidence from both iHS and XP-EHH analyses to assess candidate loci underlying recent local adaptation.

2.10 | Genome wide association studies (GWAS)

For the SNP GWAS analysis, we used the above described filtered SNP data set. For the CNV GWAS analyses, we used a gene presence–absence table indicating whether the gene was affected by a CNV or not (i.e., overlapping with a CNV > 80%). For both SNPs and CNVs, nine traits were tested (azole resistance, virulence, growth rate at 12, 18 and 22°C, temperature sensitivity, and melanization at 12, 18 and 22°C). For both SNPs and CNVs, we performed GWAS with TASSEL 20211027 using a mixed linear model (MLM; Bradbury et al., 2007). To correct for the relatedness among individuals, we calculated a kinship matrix using the centred identity-by-state method (IBS) in TASSEL (Endelman & Jannink, 2012). The *p*-values were corrected using the base R "p.adjust" function.

3 | RESULTS

3.1 | Identification of recent large segmental duplications and short deletions

To assess the impact of CNV dynamics in *Rhynchosporium commune* populations across the worldwide geographical range of the pathogen, we first performed a thorough annotation of protein-coding genes and repetitive sequences of the reference genome (Table S1). To annotate coding sequences, we integrated RNA-seq evidence both in culture medium and across the infection cycle. The improved annotation includes 11,923 genes compared to 8,847 genes in previous annotations, which were based primarily on expressed sequence tags (ESTs; Tables S2 and S3; Penselin et al., 2016). We functionally characterized all newly annotated gene models and additionally screened for candidate effector proteins and CAZymes, both of which are known to be critical for host colonization and exploitation. To analyse repetitive sequences surrounding CNVs, we performed an exhaustive *de novo* annotation of TEs. In total, we classified 10,112 high-quality TE sequences grouped into 10 superfamilies (Table S4).

We detected CNVs in 125 haploid *R. commune* isolates comparing read depth of Illumina sequencing data mapped to the reference genome (average depth of 39×, Table S1; Abyzov et al., 2011). We focused on polymorphic CNV events where both the gene presence and absence alleles are still segregating within the species. After quality filtering procedures, the total CNV count per isolate was not correlated with read coverage (see Figure S1). We retained 12,867 CNVs affecting a total of 2,195 distinct genes (18.4% of all genes). Even though we found more deletions than duplications (9191 vs. 3676), we found more duplications overlapping genes (2,160 genes

overlapping with duplications vs. 58 genes overlapping with deletions; Figure 1a; Figure S2, Table S5). Almost half of the genes deleted in some isolates were found duplicated in other isolates ($n = 23$ genes). Duplications, but not deletions, often span scaffolds, indicating large segmental duplications. In some isolates, multiple individual duplications amount to >200 kb each, suggesting that entire chromosomes may have been duplicated. Our findings of few and small deletions covering genes indicate purifying selection against the deletion of essential genes.

3.2 | Polymorphism frequency spectra across genetic clusters

To characterize the genetic differentiation across the pathogen distribution range, we performed an unsupervised Bayesian STRUCTURE analysis based on genome-wide SNPs. We identified three well-supported clusters: a first cluster including isolates from Scandinavia, Oceania and the USA, a second cluster separate from the other Northern European countries with mostly isolates from

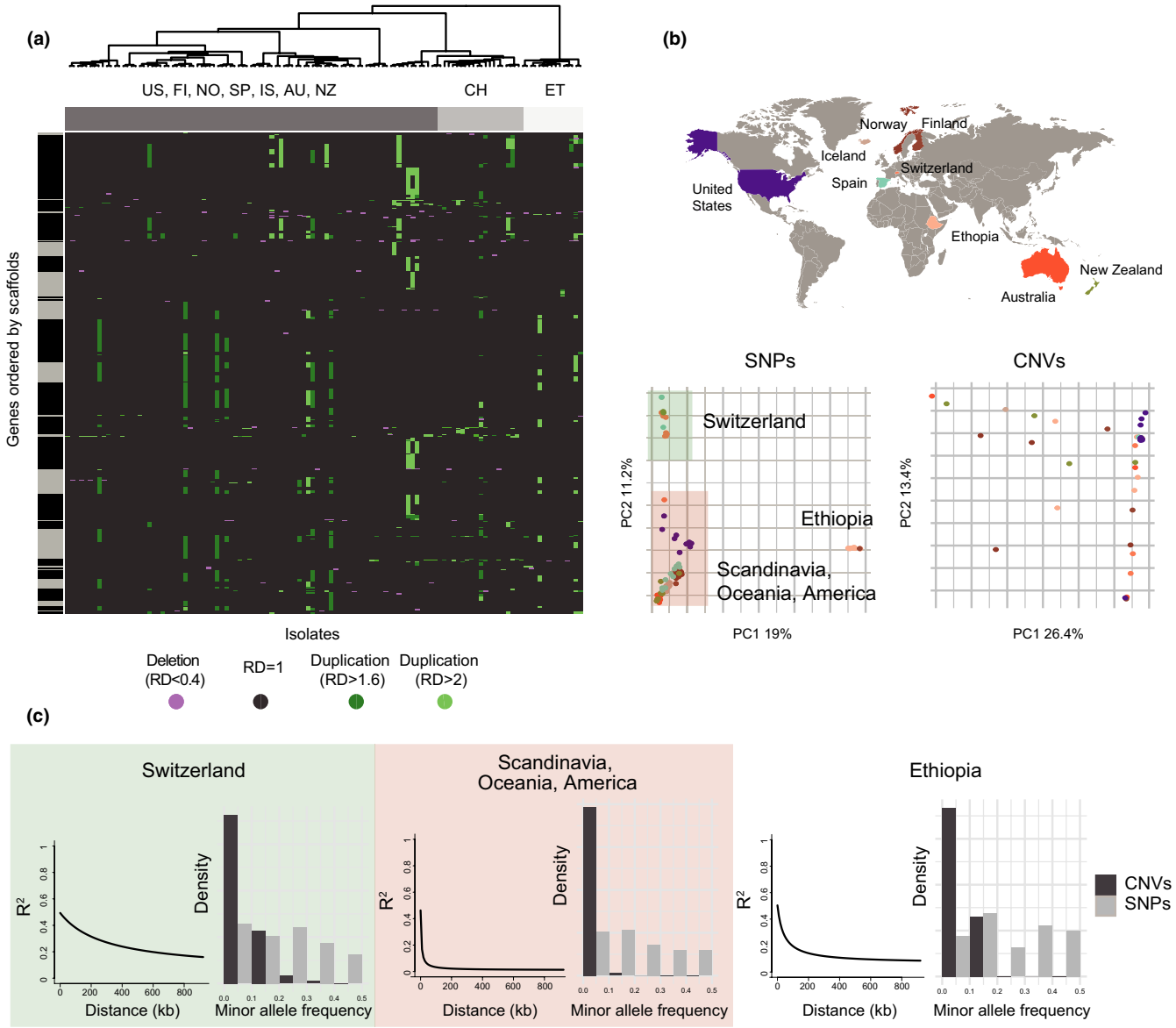


FIGURE 1 Segmental duplications and deletions cosegregate with genetic structure of *Rhynchosporium commune* populations. (a) Normalized read depth across genes and isolates. Read depth >1.6 is indicated in green, and read depth <0.4 is indicated in purple. Genes affected by copy number variations (CNVs) are sorted by scaffolds (indicated as black and grey bars). Isolates are clustered based on single nucleotide polymorphisms (SNPs). (b) Principal component analysis based on SNPs (left) and based on CNVs in genes (right). Colours indicate country of origin: New Zealand (NZ), Australia (AU), Ethiopia (ET), Switzerland (CH), Spain (SP), Norway (NO), Finland (FI), Iceland (IS) and USA (US). (c) Decay of linkage disequilibrium and minor allele frequency spectra of CNVs within the three genetic clusters. Linkage disequilibrium between pairs of SNPs, measured as r^2, is plotted against distance in kb. Minor allele frequency spectra are contrasted with the allele frequency of the derived allele at synonymous SNPs (n_{CNVs} and n_{SNPs} Swiss cluster =780 and 71,842, Scandinavian, Oceanian and American cluster = 2,061 and 219,357, Ethiopian cluster = 630 and 32,506)

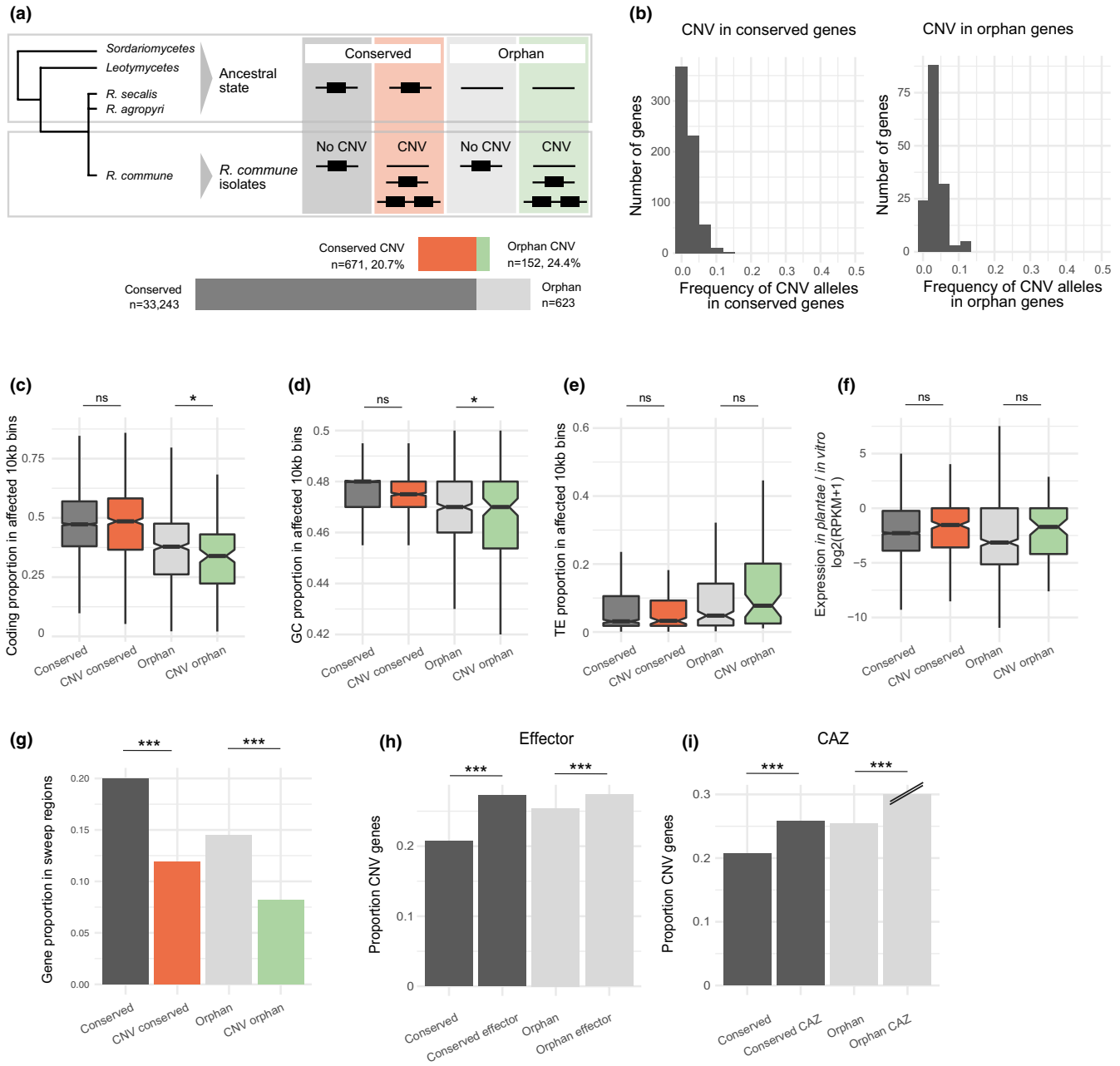


FIGURE 2 Evolutionary history and functions of genes affected by copy-number variation (CNV). (a) Phylogenetic tree representing the relationships between *Rhynchosporium commune*, Leotimycetes and Sordariomycetes. Evidence for orthologues was used to infer whether a CNV in *R. commune* is affecting a conserved or an orphan gene. (b) Frequency of CNVs in conserved and orphan genes for all isolates. (c–e) Differences between all conserved genes and conserved CNV genes, as well as between all orphan genes and orphan CNV genes regarding the coding sequence proportion (c), GC content (d) and transposable element proportion (e). All analyses were made in 10-kb bins with two-sided *t*-test (*$p < .05$). (f) Gene expression ratio between *in plantae* (9 days post-infection) and *in vitro* conditions between conserved and conserved CNV genes, and between orphan and orphan CNV genes. RPKM denotes reads per kilobase of exon per million mapped reads. (g) Enrichment analysis of conserved and orphan CNV genes in sweep regions identified by XP-EHH and iHS. The proportion of conserved CNV genes in sweep regions compared to the proportion of all conserved genes in sweep regions. Similarly, the proportion of orphan CNV genes in sweep regions compared to the proportion of orphan genes in sweep regions (Fisher's test conserved CNV vs. conserved $p < .001$, orphan CNV vs. orphan $p < .001$, number of sweeps in conserved genes =541, in recently lost genes = 72, in orphan genes = 9 and in recently gained genes = 90). (h) Enrichment analysis for effector genes in CNV genes. The proportion of effectors in conserved CNV genes is compared to the proportion of effectors among all conserved genes. Similarly, the proportion of effectors in orphan CNV genes is compared to orphan genes in general (Fisher's test conserved CNV vs. conserved $p < .001$, orphan CNV vs orphan $p < .001$, n_{Eff} in conserved genes = 22, in conserved CNV genes =6, in orphan genes =73, in orphan CNV genes = 20). (i) Enrichment analysis for cell-wall-degrading enzymes (CAZymes). Proportions are calculated as above (Fisher's test conserved CNV vs. conserved $p < .001$, orphan CNV vs. orphan $p < .001$, n_{CAZ} in conserved genes = 62, in conserved CNV genes = 16, in orphan genes = 0, in orphan CNV genes = 1; ***$p < .001$)

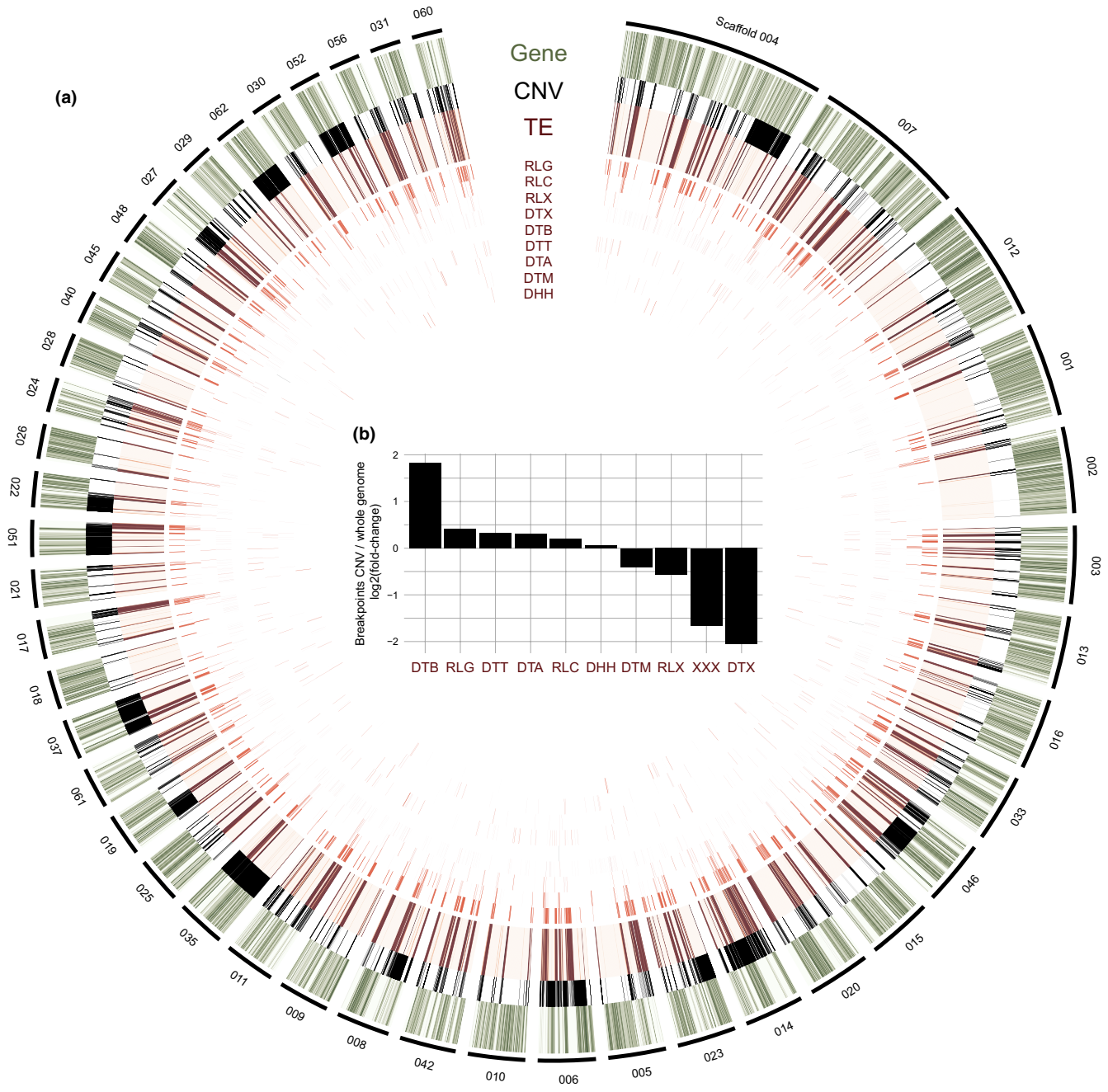


FIGURE 3 Genomic landscape of CNVs, transposable element (TE) families and genes. (a) Circos plot showing all scaffolds with length >500 kb. Indicated are the location of genes, CNVs, TEs and individual TE superfamilies: LTR retrotransposon superfamilies *Copia* (RLC), *Gypsy* (RLG) and unclassified LTR (RLX), DNA-TE superfamilies MITES-TIR (DTX), *PiggyBac* (DTB), *Tc1-Mariner* (DTT), *hAT* (DTA), *Mutator* (DTM) and *Helitron* (DHH). (b) Enrichment of TE superfamilies ±200 bp near the breakpoint of CNVs compared to their frequency in the genomic background. The fold change denotes the frequency of the TE family within the breakpoint point regions compared to the frequency of the TE family in the background

Switzerland, and a third cluster with African isolates from Ethiopia. The clustering is consistent with a PCA (Figure 1b, left panel). We analysed linkage disequilibrium decay within each cluster. The three clusters showed different rates of linkage disequilibrium (r^2) decay, indicating inbreeding among isolates or different degrees of clonality and demographic histories (Figure 1c). The Scandinavian, Oceanian and North American cluster showed a very rapid linkage disequilibrium decay to r^2 = .25 at an average distance of 2 kb, whereas the Swiss cluster showed the slowest decay to r^2 = .25 at an average distance of 580 kb. To elucidate the genetic structure of CNV polymorphisms, we performed a PCA based on CNV genotypes. The first principal component separated isolates primarily based on the total number of CNVs, with isolate AU118 showing the highest number of CNVs compared to the reference genome (Figure S3a,b). The PCA revealed no distinct clusters and shared little similarity with the PCA based on genome-wide SNPs, indicating that the majority

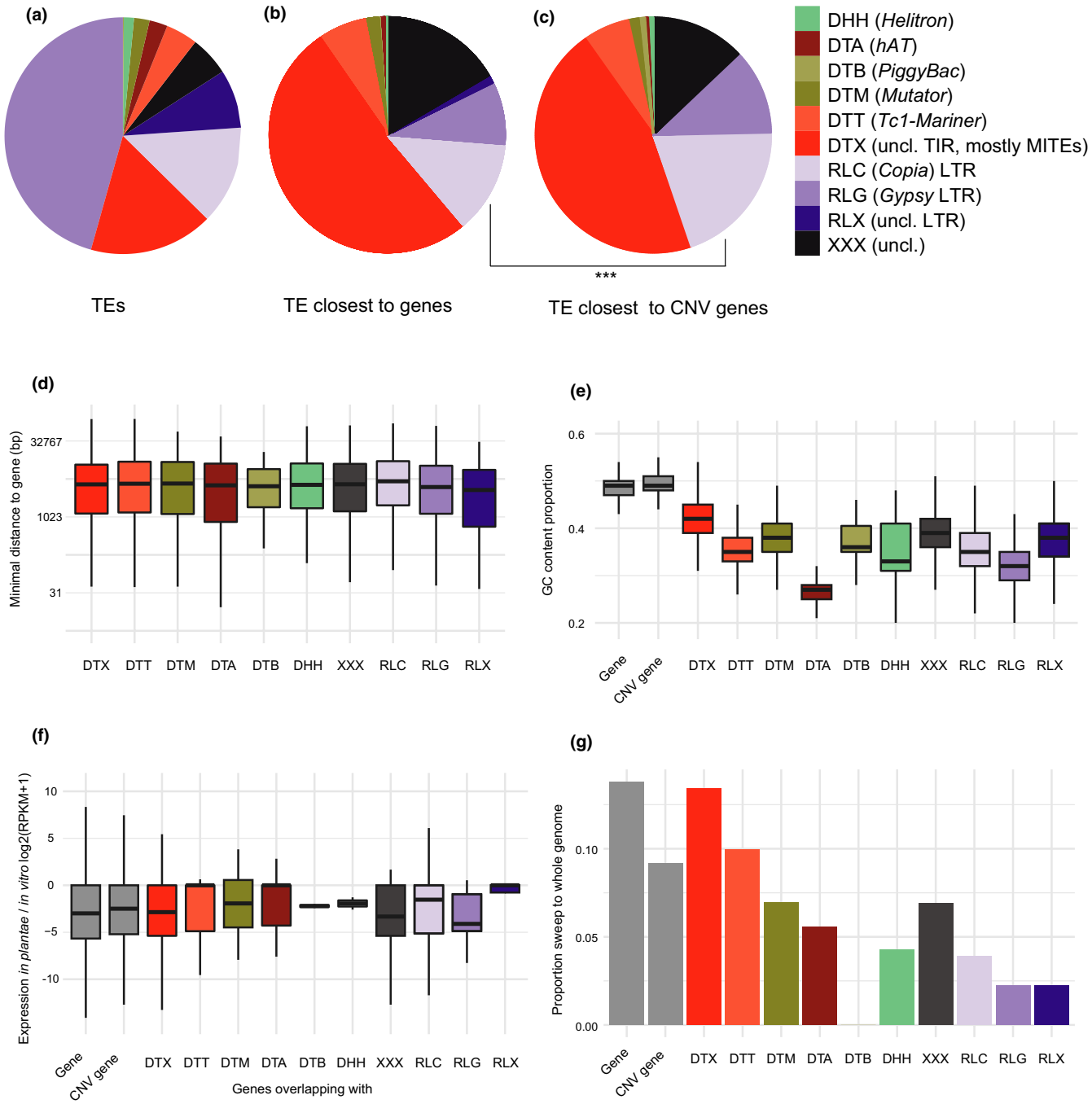


FIGURE 4 Transposable element (TE) families associated with CNV genes. (a) TE superfamily distribution in the *Rhynchosporium commune* genome, specifically DNA transposons of the superfamilies *hAT* (DTA), *PiggyBac* (DTB), *Tc1-Mariner* (DTT), *Helitron* (DHH), *Mutator* (DTM) and unclassified TIR (DTX), as well as LTR retrotransposon superfamilies *Copia* (RLC), *Gypsy* (RLG) and unclassified LTR (RLX), as well as unclassified TEs (XXX). (b) TE superfamily distribution of TEs closest to genes and (c) TEs closest to CNV genes. *Copia* elements are more often the closest CNV genes compared to other elements (Fisher's exact test $p < .001$). (d) The distance to genes of the different TE superfamilies. (e) GC content of by TE superfamilies. (f) Differential regulation upon infection of genes overlapping with different TE superfamilies. Differential regulation is shown as the difference in RPKM between 9 days post-infection *in plantae* and *in vitro* conditions. (g) TE superfamily distribution in *R. commune* sweep regions identified by XP-EHH and iHS. Bar diagram indicates the proportion of TE superfamilies in sweep regions compared to the whole genome

of the detected CNVs have been acquired recently (Figure 1b, right panel). Next, we analysed how selection acted on gene duplications and gene deletions in the different clusters. For this, we contrasted the MAF spectrum of CNV alleles (i.e., gene duplication or gene deletion) and synonymous SNPs. We found a strong excess of low-frequency CNV alleles compared with synonymous SNPs in all clusters (Figure 1c; Figure S3c, minor allele frequency ≤10%, Fisher's exact test p-values < .001). In addition, we identified regions under

positive selection using XP-EHH and iHS statistics and found that genic CNVs were depleted in positively selected regions (Table S6). This pattern suggests that both gene duplications and deletions are under strong negative selection in all clusters.

3.3 | Conserved and orphan gene CNVs are enriched in distinct adaptive functions

To investigate the evolutionary history of genes associated with CNVs in *R. commune*, we systematically analysed the presence of orthologues in other species. We defined orthogroups across sister species, and 68 other Leotiomycetes species as well as four more distantly related Sordariomycetes (Figure S4). Overall, 27.2% of *R. commune* genes ($n = 3,243$) share an orthologue in all Leotiomycetes and Sordariomycetes species, subsequently referred to as conserved genes. In contrast, 5.2% of *R. commune* genes ($n = 623$) have no orthologue in any other of the analysed species, subsequently referred to as orphan genes (Figure 2a; Figure S4). We found that 20.7% of the conserved and 24.4% of the orphan genes are affected by CNVs in *R. commune* (Figure 2a). CNV alleles in conserved genes are overall at lower frequencies in the species compared to CNV alleles in orphan genes, suggesting stronger purifying selection against CNVs in conserved regions (Figure 2b). To characterize the genomic context of genic CNVs, we compared CNV genes against their respective genomic background. We found that CNVs in conserved genes are in regions with similar coding density, GC and TE content, and show similar expression patterns as conserved genes (Figure 2c–f; Figure S5). In contrast, we found that CNVs in orphan genes are located in regions with lower coding density and GC content (Figure 2c–f; Figure S5). Subsequently, we investigated whether CNVs affecting conserved or orphan genes might impact host colonization. For this, we tested whether conserved and orphan CNV genes are enriched in positively selected regions in general, and whether CNV genes are enriched specifically in genes encoding candidate effector proteins and CAZymes. Effectors are small proteins secreted by the pathogen that manipulate the immune system and metabolism of the host to surmount plant resistance, whereas CAZymes are involved in the degradation of plant cell walls and nutrient acquisition (Esquerré-Tugayé et al., 2000; Fouché et al., 2018; Langner & Göhre, 2016; Sánchez-Vallet et al., 2018). To analyse CAZymes and effector candidates relevant for *R. commune* more comprehensively, we defined orphan genes at the genus level. We found that although conserved and orphan genes affected by CNV are depleted in sweep regions, they are strongly enriched in effector candidates and CAZymes, suggesting that CNVs contribute to host exploitation (Figure 2g–i).

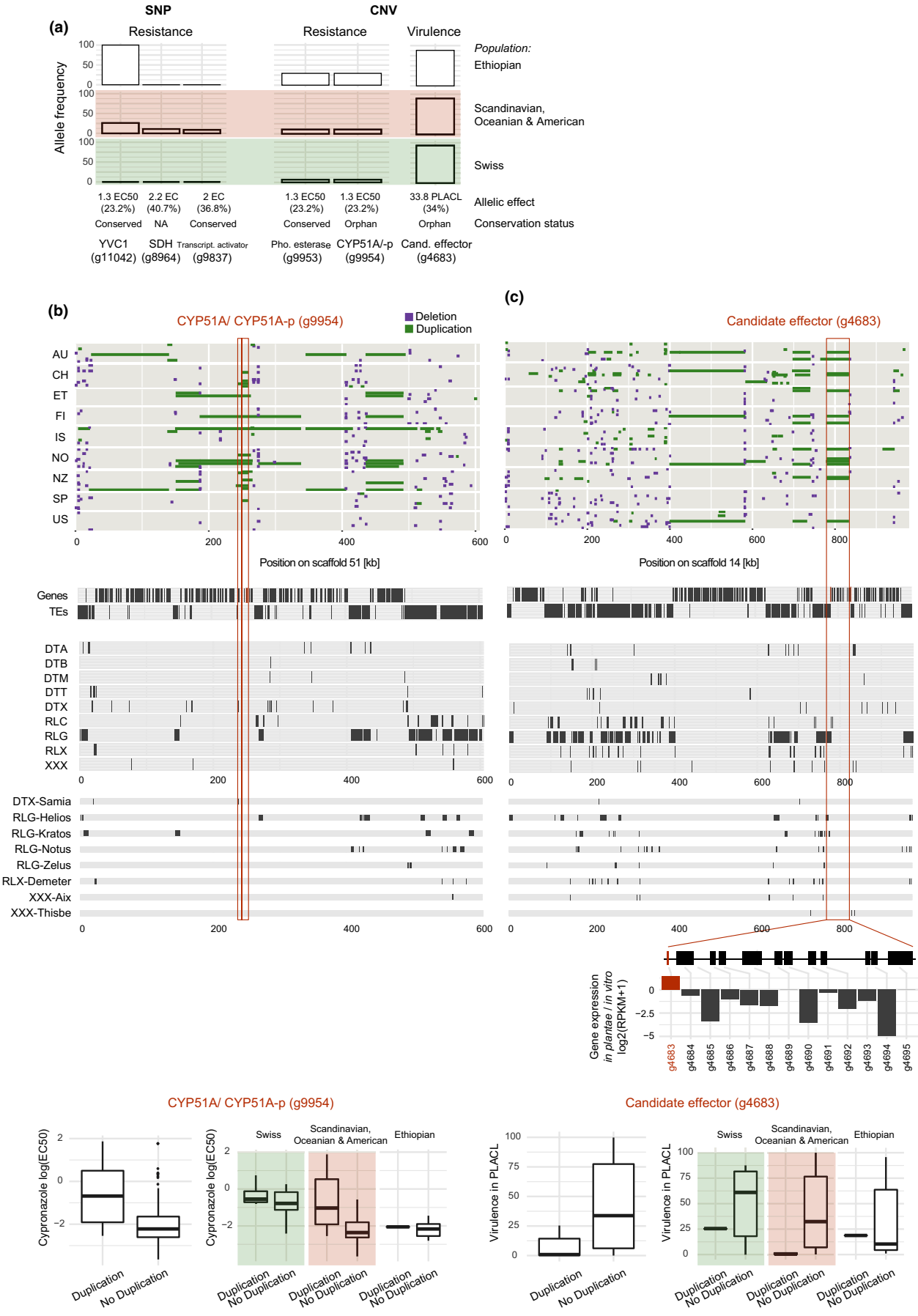
3.4 | Roles of transposable elements in CNV generation

CNVs in genomes often arise through the presence of repetitive sequences (Hastings et al., 2009; Wells & Feschotte, 2020). In agreement with this, we found that CNVs heavily overlap with TEs (Figure 3a). To investigate whether specific TE superfamilies are associated with CNV loci, we analysed whether individual TE superfamilies are enriched near CNV breakpoints compared to the genomic background. We found that the superfamilies *Piggy-Bag* (DTB) and *Gipsy* (RLG) are the most enriched, whereas the miniature inverted-repeat transposable elements (MITEs) DNA transposons (DTX) are the most depleted around CNV breakpoints (Figure 3b). We then focused specifically on genic CNVs. We found that even though the distance between genes and TEs is similar across TE superfamilies, MITEs (DTX) tend to be the closest to genes and show the highest GC content (Figure 4a–e). Interestingly, *Copia* elements (RLC) are most often closest to genes affected by CNV (Figure 4c). We found no meaningful *in plantae* upregulation of TEs overlapping with genes that could have indicated stress-induced transcriptional depression (Figure 4f). Additionally, we found that copies of MITEs, but not *Gypsy* and *Copia* elements, are frequent in regions with signatures of recent positive selection (Figure 4g; Figure S6). Together, this suggests that the creation of recent CNVs in the species was probably driven by a set of specific TE superfamilies.

3.5 | Large segmental duplications are associated with azole resistance and virulence

Next, we investigated the association of CNVs with phenotypic trait variation of *R. commune* isolates. For this, we performed a GWAS of CNV gene loci. We identified an association of CNV genes with

FIGURE 5 Associations of segmental duplications with azole resistance and virulence. (a) Frequency of the trait-associated genic CNV or genic SNP alleles identified with association analysis per genetic cluster. Alleles are associated with higher cypronazole resistance (measured as EC50) and higher virulence (measured as percentage leaf area covered by lesions, PLACL). Reported CNV and SNP alleles have FDR-corrected $p < .06$. (b) Duplications (green) and deletions (grey) on scaffold 51 encoding the azole target gene *CYP51A*/*CYP51A-p* (marked with a rectangle). The *CYP51A*/*CYP51A-p* gene is spanned by multiple large segmental duplications. At the bottom, the positions of genes and transposable element (TE) families are shown. The bar plots show cypronazole resistance depicted as log(EC50) of *Rhynchosporium commune* isolates with and without duplication of the *CYP51A*/ *CYP51A-p* gene, grouped by genetic cluster ($n_{\text{Duplication}}$ and $n_{\text{No Duplication}}$ for the Swiss cluster = 4 and 10, Scandinavian, Oceanian and American cluster =10 and 84, Ethiopian cluster = 1 and 12, *CYP51A*/*CYP51A-p* association overall $p = .01$). (c) Duplications (green) and deletions (grey) on scaffold 14 encoding the virulence-associated candidate effector gene g4683 (marked with a rectangle). The barplots show the percentage leaf area covered by lesions (PLACL) of *R. commune* isolates distinguished by the duplication of the candidate effector gene and genetic cluster ($n_{\text{Duplication}}$ and $n_{\text{No Duplication}}$ for the Swiss cluster = 1 and 11, Scandinavian, Oceanian and American cluster =9 and 81, Ethiopian cluster = 1 and 11, association overall $p = .06$)



both azole resistance and virulence (lesion coverage on leaves) (Table S7). However, we found no significant associations with pathogenesis-related colony melanization or with colony growth rates at different temperatures (FDR-corrected $p > .05$, Table S7). To provide a genomic context to the CNV-based association analyses, we repeated the association mapping with genome-wide SNPs. The GWAS with SNPs revealed significant genic associations for azole resistance, but not for virulence (FDR-corrected $p < .05$, Table S8). Among the significant associations for azole resistance, we found no overlap between SNPs and CNVs. We found considerable variation in the frequencies of the positively selected alleles among the three genetic clusters identified through CNV and SNP association mapping for azole resistance (Figure 5a). SNP and CNV alleles under positive selection for azole resistance are frequent in the Swiss and the Scandinavian, Oceanian and North American cluster, but absent in the Ethiopian cluster, which is consistent with the largely nonexistant use of azoles in Africa ("Food and Agriculture Organization of the United Nations [FAO] Pesticides Use Database" 2021).

Interestingly, the strongest CNV association with azole resistance is the duplication of a sterol 14a-demethylase gene *CYP51A*/*CYP51A-p* (g9954) and the adjacent gene encoding a calcineurin phosphoesterase (g9953) (Figure 5b; Figure S7A). The protein encoded by *CYP51A* is the direct target of azole fungicides, and overexpression and point mutations of *CYP51A* have been shown to contribute to azole resistance in many crop pathogens (Lucas et al., 2015). Here, we observed that a large duplication of the region (2–200 kb) harbouring a *CYP51A* paralogue significantly contributes to variation in azole resistance (Figure 5b). We found that the association between resistance and *CYP51A* paralogue duplication is strong in the Scandinavian, Oceanian and American cluster and the Switzerland cluster but not in the Ethiopian cluster. Only a single Ethiopian isolate carried a *CYP51A* duplication, preventing a direct test for association. The low frequency in the Ethiopian cluster is consistent with the lack of azole application in this region compared to high azole usage in European or North American regions ("Food and Agriculture Organization of the United Nations [FAO] Pesticides Use Database" 2021). CNV calling based on short read sequences is not able to confidently distinguish between the *CYP51A* and *CYP51A-p* pseudogene paralogues due to their sequence similarity. Hence, the predicted duplication could indicate either a duplication of *CYP51A* or the pseudogene. Alternatively, indications for a duplication of *CYP51A* could also stem from a genotype carrying one copy of *CYP51A* and *CYP51A-p* each. We also identified a frequent segmental duplication on scaffold 14 harbouring 13 genes, which is associated with lower virulence (Figure 5c). Interestingly, the first gene (g4683) of the duplication is a candidate effector putatively interacting with the host immune system and the gene was upregulated on the host compared to *in vitro* conditions (Figure 5c).

Analysing the genomic context of the virulence-associated duplication showed that the *Gypsy* families Helios and Kratos are located at the CNV breakpoints. In the flanking regions, we found the *Gypsy* families Notus and Zelus, and the unclassified Aix, Thisbe and Demeter elements (Figure 5c). To analyse whether individual TE families were recently active, we calculated Kimura distances for each TE family (Figure 6). Both the Helios and Kratos family are within the third percentile of the Kimura distance distribution of all TE families. This shows that these TEs are among the most recently active elements in the genome. Hence, the

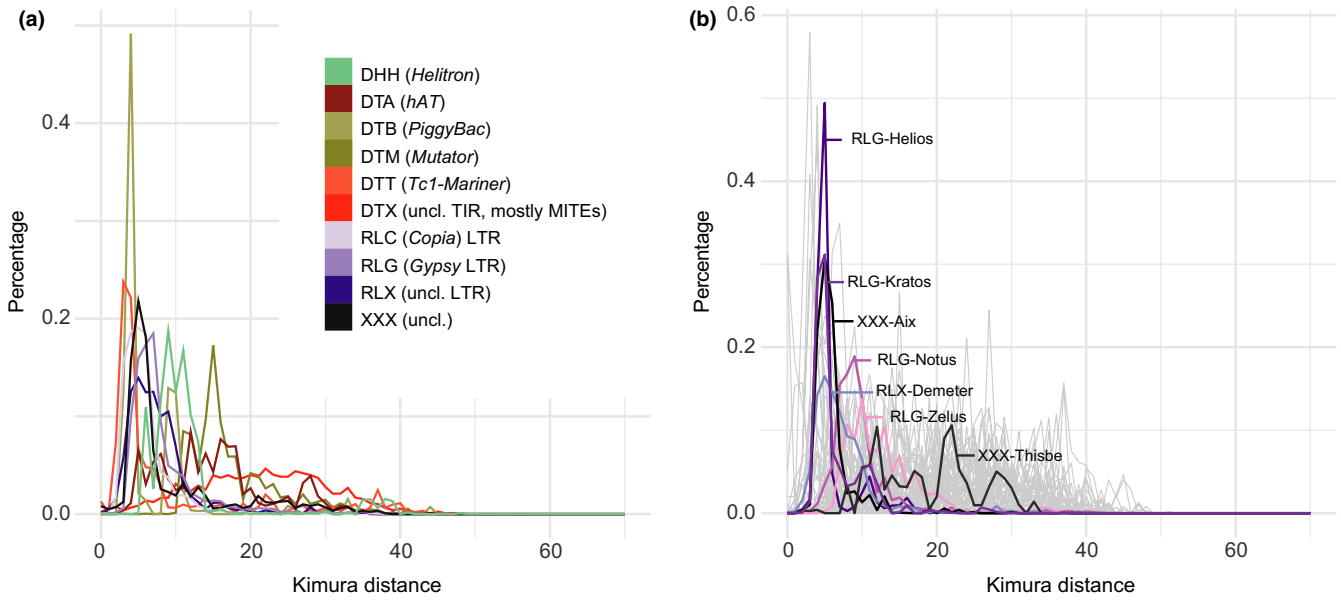


FIGURE 6 Transposition history of CNV-associated TE families. (a) Kimura distance of different TE superfamilies, specifically DNA transposons of the superfamilies *hAT* (DTA), *PiggyBac* (DTB), *Tc1-Mariner* (DTT), *Helitron* (DHH), *Mutator* (DTM) and unclassified TIR (DTX), as well as LTR retrotransposon superfamilies *Copia* (RLC), *Gypsy* (RLG) and unclassified LTR (RLX), as well as unclassified TEs (XXX). (b) Kimura distance of individual TE families. Highlighted families include RLG-Helios, RLG-Kratos, RLG-Notus, RLG-Zelus, RLX-Demeter, XXX-Aix and XXX-Thisbe

segmental duplications associated with major phenotypic trait variation across environments probably originated through the recent activity of specific TEs.

4 | DISCUSSION

Pathogens in agricultural environments are under strong selection pressure to adapt to changing environments (Fisher et al., 2018). CNVs have been increasingly recognized as important raw material for adaptive evolution (Mérot et al., 2020; Peter et al., 2018; Wellenreuther et al., 2019). We identified CNV loci associated with azole resistance and virulence variation across genetically differentiated global populations of the barley pathogen *Rhynchosporium commune*. We showed that adaptation through CNVs was probably facilitated by the action of a specific set of TE families.

We used a broad phylogenomic framework of Sordariomycetes and Leotiomycetes fungi to distinguish conserved and orphan genes affected by CNVs. We found that purifying selection acted more strongly against CNVs in conserved genes than on CNVs in orphan genes. This matches the expectation that conserved genes tend to encode essential protein functions, whereas orphan genes are selectively mostly neutral and lack essential functions at early stages of gain-of-function processes (McLysaght & Guerzoni, 2015). Interestingly, even though both conserved and orphan genes affected by CNV were depleted in regions under positive selection, both CNV gene categories are enriched in functions related to host interactions including effector candidates, which may enable the pathogen to manipulate the host physiology. Similar findings have been made in the crop pathogen *Zymoseptoria tritici*, but not in the castrating anther smut fungi *Microbotryum*, underpinning the diversity in host adaptation strategies among pathogens (Beckerson et al., 2019; Hartmann & Croll, 2017; Hartmann et al., 2018). In addition to effector genes, CNV genes were also enriched in genes encoding CAZymes involved in the degradation of plant cell walls and nutrient acquisition (Esquerré-Tugayé et al., 2000; Langner & Göhre, 2016). The increased genic CNV rates for functions related to host exploitation may have been driven by strong selection on pathogen populations to gain new functions and evade detection of proteins exposed to the host immune system.

Consistent with the over-representation of host exploitation functions at CNV loci, we found strong support for the role of individual CNV genes in association with pathogenesis-related trait variation. The restriction to genic CNV events reduces the power of the association mapping but focuses the analyses on the most robustly called set of CNV events. We found that the CNVs underlying phenotypic trait variation are large segmental duplications affecting multiple genes. The most striking duplication associated with fungicide resistance included the *CYP51A* gene known to encode the target of azoles. Using targeted polymerase chain reaction (PCR) assays, previous studies revealed a correlation between azole sensitivity and the presence of the *CYP51A* paralogue in *R. commune* (Brunner et al., 2016; Hawkins et al., 2014). Here, we found that CNV variation at the *CYP51A*/*CYP51A-p* loci is relevant for azole resistance in a global population genetic context. However, the short read mapping approach for CNV calling cannot confidently distinguish between the close paralogues *CYP51A* and *CYP51A-p* due to the high degree of sequence similarity. Hence, we could not pinpoint which of the two paralogues is causal for the variation in resistance. However, based on previous analyses, *CYP51A-p* contains premature stop codons and mutations do not contribute to azole resistance, and therefore at least part of the observed CNV variation is related to *CYP51A* (Brunner et al., 2016; Hawkins et al., 2014). In human pathogenic yeasts such as *Cryptococcus neoformans* and *Candida albicans*, several chromosomal duplications have been associated with *ERG11*, the *CYP51* orthologue (Coste et al., 2007; Sionov et al., 2010). Amplification of the locus is typically observed over the course of a single infection under strong selection pressure. Chromosomal aberrations are often accompanied by significant fitness reductions and are lost in the absence of fungicide use (Morschhäuser, 2016; Zhang et al., 2019). In plant pathogens, complete chromosomal duplications of *CYP51* have not been observed so far. However, many ascomycetes show several deeply divergent *CYP51* paralogues, whereas *Ca. albicans* and *Cr. neoformans* have only a single copy of *CYP51* (Zhang et al., 2019). In *R. commune*, we found that the duplication affecting *CYP51A* ranges from 2 to 200 kb in different isolates. Hence, it is plausible that in some *R. commune* isolates, the *CYP51A* amplification arose analogously as in human pathogens from partial chromosomal duplications. As we lacked a chromosome-scale assembly of the *R. commune* genome, we were unable to identify aneuploidy or near chromosome-scale duplication events. However, the large observed duplications share similarities with the process of intrahost antifungal resistance evolution in human pathogens (Coste et al., 2007; Selmecki et al., 2006; Sionov et al., 2010).

Our CNV association mapping detected no virulence associations at the *NIP* effector gene loci, but this can largely be explained by the fact that the tested barley cultivar lacks the corresponding receptor (Mohd-Assaad et al., 2019). We identified a previously unknown virulence locus encoding an effector candidate gene upregulated *in plantae* compared to *in vitro* conditions. Isolates carrying duplicated sequences showed reduced symptoms suggesting that the duplication increases the risk of recognition by the host immune system. Gene duplications can act on the dosage of a gene and higher transcription may indeed underpin maladaptive expression of the effector. Rapid sequence turnover at loci encoding effectors is widely documented and often driven by rapid adaptation to host resistance mechanisms (Fouché et al., 2018; Sánchez-Vallet et al., 2018; Rouxel & Balesdent, 2017). For example, the RXLR effector reservoir in *Phytophthora* species probably evolved from a common ancestor by rapid duplication and divergence (Jiang et al., 2008). Here, we pinpointed the specific TE families constituting the breakpoints of the duplication. Genome-wide analyses showed that the TE families are among the most recently active families and belong to the abundant *Gypsy* superfamily. The recent transpositions of the TEs within the species suggest that the duplication was of recent origin and generated during an episode of high TE activity.

Complex sequence rearrangements have been repeatedly associated with specific TEs inducing nonhomologues recombination events (Bourque et al., 2018; Wells & Feschotte, 2020). Specifically in the absence of meiosis, TE-mediated recombination can provide an important source of adaptive variability. In the partially clonal pathogen *R. commune*, we found that *Gypsy* retrotransposons make up about half of the TE copies in the genome, similar to many other fungal pathogens including *Blumeria graminis* and *Z. tritici*, where retrotransposons contributed even to recent genome expansions (Frantzeskakis et al., 2018; Oggenfuss et al., 2021). We found that *Piggy-Bag* DNA-transposons and *Gypsy* retrotransposons were particularly enriched at CNV break points, strongly suggesting a causal link with the generation of CNVs. In addition to *Gypsy* elements, the *R. commune* contains a large proportion of DNA transposons with terminal inverted repeats (TIRs), mostly from the MITEs and *Tc1-Mariner* families. As previously observed, we found that MITEs and *Copia* elements are located preferentially close to genes (Casacuberta et al., 1998; Castanera et al., 2021; Oki et al., 2008; Santiago et al., 2002). Interestingly, *Copia* elements were overrepresented near genes affected by CNVs. In plants, adaptive *Copia* element insertions have frequently been observed, and the elements often have regulatory functions (Baduel et al., 2021; Kanazawa et al., 2009; White et al., 1994). In plant pathogens, retrotransposon elements were found to be de-repressed upon stress with consequences for virulence on the host (Dubin et al., 2018; Fouché et al., 2020). For example, in *Z. tritici*, *Copia* elements are upregulated during early infection, whereas other TE elements are only upregulated during the less stressful saprophytic stages of the pathogen life cycle (Fouché et al., 2020). However, we found no indications for *Copia* related gene de-repression. The retention of MITEs in proximity to genes probably stems from the comparatively minor impact of short element insertions near coding sequences under purifying selection. However, the retention of some TEs near genes probably reflects also positive selection favouring, for example, an adaptive regulatory variant triggered by the insertion event (Castanera et al., 2021). MITEs are known to have expression patterns that facilitate the infection process in several pathogens (Schmidt et al., 2013). The presence of MITEs has even allowed the prediction of previously unknown virulence genes in *Fusarium oxysporum* (Schmidt et al., 2013). In line with this, we found that MITEs are among the most common elements in selective sweep regions. Together, our results suggest an involvement of specific TE families in adaptive CNV generation and pinpoint specific TE families at breakpoints of trait-associated CNVs.

Adaptive genetic variation is typically mapped using genome-wide SNPs because such markers can densely cover most chromosomal regions. At high marker density, SNPs tend to associate with additional types of polymorphisms (e.g., insertions, deletions and duplications) segregating in populations through linkage disequilibrium (Tam et al., 2019). Hence, association mapping performed on SNPs is expected to reveal associated polymorphisms even if the causal variant is not an SNP. Here, we identified no associated gene both in SNP- and in CNV-based association mapping. Overall, our comparison of SNP- and CNV-based association mapping showed how including different types of polymorphisms improves the coverage of genome-wide genetic variation relevant for phenotypic trait variation. Our findings affirm that a comprehensive analysis of polymorphisms is necessary to unravel the genetic architecture of complex traits.

Our analyses of populations spanning the global distribution range of a crop pathogen allowed us to track trait-associated CNV genes across geographical regions. We found that trait-associated CNVs are in close proximity to recently active TE families and show high degrees of variation in adaptive allele frequencies. This possibly reflects the recency of the onset of selection, the time since the CNV first appeared in the cluster or local adaptation. Understanding the population genetic context of microbial adaptation in agricultural ecosystems at a global scale is a critical foundation to address risks of pathogen spread. Locally adapted pathogen populations may not benefit from gene flow to different regions in the world and, hence, may be more easily contained (Croll & McDonald, 2017). However, homogenization of global agricultural ecosystems at the level of applied fungicides or deployed host genotypes reduces the constraints on locally adapted pathogen populations (Croll & McDonald, 2017). Under such circumstances, local adaptation may favour gene flow to locations sharing similar host or abiotic environments. Furthermore, a mechanistic understanding of microbial adaptation improves our ability to predict the speed at which pathogen populations can cope with environmental challenges. Adaptation mediated by CNVs generated in repetitive regions of the genome can progress particularly fast (Mérot et al., 2020). Deprioritizing easily surmountable abiotic or biotic pressures on pathogen populations can improve the sustainability of pathogen control.

ACKNOWLEDGEMENTS

We thank Nikhil Kumar Singh and Thomas Badet for helpful discussions and comments on previous versions of the manuscript. Open access funding provided by Universite de Neuchatel.

AUTHOR CONTRIBUTIONS

L.S. and D.C. conceived the study and designed analyses, L.S. performed analyses, U.O. annotated and classified TEs, N.M.A. provided samples/data sets, D.C. provided funding, and L.S. and D.C. wrote the manuscript with input from co-authors. All authors reviewed the manuscript and agreed on its submission.

DATA AVAILABILITY STATEMENT

The genome sequencing data for *R. commune* are available at the NCBI Sequence Read Archive under BioProject accession no. PRJNA327656. The RNA sequencing data are available at the NCBI Sequence Read Archive under BioProject accession no. PRJNA804666. The updated genome annotation is available at https://doi.org/10.5281/zenodo.5730007. The transposable element annotation is available at https://doi.org/10.5281/zenodo.6010023.

ORCID

Daniel Croll https://orcid.org/0000-0002-2072-380X

REFERENCES

Abyzov, A., Urban, A. E., Snyder, M., & Gerstein, M. (2011). CNVnator: An approach to discover, genotype, and characterize typical and atypical CNVs from family and population genome sequencing. *Genome Research*, *21*(6), 974–984. https://doi.org/10.1101/gr.114876.110

Altschul, S. F., Madden, T. L., Schäffer, A. A., Zhang, J., Zhang, Z., Miller, W., & Lipman, D. J. (1997). Gapped BLAST and PSI-BLAST: A new generation of protein database search programs. *Nucleic Acids Research*, *25*(17), 3389–3402. https://doi.org/10.1016/B978-1-4832-3211-9.50009-7

Anders, S., Pyl, P. T., & Huber, W. (2015). HTSeq-a python framework to work with high-throughput sequencing data. *Bioinformatics*, *31*(2), 166–169. https://doi.org/10.1093/bioinformatics/btu638

Azevedo, M. M., Faria-Ramos, I., Cruz, L. C., Pina-Vaz, C., & Rodrigues, A. G. (2015). Genesis of azole antifungal resistance from agriculture to clinical settings. *Journal of Agricultural and Food Chemistry*, *63*(34), 7463–7468. https://doi.org/10.1021/acs.jafc.5b02728

Baduel, P., Leduque, B., Ignace, A., Gy, I., Gil, J., Loudet, O., Colot, V., & Quadrana, L. (2021). Genetic and environmental modulation of transposition shapes the evolutionary potential of Arabidopsis thaliana. *Genome Biology*, *22*(1). https://doi.org/10.1186/s13059-021-02348-5

Bao, W., Kojima, K. K., & Kohany, O. (2015). Repbase update, a database of repetitive elements in eukaryotic genomes. *Mobile DNA*, *6*(1), 4–9. https://doi.org/10.1186/s13100-015-0041-9

Beckerson, W. C., Rodríguez, R. C., La Vega, D. E., Hartmann, F. E., Duhamel, M., Giraud, T., & Perlin, M. H. (2019). Cause and effectors: Whole-genome comparisons reveal shared but rapidly evolving effector sets among host-specific plant-castrating fungi. *MBio*, *10*(6), 1–17. https://doi.org/10.1128/mBio.02391-19

Bourque, G., Burns, K. H., Gehring, M., Gorbunova, V., Seluanov, A., Hammell, M., Imbeault, M., Izsvák, Z., Levin, H. L., Macfarlan, T. S., Mager, D. L., & Feschotte, C. (2018). Ten things you should know about transposable elements. *Genome Biology*, *19*(1). https://doi.org/10.1186/s13059-018-1577-z

Bradbury, P. J., Zhang, Z., Kroon, D. E., Casstevens, T. M., Ramdoss, Y., & Buckler, E. S. (2007). TASSEL: Software for association mapping of complex traits in diverse samples. *Bioinformatics*, *23*(19), 2633–2635. https://doi.org/10.1093/bioinformatics/btm308

Breen, J., Wicker, T., Kong, X., Zhang, J., Ma, W., Paux, E., Feuillet, C., Appels, R., & Bellgard, M. (2010). A highly conserved gene island of three genes on chromosome 3B of hexaploid wheat: diverse gene function and genomic structure maintained in a tightly linked block. *BMC Plant Biology*, *10*. https://doi.org/10.1186/1471-2229-10-98

Brown, G. D., Denning, D. W., Gow, N. A. R., Levitz, S. M., Netea, M. G., & White, T. C. (2012). Hidden killers: Human fungal infections. *Science Translational Medicine*, *4*(165). https://doi.org/10.1126/scitranslmed.3004404

Brunner, P. C., Stefansson, T. S., Fountaine, J., Richina, V., & McDonald, B. A. (2016). A global analysis of CYP51 diversity and azole sensitivity in *Rhynchosporium* commune. *Phytopathology*, *106*(4), 355–361. https://doi.org/10.1094/PHYTO-07-15-0158-R

Casacuberta, E., Casacuberta, J. M., Puigdomènech, P., & Monfort, A. (1998). Presence of miniature inverted-repeat transposable elements (MITEs) in the genome of *Arabidopsis thaliana*: Characterisation of the emigrant family of elements. *Plant Journal*, *16*(1), 79–85. https://doi.org/10.1046/j.1365-313X.1998.00267.x

Castanera, R., Vendrell-Mir, P., Bardil, A., Carpentier, M. C., Panaud, O., & Casacuberta, J. (2021). Amplification dynamics of miniature inverted-repeat transposable elements and their impact on rice trait variability. *The Plant Journal*, *107*, 118–135. https://doi.org/10.1111/tpj.15277

Cools, H. J., & Fraaije, B. A. (2013). Update on mechanisms of azole resistance in *Mycosphaerella graminicola* and implications for future control. *Pest Management Science*, *69*(2), 150–155. https://doi.org/10.1002/ps.3348

Coste, A., Selmecki, A., Forche, A., Diogo, D., Bougnoux, M. E., D'Enfert, C., Berman, J., & Sanglard, D. (2007). Genotypic evolution of azole resistance mechanisms in sequential *Candida albicans* isolates. *Eukaryotic Cell*, *6*(10), 1889–1904. https://doi.org/10.1128/EC.00151-07

Croll, D., & McDonald, B. A. (2017). The genetic basis of local adaptation for pathogenic fungi in agricultural ecosystems. *Molecular Ecology*, *26*(7), 2027–2040. https://doi.org/10.1111/mec.13870

Danecek, P., Auton, A., Abecasis, G., Albers, C. A., Banks, E., DePristo, M. A., Handsaker, R. E. Lunter, G., Marth, G. T., Sherry, S. T., McVean, G., & Durbin, R. (2011). The variant call format and VCFtools. *Bioinformatics*, *27*(15), 2156–2158. https://doi.org/10.1093/bioinformatics/btr330

Dubin, M. J., Scheid, O. M., & Becker, C. (2018). Transposons: A blessing curse. *Current Opinion in Plant Biology*, *42*, 23–29. https://doi.org/10.1016/j.pbi.2018.01.003

Emms, D. M., & Kelly, S. (2019). OrthoFinder: Phylogenetic orthology inference for comparative genomics. *Genome Biology*, *20*(238), https://doi.org/10.1101/466201

Endelman, J. B., & Jannink, J. L. (2012). Shrinkage Estimation of the Realized Relationship Matrix. *G3 Genes|Genomes|Genetics*, *2*(11), 1405–1413. https://doi.org/10.1534/g3.112.004259

Esquerré-Tugayé, M. T., Boudart, G., & Dumas, B. (2000). Cell wall degrading enzymes, inhibitory proteins, and oligosaccharides participate in the molecular dialogue between plants and pathogens. *Plant Physiology and Biochemistry*, *38*(1–2), 157–163. https://doi.org/10.1016/S0981-9428(00)00161-3

Feschotte, C. (2008). Transposable elements and the evolution of regulatory networks. *Nature Reviews. Genetics*, *9*(May), 397–405. https://doi.org/10.1038/nrg2337

Fisher, M. C., Hawkins, N. J., Sanglard, D., & Gurr, S. J. (2018). Worldwide emergence of resistance to antifungal drugs challenges human health and food security. *Science*, *360*(6390), 739–742. https://doi.org/10.1126/science.aap7999

Fisher, M. C., Henk, D. A., Briggs, C. J., Brownstein, J. S., Madoff, L. C., McCraw, S. L., & Gurr, S. J. (2012). Emerging fungal threats to animal, plant and ecosystem health. *Nature*, *484*(7393), 186–194. https://doi.org/10.1038/nature10947

Fones, H. N., Bebber, D. P., Chaloner, T. M., Kay, W. T., Steinberg, G., & Gurr, S. J. (2020). Threats to global food security from emerging fungal and oomycete crop pathogens. *Nature Food*, *1*(6), 332–342. https://doi.org/10.1038/s43016-020-0075-0

Food and Agriculture Organization of the United Nations (FAO). Pesticides Use Database. Available from: http://www.fao.org/faostat/ 2021

Fouché, S., Badet, T., Oggenfuss, U., Plissonneau, C., Francisco, C. S., & Croll, D. (2020). Stress-driven transposable element de-repression dynamics and virulence evolution in a fungal pathogen. *Molecular Biology and Evolution*, *37*(1), 221–239. https://doi.org/10.1093/molbev/msz216

Fouché, S., Plissonneau, C., & Croll, D. (2018). The birth and death of effectors in rapidly evolving filamentous pathogen genomes. *Current Opinion in Microbiology*, *46*, 34–42. https://doi.org/10.1016/j.mib.2018.01.020

Frantzeskakis, L., Kracher, B., Kusch, S., Yoshikawa-Maekawa, M., Bauer, S., Pedersen, C., Spanu, P. D., Maekawa, T., Schulze-Lefert, P., & Panstruga, R. (2018). Signatures of host specialization and a recent transposable element burst in the dynamic one-speed genome of the fungal barley powdery mildew pathogen. *BMC Genomics*, *19*(1), 381. https://doi.org/10.1186/s12864-018-4750-6

Gautier, M., Klassmann, A., & Vitalis, R. (2017). Rehh 2.0: A reimplementation of the r package rehh to detect positive selection from

haplotype structure. *Molecular Ecology Resources*, *17*(1), 78–90. https://doi.org/10.1111/1755-0998.12634

Hartmann, F. E., & Croll, D. (2017). Distinct trajectories of massive recent gene gains and losses in populations of a microbial eukaryotic pathogen. *Molecular Biology and Evolution*, *34*(11), 2808–2822. https://doi.org/10.1093/molbev/msx208

Hartmann, F. E., Rodríguez, R. C., La Vega, D. E., Brandenburg, J. T., Carpentier, F., & Giraud, T. (2018). Gene presence-absence polymorphism in castrating anther-smut fungi: recent gene gains and Phylogeographic structure. *Genome Biology and Evolution*, *10*(5), 1298–1314. https://doi.org/10.1093/gbe/evy089

Hastings, P. J., Lupski, J. R., Rosenberg, S. M., & Ira, G. (2009). Mechanisms of change in gene copy number. *Nature Reviews Genetics*, *10*(8), 551–564. https://doi.org/10.1038/nrg2593

Hawkins, J. N., Cools, H. J., Sierotzki, H., Shaw, M. W., Knogge, W., Kelly, S. L., Kelly, D. E., & Fraaije, B. A. (2014). Paralog re-emergence: A novel, historically contingent mechanism in the evolution of antimicrobial resistance. *Molecular Biology and Evolution*, *31*(7), 1793–1802. https://pubmed.ncbi.nlm.nih.gov/24732957/?from_term=-rhynchosporium+commune&from_page=2&from_pos=10

Hoff, K. J., Lange, S., Lomsadze, A., Borodovsky, M., & Stanke, M. (2016). BRAKER1: Unsupervised RNA-seq-based genome annotation with GeneMark-ET and Augustus. *Bioinformatics*, *32*(5), 767–769. https://doi.org/10.1093/bioinformatics/btv661

Jiang, R. H. Y., Tripathy, S., Govers, F., & Tyler, B. M. (2008). RXLR effector reservoir in two Phytophthora species is dominated by a single rapidly evolving superfamily with more than 700 members. *PNAS*, *105*(12), 4874–4879.

Jones, P., Binns, D., Chang, H. Y., Fraser, M., Li, W., McAnulla, C., McWilliam, H., Maslen, J., Mitchell, A., Nuka, G., Pesseat, S., Quinn, A. F., Sangrador-Vegas, A., Scheremetjew, M., Yong, S.-Y., Lopez, R., & Hunter, S. (2014). InterProScan 5: Genome-scale protein function classification. *Bioinformatics*, *30*(9), 1236–1240. https://doi.org/10.1093/bioinformatics/btu031

Kanazawa, A., Liu, B., Kong, F., Arase, S., & Abe, J. (2009). Adaptive evolution involving gene duplication and insertion of a novel Ty1/Copia-like Retrotransposon in soybean. *Journal of Molecular Evolution*, *69*(2), 164–175. https://doi.org/10.1007/s00239-009-9262-1

Kirsten, S., Navarro-Quezada, A., Penselin, D., Wenzel, C., Matern, A., Leitner, A., Baum, T., Seiffert, U., & Knogge, W. (2012). Necrosis-inducing proteins of *Rhynchosporium* commune, effectors in quantitative disease resistance. *Molecular Plant-Microbe Interactions*, *25*(10), 1314–1325. https://doi.org/10.1094/MPMI-03-12-0065-R

Krzywinski, M., Schein, J., Birol, I., Connors, J., Gascoyne, R., Horsman, D., Jones, S. J., & Marra, M. A. (2009). Circos: An information aesthetic for comparative genomics. *Genome Research*, *19*(9), 1639–1645. https://doi.org/10.1101/gr.092759.109

Lamari, L. (2002). Assess: Image analysis software for plant disease quantification.

Langner, T., & Göhre, V. (2016). Fungal chitinases: Function, regulation, and potential roles in plant/pathogen interactions. *Current Genetics*, *62*(2), 243–254. https://doi.org/10.1007/s00294-015-0530-x

Lendenmann, M. H., Croll, D., & McDonald, B. A. (2015). QTL mapping of fungicide sensitivity reveals novel genes and pleiotropy with melanization in the pathogen *Zymoseptoria tritici*. *Fungal Genetics and Biology*, *80*, 53–67. https://doi.org/10.1016/j.fgb.2015.05.001

Leroux, P., & Walker, A. S. (2013). Activity of fungicides and modulators of membrane drug transporters in field strains of *Botrytis cinerea* displaying multidrug resistance. *European Journal of Plant Pathology*, *135*(4), 683–693. https://doi.org/10.1007/s10658-012-0105-3

Lucas, J. A., Hawkins, N. J., & Fraaije, B. A. (2015). *The evolution of fungicide resistance. Advances in applied microbiology*, Vol. 90. Elsevier Ltd. https://doi.org/10.1016/bs.aambs.2014.09.001

Luzia, S., Ursula, O., Norfarhan, M.-A., & Daniel, C.. (2021) Improved genome annotation of *Rhynchosporium commune* isolate UK7 using Illumina short reads of in vitro and in plantae conditions. https://doi.org/10.5281/zenodo.5730007

Luzia, S., Ursula, O., Norfarhan, M.-A., & Daniel, C. (2021) Illumina TruSeq stranded mRNA sequences of *Rhynchosporium commune* isolate UK7 in plantae; https://doi.org/10.5281/zenodo.5729968

Luzia, S., Ursula, O., Norfarhan, M.-A., & Daniel, C. (2021) Illumina TruSeq stranded mRNA sequences of *Rhynchosporium commune* isolate UK7 in vitro; https://doi.org/10.5281/zenodo.5729863

Luzia, S., Ursula, O., Norfarhan, M.-A., & Daniel, C. (2021). Transposable element annotation *Rhynchosporium commune* isolate UK7. Zenodo. https://doi.org/10.5281/zenodo.6010024

Ma, Z., Proffer, T. J., Jacobs, J. L., & Sundin, G. W. (2006). Overexpression of the 14α-demethylase target gene (CYP51) mediates fungicide resistance in *Blumeriella jaapii*. *Applied and Environmental Microbiology*, *72*(4), 2581–2585. https://doi.org/10.1128/AEM.72.4.2581-2585.2006

McKenna, A., Hanna, M., Banks, E., Sivachenko, A., Cibulskis, K., Kernytsky, A., Garimella, K., Altshuler, D., Gabriel, S., Daly, M., & DePristo, M. A. (2010). The genome analysis toolkit: A MapReduce framework for analyzing next-generation DNA sequencing data. *Genome Research*, *20*, 1297–1303. https://doi.org/10.1101/gr.107524.110.20

McLysaght, A., & Guerzoni, D.. (2015). New genes from non-coding sequence: the role of de novo protein-coding genes in eukaryotic evolutionary innovation. *Philosophical Transactions of the Royal Society B: Biological Sciences*, *370*(1678), 20140332. https://doi.org/10.1098/rstb.2014.0332

Mérot, C., Oomen, R. A., Tigano, A., & Wellenreuther, M. (2020). A roadmap for understanding the evolutionary significance of structural genomic variation. *Trends in Ecology and Evolution*, *35*(7), 561–572. https://doi.org/10.1016/j.tree.2020.03.002

Mishra, S., & Whetstine, J. R. (2016). Different facets of copy number changes: permanent, transient, and adaptive. *Molecular and Cellular Biology*, *36*(7), 1050–1063. https://doi.org/10.1128/mcb.00652-15

Mohd-Assaad, N., McDonald, B. A., & Croll, D. (2016). Multilocus resistance evolution to azole fungicides in fungal plant pathogen populations. *Molecular Ecology*, *25*(24), 6124–6142. https://doi.org/10.1111/mec.13916

Mohd-Assaad, N., Mcdonald, B. A., & Croll, D. (2018). Genome-wide detection of genes under positive selection in worldwide populations of the barley scald pathogen. *Genome Biol. Evol*, *5*, 1315–1332. https://doi.org/10.1093/gbe/evy087

Mohd-Assaad, N., McDonald, B. A., & Croll, D. (2019). The emergence of the multi-species NIP1 effector in Rhynchosporium was accompanied by high rates of gene duplications and losses. *Environmental Microbiology*, *21*(8), 2677–2695. https://doi.org/10.1111/1462-2920.14583

Morris, J. J., Lenski, R. E., & Zinser, E. R. (2012). The black queen hypothesis: evolution of dependencies through adaptive gene loss. *MBio*, *3*(2), https://doi.org/10.1128/mBio.00036-12

Morschhäuser, J. (2016). The development of fluconazole resistance in *Candida albicans* – an example of microevolution of a fungal pathogen. *Journal of Microbiology*, *54*(3), 192–201. https://doi.org/10.1007/s12275-016-5628-4

Norfarhan, M.-A., & Daniel, C. (2016) Population genomics of *Rhynchosporium commune*; NCBI Sequence Read Archive; PRJNA327656.

Oggenfuss, U., Badet, T., Wicker, T., Hartmann, F. E., Singh, N. K., Abraham, L., Karisto, P., Vonlanthen, T., Mundt, C., McDonald, B. A., & Croll, D. (2021). A population-level invasion by transposable elements triggers genome expansion in a fungal pathogen. *eLife*, *10*, e69249. https://doi.org/10.7554/eLife.69249

Oki, N., Yano, K., Okumoto, Y., Tsukiyama, T., Teraishi, M., & Tanisaka, T. (2008). A genome-wide view of miniature inverted-repeat transposable elements (MITEs) in rice, *Oryza sativa* Ssp. Japonica. *Genes and Genetic Systems*, *83*(4), 321–329. https://doi.org/10.1266/ggs.83.321

Olson, M. V. (1999). When less is more: gene loss as an engine of evolutionary change. *American Journal of Human Genetics*, *64*, 18–23. https://doi.org/10.1086/302219

Penselin, D., Münsterkötter, M., Kirsten, S., Felder, M., Taudien, S., Platzer, M., Ashelford, K., Paskiewicz, K. H., Harrison, R. J., Hughes, D. J., Wolf, T., Shelest, E., Graap, J., Hoffmann, J., Wenzel, C., Wöltje, N., King, K. M., Fitt, B. D. L., Güldener, U., … Knogge, W. (2016). Comparative genomics to explore phylogenetic relationship, cryptic sexual potential and host specificity of Rhynchosporium species on grasses. *BMC Genomics*, *17*(1). https://doi.org/10.1186/s12864-016-3299-5

Peter, J., De Chiara, M., Friedrich, A., Yue, J.-X., Pflieger, D., Bergström, A., Sigwalt, A., Barre, B., Freel, K., Llored, A., Cruaud, C., Labadie, K., Aury, J.-M., Istace, B., Lebrigand, K., Barbry, P., Engelen, S., Lemainque, A., Wincker, P., … Schacherer, J. (2018). Genome evolution across 1,011 *Saccharomyces cerevisiae* isolates. *Nature*, *556*(7701), 339–344. https://doi.org/10.1038/s41586-018-0030-5

Petersen, T. N., Brunak, S., Von Heijne, G., & Nielsen, H. (2011). SignalP 4.0: Discriminating signal peptides from transmembrane regions. *Nature Methods*, *8*(10), 785–786. https://doi.org/10.1038/nmeth.1701

Pritchard, J. K., Stephens, M., & Donnelly, P. (2008). Inference of population structure using multilocus genotype Data. *Genetics*, *155*, 945–959. https://doi.org/10.1007/s10681-008-9788-0

Purcell, S., Neale, B., Todd-Brown, K., Thomas, L., Ferreira, M. A. R., Bender, D., Maller, J., Sklar, P., de Bakker, P. I. W., Daly, M. J., & Sham, P. C. (2007). PLINK: A tool set for whole-genome association and population-based linkage analyses. *American Journal of Human Genetics*, *81*(3), 559–575. https://doi.org/10.1086/519795

Quinlan, A. R., & Hall, I. M. (2010). BEDTools: A flexible suite of utilities for comparing genomic features. *Bioinformatics*, *26*(6), 841–842. https://doi.org/10.1093/bioinformatics/btq033

Rice, P., Longden, L., & Bleasby, A. (2000). EMBOSS: The European molecular biology open software suite. *Trends in Genetics*, *16*(6), 276–277. https://doi.org/10.1016/S0168-9525(00)02024-2

Ritz, C., Baty, F., Streibig, J. C., & Gerhard, D. (2015). Dose-response analysis using R. *PLoS One*, *10*(12), e0146021. https://doi.org/10.1371/journal.pone.0146021

Robinson, M. D., McCarthy, D. J., & Smyth, G. K. (2009). EdgeR: A bioconductor package for differential expression analysis of digital gene expression data. *Bioinformatics*, *26*(1), 139–140. https://doi.org/10.1093/bioinformatics/btp616

Rohe, M., Gierlich, A., Hermann, H., Hahn, M., Schmidt, B., Rosahl, S., & Knogge, W. (1995). The race-specific elicitor, NIP1, from the barley pathogen, *Rhynchosporium secalis*, determines avirulence on host plants of the Rrs1 resistance genotype. *EMBO Journal*, *14*(17), 4168–4177. https://doi.org/10.1002/j.1460-2075.1995.tb00090.x

Rouxel, T., & Balesdent, M. H. (2017). Life, death and rebirth of avirulence effectors in a fungal pathogen of brassica crops, *Leptosphaeria maculans*. *New Phytologist*, *214*(2), 526–532. https://doi.org/10.1111/nph.14411

Sabeti, P. C., Reich, D. E., Higgins, J. M., Levine, H. Z. P., Richter, D. J., Schaffner, S. F., Gabriel, S. B., Platko, J. V., Patterson, N. J., McDonald, G. J., Ackerman, H. C., Campbell, S. J., Altshuler, D., Cooper, R., Kwiatkowski, D., Ward, R., & Lander, E. S. (2002). Detecting recent positive selection in the human genome from haplotype structure. *Nature*, *419*, 832–837. https://doi.org/10.1038/nature01140

Sánchez-Torres, P. (2021). Molecular mechanisms underlying fungicide resistance in citrus postharvest green mold. *Journal of Fungi*, *7*(9), https://doi.org/10.3390/jof7090783

Sánchez-Vallet, A., Fouché, S., Fudal, I., Hartmann, F. E., Soyer, J. L., Tellier, A., & Croll, D. (2018). The genome biology of effector gene evolution in filamentous plant pathogens andrea. *Annual Review of Phytopathology*, *56*, 12–52. https://doi.org/10.1146/annurev-phyto-080516-

Santiago, N., Cristina Herráiz, J., Goñi, R., Messeguer, X., & Casacuberta, J. M. (2002). Genome-wide analysis of the emigrant family of MITEs of *Arabidopsis thaliana*. *Molecular Biology and Evolution*, *19*(12), 2285–2293. https://doi.org/10.1093/oxfordjournals.molbev.a004052

Schmidt, S. M., Houterman, P. M., Schreiver, I., Ma, L., Amyotte, S., Chellappan, B., Boeren, S., Takken, F. L. W., & Rep, M. (2013). MITEs in the promoters of effector genes allow prediction of novel virulence genes in *Fusarium oxysporum*. *BMC Genomics*, *14*(1), 119. https://doi.org/10.1186/1471-2164-14-119

Schneider, C. A., Rasband, W. S., & Eliceiri, K. W. (2012). NIH image to ImageJ: 25 years of image analysis. *Nature Methods*, *9*(7), 671–675. https://doi.org/10.1007/978-1-84882-087-6_9

Schürch, S., Linde, C. C., Knogge, W., Jackson, L. F., & McDonald, B. A. (2004). Molecular population genetic analysis differentiates two virulence mechanisms of the fungal avirulence gene NIP1. *Molecular Plant-Microbe Interactions*, *17*(10), 1114–1125. https://doi.org/10.1094/MPMI.2004.17.10.1114

Selmecki, A., Forche, A., & Berman, J. (2006). Aneuploidy and isochromosome formation in drug-resistant *Candida albicans*. *Science*, *313*(5785), 367–370.

Sionov, E., Lee, H., Chang, Y. C., & Kwon-Chung, K. J. (2010). *Cryptococcus neoformans* overcomes stress of azole drugs by formation of disomy in specific multiple chromosomes. *PLoS Path*, *6*(4), e1000848. https://doi.org/10.1371/journal.ppat.1000848

Smit, A. F. A., Hubley, R., & Green, P. (2015). RepeatMasker Open-4.0. Available from: http://repeatmasker.org

Stanke, M., Keller, O., Gunduz, I., Hayes, A., Waack, S., & Morgenstern, B. (2006). AUGUSTUS: A b initio prediction of alternative transcripts. *Nucleic Acids Research*, *34*(WEB. SERV. ISS.), 435–439. https://doi.org/10.1093/nar/gkl200

Steenwyk, J. L., & Rokas, A. (2018). Copy number variation in fungi and its implications for wine yeast genetic diversity and adaptation. *Frontiers in Microbiology*, *9*(288). https://doi.org/10.3389/fmicb.2018.00288

Stefansson, T. S., Mcdonald, B. A., & Willi, Y. (2013). Local adaptation and evolutionary potential along a temperature gradient in the fungal pathogen *Rhynchosporium commune*. *Evolutionary Applications*, *6*(3), 524–534. https://doi.org/10.1111/eva.12039

Stefansson, T. S., McDonald, B. A., & Willi, Y. (2014). The Influence of genetic drift and selection on quantitative traits in a plant pathogenic fungus. *PLoS One*, *9*(11), e112523. https://doi.org/10.1371/journal.pone.0112523

Stefansson, T. S., Willi, Y., Croll, D., & McDonald, B. A. (2014). An assay for quantitative virulence in *Rhynchosporium commune* reveals an association between effector genotype and virulence. *Plant Pathology*, *63*(2), 405–414. https://doi.org/10.1111/ppa.12111

Tam, V., Patel, N., Turcotte, M., Bossé, Y., Paré, G., & Meyre, D. (2019). Benefits and limitations of genome-wide association studies. *Nature Reviews Genetics*, *20*(8), 467–484. https://doi.org/10.1038/s41576-019-0127-1

Tang, Y. C., & Amon, A. (2013). Gene copy-number alterations: A cost-benefit analysis. *Cell*, *152*(3), 394–405. https://doi.org/10.1016/j.cell.2012.11.043

Thompson, J. D., Higgins, D. G., & Gibson, T. J. (1994). CLUSTAL W: Improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice. *Nucleic Acids Research*, *22*(22), 4673–4680. https://doi.org/10.1093/nar/22.22.4673

Todd, R. T., & Selmecki, A. (2020). Expandable and reversible copy number amplification drives rapid adaptation to antifungal drugs. *eLife*, *9*, 1–33. https://doi.org/10.7554/eLife.58349

Trapnell, C., Pachter, L., & Salzberg, S. L. (2009). TopHat: Discovering splice junctions with RNA-seq. *Bioinformatics*, *25*(9), 1105–1111. https://doi.org/10.1093/bioinformatics/btp120

Urban, M., Cuzick, A., Seager, J., Wood, V., Rutherford, K., Venkatesh, S. Y., De Silva, N., Martinez, M. C., Pedro, H., Yates, A. D., Hassani-Pak,

K., & Hammond-Kosack, K. E. (2020). PHI-base: The pathogen-host interactions database. *Nucleic Acids Research*, *48*, 613–620. https://doi.org/10.1093/nar/gkz904

Vega, B., Liberti, D., Harmon, P. F., & Dewdney, M. M. (2012). A Rapid resazurin-based microtiter assay to evaluate QoI sensitivity for *Alternaria alternata* isolates and their molecular characterization. *Plant Disease*, *96*(9), 1262–1270. https://doi.org/10.1094/PDIS-12-11-1037-RE

Voight, B. F., Kudaravalli, S., Wen, X., & Pritchard, J. K. (2006). A Map of recent positive selection in the human genome. *PLoS Biology*, *4*(3), 446–458. https://doi.org/10.1371/journal.pbio.0040072

Wellenreuther, M., Mérot, C., Berdan, E., & Bernatchez, L. (2019). Going beyond SNPs: The Role of Structural Genomic Variants in Adaptive Evolution and Species Diversification. *Molecular Ecology*, *28*(6), 1203–1209. https://doi.org/10.1111/mec.15066

Wells, J. N., & Feschotte, C. (2020). A field guide to eukaryotic transposable elements. *Annual Review of Genetics*, *54*, 539–561. https://doi.org/10.1146/annurev-genet-040620-022145

White, S. E., Habera, L. F., & Susan, R. (1994). *Retrotransposons* in the flanking regions of normal plant Genes: A role for copia-like elements in the evolution of gene structure and expression. *Proceedings of the National Academy of Sciences of the United States of America*, *91*, 11792–11796. https://doi.org/10.1073/pnas.91.25.11792

Wicker, T., Sabot, F., Hua-Van, A., Bennetzen, J. L., Capy, P., Chalhoub, B., Flavell, A., Leroy, P., Morgante, M., Panaud, O., Paux, E., SanMiguel, P., & Schulman, A. H. (2007). A unified classification system for eukaryotic transposable elements. *Nature Genetics*, *8*, 973–982. https://doi.org/10.2118/36071-JPT

Yin, Y., Mao, X., Yang, J., Chen, X., Mao, F., & Ying, X. U. (2012). DbCAN: A web resource for automated carbohydrate-active enzyme annotation. *Nucleic Acids Research*, *40*(W1), 445–451. https://doi.org/10.1093/nar/gks479

Zhang, J., Li, L., Lv, Q., Yan, L., Wang, Y., & Jiang, Y. (2019). The fungal CYP51s: Their functions, structures, related drug resistance, and inhibitors. *Frontiers in Microbiology*, *10*(April). https://doi.org/10.3389/fmicb.2019.00691

SUPPORTING INFORMATION

Additional supporting information may be found in the online version of the article at the publisher's website.

How to cite this article: Stalder, L., Oggenfuss, U., Mohd-Assaad, N., & Croll, D. (2023). The population genetics of adaptation through copy number variation in a fungal plant pathogen. *Molecular Ecology*, *32*, 2443–2460. https://doi.org/10.1111/mec.16435