environmental microbiology

Environmental Microbiology (2021) **00**(00), 00–00 doi:10.1111/1462-2920.15605

Climate-specific biosynthetic gene clusters in populations of a lichen-forming fungus

Garima Singh,[1,2] Anjuli Calchera,[1] Meike Schulz,[1] Moritz Drechsler,[3,4] Helge B. Bode,[1,2,3,4,5] Imke Schmitt[1,2,6] and Francesco Dal Grande[1,2*]

[1]*Senckenberg Biodiversity and Climate Research Centre (SBiK-F), Frankfurt, 60325, Germany.*

[2]*LOEWE Center for Translational Biodiversity Genomics (TBG), Frankfurt, 60325, Germany.*

[3]*Molekulare Biotechnologie, Fachbereich Biowissenschaften, Goethe Universität Frankfurt, Frankfurt, 60438, Germany.*

[4]*Department Natural Products in Organismic Interactions, Max-Planck-Institute for Terrestrial Microbiology, Marburg, 35043, Germany.*

[5]*Buchmann Institute for Molecular Life Sciences, Goethe Universität Frankfurt, Frankfurt, 60438, Germany.*

[6]*Institute of Ecology, Evolution and Diversity, Fachbereich Biowissenschaften, Goethe Universität Frankfurt, Frankfurt, 60438, Germany.*

Summary

Natural products can contribute to abiotic stress tolerance in plants and fungi. We hypothesize that biosynthetic gene clusters (BGCs), the genomic elements that underlie natural product biosynthesis, display structured differences along elevation gradients. We analysed biosynthetic gene variation in natural populations of the lichen-forming fungus *Umbilicaria pustulata*. We collected a total of 600 individuals from the Mediterranean and cold-temperate climates. Population genomic analyses indicate that *U. pustulata* contains three clusters that are highly differentiated between the Mediterranean and cold-temperate populations. One entire cluster is exclusively present in cold-temperate populations, and a second cluster is putatively dysfunctional in all cold-temperate populations. In the third cluster variation is fixed in all cold-temperate populations due to hitchhiking. In these two clusters the presence of consistent allele frequency differences among replicate populations/gradients suggests that selection rather than drift is driving the pattern. We advocate that the landscape of fungal biosynthetic genes is shaped by both positive and hitchhiking selection. We demonstrate, for the first time, the presence of climate-associated BGCs and BGC variations in lichen-forming fungi. While the associated secondary metabolites of the candidate clusters are presently unknown, our study paves the way for targeted discovery of natural products with ecological significance.

Introduction

Natural products (secondary metabolites) are known to play an important role in environmental adaptation and abiotic stress tolerance in vascular plants (Ramakrishna and Ravishankar, 2011) and fungi (Medina *et al*., 2015). For instance, particular flavonoids, phenolic acids and alkaloids are linked to drought conditions and protection against excessive light, while certain terpenes may increase in response to high temperature (Yang *et al*., 2018). Xerophilic fungi may respond to abiotic stress with toxin production (Medina *et al*., 2015), while endophytic fungi and bacteria may synthesize bioactive secondary metabolites that influence the growth and stress tolerance of their host plants (e.g., Camarena-Pozos *et al*., 2019). Lichens, symbiotic associations consisting of fungal and photosynthetic partners, are particularly rich in natural products (Huneck, 1999). As long-lived, sessile organisms, often growing in exposed habitats, lichens must feature adaptations to abiotic stressors including heat, cold, desiccation, UV radiation, or oversaturation with water. Some of the environmental adaptations in lichens are conferred by secondary metabolites. For example, biosynthesis of the orange pigment parietin has been linked to UV radiation (Solhaug and Gauslaa, 1996; Gauslaa and Ustvedt, 2003), and water repellent depsides and depsidones deposited on internal fungal hyphae have been suggested to facilitate gas exchange when the lichen thallus is saturated with water (Huneck, 2003). However, the function of the majority of lichen compounds, as well as the molecular basis of

Received 7 January, 2021; revised 17 May, 2021; accepted 18 May, 2021. *For correspondence. E-mail francesco.dalgrande@senckenberg.de; Tel. +49 (0)69 7542 1856.

natural product biosynthesis in relation to climatic factors, is poorly understood.

The genomes of lichenized fungi contain large numbers of biosynthetic genes (Calchera *et al*., 2019). It would require functional evaluation of each cluster to detect those clusters involved in environmental stress tolerance. A shortcut to finding natural products putatively important in environmental adaptation is by identifying climate-specific differences in biosynthetic genes and their allele frequencies in natural populations. Such assessments of biosynthetic gene variation at the intraspecific level can help us to identify clusters responding to abiotic selection pressures, and to better understand the genetic underpinnings of climate adaptation mediated by secondary metabolism in fungi.

In fungi, genes associated with natural product biosynthesis are organized in contiguous genomic regions called biosynthetic gene clusters (BGCs) (Keller *et al*., 2005; Calcott *et al*., 2018). The clusters include genes encoding backbone enzymes (that synthesize the core structure of the compound), tailoring enzymes (that modify the core structure) and additional proteins, such as transcription factors and transporters. The most common backbone enzymes in lichenized fungi are polyketide synthases (PKSs) and non-ribosomal peptide synthetases (NRPSs) (Keller *et al*., 2005). Conserved sequence motifs in the core genes facilitate efficient genome mining of fungal BGCs. Recent advances in genomic techniques have facilitated the characterization of BGCs from filamentous fungi including lichen-forming fungi (e.g. Dal Grande *et al*., 2018a; Calchera *et al*., 2019; Pizarro *et al*., 2020). These studies show that the number of BGCs far exceeds the number of detectable compounds indicating that the biosynthetic potential of lichenized fungi is vastly underestimated. Furthermore, although the experimental evidence suggests that the biosynthetic potential of species varies under different environmental conditions, we do not know how climate shapes biosynthetic gene variations in natural populations. Comparing BGCs across populations of a lichen inhabiting different environmental conditions will facilitate the identification of candidate clusters of potential ecological relevance.

In this study we explore the genomic landscape of BGCs in natural populations of a lichen-forming fungus in the Mediterranean and cold-temperate climate zones along replicated elevational gradients. *Umbilicaria* ('*Lasallia*') (Davydov *et al*., 2017) *pustulata* is a lichen-forming fungus with a Europe-wide distribution. A previous study by Dal Grande *et al*. (2017), based on a detailed sampling of populations along a mountain gradient, analysed whole-genome differentiation and demographic processes of *U. pustulata* at a regional scale and showed that there is high genome-wide differentiation and limited gene flow between cold-temperate and Mediterranean populations in the fungal host of the symbiosis. Based on the strong genome-wide differentiation and ecophysiological evidence, the authors showed that the two population clusters represent distinct ecotypes differentially adapted to the contrasting environmental conditions of the two climatic zones (Dal Grande *et al*., 2017). The species' broad geographic distribution and local abundance allow sampling of replicated populations of both ecotypes from independent elevational transects (Dal Grande *et al*., 2017; Rolshausen *et al*., 2018). Analysing replicate gradients helps eliminating potential confounding effects of local population structure when identifying potential targets of selection.

Specifically, we address the following research questions: (i) Are there BGCs in *U. pustulata*, which occur exclusively in either the Mediterranean or cold-temperate climate? (ii) Are there climate-specific patterns of biosynthetic gene variation? (iii) Can differences in biosynthetic genes in the Mediterranean and cold-temperate populations be linked to measurable chemical differences of the lichen thalli?

Materials and methods

Sampling

Umbilicaria pustulata is a foliose, rock-inhabiting lichen, attached to the substrate with a central holdfast. We collected populations from the Mediterranean and cold-temperate climate zones from three elevational gradients (Dal Grande *et al.*, 2017; Dal Grande *et al.*, 2018). *Umbilicaria pustulata* is a predominantly asexual species, only rarely forming apothecia. Two gradients are located in Spain (Puerto de Pico - ESi and Sierra de Gredos - ESii, Sistema Central) and one in Italy (Mount Limbara, Sardinia- IT) (Dal Grande *et al*., 2017; Dal Grande *et al*., 2018b). - Per population, one hundred individuals were collected, pooled and sequenced on Illumina HiSeq (Pool-seq) (Dal Grande *et al.*, 2017). In total we analysed six populations (six libraries), including three Mediterranean and three cold-temperate ones (Fig. 1A).

Ecotypic characterization of the populations

To group the Pool-seq datasets into the Mediterranean and cold-temperate *U. pustulata* ecotypes *sensu* Dal Grande *et al*. (2017), we visualized the positional relations among populations using a three-way generalization of classical multidimensional scaling (DISTATIS) (Fig. 1B). We first calculated pairwise F_{ST} values using the script fst-sliding.pl in PoPoolation2 (Kofler *et al*., 2011b). For this, we retained SNPs with a minimum coverage of 30, a maximum coverage below the 2% of the empirical coverage distribution of each pool and a minimum count of two reads supporting the minor allele

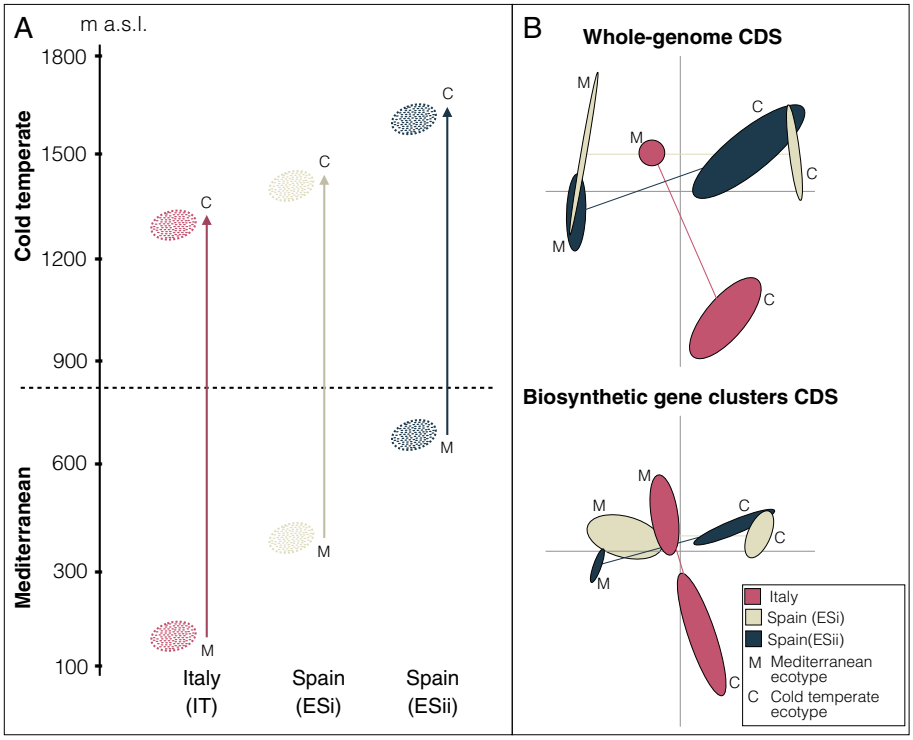

Fig. 1. A. Schematic representation of the three elevational gradients; (B) compromise distance space among populations based on F_{ST} quantile distance matrix for (i) all, genome-wide polymorphic coding SNPs (top), (ii) coding SNPs in BGCs (bottom), with 95% tolerance ellipses. 'M' and 'C' correspond to the Mediterranean and cold-temperate *U. pustulata* ecotypes *sensu* Dal Grande *et al*. (2017) respectively. [Color figure can be viewed at wileyonlinelibrary.com]

across populations. These parameters were chosen to keep variants supported by a large number of reads, to exclude variants likely resulting from sequencing errors and to exclude spurious SNPs likely falling within copy number variations or repetitive regions characterized by extremely high coverage respectively. We then obtained a reduced set of pairwise F_{ST} distance matrices based on sample quantiles (0.975, 0.75, 0.5, 0.25, 0.025) on two datasets: the full set of distances across all, genome-wide polymorphic coding SNPs ($N = 429{,}876$) and (ii) the set of distances across all polymorphic coding SNPs present in BGC regions ($N = 11{,}008$). Finally, we obtained 95% confidence intervals around each population's position in the resulting compromise distance space using bootstrap resampling ($N = 1000$). Populations grouping with 'Italy - Mediterranean population' represent the warm-adapted fungal ecotype, and those grouping with 'Italy – cold-temperate population' the cold-adapted ecotype *sensu* Dal Grande *et al*. (2017) (Fig. 1B). All analyses were performed in R (R Core Team 2019) as described in Dal Grande *et al*. (2017).

Identification and annotations of biosynthetic gene clusters

As reference genomes for the Mediterranean- and cold-temperate ecotypes of *U. pustulata* we used two published *U. pustulata* genomes (ENA accession no. GCA_000938525 for the Mediterranean ecotype (Dal Grande *et al*., 2017) and GenBank accession no. VXIT00000000 (Greshake-Tzovaras *et al*., 2020) for the cold-temperate ecotype). The latter genome was derived from a PacBio metagenome assembly and was used as a reference in all subsequent Pool-seq analyses due to its higher contiguity. BGCs were annotated in both reference genomes using antiSMASH fungal version (Table 1) (antibiotics & SM Analysis Shell, v5.0) (Medema *et al*., 2011; Blin *et al*., 2019).

Homologous BGCs between warm and cold ecotype

Homologous clusters were identified by performing reciprocal BLAST between the BGCs inferred from the reference genomes of the Mediterranean and cold-temperate *U. pustulata*. Homology between candidate clusters was visualized via synteny plots using Easyfig v2.2.3 (Sullivan *et al*., 2011). The GBK input files for Easyfig were generated with seqkit v0.10.1 (Shen *et al*., 2016) and the seqret tool from EMBOSS v6.6.0.0 (Rice *et al*., 2000). Easyfig was run with tblastx v2.6.0+, a minimum identity value of 90 and a minimum length of 50 to draw the blast hits (Kjærbølling *et al*., 2018). Clusters were manually matched for direction and organized to have the core gene with the same orientation.

Verification of presence/absence patterns

For the clusters which were predicted only in one reference genome (six clusters, Table 2), we verified the presence/absence pattern with the help of the Pool-seq reads in two ways: (i) we mapped the raw, quality filtered Pool-seq reads of the six populations along the Italian elevational gradient (ENA accession nos. ERR1110280 to ERR1110285) against the core gene of the candidate cluster (Table 2), (ii) for the clusters present only in one climate group (Mediterranean or cold temperate) in the above-mentioned analysis (i.e. in the Pool-seq data of the Italian elevation gradient) (cluster 22 in Table 2, we further verified the presence in only one climate group by blasting (blastN) its core PKS gene against the mycobiont genomes reconstructed from the Pool-seq data of the Mediterranean and cold-temperate populations of *U. pustulata* from the three elevation gradients (Table 3; Pool-seq assemblies are available at: https://doi.org/10.6084/m9.figshare.14611191.v2.). Details of the approach are described in Supporting Information S1.

To further evaluate the presence/absence pattern of candidate genes we investigated their presence in four additional Pool-seq datasets representing mid-elevation populations (populations 2–5) from Dal Grande *et al*. (2017) (Table 4). These were directly linked to *U. pustulata* ecotypes which were identified via ecophysiological analyses. Also for these additional populations we studied raw reads alignments and presence/absence patterns in reconstructed mycobiont genomes (Supporting Information 7B) as described above. Pool-seq assemblies are available at: https://doi.org/10.6084/m9.figshare.14611191.v2.

Table 1. Summary of biosynthetic gene clusters predicted by antiSMASH in the reference genomes of Mediterranean and cold-temperate ecotypes of *U. pustulata*.

Category	Mediterranean ecotype	Cold-temperate ecotype
#clusters (in reference genome)	26	28
Type I R-PKS	10	10
Type I NR-PKS	5	5
Type III PKS	1	1
Hybrid PKS-NRPS	2	2
NRPS-like	2	4
Terpene synthases	6	6
Homologous clusters	24	
Unique clusters	2	4
Unique clusters confirmed by Pool-seq data	None	1

PKS, Polyketide Synthase; R-PKS, reducing PKS; NR-PKS, non-reducing PKS.

SNP analysis

We used Pool-seq data of the six populations (100 individuals/pool) to identify significantly differentiated coding SNPs in BGCs between the Mediterranean and cold-temperate ecotypes. SNPs were identified using the PoPoolation2 pipeline (Kofler *et al*., 2011b). Considering the comparatively high cost of resequencing individual samples for populations genomic analyses in order to obtain a reliable population-wide allele frequency estimation, the Pool-seq method, i.e. sequencing pooled DNA samples from populations, represents a cost-effective means to infer genome-wide patterns of polymorphism from natural populations (Kofler *et al*., 2011b; Raineri *et al*., 2012; Guirao-Rico and González, 2021).

Details of the approach are described in Supporting Information S1.

Table 2. Verification of ecotype-specific BGCs present in the reference genome using Pool-seq data. Summary of reciprocal BLAST hits of six BGCs which were detected by antiSMASH in the reference genome of only one ecotype. Pool 1 to pool 4 represent the Mediterranean ecotype, pool 5 is an admixed population and contains individuals of both ecotypes and pool 6 represents the cold-temperate ecotype. The only BGC fixed in cold-temperate populations and absent from Mediterranean populations, cluster 22, is highlighted in bold.

Cluster number	Reference used	BGC core gene	Length of core gene	Number of hits					
				Pool 1	Pool 2	Pool 3	Pool 4	Pool 5	Pool 6
Cluster 4	Cold-temperate ecotype	Type I R-PKS	6354	1266	1198	1046	1546	4058	4644
Cluster 9	Cold-temperate ecotype	NRPS-like	4089	4302	3706	3342	4448	4514	4696
Cluster 12	Cold-temperate ecotype	NRPS-like	2409	2885	2554	2393	3470	3131	3032
Cluster 22	**Cold-temperate ecotype**	**Type I NR-PKS**	**8100**	**82**	**74**	**125**	**192**	**5576**	**7498**
Cluster 29	Mediterranean ecotype	Type I R-PKS	6597	6828	6344	5224	7969	7363	8020
Cluster 30	Mediterranean ecotype	Type I R-PKS	5073	5284	4448	4061	4422	2418	545

Table 3. Summary of reciprocal BLAST hits of cluster 22 against the Pool-seq raw reads of Mediterranean and cold-temperate populations of *U. pustulata* from the three elevation gradients.

	Mediterranean ecotype			Cold-temperate ecotype		
	IT1	ESii1	ESi1	IT6	ESii6	ESi3
Accession of Pool-seq data	ERR1110280 (ENA accession)	SRR14573781 (SRA accession)	SRR14573779 (SRA accession)	ERR1110285 (ENA accession)	SRR14573780 (SRA accession)	SRR14573778 (SRA accession)
Total number of reads in the pool	29,162,770	26,758,269	28,862,057	32,064,853	26,690,508	35,351,050
# Paired-end reads correctly mapped	1	49	266	3,717	4,235	5990
% Mapped reads	0.00%	0.02%	0.09%	1.16%	1.59%	1.69%
# Bits score Blastn metagenome vs. genes	No hits	2244	7925	13,771	13,771	13,771
# Coverage scaffold with highest Bits score	–	1	1.5	21	69.3	35.1

Table 4. Abundance of the cluster 22 (cluster fixed in cold-temperate populations and absent from Mediterranean populations), in multiple populations along an elevation gradient in Italy encompassing the Mediterranean to cold-temperate climate (Dal Grande *et al*., 2017).

	IT1	IT2	IT3	IT4	IT5	IT6
# Bits score Blastn metagenome vs. genes	No hits	No hits	580	822	13,771	13,771
# coverage scaffold with highest Bits score	–	–	0.6	0.6	15.03	21
ENA accession number of Pool-seq raw data	ERR1110280	ERR1110281	ERR1110282	ERR1110283	ERR1110284	ERR1110285

IT1 to IT4 represents Mediterranean populations, IT5 an admixed population and IT6 a cold-temperate population.

Identifying consistent allele frequency differences in populations: Quasibinomial GLMs

The allele frequency differences between populations can be used to infer the action of selection. However, alleles fixed in a population could also result from genetic drift. To differentiate between allele frequency differences resulting from selection and genetic drift, we used Generalized Linear Models (GLMs) with a quasibinomial error structure, as implemented in the program *poolFreqDiff* (-mincnt 6 -minc 30 -maxc 600 (Wiberg *et al*., 2017). This method rescales all allele counts to the effective sample size n_{eff} (Wiberg *et al*., 2017), thus distinguishing between heterogeneity in allele frequency differences (e.g., arising from genetic drift) among replicates and main effects. This allows for a more reliable identification of true positives (i.e., loci under selection) while keeping the false positives (i.e., variants with idiosyncratic responses or resulting from population drift) to a minimum. Only SNPs passing a Bonferroni correction at $\alpha = 0.0007$ were considered significantly differentiated between the Mediterranean and cold-temperate populations as they showed consistent variation across replicate populations (Wiberg *et al*., 2017).

Identifying functionally deleterious SNPs: PROVEAN

Significantly differentiated, non-synonymous SNPs were subsequently identified using an in-house python script (Supporting Information S2). For this, we first reconstructed CDS sequences of each BGC gene for each population by inserting the most common allele found in the population at each variable position using a custom perl script (Supporting Information S3).

We then used PROVEAN to distinguish non-synonymous SNPs which were functionally deleterious from functionally/structurally neutral SNPs (Choi and Chan, 2015). Details of the approach are described in the Supporting Information S1.

Filtering ecologically relevant SNPs: frequencies of SNPs

For each SNP varying consistently across replicate gradients and having a low PROVEAN score, we calculated the frequency at the population level, i.e., the combined frequency of that SNP in three Mediterranean and three cold-temperate populations. We did this by calculating the percentage of occurrence of an allele out of the total number of occurrences.

To finally select the clusters which are significantly differentiated between the Mediterranean and cold-temperate populations, the following criteria were applied: (i) low PROVEAN score (≤−4.0) of the changed amino acid (PROVEAN score lower than −4.0 suggests highly significant functionally deleterious variation, whereas a score ≥−4.0 is considered neutral variation), (ii) >90% frequency of a SNP in all three populations of an elevational group (Mediterranean or cold temperate) and low frequency (<5%) in all three populations of the other group. This was done to exclude locally differentiated SNPs, i.e., significantly differentiated SNPs not consistent across all three replicate gradients.

Testing the possibility of selective sweeps: Tajima's D*,* F_{ST} *and nucleotide diversity* π

The allele frequency differences between the Mediterranean and cold-temperate populations might result from selective sweeps driven by positive selection in the surrounding regions. To test if the fixed alleles in BGCs are a result of selective sweeps affecting neighbouring genomic regions we calculated corrected Tajima's *D* (sensu Kofler *et al*. (2011a)), nucleotide diversity π, and F_{ST} (Supporting Information S4, S5), and evaluated whether these statistics were significantly different between BGCs and 5 kb non-overlapping windows in a region of 100 kb up- and downstream of each BGC across all populations using non-parametric Wilcoxon rank-sum tests. BGCs with significantly higher π and lower F_{ST} than their surrounding regions were considered hitchhiking regions. In addition, for these regions, we tested whether Tajima's *D* values in their neighbouring 100 kb were significantly lower than the overall Tajima's *D* distribution across BGCs and surrounding regions combined (Supporting Information S4).

Calculations were performed on mpileup files - subsampled to a uniform depths (30× per pool), excluding positions with a maximum coverage above the 2% of the empirical coverage distribution of each pool (*subsample-pileup.pl*) and filtered for indels (*filter-pileup-by-gtf.pl*) - using the scripts *variance-at-position.pl* (for both Tajima's *D* and nucleotide diversity π; PoPoolation; Kofler *et al*., 2011a) and *fst-sliding.pl* (PoPoolation2; (Kofler *et al*., 2011b) (*--min-count 2*, *--min-covered-fraction 0.5*). Further details of the analyses are given in Supporting Information S1.

Finally, we tested whether the differences in allele frequencies between the Mediterranean and cold-temperate populations might simply be a result of an increased accumulation on mutations due to increased UV radiation at higher elevations. For this, we compared the distribution of nucleotide diversity π as well as the number of SNPs in the genomic regions surrounding BGCs between the Mediterranean and cold-temperate populations using non-parametric Wilcoxon rank-sum tests.

Another index used to infer positive selection on genes and proteins is the ratio of substitution rates at non-synonymous and synonymous sites, i.e., dN/dS. This method was developed for distantly related taxa/sequences to infer signatures of selection from fixed polymorphisms over divergent lineages. However, this measure is not suitable for inferring selection from population data such as the Pool-seq data used in our study, as the typical signature of positive selection, dN/dS >1, is violated within a population (Kryazhimskiy and Plotkin, 2008). For this reason, we refrained from using the dN/dS ratio to identify positive selection in our study.

HPLC analysis

We used high-performance liquid chromatography (HPLC) to investigate the chemical profiles of *U. pustulata* from the Mediterranean and cold-temperate populations. For this, we took four samples of *U. pustulata* from the Mediterranean and cold-temperate zones of Mount Limbara (Italy - IT) and Sierra de Gredos (Spain - ESii) gradients each (Supporting Information S6) and performed HPLC analysis using a protocol based on Feige *et al*. (1993), described in more detail in Benatti *et al*. (2013) and Supporting Information S1.

UPLC-high resolution mass spectrometry analysis and GNPS network analysis

Ultra-performance liquid chromatography high-resolution mass spectrometry analyses were performed with an UltiMate 3000 system (Thermo Fisher) coupled to an Impact II qToF mass spectrometer (Bruker) as described previously (Tobias *et al*., 2017). Small fragments of the same 16 samples used for HPLC were also screened using UV HPLC. The raw MS-data acquired was then converted to .mzXML format using DataAnalysis 4.3 (Bruker). MS^2-based network analysis was performed using the Molecular Networking workflow of the GNPS platform (Wang *et al*., 2016). Details are provided in Supporting Information S1.

Results

Reference genomes and Pool-seq data

The reference genome of the PacBio-derived cold-temperate ecotype (GenBank acc. no. VXIT00000000) is ~33 Mb in length (N50 = 1.8 Mb, number of scaffolds = 43) and has a completeness of ~95% according to BUSCO (Greshake-Tzovaras *et al.*, 2020). The Illumina-derived reference genome of the Mediterranean ecotype (ENA acc. no. GCA_000938525) is ~39 Mb in length (N50 = 104.3 kb, number of scaffold = 3891) and has a completeness of ~92% according to BUSCO (Dal Grande *et al*., 2017).

The number of reads for each population retained after quality and length filtering is given in Supporting Information S7 (A: three cold-temperate and three Mediterranean populations; B: four additional populations from the Italian gradient, Dal Grande *et al*. (2017); the Pool-seq assemblies are available at https://doi.org/10.6084/m9.figshare.14611191.v2). We detected 11 008 coding SNPs in all BGC regions, consisting of 335 genes. The genome-wide DISTATIS analysis of the coding SNPs clearly shows that the studied populations represent the two previously determined *U*. *pustulata* ecotype groups *sensu* Dal Grande *et al*. (2017), i.e. Mediterranean = warm-adapted, cold temperate = cold-adapted ecotype (Fig. 1B). The population groups resulting from a DISTATIS analysis of the coding BGC SNPs were remarkably similar to the groups inferred using the genome-wide set of coding SNPs (Fig. 1).

Predicted BGCs

The fungal version of antiSMASH (v5.0) predicted 26 and 28 clusters for the Mediterranean and cold-temperate ecotype reference genome respectively (Table 1). Of these 24 were homologous (Table 1) and six were detected only in the reference genome of one ecotype (Table 2). However, when checked for presence–absence patterns in the Pool-seq data (by assessing the number of raw reads aligning to the core gene of the cluster of interest), only one cluster, cluster 22, was specific to a climate group, and the other five were present in all populations (Table 2–4. Therefore, although antiSMASH (v5.0) predicted only 26 and 28 clusters in the Mediterranean and cold-temperate reference genome, Pool-seq data suggest that the total number of BGCs present in *U*. *pustulata* is 30. Type I PKSs were the most abundant family of natural product enzymes identified in both reference genomes (14 and 16 clusters respectively from the Mediterranean and cold-temperate ecotypes), followed by terpenes (six clusters) and NRPS-like clusters (four) and hybrid type I PKS-NRPS-like clusters (two).

The cluster 22, specific to cold-temperate populations, contains a non-reducing type I PKS and shows 67% similarity to the BGC associated with the yellow fungal pigment ankaflavin in *Monascus pilosus* (Ascomycota) (Fig. 2A). The absence of cluster 22 was confirmed in all the populations using both raw and assembled Pool-seq reads (Tables 2–4). This cluster has 13 genes (Fig. 2A), of which nine, including the core genes, are entirely absent in the Mediterranean populations.

Identifying consistent allele frequency differences in populations: Quasibinomial GLM

We detected 7353 SNPs (in 325 genes) that were then used in the quasibinomial GLM analysis. Quasibinomial GLMs recovered 1516 SNPs (in 249 genes) as significantly differentiated (at Bonferroni-type $\alpha = 0.0007$) between the Mediterranean and cold-temperate climate populations. These significant allele frequency differences are likely a result of selection rather than drift. Of these, 47.6% (721 SNPs in 201 genes, representing 20 BGCs) were non-synonymous SNPs. From these SNPs we then selected those with potentially functionally deleterious effect on the synthesized protein using PROVEAN.

Identifying functionally deleterious SNPs

We identified 109 non-synonymous, functionally deleterious mutations. Using the conservative approach to filter out any potential false positives we implemented a stricter threshold (PROVEAN score <-4.0), and selected SNPs that had a PROVEAN score lower than -4.0. Finally, we zoomed in on those SNPs that varied consistently in all three Mediterranean–cold-temperate population pairs. Twenty-eight SNPs (in eight BGCs) varied consistently across replicate populations (Table 5).

Differentiated BGCs between the Mediterranean and cold-temperate populations

Eight BGCs contained highly differentiated, functionally deleterious, variations between the Mediterranean and cold-temperate populations (Table 5). The high genetic variation found in eight homologous BGCs between ecotypes stems from missense and functionally deleterious SNPs, including four STOP codons (Table 5). As candidate SNPs we then considered only those which had an allele frequency difference of >90% between the two climate groups.

Out of the eight differentiated BGCs, six clusters had functionally deleterious SNPs both in the Mediterranean

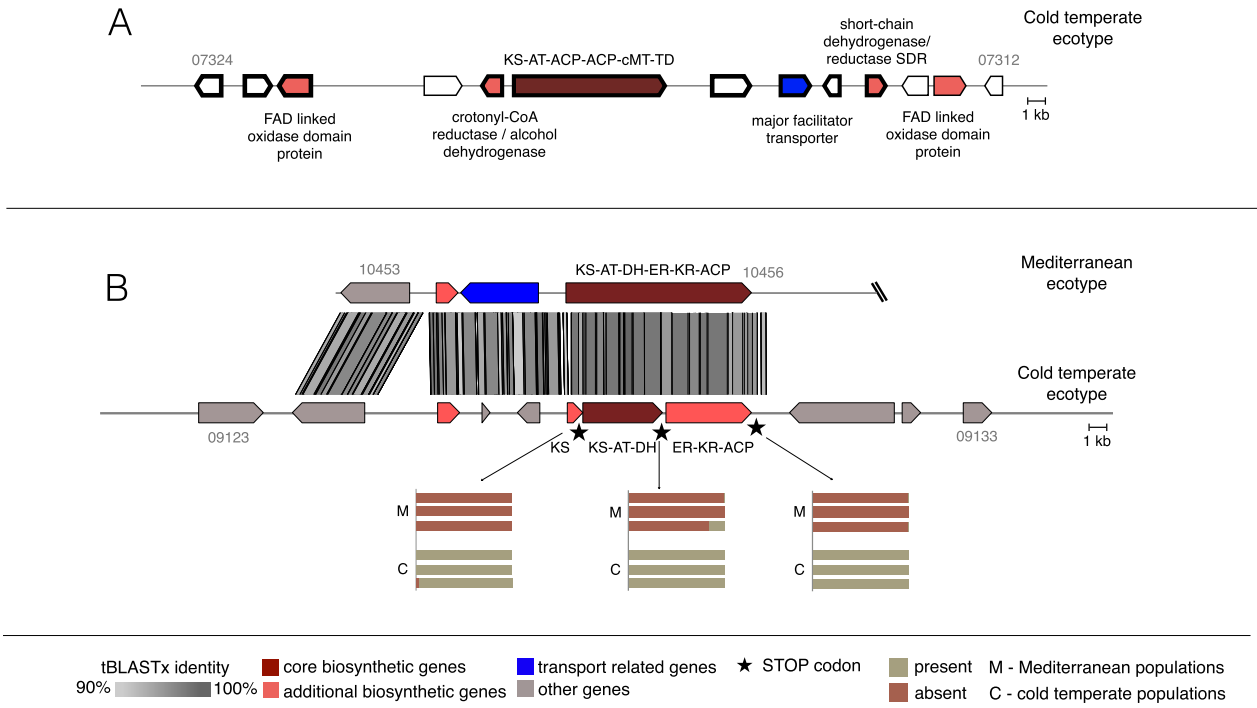

Fig. 2. Differentiated BGCs between the Mediterranean and cold-temperate *U. pustulata* populations.
A. Cluster architecture of cluster 22. Rectangular boxes represent genes belonging to the cluster and correspond to gene numbers 07312 to 07324 from right to left. Thick boxes around the genes represent genes lost in all Mediterranean populations.
B. Synteny plot depicting the homology of cluster 27 between the Mediterranean and cold-temperate ecotypes. The figure depicts the core gene in the PKS cluster in the cold-temperate ecotype interrupted by three STOP codons. Rectangular boxes in the cluster correspond to genes, from gene number 09123 to 09133 (cold-temperate ecotype) and 10453 to 10456 (Mediterranean ecotype) from left to right. Different domains identified by antiSMASH within a cluster are highlighted with different colours. STOP codons in the core biosynthetic gene are marked with stars. The combined frequency of STOP codons in the core biosynthetic gene in three Mediterranean and cold-temperate populations is represented by bar plots below the synteny plot. The M and C bars correspond to the three Mediterranean and three cold-temperate populations respectively. Frequency of SNPs present in each domain for each population is given in Table 5. [Color figure can be viewed at wileyonlinelibrary.com]

and cold-temperate populations (Table 5). Examples include three clusters containing a type-I PKS: (i) cluster 19 (53% similarity to the grayanic acid cluster), (ii) cluster 3 (47% similarity to the squalestatin cluster) and (iii) cluster 11 (47% similarity to the cichorine cluster) (Table 5). The identity to the most similar compound does not mean that the cluster is involved in the synthesis of the same compound.

Two clusters had all deleterious SNPs concentrated in one climate group (Table 5). Of these, cluster 27 (Table 5; Fig. 2B) contained 15 non-synonymous, functionally deleterious SNPs in the cold-temperate populations. Four of these SNPs induced a premature stop codon: two stop codons are in the core biosynthetic gene (PKS), one in a transporter gene and one in an adjacent, uncharacterized gene (Fig. 2B). The SNPs are accumulated in the cold-temperate populations (99%–93% frequency in the three gradients, 300 samples, vs. 0%–1.8% frequency in the Mediterranean populations). This cluster showed 47% similarity to the cluster associated with the biosynthesis of 4-epi-15-epi-brefeldin A from *Penicillium brefeldianum*. The second highly differentiated cluster, cluster 13, had one significantly deleterious mutation in an oxidoreductase gene (PROVEAN score −10.8), which was detected at 94% frequency in the cold-temperate, and 0% in the Mediterranean populations (Table 5). This cluster had 66% similarity to the emodin cluster.

Testing the possibility of selective sweeps: Tajima's D, F_{ST} *and nucleotide diversity* π

To exclude the possibility of selective sweeps driving the fixation of alleles or accumulation of deleterious mutations in BGCs, we compared Tajima's *D*, π and F_{ST} distribution within BGCs and in 100 kb regions surrounding each BGC using a sliding window approach, for a total of 5.25 Mb (15.9% of the genome) (Supporting Information S4, S5). Cluster 13 showed strong evidence of hitchhiking, i.e. the 100 kb upstream region displayed on average significantly lower π and higher F_{ST} values than the corresponding BGC (both $p < 0.01$) (Supporting Information S8 and S9). Although the difference in

Table 5. Biosynthetic gene clusters differentiated between the Mediterranean and cold-temperate populations of *U. pustulata* along with the information on the most similar cluster, SNP-containing domain, amino acid change and PROVEAN score for the variation.

Cluster number	Scaffold and location	Core gene	Most similar cluster	Affected domain	Amino acid change	PROVEAN score	Combined frequency in three Mediterranean pools (300 samples)	Combined frequency in three cold-temperate pools (300 samples)
Cluster 3	Scaffold 1 (2 549 459–2 609 273 bp)	T1PKS	Squalestatin (alernapyrone)	**PKS-DH**	P991S	−5.800	6.3%	99.4%
				Others	P240T	−6.050	95.2%	0.9%
Cluster 5	Scaffold 2 (1 210 936–1 252 962 bp)	NRPS-like		Other genes	T481I	−4.400	3.8%	87.2
				Other genes	P500A	−5.792	78.00%	0
Cluster 11	Scaffold 5 (553 260–605 232 bp)	T1PKS, NRPS-like	Cichorine	cyt p450	G218W	−4.387	95.00%	2.9%
				cyt p450	P334A	−5.058	4.6%	97.3
Cluster 13	Scaffold 6 (975035–1,020,635 bp)	T1PKS	Emodin	Oxidoreductase	C116R	−10.833	0%	94.00%
Cluster 18	Scaffold 10 (34 379–77 533 bp)	T1PKS	Pestheic acid	Additional bgs	**Y352--**	NA	82.6	7.3%
Cluster 19	Scaffold 10 (357 565–402 103 bp)	T1PKS	Gryanic acid	PKS	P837Q	−4.458	3%	98%
				Additional gene	Q110--	NA	100.00%	0
				Other genes	P102Q	−4.317	93.3%	0.3%
				Other genes	P224L	−6.025	94.4%	1.3%
Cluster 26	Scaffold 19 (281 908–334 111 bp)	NRPS, T1PKS		Transport-related	**Y564--**	NA	48.5%	0.9
Cluster 27	**Scaffold 21 (49 578–93 103 bp)**	**T1PKS**	**4-Epi-15-epi-brefeldin A**	**PKS-KS**	**R230--**	**NA**	**0.35%**	95.00%
				PKS-KS	Y86C	−8.925	1.8%	95.00%
				PKS-AT	T421N	−4.858	0.4%	95.00%
				PKS	F724S	−5.267	0.7%	96.00%
				PKS	**S1044--**	**NA**	**1.00%**	99.2%
				PKS	W252S	−14. 000	0.7%	99.00%
				PKS-ER	G454E	−4.622	0.4%	99.00%
				PKS-ER	S623F	−5.267	0.6%	99.7%
				Dehydrogenase/ Reductase	F10L	−4.667	0.8%	93.00%
				Drug resistance transporter	G110S	−3.567	0.4%	98.00%
					G271R	−6.314	0.3%	98.00%
					S306--	**NA**	**1.3%**	**97.3%**
				Other genes	R394M	−4.356	0	95.00%
				Other genes	W703S	−10.989	45%	0%
				Other genes	--24K		28.5%	100.00%

Combined frequencies of the significantly differentiated SNPs between the Mediterranean and cold-temperate populations from three elevation gradients are listed. Only differentiated, non-synonymous SNPs which are major alleles with significant PROVEAN score (<−4.0) were considered significant. BGCs with STOP codon as major allele in core PKS genes are represented in Fig. 2B along with the frequency of STOP codons. SNPs present predominantly in the Mediterranean populations are highlighted in bold italic. PKS, polyketide synthase gene; PKS-KS, keto-synthase; PKS-AT, acyltransferase; ER, Enoylreductase; PKS-DH, dehydratase domain of PKS gene; T1PKS, type I PKS; NRPS, non-ribosomal peptide synthetase. The cluster, genes and STOP codons highlighted in bold in cluster 27 are represented in Fig. 2 along with the frequencies of STOP codons in each population.

Tajima's *D* value distribution between BGC and its upstream region was not statistically significant, the upstream region had negative mean Tajima's *D* (−0.38), significantly lower than the overall Tajima's *D* value distribution ($p < 0.01$).

Conversely, cluster 19 displayed evidence of selection with significantly lower π and higher F_{ST} values (both $p < 0.01$) compared to its surrounding 100 kb up- and downstream regions, as well as negative mean Tajima's *D* (−0.3) (Supporting Information S4, S5, and S8). In this cluster deleterious mutations were present and almost fixed in both cold-temperate and Mediterranean populations; in the PKS (the core gene involved in the synthesis of the backbone of the compound) in the cold-temperate populations, and three mutations in 'other' genes (coding for unknown proteins) in the cluster in the Mediterranean populations. This cluster was not selected as a candidate cluster displaying climate-specific SNP patterns, because it showed an accumulation of deleterious mutations in both climates (in different genes).

We found no evidence for increased accumulation of mutations at higher elevation (Wilcoxon Rank Sum Test, $p = 0.2$), with the Mediterranean and cold-temperate populations displaying similar levels of nucleotide diversity π in the analysed ~5 Mb of regions surrounding BGCs (Supporting Information S8). The difference in terms of number of SNPs between low and high elevation populations was also not significant ($p = 0.963$).

HPLC

We found three major compounds present in both, the Mediterranean and cold-temperate, samples of *U. pustulata*: gyrophoric acid, lecanoric acid and hiascic acid (Supporting Information S10). The relative proportion of the natural products was similar in all samples, and no elevation-specific differences were detected. Gyrophoric acid was the major compound in all samples (about 85%–95%), and there were traces of lecanoric acid (5%–10%) and hiascic acid (0%–10%) (Supporting Information S10).

Network analysis

The MS^2-based network analysis showed that there are indeed chemical differences between the Mediterranean and cold-temperate lineages (Supporting Inform ation S11). We found that one subnetwork was specific to the Mediterranean samples. Two natural products from this subnetwork were identified as the O-methylated flavonols jaceidin (**1**) and myricetin-3,7,3′,4′-tetrameth ylether (**2**) respectively, via library search against the GNPS spectral library. Albeit not being present in all of the Mediterranean samples, none of the O-methylated flavonols was present in the cold-temperate samples. Three compounds of this class were specific to the Italian samples (jaceidin (**1**), myricetin-3,7,3′,4′-tetramethylether (**2**) and *m/z* 331), while one (*m/z* 345) also occurred in samples from Spain. All of the other subnetworks that seemed to be climate-specific were identified as false positives.

Discussion

Natural products are suggested to provide cold/desiccation tolerance to species facing a broad range of abiotic conditions (Ramakrishna and Ravishankar, 2011; Medina *et al*., 2015). This tolerance has mainly been demonstrated on laboratory strains under controlled experimental conditions and mostly involves temporary changes in gene expression and metabolite profiles under different abiotic conditions. For instance, cold acclimation in *Arabidopsis* involves enhanced expression of biosynthetic genes involved in flavonoid metabolism which results in an increase in flavonol and anthocyanin content when exposed to cold temperature (Schulz *et al*., 2015). Ours is the first study showing climate-specific differentiation of biosynthetic genes in natural populations of a lichenized fungus. We found that BGCs of *U. pustulata* vary in presence/absence or in structure between the Mediterranean and cold-temperate populations (Tables 1–5). Specifically, we found two clusters of interest: cluster 22, which was almost entirely absent from Mediterranean populations, but always present in cold-temperate populations (Tables 2–4), and cluster 27 which had several deleterious mutations in the cold-temperate populations (Table 5). The pattern in these clusters appears to be a direct response to selection rather than an effect of genetic drift or hitchhiking (Supplementary Information S4, S5 and S8).

Our results suggest that the proposed candidate BGCs are linked to natural products that potentially allow the producer to persist in its respective climatic niche. These functional traits could be related to the abiotic conditions directly, e.g., by conferring temperature or UV tolerance, or they could be related to the interaction with a particular photosynthetic partner, which only occurs in a specific climate niche. While we cannot associate a metabolite with these clusters yet, our results suggest that the natural products associated with these highly differentiated BGCs are, among several other differences (Dal Grande *et al*., 2017), ecotype-defining traits.

A BGC present only in cold-temperate populations

Intraspecific BGC variation, including the presence or absence of whole clusters, has been shown in a few fungal species by comparing genomes of different strains

(e.g. Susca *et al*., 2016; Lind *et al*., 2017; Villani *et al*., 2019). However, we still have limited knowledge about the evolution of BGCs in natural populations. Whether BGC occurrence is associated with particular ecologies is an open question in fungal evolution. Here we report a cluster, cluster 22, that is apparently fixed in cold-temperate populations, while being absent from Mediterranean populations (Fig. 2A; Tables 2–4). The abundance of cluster 22 progressively increases with elevation when traced in multiple populations along an elevation gradient (Tables 2–4). This suggests a strong positive selection for maintaining the function of the encoded product in the cold-temperate climate.

The presence or absence of whole clusters associated with natural product biosynthesis is a genetic driver of chemical variation among fungal strains (Lackner *et al*., 2012; Zeng *et al*., 2018; Olarte *et al*., 2019). For instance, in the filamentous fungus *Aspergillus nidulans* nine BGCs vary in presence/absence and this has been suggested to contribute to isolate-specific virulence (Drott *et al*., 2020). Here we show that the presence of an entire cluster can be restricted to a particular ecological niche.

Based on the strong evidence of climate-related presence/absence patterns across replicate populations, it is tempting to speculate that the BGC found exclusively in cold-temperate populations is involved in temperature-related or UV-related stress tolerance in *U. pustulata*. One major difference between the studied biomes is the regular occurrence of frost in the cold-temperate biome. Cold tolerance in fungi often involves the synthesis of frost-resistance compounds, such as trehalose and other cryoprotectant sugars, polyols and lipids/fatty acids, and conidial pigments such as melanin (Robinson, 2001; Hagiwara *et al*., 2014). Despite this, the molecular mechanisms and the genomic basis of frost resistance in fungi are not clear. Biosynthetic gene clusters, such as cluster 22, only found in cold-temperate populations that experience regular frost events are candidates for encoding functions related to cryoprotection.

Cluster 22 is similar to the cluster associated with the conidial pigment ankaflavin (*Monascus pilosus*). The role of ankaflavin has not been ascertained in nature, but conidial pigments have been shown to play a role in cold- and UV-related stress resistance (Cordero and Casadevall, 2017; Blachowicz *et al*., 2020; van den Brule *et al*., 2020). For instance, intraspecific variations in oxidative- and heat-resistance properties have been attributed to differences in conidial pigment compounds in *Aspergillus fumigatus* and *A. terreus* (Low *et al*., 2011; Geib *et al*., 2016; Hagiwara *et al*., 2017). Similarity to a cluster does not necessarily indicate synthesis of the same compound. However, the observed similarity to several clusters producing compounds with similar absorption spectra suggests that the cluster might be involved in the synthesis of a compound with similar function or properties. Therefore, we speculate that cluster 22 might be involved in UV-related stress tolerance.

A BGC with deleterious mutations only in cold-temperate populations

We identify deleterious mutations as the second genetic driver that leads to environment-specific variation in BGCs. Specifically, we found multiple, functionally deleterious SNPs in core genes of cluster 27 that nearly reach fixation in all cold-temperate populations (Fig. 2B; Table 5). The occurrence of these SNPs in similar environments but at geographically distant sites potentially indicates the presence of strong selection pressures – and a functional role of the related compounds – rather than random heterogeneity or drift. The associated compounds might therefore provide beneficial functions to the Mediterranean ecotype or alternatively be detrimental to the cold-temperate ecotype.

One environmental stressor that organisms encounter in the Mediterranean climate is extended drought. In lichens, the exact mechanism of drought resistance is not known, however, it is suggested that it involves activation of three major stress-response pathways in the fungal partner: osmoregulation, DNA and protein damage repair and antioxidation (Wang *et al*., 2015). In fact, lichens such as *Umbilicaria antarctica*, *Ramalina terebrata* and *Stereocaulon alpinum*, which thrive under extreme drought and temperature conditions, contain high amounts of antioxidant substances (Bhattarai *et al*., 2008; Paudel *et al*., 2008). Similarity searches of cluster 27 yielded high similarity with the cluster responsible for the synthesis of the protein transport inhibitor epibrefeldin A in *Penicillium brefeldianum* which has potent antifungal, antiviral and antitumor properties in cell cultures (Seehafer *et al*., 2013). Although the role of epibrefeldin A in nature is not well understood, cluster 27 is an interesting candidate for future investigations.

A climate-specific BGC showing evidence of hitchhiking selection

Apart from positive selection, we identified hitchhiking selection as an additional driver for environment-specific accumulation of variations in BGCs. Under the hitchhiking effect an allele under positive selection may increase the fixation probability in surrounding linked loci (Smith and Haigh, 1974; Hedrick, 1982). Natural genetic variation is shaped via a variety of evolutionary forces, such as background, balancing and purifying selection, directional selection, genetic drift and genetic hitchhiking (Buskirk *et al*., 2017; Collevatti *et al*., 2019). In our study, cluster 13 (the BGC with the highest similarity to the

emodin cluster) has one deleterious mutation in the transporter gene which is almost fixed in the cold-temperate populations. We found, however, evidence of a selective sweep in the surrounding regions of this cluster (Supporting Information S8 and S9), suggesting that the fixed mutations in the cluster might be the result of an indirect response to selection or hitchhiking selection.

The precursor of emodin, trypacidin, serves important ecological functions related to interactions between organisms and environmental factors, such as light intensity and osmotic stress (drought and salinity) (Inderjit and Nishimura, 1999; Izhaki, 2002). Interestingly, Blachowicz *et al*. (2020) showed that the loss of the emodin precursor enhances UV-C resistance in *Aspergillus fumigatus*. The loss of function of this cluster could thus provide a selective advantage to cold-temperate populations which occur at higher altitude and might experience higher levels of UV radiation. Although emodin has not been reported from *U*. *pustulata*, it has been reported from *U*. *papulosa* - a congeneric species with North American distribution preferentially at high altitudes (Boh man, 1969; Cohen, 2002). The fixation of deleterious mutations specific to cold-temperate populations likely resulting from selective sweeps shows that genetic hitchhiking is another powerful force driving the evolution of the biosynthetic landscape in fungi.

Natural products and fungal–algal interactions

To understand potential links between genotype and phenotype we analysed chemical differences between samples from the Mediterranean and cold-temperate populations (Legaz *et al*., 1986; Molnár and Farkas, 2010; Gauslaa *et al*., 2013; Shrestha and St Clair, 2014; Paukov *et al*., 2019). Natural products in lichenized fungi have been implicated in environmental stress tolerance (Legaz *et al*., 1986; Gauslaa *et al*., 2013; Paukov *et al*., 2019). Consequently, the chemical profile of an organism may show seasonal variations and vary under different environmental conditions (Legaz *et al*., 1986; McEvoy *et al*., 2006; Solhaug *et al*., 2009; Shrestha and St Clair, 2014). Hence when sampling across gradients and from different geographic locations it is quite likely that samples have different chemical profiles. However, in our study typical lichen compounds, such as the depside gyrophoric acid, did not differ in content and quantity between the two climate groups (Supporting Information S10). This is in accordance with our genome-wide BGC analyses, which showed that the five clusters putatively associated with depside and depsidone biosynthesis, i.e. BGCs containing non-reducing PKSs (Armaleo *et al*., 2011), do not show features specific for either climate zone. The three environment-associated clusters identified here carry reducing *PKSs* and show no similarity to the *PKSs* linked to depside biosynthesis.

Using HPLC-MS we detected slight chemical differences between the samples from the Mediterranean and cold-temperate regions, for example a 3-O-methylated flavonol, myricetin, which occurs only in samples from Mediterranean populations (Supporting Information S11). Flavonols can most likely be attributed to the algal partner in the symbiosis, because they are well known from the green plant lineage, including green algal lichen photobionts (Jorgensen, 1993; Falcone Ferreyra *et al*., 2012). While this substance is not linked to our findings of BGC differences in the fungal partner, its detection solely in Mediterranean samples (Supporting Information S11) is nonetheless interesting from an ecological point of view. First, flavonol concentration in lichens varies seasonally, suggesting that flavonol biosynthesis may be directly affected by environmental cues (Aoussar *et al*., 2018), and second, myricetin is an antioxidant that reduces UV-induced cell damage (Ruhland *et al*., 2005; Falcone Ferreyra *et al*., 2012). It is thus possible that the specific lineages of *Trebouxia* algae associated with Mediterranean *U*. *pustulata* samples provide a chemical trait that contributes to environmental stress tolerance of the lichen. This is in line with a study showing that the presence of photobionts alleviates desiccation-related stress in mycobionts (Kranner *et al*., 2005).

The variation in the distribution of BGCs and natural products along elevation might be related to environmental factors directly (as discussed above), or it could be related to environmentally driven changes in mycobiont–photobiont interactions. Photobiont communities associated with lichens may change along climatic gradients across fungal host species (Vargas Castillo and Beck, 2012; Singh *et al*., 2017), and within fungal host species (Dal Grande *et al*., 2018b; Rolshausen *et al*., 2020). Some lineages of symbiotic algae have explicit climatic preferences (Rolshausen *et al*., 2018). In *U*. *pustulata* a turnover of algal communities takes place at the transition between the Mediterranean and the cold-temperate climate zone (Dal Grande *et al*., 2018b; Rolshausen *et al*., 2020). Therefore, it is possible that the natural product portfolio of the fungal partner is directly related to the interaction with a specific lineage of *Trebouxia* photobiont. Mycobionts confer stress tolerance (against, e.g., desiccation, high light) to photobionts, as shown in physiological comparisons of intact lichen thalli with isolated and cultured strains of lichen photobionts (Kranner *et al*., 2005). The mechanism behind this is at least partially mediated by natural products, such as the fungal phytohormone indole-3-acetic acid, which has been shown to increase the photobiont's water content (Pichler *et al*., 2020). It is an area of further investigation

to assess whether Mediterranean fungal ecotypes synthesize compounds specifically tailored to the needs of Mediterranean algae.

Conclusions

Our study contributes to the understanding of the mechanisms of environmental adaptation and the discovery of fungal metabolites with potentially valuable biological activities. We demonstrated that natural populations of *U. pustulata* from the Mediterranean and cold-temperate environments are equipped with the genetic potential to produce different, environment-specific metabolites. We identified whole cluster loss and deleterious point mutations as genetic drivers of ecotype-specific variation in BGCs. Evolutionary mechanisms that drive the observed patterns of point mutations are intra-cluster (positive) selection, and - in one case - a selective sweep (hitchhiking selection). We advocate that the biosynthetic landscape of organisms is shaped by positive and hitchhiking selection and that both play an important role in the ecology and evolution of BGCs in *U. pustulata*. The population genomic approach presented here constitutes a way forward in pinpointing BGCs putatively linked to important biological functions.

Acknowledgements

We thank the LOEWE-Centre TBG funded by the Hessen State Ministry of Higher Education, Research and the Arts (HMWK). We are grateful to the developers of Provean and antiSMASH, Yongwook Choi (The J. Craig Venter Institute, Maryland, USA), and Marnix H Medema (Wageningen University, Wageningen, the Netherlands) respectively, for their suggestions on the interpretation on some of the analyses, Daniele Armaleo (Duke University, NC, USA) for his helpful insights on our results and Jürgen Otte (Senckenberg Biodiversity and Climate Research Centre, Frankfurt, Germany) for help in the lab. We thank three anonymous reviewers for the critical and thorough evaluation of the manuscript.

References

Aoussar, N., Rhallabi, N., Ait Mhand, R., Manzali, R., Bouksaim, M., Douira, A., and Mellouki, F. (2018) Seasonal variation of antioxidant activity and phenolic content of *Pseudevernia furfuracea*, *Evernia prunastri* and *Ramalina farinacea* from Morocco. *J Saudi Soc Agric Sci* **19**: 1–6.

Armaleo, D., Sun, X., and Culberson, C. (2011) Insights from the first putative biosynthetic gene cluster for a lichen depside and depsidone. *Mycologia* **103**: 741–754.

Benatti, M.N., Gernert, M., and Schmitt, I. (2013) *Parmotrema hydrium*, a new species of Parmeliaceae in southeastern Brazil. *Acta Bot Brasilica* **27**: 810–814.

Bhattarai, H.D., Paudel, B., Hong, S.G., Lee, H.K., and Yim, J.H. (2008) Thin layer chromatography analysis of antioxidant constituents of lichens from Antarctica. *J Nat Med* **62**: 481–484.

Blachowicz, A., Raffa, N., Bok, J.W., Choera, T., Knox, B., Lim, F.Y., *et al*. (2020) Contributions of spore secondary metabolites to UV-C protection and virulence vary in different *Aspergillus fumigatus* strains. *MBio* **11**: e03415-19.

Blin, K., Shaw, S., Steinke, K., Villebro, R., Ziemert, N., Lee, S.Y., *et al*. (2019) antiSMASH 5.0: updates to the secondary metabolite genome mining pipeline. *Nucleic Acids Res* **47**: W81–W87.

Bohman, G. (1969) Chemical studies on lichens. 22. Anthraquinones from the lichen *Lasallia papulosa* var. *rubiginosa* and the fungus *Valsaria rubricosa*. *Acta Chem Scand* **23**: 2241–2244.

Buskirk, S.W., Peace, R.E., and Lang, G.I. (2017) Hitchhiking and epistasis give rise to cohort dynamics in adapting populations. *Proc Natl Acad Sci U S A* **114**: 8330–8335.

Calchera, A., Dal Grande, F., Bode, H.B., and Schmitt, I. (2019) Biosynthetic gene content of the "perfume lichens" *Evernia prunastri* and *Pseudevernia furfuracea*. *Molecules* **24**: 203.

Calcott, M.J., Ackerley, D.F., Knight, A., Keyzers, R.A., and Owen, J.G. (2018) Secondary metabolism in the lichen symbiosis. *Chem Soc Rev* **47**: 1730–1760.

Camarena-Pozos, D.A., Flores-Núñez, V.M., López, M.G., López-Bucio, J., and Partida-Martínez, L.P. (2019) Smells from the desert: microbial volatiles that affect plant growth and development of native and non-native plant species. *Plant Cell Environ* **42**: 1368–1380.

Choi, Y., and Chan, A.P. (2015) PROVEAN web server: a tool to predict the functional effect of amino acid substitutions and indels. *Bioinformatics* **31**: 2745–2747.

Cohen, P.A. (2002) Halogenated anthraquinones from the rare southern Illinois lichen *Lasallia papulosa*. *Lichenologist* **34**: 521–525.

Collevatti, R.G., Novaes, E., Silva-Junior, O.B., Vieira, L.D., Lima-Ribeiro, M.S., and Grattapaglia, D. (2019) A genome-wide scan shows evidence for local adaptation in a widespread keystone Neotropical forest tree. *Heredity (Edinb)* **123**: 117–137.

Cordero, R.J.B., and Casadevall, A. (2017) Functions of fungal melanin beyond virulence. *Fungal Biol Rev* **31**: 99–112.

Dal Grande, F., Meiser, A., Greshake Tzovaras, B., Otte, J., Ebersberger, I., and Schmitt, I. (2018a) The draft genome of the lichen-forming fungus *Lasallia hispanica* (Frey) Sancho & A. Crespo. *The Lichenologist* **50**: 329–340.

Dal Grande, F., Rolshausen, G., Divakar, P.K., Crespo, A., Otte, J., Schleuning, M., and Schmitt, I. (2018b) Environment and host identity structure communities of green algal symbionts in lichens. *New Phytol* **217**: 277–289.

Dal Grande, F., Sharma, R., Meiser, A., Rolshausen, G., Büdel, B., Mishra, B., *et al*. (2017) Adaptive differentiation coincides with local bioclimatic conditions along an elevational cline in populations of a lichen-forming fungus. *BMC Evol Biol* **17**: 93.

Davydov, E.A., Peršoh, D., and Rambold, G. (2017) Umbilicariaceae (lichenized Ascomycota) – trait evolution and a new generic concept. *Taxon* **66**: 1282–1303.

Drott, M.T., Bastos, R.W., Rokas, A., Ries, L.N.A., Gabaldón, T., Goldman, G.H., *et al*. (2020) Diversity of

secondary metabolism in *Aspergillus nidulans* clinical isolates. *mSphere* **5**: e00156-20.

Falcone Ferreyra, M.L., Rius, S.P., and Casati, P. (2012) Flavonoids: biosynthesis, biological functions, and biotechnological applications. *Front Plant Sci* **3**: 222.

Feige, G.B., Lumbsch, H.T., Huneck, S., and Elix, J.A. (1993) Identification of lichen substances by a standardized high-performance liquid chromatographic method. *J Chromatogr A* **646**: 417–427.

Gauslaa, Y., Bidussi, M., Solhaug, K.A., Asplund, J., and Larsson, P. (2013) Seasonal and spatial variation in carbon based secondary compounds in green algal and cyanobacterial members of the epiphytic lichen genus *Lobaria*. *Phytochemistry* **94**: 91–98.

Gauslaa, Y., and Ustvedt, E.M. (2003) Is parietin a UV-B or a blue-light screening pigment in the lichen *Xanthoria parietina*? *Photochem Photobiol Sci* **2**: 424–432.

Geib, E., Gressler, M., Viediernikova, I., Hillmann, F., Jacobsen, I.D., Nietzsche, S., *et al*. (2016) A non-canonical melanin biosynthesis pathway protects *Aspergillus terreus* conidia from environmental stress. *Cell Chem Biol* **23**: 587–597.

Greshake Tzovaras, B., Segers, F., Bicker, A., Grande, F.D., Otte, J., Anvar, S.Y., *et al*. (2020) What is in Umbilicaria pustulata? A metagenomic approach to reconstruct the holo-genome of lichen. *Genome Biol Evol* **12**: 309–324.

Guirao-Rico, S., and González, J. (2021) Benchmarking the performance of Pool-seq SNP callers using simulated and real sequencing data. *Mol Ecol Resour* **21**: 1216–1229.

Hagiwara, D., Sakai, K., Suzuki, S., Umemura, M., Nogawa, T., Kato, N., *et al*. (2017) Temperature during conidiation affects stress tolerance, pigmentation, and trypacidin accumulation in the conidia of the airborne pathogen *Aspergillus fumigatus*. *PLoS One* **12**: e0177050.

Hagiwara, D., Suzuki, S., Kamei, K., Gonoi, T., and Kawamoto, S. (2014) The role of AtfA and HOG MAPK pathway in stress tolerance in conidia of *Aspergillus fumigatus*. *Fungal Genet Biol* **73**: 138–149.

Hedrick, P.W. (1982) Genetic hitchhiking: a new factor in evolution? *Bioscience* **32**: 845–853.

Huneck, S. (1999) The significance of lichens and their metabolites. *Naturwissenschaften* **86**: 559–570.

Huneck, S. (2003) Die wasserabweisende Eigenschaft von Flechtenstoff. In *Lichenological Contributions in Honour of G.B. Feige*, Jensen, M. (ed). Berlin, Stuttgart: Bibliotheca Lichenologica, J. Cramer, pp. 9–12.

Inderjit and Nishimura, H. (1999) Effect of the anthraquinones emodin and physcion on availability of selected soil inorganic ions. *Ann Appl Biol* **135**: 425–429.

Izhaki, I. (2002) Emodin - a secondary metabolite with multiple ecological functions in higher plants. *New Phytol* **155**: 205–217.

Jorgensen, R. (1993) The origin of land plants: a union of alga and fungus advanced by flavonoids? *Biosystems* **31**: 193–207.

Keller, N.P., Turner, G., and Bennett, J.W. (2005) Fungal secondary metabolism - from biochemistry to genomics. *Nat Rev Microbiol* **3**: 937–947.

Kjærbølling, I., Vesth, T.C., Frisvad, J.C., Nybo, J.L., Theobald, S., Kuo, A., *et al*. (2018) Linking secondary metabolites to gene clusters through genome sequencing of six diverse *Aspergillus* species. *Proc Natl Acad Sci U S A* **115**: E753–E761.

Kofler, R., Orozco-terWengel, P., De Maio, N., Pandey, R.V., Nolte, V., Futschik, A., *et al*. (2011a) PoPoolation: a toolbox for population genetic analysis of next generation sequencing data from pooled individuals. *PLoS One* **6**: e15925.

Kofler, R., Pandey, R.V., and Schlötterer, C. (2011b) PoPoolation2: identifying differentiation between populations using sequencing of pooled DNA samples (Pool-Seq). *Bioinformatics* **27**: 3435–3436.

Kranner, I., Cram, W.J., Zorn, M., Wornik, S., Yoshimura, I., Stabentheiner, E., and Pfeifhofer, H.W. (2005) Antioxidants and photoprotection in a lichen as compared with its isolated symbiotic partners. *Proc Natl Acad Sci U S A* **102**: 3141–3146.

Kryazhimskiy, S., and Plotkin, J.B. (2008) The population genetics of dN/dS. *PLoS Genet* **4**: e1000304.

Lackner, G., Misiek, M., Braesel, J., and Hoffmeister, D. (2012) Genome mining reveals the evolutionary origin and biosynthetic potential of basidiomycete polyketide synthases. *Fungal Genet Biol* **49**: 996–1003.

Legaz, M.E., Avalos, A., de Torres, M., Escribano, M.I., González, A., Martin-Falquina, A., *et al*. (1986) Annual variations in arginine metabolism and phenolic content of *Evernia prunastri*. *Environ Exp Bot* **26**: 385–396.

Lind, A.L., Wisecaver, J.H., Lameiras, C., Wiemann, P., Palmer, J.M., Keller, N.P., *et al*. (2017) Drivers of genetic diversity in secondary metabolic gene clusters within a fungal species. *PLoS Biol* **15**: e2003583.

Low, S.Y., Dannemiller, K., Yao, M., Yamamoto, N., and Peccia, J. (2011) The allergenicity of *Aspergillus fumigatus* conidia is influenced by growth temperature. *Fungal Biol* **115**: 625–632.

McEvoy, M., Nybakken, L., Solhaug, K.A., and Gauslaa, Y. (2006) UV triggers the synthesis of the widely distributed secondary lichen compound usnic acid. *Mycol Prog* **5**: 221–229.

Medema, M.H., Blin, K., Cimermancic, P., de Jager, V., Zakrzewski, P., Fischbach, M.A., *et al*. (2011) antiSMASH: rapid identification, annotation and analysis of secondary metabolite biosynthesis gene clusters in bacterial and fungal genome sequences. *Nucleic Acids Res* **39**: W339–W346.

Medina, A., Schmidt-Heydt, M., Rodríguez, A., Parra, R., Geisen, R., and Magan, N. (2015) Impacts of environmental stress on growth, secondary metabolite biosynthetic gene clusters and metabolite production of xerotolerant/xerophilic fungi. *Curr Genet* **61**: 325–334.

Molnár, K., and Farkas, E. (2010) Current results on biological activities of lichen secondary metabolites: a review. *Z Naturforsch - Sect C J Biosci* **65**: 157–173.

Olarte, R.A., Menke, J., Zhang, Y., Sullivan, S., Slot, J.C., Huang, Y., *et al*. (2019) Chromosome rearrangements shape the diversification of secondary metabolism in the cyclosporin producing fungus *Tolypocladium inflatum*. *BMC Genomics* **20**: 120.

Paudel, B., Bhattarai, H.D., Lee, J.S., Hong, S.G., Shin, H. W., and Yim, J.H. (2008) Antioxidant activity of polar lichens from King George Island (Antarctica). *Polar Biol* **31**: 605–608.

Paukov, A., Teptina, A., Morozova, M., Kruglova, E., Favero-Longo, S.E., Bishop, C., and Rajakaruna, N. (2019) The effects of edaphic and climatic factors on secondary lichen chemistry: a case study using saxicolous lichens. *Diversity* **11**: 94.

Pichler, G., Stöggl, W., Trippel, D., Candotto Carniel, F., Muggia, L., Ametrano, C.G., *et al*. (2020) Phytohormone release by three isolated lichen mycobionts and the effects of indole-3-acetic acid on their compatible photobionts. *Symbiosis* **82**: 95–108.

Pizarro, D., Divakar, P.K., Grewe, F., Crespo, A., Dal Grande, F., and Lumbsch, H.T. (2020) Genome-wide analysis of biosynthetic gene cluster reveals correlated gene loss with absence of Usnic acid in lichen-forming fungi. *Genome Biol Evol* **12**: 1858–1868.

Raineri, E., Ferretti, L., Esteve-Codina, A., Nevado, B., Heath, S., and Pérez-Enciso, M. (2012) SNP calling by sequencing pooled samples. *BMC Bioinformatics* **13**: 239.

Ramakrishna, A., and Ravishankar, G.A. (2011) Influence of abiotic stress signals on secondary metabolites in plants. *Plant Signal Behav* **6**: 1720–1731.

R Core Team (2019). R: A Language and Environment for Statistical Computing. R Foundation for Statistical Computing, Vienna, Austria. https://www.R-project.org/.

Rice, P., Longden, L., and Bleasby, A. (2000) EMBOSS: the European molecular biology open software suite. *Trends Genet* **16**: 276–277.

Robinson, C.H. (2001) Cold adaptation in Arctic and Antarctic fungi. *New Phytol* **151**: 341–353.

Rolshausen, G., Dal Grande, F., Sadowska-Deś, A.D., Otte, J., and Schmitt, I. (2018) Quantifying the climatic niche of symbiont partners in a lichen symbiosis indicates mutualist-mediated niche expansions. *Ecography (Cop)* **41**: 1380–1392.

Rolshausen, G., Hallman, U., Grande, F.D., Otte, J., Knudsen, K., and Schmitt, I. (2020) Expanding the mutualistic niche: parallel symbiont turnover along climatic gradients. *Proc R Soc B Biol Sci* **287**: 20192311.

Ruhland, C.T., Xiong, F.S., Clark, W.D., and Day, T.A. (2005) The influence of ultraviolet-B radiation on growth, hydroxycinnamic acids and flavonoids of *Deschampsia Antarctica* during springtime ozone depletion in Antarctica. *Photochem Photobiol* **81**: 1086–1093.

Schulz, E., Tohge, T., Zuther, E., Fernie, A.R., and Hincha, D. K. (2015) Natural variation in flavonol and anthocyanin metabolism during cold acclimation in *Arabidopsis thaliana* accessions. *Plant Cell Environ* **38**: 1658–1672.

Seehafer, K., Rominger, F., Helmchen, G., Langhans, M., Robinson, D.G., Özata, B., *et al*. (2013) Synthesis and biological properties of novel brefeldin a analogues. *J Med Chem* **56**: 5872–5884.

Shen, W., Le, S., Li, Y., and Hu, F. (2016) SeqKit: a cross-platform and ultrafast toolkit for FASTA/Q file manipulation. *PLoS One* **11**: e0163962.

Shrestha, G., and St Clair, L.L. (2014) Chapter 4 – Lichen phenolics: environmental effects. In *Polyphenols in Plants*, pp. 53–62. UT, USA: Brigham Young University, Provo.

Singh, G., Dal Grande, F., Divakar, P.K., Otte, J., Crespo, A., and Schmitt, I. (2017) Fungal-algal association patterns in lichen symbiosis linked to macroclimate. *New Phytol* **214**: 317–329.

Smith, J.M., and Haigh, J. (1974) The hitch-hiking effect of a favourable gene. *Genet Res* **23**: 23–35.

Solhaug, K.A., and Gauslaa, Y. (1996) Parietin, a photoprotective secondary product of the lichen *Xanthoria parietina*. *Oecologia* **108**: 412–418.

Solhaug, K.A., Lind, M., Nybakken, L., and Gauslaa, Y. (2009) Possible functional roles of cortical depsides and medullary depsidones in the foliose lichen *Hypogymnia physodes*. *Flora Morphol Distrib Funct Ecol Plants* **204**: 40–48.

Sullivan, M.J., Petty, N.K., and Beatson, S.A. (2011) Easyfig: A genome comparison visualizer. *Bioinformatics* **27**: 1009–1010.

Susca, A., Proctor, R.H., Morelli, M., Haidukowski, M., Gallo, A., Logrieco, A.F., and Moretti, A. (2016) Variation in fumonisin and ochratoxin production associated with differences in biosynthetic gene content in *Aspergillus Niger* and *A. welwitschiae* isolates from multiple crop and geographic origins. *Front Microbiol* **7**: 1664–302X.

Tobias, N.J., Wolff, H., Djahanschiri, B., Grundmann, F., Kronenwerth, M., Shi, Y-M., *et al*. (2017). Natural product diversity associated with the nematode symbionts Photorhabdus and Xenorhabdus. *Nat Microbiol* **2**: 1676–1685. http://doi.org/10.1038/s41564-017-0039-9.

van den Brule, T., Punt, M., Teertstra, W., Houbraken, J., Wösten, H., and Dijksterhuis, J. (2020) The most heat-resistant conidia observed to date are formed by distinct strains of *Paecilomyces variotii*. *Environ Microbiol* **22**: 986–999.

Vargas Castillo, R., and Beck, A. (2012) Photobiont selectivity and specificity in *Caloplaca* species in a fog-induced community in the Atacama Desert, northern Chile. *Fungal Biol* **116**: 665–676.

Villani, A., Proctor, R.H., Kim, H.S., Brown, D.W., Logrieco, A.F., Amatulli, M.T., *et al*. (2019) Variation in secondary metabolite production potential in the *Fusarium incarnatum-equiseti* species complex revealed by comparative analysis of 13 genomes. *BMC Genomics* **20**: 314.

Wang, Y., Zhang, X.Y., Zhou, Q., Zhang, X.L., and Wei, J. (2015) Comparative transcriptome analysis of the lichen-forming fungus *Endocarpon pusillum* elucidates its drought adaptation mechanisms. *Sci China Life Sci* **58**: 89–100.

Wang, M., Carver, J.J., Phelan, V.V., Sanchez, L.M., Garg, N., Peng, Y., *et al.* (2016). Sharing and community curation of mass spectrometry data with Global Natural Products Social Molecular Networking. *Nat Biotechnol* **34**: 828–837. http://doi.org/10.1038/nbt.3597.

Wiberg, R.A.W., Gaggiotti, O.E., Morrissey, M.B., and Ritchie, M.G. (2017) Identifying consistent allele frequency differences in studies of stratified populations. *Methods Ecol Evol* **8**: 1899–1909.

Yang, L., Wen, K.S., Ruan, X., Zhao, Y.X., Wei, F., and Wang, Q. (2018) Response of plant secondary metabolites to environmental factors. *Molecules* **23**: 762.

Zeng, G., Zhang, P., Zhang, Q., Zhao, H., Li, Z., Zhang, X., *et al*. (2018) Duplication of a PKS gene cluster and subsequent functional diversification facilitate environmental adaptation in *Metarhizium* species. *PLoS Genet* **14**: e1007472.

Supporting Information

Additional Supporting Information may be found in the online version of this article at the publisher's web-site:

Appendix S1. Details of materials and methods.

Appendix S2. Scripts used to predict synonymous and non-synonymous Single nucleotide polymorphisms.

Appendix S3. Scripts used to extract the coding (CDS) sequence for SNP from the co-ordinates files.

Appendix S4. Corrected Tajima's *D* (Kofler *et al*., 2011a) and nucleotide diversity π values in a region of 100 Kb up- and downstream of each BGC (highlighted in grey) and their mean across all populations. SNPs were identified based on a minimum count of 2, and a minimum covered fraction of 0.5 from pileup files subsampled to a uniform coverage of 30 (per population), excluding positions with a maximum coverage above the 2% of the empirical coverage distribution of each pool and filtered for indels, using the PoPoolation pipeline (Kofler, *et al*., 2011a).

Appendix S5. F_{ST} in sliding windows within BGCs and 100 Kb upstream and downstream. Pairwise F_{ST} in 5 Kb non-overlapping windows in a region of 100 Kb up- and downstream of each BGC (highlighted in grey) between all pairs of populations and their mean across all populations. SNPs were identified based on a minimum count of 2 from mpileup files subsampled to a uniform coverage of 30 (per population), excluding positions with a maximum coverage above the 2% of the empirical coverage distribution of each pool and filtered for indels, using the PoPoolation2 pipeline (Kofler, *et al*., 2011b).

Appendix S6. Voucher information of the samples used for HPLC and UV-HPLC

Appendix S7. Summary of Pool-seq assemblies. **A)** Summary of Pool-seq assemblies representing the Mediterranean and cold-temperate ecotype of *U*. *pustulata* from three elevational gradients. Each pool is derived from 100 individuals. IT1 = pool 1, Italy (Mediterranean ecotype), pool 6, Italy (cold temperate ecotype), ESii1 = Spain 2 (Gredos, Spain; Mediterranean ecotype), ESii6 = Spain 2 (Gredos, Spain; cold temperate ecotype), ESi1 = Spain 1 (Puerto de Pico. Spain; Mediterranean ecotype), ESi6 = Spain 1 (Talavera, Spain; cold temperate ecotype). **B)** Summary of Pool-seq assemblies representing the Mediterranean and cold temperate populations of *U*. *pustulata* from elevational gradient from Italy. Each pool is derived from 100 individuals. IT2 = pool 1, Italy (Mediterranean ecotype), IT3 = pool 3, Italy (Mediterranean ecotype), IT4 = pool 4, Italy (Mediterranean ecotype), pool 5, Italy (admixed population, individuals of both cold temperate and Mediterranean ecotype).

Appendix S8. Results of non-parametric Wilcoxon rank-sum tests comparing the distribution of corrected Tajima's *D*, nucleotide diversity π, and pairwise F_{ST} values in 5 Kb non-overlapping windows between BGCs and their corresponding 100 Kb surrounding regions. Significant P values (*$p < 0.01$, **$p < 0.001$) are highlighted in red if the comparison suggests hitchhiking effect on the genetic variation observed in the BGC (i.e. lower Tajima's *D*, lower π, higher F_{ST} in the surrounding regions), and in green if the comparison suggests intra-cluster selection.

Appendix S9. General pattern of polymorphism in 5 Kb non-overlapping windows in a region of 100 Kb up- and downstream of the cluster 13 (highlighted in grey; core biosynthetic PKS highlighted in green) based on mean F_{ST} among all pairs of populations (top), mean nucleotide diversity π across all populations (centre), and mean Tajima's *D* across all populations (bottom). Missing values: windows with insufficient coverage to compute the given index. The plot was generated using the *Visualize-output.pl* script from the PoPoolation pipeline (Kofler, *et al*., 2011a).

Appendix S10. HPLC results demonstrating the relative proportions of the natural compounds in *U*. *pustulata*. Yellow, green and blue colours represent gyrophoric acid, hiascic acid and lecanoric acid respectively.

Appendix S11. MS^2-based network analysis of selected samples of warm- and cold-adapted *U*. *pustulata* strains. A) Complete MS^2-based network analysis. Red diamonds: cold-adapted. Green circles: warm-adapted. Yellow squares: both. B) subnetwork containing O-methylated flavonols including jaceidin (**1**) and myricetin-3,7,3′,4′-tetramethylether (**2**). C) chemical structures of compounds (**1**) and (**2**).