High sorbic acid resistance of Penicillium roqueforti is mediated by the SORBUS gene cluster

Punt, M.; Seekles, S.J.; Dam, J.L. van; Adelhart Toorop, C. de; Martina, R.R.; Houbraken, J.; ... ; Butler Geraldine

PLOS GENETICS

High sorbic acid resistance of *Penicillium roqueforti* is mediated by the SORBUS gene cluster

Maarten Punt[1,2], **Sjoerd J. Seekles**[1,3], **Jisca L. van Dam**[3], **Connor de Adelhart Toorop**[2], **Raithel R. Martina**[2], **Jos Houbraken**[1,4], **Arthur F. J. Ram**[1,3], **Han A. B. Wösten**[1,2], **Robin A. Ohm**[1,2]*

1 TiFN, Wageningen, The Netherlands, 2 Microbiology, Department of Biology, Utrecht University, Utrecht, The Netherlands, 3 Department Molecular Microbiology and Biotechnology, Institute of Biology Leiden, Leiden University, Leiden, The Netherlands, 4 Westerdijk Fungal Biodiversity Institute, Utrecht, The Netherlands

* r.a.ohm@uu.nl

Abstract

Penicillium roqueforti is a major food-spoilage fungus known for its high resistance to the food preservative sorbic acid. Here, we demonstrate that the minimum inhibitory concentration of undissociated sorbic acid (MIC_u) ranges between 4.2 and 21.2 mM when 34 *P. roqueforti* strains were grown on malt extract broth. A genome-wide association study revealed that the six most resistant strains contained the 180 kbp gene cluster SORBUS, which was absent in the other 28 strains. In addition, a SNP analysis revealed five genes outside the SORBUS cluster that may be linked to sorbic acid resistance. A partial SORBUS knock-out (>100 of 180 kbp) in a resistant strain reduced sorbic acid resistance to similar levels as observed in the sensitive strains. Whole genome transcriptome analysis revealed a small set of genes present in both resistant and sensitive *P. roqueforti* strains that were differentially expressed in the presence of the weak acid. These genes could explain why *P. roqueforti* is more resistant to sorbic acid when compared to other fungi, even in the absence of the SORBUS cluster. Together, the MIC_u of 21.2 mM makes *P. roqueforti* among the most sorbic acid-resistant fungi, if not the most resistant fungus, which is mediated by the SORBUS gene cluster.

Author summary

Chemical preservatives, such as sorbic acid, are often used in food to prevent spoilage by fungi, yet some fungi are particularly well-suited to deal with these preservatives. First, we investigated the resistance of 34 *Penicillium roqueforti* strains to various food preservatives. This revealed that some strains were highly resistant to sorbic acid, while others are more sensitive. Next, we used DNA sequencing to compare the genetic variation between these strains and discovered a specific genetic region (SORBUS) that is unique to the resistant strains. Through comparative analysis with other fungal species the SORBUS region was studied in more detail and with the use of genetic engineering tools we removed this

unique region. Finally, the mutant lacking the SORBUS region was confirmed to have lost its sorbic acid resistance. This finding is of particular interest as it suggests that only some, not all, *P. roqueforti* strains are potent spoilers and that specific genetic markers could help in the identification of resistant strains.

Introduction

Fungi are responsible for 5–10% of all food spoilage [1] resulting in the production of off-flavors, discoloration and acidification [1,2]. Some species also produce mycotoxins, such as Penicillin Roquefort Toxin and roquefortine C in the case of *Penicillium roqueforti* [3]. Toxins partially contribute to the high incidence of food-borne diseases affecting up to 30% of the people in industrialized countries every year [4].

Food spoilage by filamentous fungi is believed to be mainly caused by spores that are spread through the air, water or other vectors like insects [5]. Preservation techniques such as pasteurization, fermentation, cooling or the addition of preservatives are used to reduce spoilage [6]. Some of the most applied preservatives are weak organic acids such as benzoic, propionic and sorbic acid. *Paecilomyces variotii*, *Penicillium paneum*, *Penicillium carneum* and *P. roqueforti* are among the few filamentous fungi capable of spoiling products containing weak acids, and are therefore called preservative-resistant moulds [7]. Weak-acid preservatives inhibit microbial growth, but their mode-of-action is not completely understood. According to the classical 'weak-acid preservative theory' the antimicrobial activity of weak acids is derived from their undissociated form that can pass the plasma membrane. These weak acids dissociate in the cytosol due to its neutral pH, and inhibit growth through acidification of the cytoplasm [8]. The inhibitory activity of sorbic acid at pH 6.5 and the correlation of sorbic acid resistance with ethanol tolerance in *Saccharomyces cerevisiae* suggest that this weak acid can also act as a membrane-active compound [8].

In *Aspergillus niger*, sorbic acid resistance is mediated by the phenylacrylic acid decarboxylase gene *padA*, and the putative 4-hydroxybenzoate decarboxylase gene, known as the cinnamic acid decarboxylase (*cdcA*) gene [9,10]. These genes encode proteins that catalyze the conversion of sorbic acid into 1,3-pentadiene. Genes *padA* and *cdcA* are regulated by the sorbic acid decarboxylase regulator (SdrA), which is a Zn2Cys6-finger transcription factor [10]. These three genes are present on the same genetic locus in *A. niger* with *sdrA* being flanked by *cdcA* and *padA*. Orthologs of *cdcA* and *padA* are also clustered in *S. cerevisiae*, but this is not the case for the *sdrA* homologue [10,11]. Inactivation of *cdcA*, *padA* or *sdrA* results in reduced growth of *A. niger* on sorbic acid and cinnamic acid [9]. The fact that growth is not abolished and the fact that *padA* transcript levels are less affected than that of *cdcA* on sorbic acid upon deletion of *sdrA* suggests that an additional regulator is involved [9]. Indeed, the weak-acid regulator WarA is also involved in sorbic acid resistance as well as other weak acids such as propionic and benzoic acid [12].

Weak-acid resistance in fungi is subject to intra- and inter-strain heterogeneity. Intra-strain heterogeneity has been observed in *Zygosaccharomyces bailii* with the existence of a small sub-population of cells that are more resistant (MIC = 7.6 mM) to sorbic acid than the sensitive population (MIC = 3 mM) [13], while inter-strain resistance has been observed in *P. roqueforti* with the existence of sorbic acid-resistant and sorbic acid-sensitive strains [14]. Genome analysis of *P. roqueforti* strains have yielded clues about adaptive divergence in this species. Genomic islands with high identity are present in distant *Penicillium* species, while they are absent in closely related species, supporting the hypothesis of recent horizontal gene transfer events

[15,16]. For instance, the presence of two large genomic regions, *Wallaby* and *CheesyTer* correlates with faster growth and functions relevant in a cheese matrix, respectively [16,17]. Noteworthy, the *CheesyTer* and *Wallaby* regions are only present in cheese isolates other than Roquefort, indicating that the Roquefort isolate population is potentially more closely related to the 'ancestral' *P. roqueforti* populations [18].

In this study, sorbic acid resistance of 34 *P. roqueforti* strains was assessed. A genome-wide association analysis revealed the presence of the 180 kbp gene cluster SORBUS in the six most resistant strains. A partial SORBUS knockout in such a strain showed a reduced sorbic acid resistance similar to that of the other 28 sensitive *P. roqueforti* strains.

Results

Weak-acid sensitivity screening

Weak-acid sensitivity of 34 *P. roqueforti* wild-type strains was assessed on MEA plates supplemented with 5 mM propionic, sorbic or benzoic acid, which corresponds to 4.42, 4.25 and 3.07 mM undissociated acid, respectively. Three strains had been isolated from blue-veined cheeses such as Roquefort and the other 31 strains had been isolated from non-cheese environments (mostly related to spoiled food) (S1 Table). The colony surface area was determined after five days of growth (Fig 1). The inhibitory effect of propionic acid at the tested concentration was limited for most strains, since 26 out of 34 strains grew to > 80% of the colony surface area reached under control conditions. Strains DTO012A1 and DTO012A8 even showed an increased colony size (up to 120%) when compared to the control. The inhibitory effects of sorbic and benzoic acid were more pronounced. MEA supplemented with potassium sorbate reduced colony area for all 34 *P. roqueforti* strains. The surface area under sorbic acid stress ranged from 0 to 80% of the surface area reached under control conditions. Strains DTO006G1, DTO006G7, DTO013E5 and DTO013F2 showed the highest sorbic acid resistance, followed by DTO046C5. Benzoic acid was the most inhibitory compound, resulting in a
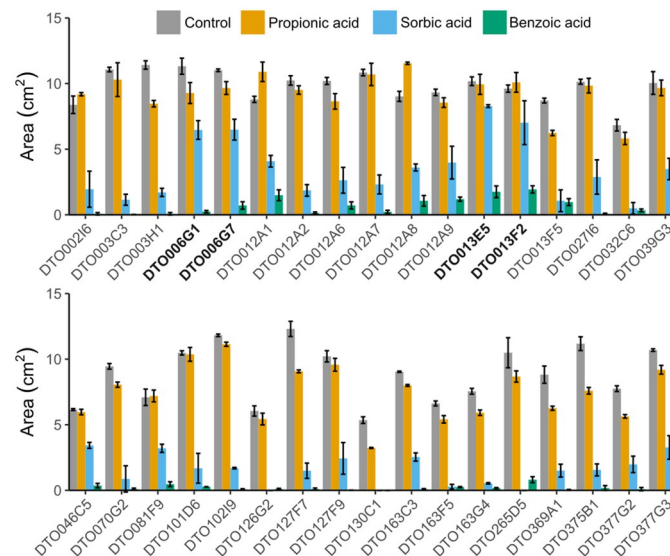
Fig 1. Weak-acid screening shows variable resistances to multiple acids. Average colony size (cm^2) of 34 *P. roqueforti* strains after five days of growth on MEA (pH 4.0) (grey) or MEA (pH 4.0) supplemented with 5 mM propionic acid (orange), sorbic acid (blue) or benzoic acid (green). Strains DTO006G1, DTO006G7, DTO013E5 and DTO013F2 (indicated in bold) are relatively resistant to sorbic acid. Error bars indicate standard deviation of biologically independent replicates.

https://doi.org/10.1371/journal.pgen.1010086.g001

maximum colony surface area between 0 and 20% of the control. The most benzoic acid-resistant strains were DTO013F2 and DTO013E5. As these strains were also among the most sorbic acid-resistant strains, similar resistance mechanisms may be involved to cope with benzoic and sorbic acid stress.

Sorbic acid resistance was further analyzed by determining the MIC_u values of sorbic acid for the 34 *P. roqueforti* strains. The four strains with the highest sorbic acid resistance on MEA (Fig 1), were also among the strains (DTO006G1, DTO006G7, DTO012A1, DTO012A8, DTO013E5 and DTO013F2) that showed the highest MIC_u (Fig 2). Although strains DTO012A1 and DTO012A8 did not exhibit increased colony growth on agar (Fig 1), they were among the resistant strains. This might be explained by the limited growth period (5 days) or their overall relative low growth rate compared to the other resistant strains (See controls Fig 1; DTO006G1, DTO006G7, DTO013E5 and DTO013F2). DTO013E5 and DTO013F2 were the most resistant and even showed growth at the highest tested undissociated sorbate concentration of 21.2 mM, indicating a $MIC_u > 21.2$ mM. The other strains showed a distinctly lower MIC_u, ranging between 4.2 mM and 9.95 mM. Only strain DTO012A9 showed an intermediate resistance to sorbic acid with an average MIC_u of 13.72 mM undissociated sorbic acid.

Genome statistics and phylogeny

The genomes of the 34 *P. roqueforti* strains were sequenced. Scaffold count varied between 45 and 1358, assembly length between 26.53 and 31.74 Mb and GC content between 46.85 and 48.44% (Table 1). The number of predicted genes varied between 9633 and 10644, the number of genes with PFAM domains between 73.11 and 75.38%, and the number of secondary metabolism gene clusters between 32 and 36. All strains had a BUSCO completeness of >99%, indicating high quality assembly and gene predictions, except for DTO012A8 with a completeness of 94.83%. This and the high scaffold count of the DTO012A8 assembly (1358 scaffolds) indicates that its genome assembly is not complete. The strain was kept in the downstream analysis as most other metrics did not differ much compared to the other strains (Table 1). Fig 3 presents a phylogenetic tree of the 34 *P. roqueforti* strains based on 6923 single-copy orthologous genes. Three of the six sorbic acid-resistant strains (DTO012A2, DTO013F2, DTO013E5 and DTO012A8) are similar with DTO012A8 being closely related to those strains. The two other resistant strains are also similar and more distantly related.
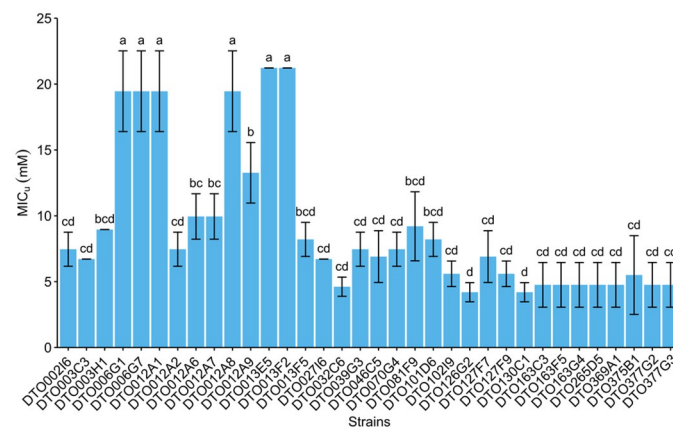

Fig 2. Sorbic acid resistance screening reveals 6 resistant *P. roqueforti* strains. Average undissociated MIC_u values of sorbic acid of 34 *P. roqueforti* strains (mM ± standard deviation). Each bar graph represents biological triplicates. Error bars indicate standard deviation and letters indicate significant difference in MIC_u (p < 0.05).

Table 1. Genome assembly and annotation statistics. The number of scaffolds, assembly length, GC content and genes of the 34 sequenced *P. roqueforti* strains are listed. In addition, the number of genes with a PFAM domain, the number of secondary metabolism gene clusters and the BUSCO completeness are listed. The type column indicates if the strain is sorbic acid-resistant (R) or sorbic acid-sensitive (S).

Strain	Assembly			Annotation			Annotation quality	Type
	Scaffolds	Total assembly length (Mbp)	Assembly GC content (%)	Genes	Genes with PFAM (total, %)	Secondary metabolism gene cluster	BUSCO2 completeness (%)	R/S
DTO002I6	134	29.48	47.75	10135	7539 (74.39%)	33	99.66%	S
DTO003C3	353	30.28	47.82	10362	7612 (73.46%)	36	99.66%	S
DTO003H1	382	30.29	47.82	10349	7607 (73.5%)	36	99.66%	S
DTO006G1	109	27.97	48.18	9975	7468 (74.87%)	33	99.66%	R
DTO006G7	100	27.97	48.18	9982	7471 (74.84%)	33	99.66%	R
DTO012A1	158	29.37	48.11	10280	7557 (73.51%)	33	100%	R
DTO012A2	73	27.07	48.26	9786	7362 (75.23%)	33	99.66%	S
DTO012A6	93	27.29	48.32	9817	7376 (75.13%)	33	99.66%	S
DTO012A7	62	27.33	48.07	9748	7335 (75.25%)	34	99.66%	S
DTO012A8	1358	29.6	47.88	10253	7483 (72.98%)	32	94.83%	R
DTO012A9	581	29.84	47.89	10341	7613 (73.62%)	34	99.31%	S
DTO013E5	165	29.27	48.11	10252	7537 (73.52%)	33	100%	R
DTO013F2	657	31.74	47.49	10644	7739 (72.71%)	33	100%	R
DTO013F5	304	29.56	48.04	10313	7591 (73.61%)	35	99.66%	S
DTO027I6	177	29.3	48.12	10325	7560 (73.22%)	32	99.66%	S
DTO032C6	141	28.45	48.13	10008	7441 (74.35%)	33	100%	S
DTO039G3	45	27.22	48.14	9764	7352 (75.3%)	32	99.66%	S
DTO046C5	143	28.8	48.15	10098	7488 (74.15%)	33	99.66%	S
DTO070G2	60	26.75	48.39	9724	7324 (75.32%)	33	100%	S
DTO081F9	47	27.24	48.13	9760	7349 (75.3%)	32	99.66%	S
DTO101D6	141	28.96	48.07	10109	7497 (74.16%)	33	99.31%	S
DTO102I9	186	28.53	47.92	9978	7439 (74.55%)	32	99.66%	S
DTO126G2	174	28.06	48.02	9939	7431 (74.77%)	33	99.31%	S
DTO127F7	59	26.54	48.43	9639	7315 (75.89%)	35	99.66%	S
DTO127F9	54	26.53	48.44	9633	7311 (75.9%)	36	99.66%	S
DTO130C1	537	30.48	47.79	10375	7612 (73.37%)	36	99.66%	S
DTO163C3	239	30	48.16	10485	7653 (72.99%)	33	99.66%	S
DTO163F5	284	30.9	46.85	10010	7441 (74.34%)	33	100%	S
DTO163G4	104	28.45	48.12	10006	7441 (74.37%)	33	100%	S
DTO265D5	66	26.76	48.44	9751	7350 (75.38%)	33	100%	S
DTO369A1	156	28.29	47.97	9945	7437 (74.78%)	32	99.66%	S
DTO375B1	62	27.2	48.2	9804	7378 (75.25%)	33	99.31%	S
DTO377G2	101	28.07	48.14	9962	7444 (74.72%)	33	99.31%	S
DTO377G3	84	26.97	48.24	9762	7335 (75.14%)	33	99.66%	S
Average	217.32	28.55	48.05	10038.65				
Min	45	26.53	46.85	9633				
Max	1358	31.74	48.44	10644				

https://doi.org/10.1371/journal.pgen.1010086.t001

Sorbic acid resistance correlates with a genomic cluster containing genes regulating sorbic acid decarboxylation

Two complementary genome-wide association study (GWAS) approaches were taken to identify the genetic elements (e.g., genomic regions, genes, SNPs, etc) associated with sorbic acid resistance. The first approach is based on the presence of large genomic regions that are only
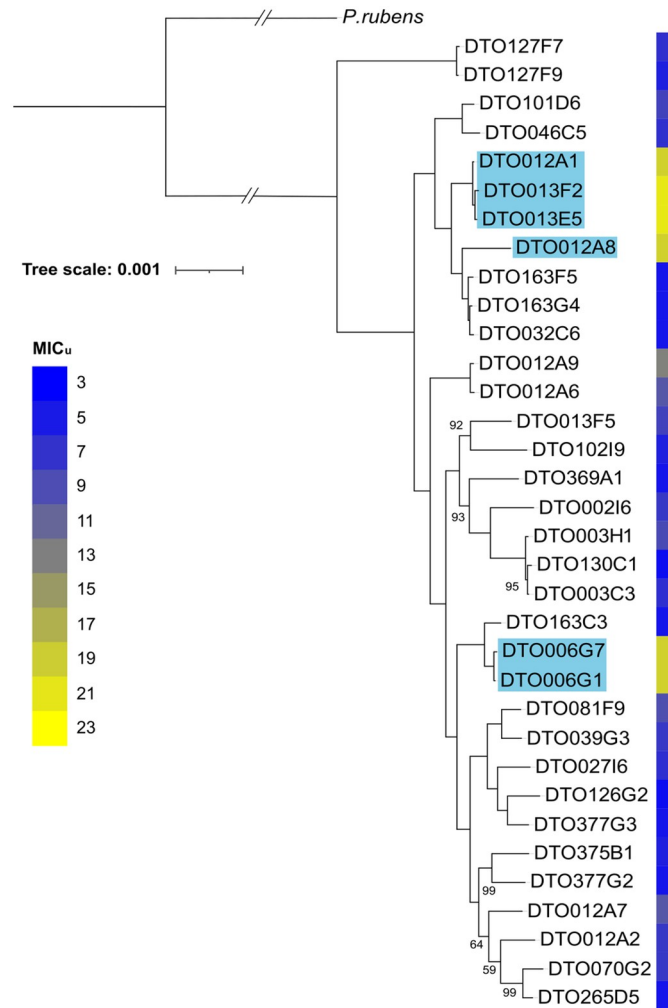
Fig 3. Phylogeny of the *P. roqueforti* strains. Phylogenetic tree of the 34 *P. roqueforti* strains used in this study. Sorbic acid resistance (MIC$_u$) is indicated in blue-yellow shading. The tree is based on 6923 single-copy orthologous genes and was constructed using RAxML. *P. rubens* [44] was used as outgroup. Bootstrap values <100 are indicated. Strains containing the SORBUS cluster are highlighted in blue.

https://doi.org/10.1371/journal.pgen.1010086.g003

found in the sorbic acid resistant strains, and the second approach is based on SNPs that are significantly correlated with the level of sorbic acid resistance.

In the first GWAS approach, the 34 *P. roqueforti* strains were divided into two groups based on their sorbic acid resistance, a group of six resistant (R-type) strains ('a', Fig 2) and a group of 28 sensitive (S-type) strains ('b-d', Fig 2). The assemblies of all strains were aligned to the assembly of DTO006G7 using MUMmer, and subsequently genomic regions (and genes encoded on those regions) were identified that were unique to the R-type strains. With this method 57 genes were identified, of which 51 were present on scaffold 43 of DTO006G7 (Table 2), and the six genes outside of scaffold 43 are g6241, g8103, g9940, g9945 and g9946. In addition to the 51 unique genes in this scaffold, it contains 19 genes which are also found completely or in part in some of the S-type strains. The genomic alignment shows the genes on scaffold 43, which is present in the R-type strains (Fig 4). The first 80 kbp of scaffold 43 (protein IDs g12000-g12029) mainly contains hypothetical proteins without predicted function, while the remaining region between 80–180 kbp (g12030–g12069) contains multiple regions

Table 2. Genes located on the SORBUS cluster in strain DTO006G7. Fold change (log₂FC) of the sorbic acid samples compared to the control is given. Underlined genes are significantly differentially expressed (adjusted p-value < 0.05) and the mean expression (FPKM) of three biological replicates is given per condition (control and sorbic acid). Rows highlighted in grey indicate genes that are not unique for the R-type strains.

GeneId	Name	Functional annotation (PFAM or description)	Expression (FPKM)		log₂FC	p-value (Adj.)
			Control	Sorbic acid		
g12000		FAM167	4	8	0.69	0.66
g12001		hypothetical protein	16	86	2.18	0.00
g12002		NEMP	2	2	-0.27	0.89
g12003		BTB/POZ domain	16	23	0.48	0.35
g12004		hypothetical protein	59	40	-0.56	0.16
g12005		hypothetical protein	0	0	1.05	0.59
g12006		Reverse transcriptase	0	0	-	-
g12007		DUF3723	0	0	-	-
g12008		hypothetical protein	0	0	-	-
g12009		Hly-III related protein	0	0	-	-
g12010		Cyclin like F-box	1	1	0.72	0.32
g12011		Cyclin like F-box	17	24	0.47	0.25
g12012		Reverse transcriptase	3	7	1.02	0.03
g12013		Endonuclease/Exonuclease/phosphatase family	1	4	1.40	0.00
g12014		Probable transposable element	1	3	0.83	0.51
g12015		Aldo/keto reductase family	463	365	-0.34	0.36
g12016		Reverse transcriptase	26	20	-0.35	0.57
g12017		Telomere-associated recq-like helicase	13	15	0.23	0.57
g12018		Cyclin like F-box	29	25	-0.23	0.41
g12019		hypothetical protein	0	0	-1.04	0.83
g12020		hypothetical protein	1	1	0.22	0.96
g12021		Centrosomin N-terminal motif 1	0	0	2.51	0.24
g12022		hypothetical protein	3	7	1.02	0.09
g12023		Cyctochrome C mitochondrial import factor	1	0	-0.84	0.73
g12024		hypothetical protein	8	10	0.21	0.86
g12025		hypothetical protein	5	6	0.25	0.60
g12026		Winged helix-turn helix	0	0	-	-
g12027		hypothetical protein	15	10	-0.57	0.23
g12028		Pronucleotidyl transferase	0	0	-0.14	0.95
g12029		Pronucleotidyl transferase	1	0	-0.83	0.64
g12030		hypothetical protein	11	15	0.42	0.58
g12031		hypothetical protein	0	0	-	-
g12032	*padB*	Flavoprotein	450	288	-0.62	0.15
g12033		PHF5-like protein	3	3	0.11	0.91
g12034		Pyridoxamine 5'-phosphate oxidase	370	287	-0.36	0.48
g12035		Mitochondrial carrier protein	11	12	0.06	0.96
g12036		Mitochondrial carrier protein	12	17	0.42	0.65
g12037		GTP cyclohydrolase II	10	2	-1.86	0.01
g12038		Potassium channel tetramerisation domain	26	30	0.18	0.82
g12039		hypothetical protein	17	10	-0.58	0.58
g12040	*cdcC*	UbiD	134	65	-1.02	0.01
g12041		Pyridoxamine 5'-phosphate oxidase	2	1	-0.54	0.90
g12042		hypothetical protein	9	11	0.27	0.76
g12043		GTP cyclohydrolase II	5	1	-1.63	0.03

(Continued)

Table 2. (Continued)

GeneId	Name	Functional annotation (PFAM or description)	Expression (FPKM) Control	Sorbic acid	log₂FC	p-value (Adj.)
g12044		GPR1/FUN34/yaaH family	47	9	-2.05	0.00
g12045		Mitochondrial carrier protein	1	1	-0.61	0.73
g12046		hypothetical protein	12	12	-0.03	0.97
g12047		Helix-turn-helix domain/endonuclease	8	8	0.03	0.95
g12048		Flavonol reductase/cinnamoyl-CoA reductase	53	56	0.07	0.93
g12049		Flavonol reductase/cinnamoyl-CoA reductase	9	8	-0.26	0.63
g12050		Tannase	427	352	-0.27	0.55
g12051		hypothetical protein	0	1	1.76	
g12052		Transposase-like protein	6	7	0.13	0.84
g12053		Reverse transcriptase	0	0	0.31	0.89
g12054		Zinc finger transcription factor	26	13	-0.98	0.07
g12055		Transposase-like protein	0	0	1.64	
g12056	*cdcB*	UbiD	211	142	-0.51	0.65
g12057	*padC*	Flavoprotein	78	75	-0.07	0.90
g12058		Pyridoxamine 5'-phosphate oxidase like	348	504	0.50	0.28
g12059		GTP cyclohydrolase II	25	24	-0.04	0.96
g12060		GPR1/FUN34/yaaH family	3	3	-0.39	0.67
g12061		Flavin reductase like domain	2028	1532	-0.40	0.42
g12062		hypothetical protein	272	295	0.08	0.86
g12063		3, 4-dihydroxy-2-butanone 4-phosphate synthase	65	34	-0.85	0.25
g12064	*padA*	Flavoprotein	92	45	-0.97	0.08
g12065	*cdcA*	UbiD	65	22	-1.53	0.00
g12066	*sdrA*	Hypothetical transcription factor	2	1	-0.76	0.71
g12067		hypothetical protein	0	0		
g12068		NACHT domain	4	8	0.85	0.09
g12069		Histone H3	3	6	0.95	0.00

homologous to genes previously reported as related to weak-acid resistance in *A. niger*. Predicted genes orthologous (based on bidirectional best BLAST hit) to *padA*, *cdcA* and *sdrA* of *A. niger* were found alongside each other (g12064-g12066) with respective identities of 87%, 83% and 53%. Moreover, two *padA* paralogs with high BLAST similarities to *padA* of *A. niger* (63% and 58%) were identified on the R-type-specific cluster and named *padB* (g12032) and *padC* (g12057). Similarly, two paralogs of *cdcA*, named *cdcB* (g12056) and *cdcC* (g12040), were identified on the same gene cluster as well, with identities of 72% and 71%, respectively, when compared to *cdcA* of *A. niger*. An additional *cdcA* paralog, *cdcD* (g2591), is not located on this cluster and is also present in S-type strains. In contrast, no homologs of *A. niger sdrA* and *padA* were found outside of the cluster. In addition, a transcription factor (g2820 in DTO006G7) orthologous (based on a bidirectional best BLAST hit) to *warA* of *A. niger* was identified outside of the cluster in the genomes of all strains. These results indicate that the R-type strains contain a gene cluster similar to the sorbic acid resistance gene cluster described in *A. niger*, but considerably expanded [9,10]. For further reference, we name this cluster (i.e., scaffold 43 of strain DTO006G7) SORBUS after the tree *Sorbus aucuparia*, as sorbic acid has been first isolated from its berries by August Hoffman [19,20]. While SORBUS as a whole is only present in the R-type strains (DTO006G1, DTO006G7, DTO012A2, DTO012A8, DTO013F2, DTO013E5), some S-type strains share up to 5 kbp parts of the sequence,
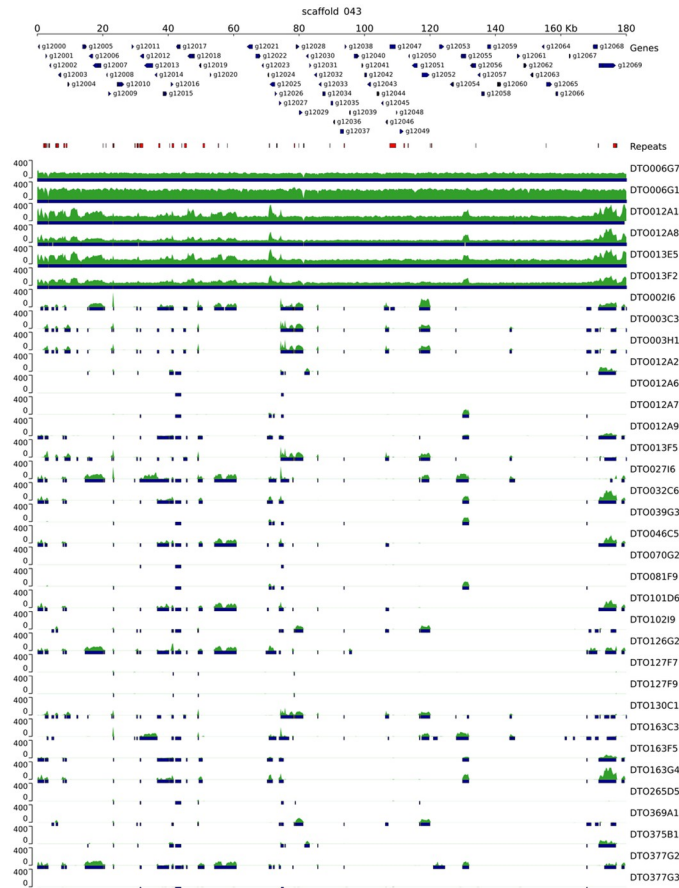
Fig 4. Genome comparison reveals unique gene cluster in R-type strains. Genomic alignment of 33 *P. roqueforti* strains to the SORBUS cluster (scaffold 43 of DTO006G7). Predicted genes are indicated with arrows, repetitive DNA is indicated in red, the sequence read coverage is indicated in the green tracks, and blue bars indicate (partial) overlap with DTO006G7. The complete SORBUS cluster is only present in the R-type strains (DTO006G1, DTO006G7, DTO012A2, DTO012A8, DTO013F2 and DTO013E5).

especially in the first 80 kbp of the cluster (Fig 4). The conservation of SORBUS in the R-type strains was visualized with Clinker, demonstrating that the genes of the SORBUS cluster are well conserved and highly syntenic in the R-type strains (S1 Fig). It should be noted that the R-type strains were all isolated from non-cheese environments. Based on alignments of the sequencing reads to DTO006G7 we confirmed that out of 35 previously sequenced *P. roqueforti* strains [18], none of the 17 cheese strains contained the SORBUS cluster, while two out of the 18 non-cheese strains contained the SORBUS cluster (S1 Table).

In the second GWAS approach, PLINK [21] was used to identify which SNPs correlate with sorbic acid resistance. This method allowed for the quantitative use of $\log_{10}(MIC_u)$ values as input for analysis, as opposed to the first approach described above. The correlation between the presence of SNPs and the sorbic acid resistance is visualized in a Manhattan plot (Fig 5). SNPs located on genes with a $-\log_{10}(P) > 5$ and either a high or moderate impact (SNPeff) were selected (Table 3). This resulted in 338 SNPs in 41 genes. Out of these SNPs, 29 had a 'high' impact according to SNPeff and were located in 17 genes. Only six out of these 17 genes with high impact variants (g7017, g8100, g8106, g9942, g9943, g9976) were not located in the SORBUS cluster, the other 11 genes were either among the non-unique genes present on SOR-BUS, or genes of which less than 90% of the sequence was found in S-type strains. Functional
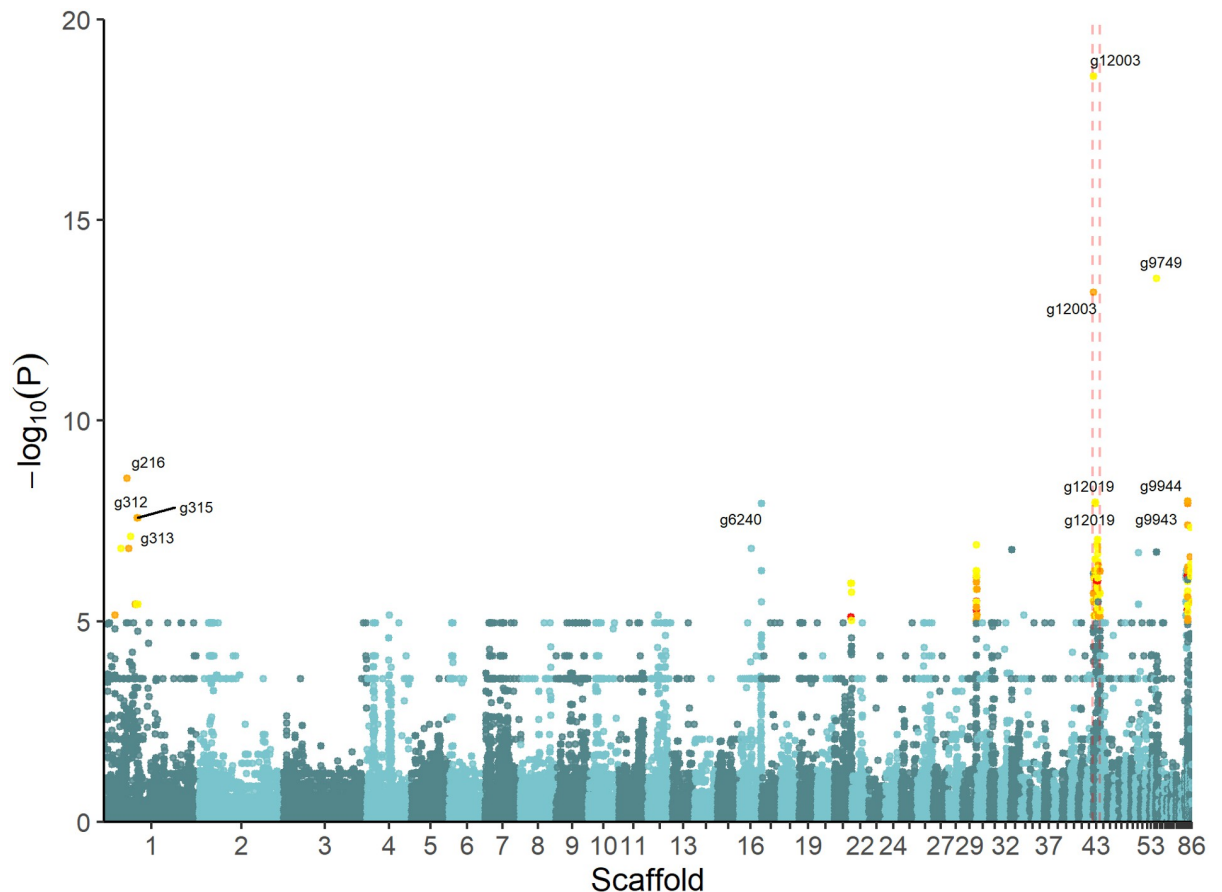
Fig 5. Manhattan plot shows SNPs associated with sorbic acid resistance. Scaffolds are listed on the x-axis, while the y-axis display the significance of the association ($-\log_{10}$(p-value)). Yellow, orange and red dots indicate 'low', 'moderate' or 'high' impact SNPs as determined by SNPeff, respectively. The GeneIDs associated with the SNPs with a $-\log_{10}$(p-value) > 7.5 are indicated. The SORBUS cluster is located between the dashed lines.

annotation revealed that protein g8100 contains an ankyrin-repeat domain, while protein g9943 is homologous to a zinc finger C3H1-type domain-containing protein. In contrast, the four remaining proteins had no functional annotations. In all cases, these six high impact variant-containing genes showed high similarity (> 99%) to genes in other *Penicillium* species. SNPs in two additional genes encoding a putative transmembrane transporter (g216) and cation transporter (g296) also correlated with increased resistance to sorbic acid.

To investigate the evolutionary origin of the SORBUS cluster, the presence of five PFAM domains (from g12060, g12061 and g12063-g12065) that are present on the SORBUS cluster was analysed in 32 Aspergilli and Penicillia as well as the 34 *P. roqueforti* strains (Table 4). These PFAM domains encode a putative GPR1/FUN34/yaaH family (g12060), a flavin reductase like domain (g12061), 3, 4-dihydroxy-2-butanone 4-phosphate synthase (g12063), a flavoprotein (g12064, *padA*) and a UbiD domain (g12065, *cdcA*). These domains are selected as they are clustered and because of their predicted role. The first domain has been associated with acetic acid sensitivity in *S. cerevisiae* [22], while the latter four domains are part of the gene cluster described in *A. niger* [9]. The SORBUS genes g12061, g12064 and g12065 are more similar to genes of several *Aspergillus* species, as determined by a gene tree approach (Table 4), whereas the PFAM domains from g12060 and g12063 did not cluster with any of the

Table 3. Genes (DTO006G7) containing SNPs associated with sorbic acid resistance. Only SNPs with a $-\log_{10}$(p-value) > 5 and a moderate or high impact as determined by SNPeff are listed. Grey shading indicates overlap with sequence repeats.

GeneID	SNP impact		Effect	PFAM annotation
	Moderate	High		
g103	1	0		hypothetical protein
g216	1	0		ABC transporter transmembrane region
g235	1	0		STAG domain
g296	5	0		Cation transporter/ATPase, N-terminus
g312	1	0		Probable molybdopterin binding domain
g313	1	0		DDHD domain
g314	1	0		hypothetical protein
g315	1	0		FAD binding domain
g7015	1	1	Stop lost & splice region variant	hypothetical protein
g8100	0	1	Stop gained	Ankyrin repeats (3 copies)
g8101	3	0		DDE superfamily endonuclease
g8104	7	0		hypothetical protein
g8105	10	0		hypothetical protein
g8106	3	1	Stop gained	hypothetical protein
g8107	26	0		Ankyrin repeats (3 copies)
g12002	2	0		NEMP
g12005	9	1	Frameshift variant	hypothetical protein
g12006	42	2	Frameshift variant & Stop gained	Reverse transcriptase
g12007	8	0		Protein of unknown function (DUF3723)
g12013	48	3	Frameshift variants	Endonuclease/Exonuclease/phosphatase family
g12014	8	1	Stop lost & splice region variant	Probable transposable element
g12016	2	0		Reverse transcriptase
g12019	11	0		hypothetical protein
g12025	9	1	Stop gained	hypothetical protein
g12028	18	1	Stop lost & splice region variant	Pronucleotidyl transferase
g12029	26	6	Stop gained, Stop lost & splice variant	Pronucleotidyl transferase
g12030	16	2	Frameshift variant & Stop lost	hypothetical protein
g12046	1	0		hypothetical protein
g12052	4	1	Stop gained	Transposase-like protein
g12053	1	0		Reverse transcriptase
g12054	4	0		hypothetical protein
g12055	6	2	Frameshift variant & Stop lost	Transposase-like protein
g12069	7	1	Frameshift variant	Histone H3
g9749	7	0		hypothetical protein
g9942	1	2	Stop gained	hypothetical protein
g9943	2	2	Stop gained	AAA domain
g9944	8	0		hypothetical protein
g9945	4	0		ATPase family associated with various cellular activities (AAA)
g9961	1	0		hypothetical protein
g9966	2	0		Endonuclease-reverse transcriptase
g9976	0	1	Start lost	hypothetical protein
TOTAL	309	29		

https://doi.org/10.1371/journal.pgen.1010086.t003

Table 4. Number of genes containing PFAM domains corresponding to g12060, g12061 and g12063-g12065 (PF01184, PF01613, PF00926, PF02441, PF01977) based on phylogenetic trees constructed with their respective PFAM domains. Top six strains are R-type *P. roqueforti* strains containing the SORBUS cluster. C (CORE) indicates if the domains aligned closely to the PFAMs not unique for the SORBUS cluster or did not align to *P. roqueforti*, S (SORBUS) indicates the number of PFAM domains which aligned closely to PFAMs originated from SORBUS.

PFAM family	g12060 PF01184		g12061 PF01613		g12063 PF00926		g12064 PF02441 (*padA*)		g12065 PF01977 (*cdcA*)	
	Gpr1_Fun34 YaaH		Flavin_Reduct		3,4-dihydroxy-2-butanone 4-phosphate synthase		Flavoprotein		UbiD domain	
Strain	C	S	C	S	C	S	C	S	C	S
DTO013F2	7	1	4	1	1	1	5	3	2	3
DTO013E5	7	1	3	1	1	1	5	3	2	3
DTO012A8	7	1	3	1	1	1	5	3	2	3
DTO012A1	7	1	3	1	1	1	5	3	2	3
DTO006G7	7	1	4	1	1	1	5	3	2	3
DTO006G1	7	1	4	1	1	1	5	3	2	3
DTO377G3	7	0	4	0	1	0	5	0	2	0
DTO377G2	7	0	4	0	1	0	5	0	2	0
DTO375B1	7	0	3	0	1	0	5	0	2	0
DTO369A1	7	0	4	0	1	0	5	0	2	0
DTO265D5	7	0	3	0	1	0	5	0	2	0
DTO163G4	7	0	3	0	1	0	5	0	2	0
DTO163F5	7	0	3	0	1	0	5	0	2	0
DTO163C3	7	0	4	0	1	0	5	0	2	0
DTO130C1	7	0	3	0	1	0	5	0	2	0
DTO127F9	7	0	3	0	1	0	5	0	2	0
DTO127F7	7	0	3	0	1	0	5	0	2	0
DTO126G2	7	0	4	0	1	0	5	0	2	0
DTO102I9	7	0	4	0	1	0	5	0	2	0
DTO101D6	7	0	3	0	1	0	5	0	2	0
DTO081F9	7	0	4	0	1	0	5	0	2	0
DTO070G2	7	0	3	0	1	0	5	0	2	0
DTO046C5	7	0	3	0	1	0	5	0	2	0
DTO039G3	7	0	3	0	1	0	5	0	2	0
DTO032C6	7	0	3	0	1	0	5	0	2	0
DTO027I6	7	0	4	0	1	0	5	0	2	0
DTO013F5	7	0	4	0	1	0	5	0	2	0
DTO012A9	7	0	3	0	1	0	5	0	2	0
DTO012A7	7	0	3	0	1	0	5	0	2	0
DTO012A6	7	0	3	0	1	0	5	0	2	0
DTO012A2	7	0	3	0	1	0	5	0	2	0
DTO003H1	7	0	3	0	1	0	5	0	2	0
DTO003C3	7	0	3	0	1	0	5	0	2	0
DTO002I6	7	0	3	0	1	0	5	0	2	0
Penicillium roqueforti FM164	8	0	3	0	1	0	5	0	2	0
Penicillium oxalicum	4	0	2	0	1	0	2	0	0	0
Penicillium digitatum	1	0	2	0	1	0	4	0	1	0
Paecilomyces variotii	8	0	6	0	1	0	4	1	2	0
Paecilomyces variotii DTO217A2	7	0	7	0	1	0	5	1	3	0
Paecilomyces niveus	9	0	8	0	1	0	5	1	2	0
Aspergillus wentii	4	0	5	0	1	0	4	1	2	0
Aspergillus violaceofuscus	4	0	5	1	1	0	4	0	2	0
Aspergillus vadensis	3	0	5	0	1	0	4	1	2	1

(*Continued*)

Table 4. (Continued)

PFAM family	g12060 PF01184		g12061 PF01613		g12063 PF00926		g12064 PF02441 (*padA*)		g12065 PF01977 (*cdcA*)	
	Gpr1_Fun34 YaaH		Flavin_Reduct		3,4-dihydroxy-2-butanone 4-phosphate synthase		Flavoprotein		UbiD domain	
Strain	**C**	**S**	**C**	**S**	**C**	**S**	**C**	**S**	**C**	**S**
Aspergillus uvarum	4	0	5	1	1	0	3	0	2	0
Aspergillus tubingensis	3	0	6	0	1	0	4	0	2	1
Aspergillus sclerotiicarbonarius	3	0	5	1	1	0	4	1	2	0
Aspergillus sclerotioniger	3	0	4	1	1	0	3	0	1	0
Aspergillus piperis	3	0	5	0	1	0	4	1	2	0
Aspergillus aureofulgens	3	0	5	0	1	0	4	1	2	0
Aspergillus niger N402	3	0	7	0	1	0	5	1	3	0
Aspergillus niger ATCC 1015	3	0	7	0	1	0	5	1	3	0
Aspergillus neoniger	4	0	5	0	1	0	4	1	2	1
Aspergillus niger (lacticoffeatus)	3	0	5	0	1	0	4	1	2	0
Aspergillus japonicus	4	0	5	1	1	0	4	0	3	0
Aspergillus indologenus	4	0	5	1	1	0	4	0	2	0
Aspergillus ibericus	3	0	5	1	1	0	4	1	1	0
Aspergillus heteromorphus	3	0	5	0	1	0	3	1	2	0
Aspergillus fumigatus	3	0	3	0	1	0	2	0	0	0
Aspergillus flavus	5	0	4	0	1	0	4	0	2	0
Aspergillus fijiensis	4	0	5	1	1	0	4	1	4	0
Aspergillus eucalypticola	3	0	5	0	1	0	4	1	2	1
Aspergillus ellipticus	3	0	5	0	1	0	6	2	7	0
Aspergillus costaricaensis	4	0	5	0	1	0	4	0	2	1
Aspergillus brunneovolaceus	4	0	6	1	1	0	4	1	4	0
Aspergillus aculeatinus	4	0	5	1	1	0	4	0	3	0
Arthroderma benhamiae	2	0	1	0	0	1	2	0	0	0

species included. In addition, the PFAMs present in the core genome (present both in S- and R-type strains) showed higher similarities to those of *P. digitatum* and *P. oxalicum*.

Furthermore, two genes with homology to transposase-like proteins (g12052 and g12055) and several reverse-transcriptase domains and other transposon-related domains were identified in encoded proteins on the SORBUS cluster (Table 2). An analysis of the repetitive content of the genomic DNA revealed that the SORBUS cluster was more repetitive than the average of the genome (6.4% and 3.8%, respectively). In particular, long interspersed nuclear elements (LINEs) of the Tad1 retrotransposon family were enriched on SORBUS, as well as unknown/unclassified repetitive elements (S2 Table).

Genome-wide expression profiles of a sorbic acid-sensitive and sorbic acid-resistant strain

A genome-wide expression profile was performed on a sorbic acid-sensitive strain (DTO377G3) and a sorbic acid-resistant strain (DTO006G7) grown on MEB in the presence or absence of 3 mM sorbic acid. The sequence reads were aligned to the assemblies of these two *P. roqueforti* strains. Gene expression values were calculated and differentially expressed genes were identified (S3 Table). The expression profiles of the biological replicates were similar, demonstrated by their clustering in the PCA plot (Fig 6). Combined, PC1 and PC2 explain
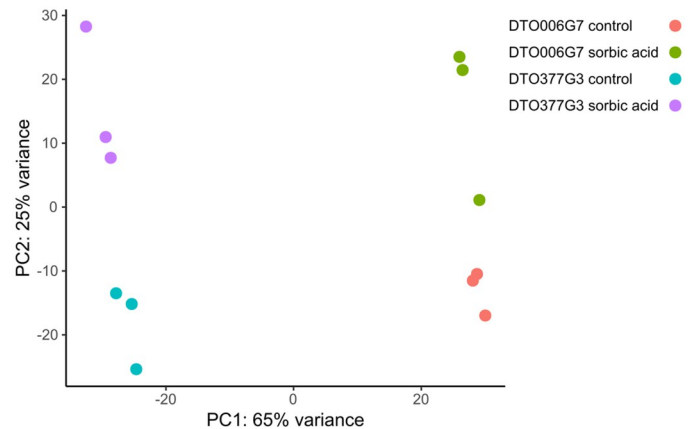
Fig 6. Principle component analysis demonstrates clustering of sorbic acid-exposed strains. Each dot represents a biological replicate. PC1 and PC2 together describe 90% of the variation. The sample grown on sorbic acid separate on the Y-axis and the two strains are separated on the X-axis.

90% of the variation observed. The samples treated with sorbic acid separate from the control samples on Y-axis while the differences between the strains are separated by on the X-axis.

Genes that are either up- or down-regulated in both the R-type and S-type strain when exposed to sorbic acid might be involved in a general response to sorbic acid stress in *P. roqueforti*. Venn diagrams were constructed revealing that 33 genes were significantly up-regulated in both the S-type and R-type strain (Fig 7A). An enrichment analysis revealed that the functional annotation terms 'secretion signal' and 'small secreted protein' are over-represented in these genes. Among the 21 shared down-regulated genes (Fig 7B) the NmrA-like family and NAD(P)H-binding domains were over-represented. S3 Table lists all genes present in both shared pools. The expression of the SORBUS genes (g12000-g12069) was analysed. This revealed that nine out of the 70 genes were significantly differentially expressed (Table 2), including two out of the three genes homologous to *cdcA* that were lower expressed ($\log_2$FC = -1.5) in the presence of sorbic acid. On the SORBUS cluster g12061 (with a flavin-reductase like domain) was the highest expressed gene, reaching 2000 FPKM in the control condition.

Partial knock-out confirms role of SORBUS in sorbic acid resistance

Strains *P. roqueforti* DTO013F2Δ*kusA* SC1 and SC2 were obtained lacking part of the SORBUS cluster (DTO013F2 was chosen since a Δ*kusA* mutant was only available for that strain). Nanopore sequencing was used to investigate how much of the SORBUS cluster was removed in these two transformants (Fig 8). This revealed that the deletions were larger than the 93 kbp
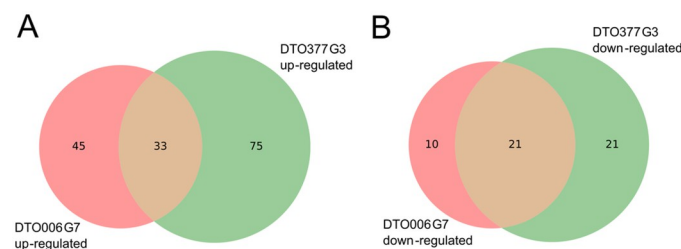

Fig 7. Venn diagram of differentially expressed genes in the presence of sorbic acid. Up- (A) or down-regulated (B) in R-type DTO006G7 and S-type DTO377G3 when exposed to sorbic acid compared to the control in the absence of this weak acid. Genes were considered up- or down-regulated when the $\log_2$FC > 2, p-value < 0.05 and FPKM >10.
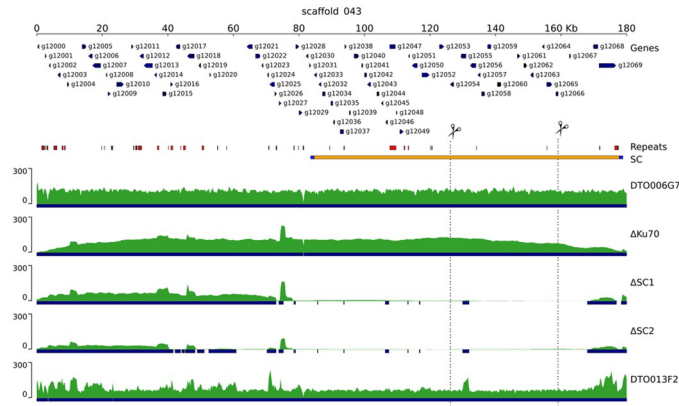
Fig 8. Schematic overview of SORBUS sequences of the knockouts strains. Strains DTO006G7, DTO013F2 and knock-out strains DTO013F2 Δ*kusA*, SC1 and SC2 are depicted. Predicted genes are indicated with arrows, repetitive DNA is indicated in red, the sequence read coverage is indicated in the green tracks, and blue bars indicate (partial) overlap with DTO006G7. The orange bar indicates the targeted knock-out region (target), flanked by the 5' and 3' ends (in blue). Dotted lines and scissors indicate the loci targeted by the sgRNAs.

region. A 105 and a 131 kbp part of the cluster was removed in DTO013F2 Δ*kusA* SC1 and SC2, respectively. Detailed analysis of the nanopore reads revealed that the smaller fragments within the SORBUS cluster present in SC1 and SC2 strains consisted of highly dissimilar sequences, indicating these are parts of repetitive DNA (Fig 8). Both SC strains had a reduced resistance to sorbic acid when compared to DTO013F2 and DTO013F2 Δ*kusA* with a MIC_u similar to the S-type strain DTO377G3 (Fig 9). This shows that part of the SORBUS cluster is involved in sorbic acid stress mitigation.
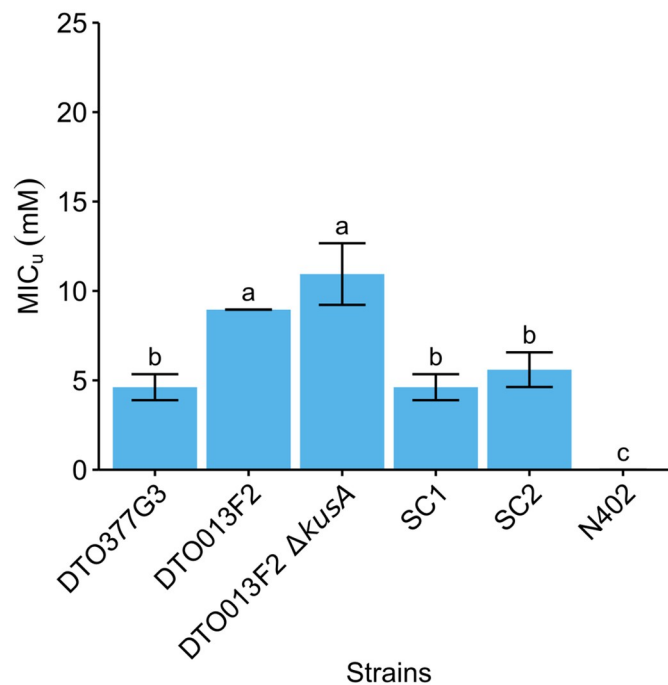

Fig 9. Sorbic acid resistance of five *P. roqueforti* strains and *A. niger* N402. The average MIC_u (mM ± standard deviation) is given for the fungal strains. Each bar graph represents average value of biological independent triplicates. Error bars indicate standard deviation and letters indicate significant difference in MIC_u ($p < 0.05$).

Discussion

P. roqueforti is often encountered as a spoilage organism of food and feed. This can partly be attributed to its ability to grow at refrigeration temperatures [23], low O_2 levels [24] and/or its resistance to preservatives, such as sorbic acid [25]. The inhibitory effect of propionic, benzoic and sorbic acid on *P. roqueforti* growth was assessed and benzoic acid was found to have the strongest inhibitory effect on 34 *P. roqueforti* strains. Previous studies report that *P. roqueforti* is resistant to benzoic acid. Growth was observed up to levels of 3000 ppm sodium benzoate [26,27], while we observed growth at 610 ppm (5 mM). Propionic acid hardly affected the growth of *P. roqueforti*, which is not surprising since it has already been reported that this fungus germinates on potato dextrose agar containing 0.5 M propionic acid (pH 5.6) with an estimated MIC of 0.79 M [23]. The sorbic acid resistance of *P. roqueforti* has also been assessed previously [14,25–30], reporting MIC values ranging from 0–40 mM sorbic acid (i.e. a MIC_u = 0–20 mM). This is similar to the range (MIC_u = 4.2–21.3 mM) found in this study. In fact, a resistant and sensitive group (R- and S-types) consisting of six and 28 strains, respectively, were found in this study. S-type and R-type strains showed a resistance up to 11.9 mM and 21.2 mM undissociated sorbic acid, respectively.

The R-type but not the S-type strains were found to contain a gene cluster (SORBUS) containing 70 genes, of which 51 genes are unique for the R-type strains. Even though the R-type strain DTO006G7 was used with the least fragmented Illumina assembly, the limits of the SORBUS cluster could not be established based on these results alone. The Oxford Nanopore of the DTO013F2 Δ*kusA* assembly revealed that SORBUS is part of a 2.3 Mb contig. Genes homologous to sorbic acid degradation-associated genes in *A. niger* (*sdrA*, *cdcA*, *padA*, *warA*) were identified in the *P. roqueforti* genome [9,12]. A total of 1, 4, 3 and 1 homologs were found, respectively, and only *warA* and one *cdcA* homolog were not located on the SORBUS cluster. Two genes with putative transmembrane transport (g216) or cation transporter (g296) function were identified in the PLINK analysis. The encoded proteins might also be involved in sorbic acid stress mediation alongside the SORBUS cluster, because in addition to decarboxylation, sorbic acid stress could be mediated by an efflux pump or through removal of protons from the plasma membrane by H^+-ATPase [12,31]. As mentioned, only six out of the 34 strains assessed in this study were found to contain the SORBUS cluster. This might be explained by the ecology of *P. roqueforti*. *P. roqueforti* is found in forest soil and wood, but is also associated with lactic acid bacteria (e.g. in silage). Frisvad & Samson [32] already speculated that these micro-organisms may have very well co-evolved, because all the metabolics produced by lactic acid bacteria (e.g. lactic and acetic acid, CO_2) are tolerated by *P. roqueforti* [33]. The elevated levels of weak acids might act as selection pressure to maintain SORBUS in *P. roqueforti* strains which grow in this niche environment. In contrast, this selection pressure is not present in cheese which might explain that none of *P. roqueforti* strains in the 'cheese' population contain the SORBUS cluster [18]. It should be noted that the S-type sequence fragments that align with SORBUS mostly consists of proteins annotated as transposase-like proteins or reverse transcriptase, which might explain why these fragments are found in the S-type strains and it suggests that SORBUS has been obtained from a different species by horizontal gene transfer. This is supported by the phylogenetic analysis on PFAM domains of five SORBUS genes, as the results show that three out of the five SORBUS specific genes are more closely aligned to *Aspergillus* species than *Penicillium* species. The gene cluster SORBUS was also present in two of 35 previously sequenced *P. roqueforti* strains [18] and their sorbic acid resistance could be determined to validate the role of SORBUS in sorbic acid resistance.

The transcriptome analysis of the R-type *P. roqueforti* strain DTO006G7 revealed that two of the three homologs of *A. niger cdcA* that are located on SORBUS (*cdcA* and *cdcC*) are

significantly down-regulated during growth in the presence of sorbic acid. This is in contrast with the results previously described [9], where the authors found a > 500-fold change for *cdcA* when *A. niger* was cultivated on sorbic acid. This difference might be caused by the difference in medium type, because that study [9] used sorbic acid as the sole carbon source, whereas in our experiments sorbic acid was used as a stressor in a nutrient-rich medium. This indicates that the sorbic acid content in the medium does not increase gene expression of the loci leading to increased resistance, suggesting that either these genes are constitutively expressed or expression is induced based on a different compound present in MEB.

Two partial SORBUS knockout strains in the DTO013F2 Δ*kusA* R-type strain showed reduced sorbic acid resistance to a level similar to that of the S-type strains. In contrast to the DTO013F2 Δ*kusA* strain, the SORBUS knockout strains were not repaired through homology directed repair using the donor DNA nor non-homologous end-joining. This might be due to the size of the fragment. Possibly, an alternative DNA repair mechanism such as microhomology-mediated end joining as described in *A. fumigatus* is employed in the DTO013F2 Δ*kusA* strain [34].

Despite the absence of a full SORBUS cluster in the S-type strains and the deletion strain, their MIC is still relatively high when compared to other fungal species such as *A. niger* (as shown in Fig 9) or *A. fumigatus* [12]. This suggests that along with the genes on the SORBUS cluster, other proteins are involved in sorbic acid stress mitigation. Our transcriptomics analysis revealed 21 down-regulated and 33 up-regulated genes that were similarly expressed both in a S-type and a R-type strain when exposed to sorbic acid. One of these up-regulated genes is the cation transporter (g1689, S4 Table) which could have H^+ ATPase activity in the plasma membrane to counteract the acidification caused by the undissociated sorbic acid in the cytosol [31].

In conclusion, the results presented in this study demonstrate that weak-acid resistance varies between *P. roqueforti* strains and the SORBUS cluster contributes to a high sorbic acid resistance. Yet, even in the absence of this cluster the resistance is still relatively high, implying that other mechanisms are also involved in resistance to this weak acid.

Methods

Strain and cultivation conditions

All fungal strains (S1 Table) were provided by the Westerdijk Fungal Biodiversity Institute. Conidia were harvested with a cotton swab after seven days of growth at 25°C on malt extract agar (MEA, Oxoid, Hampshire, UK) and suspended in 10 mL ice-cold ACES buffer (10 mM N (2-acetamido)-2-aminoethanesulfonic acid, 0.02% Tween 80, pH 6.8). The conidia suspension was passed through a syringe containing sterilized glass wool and washed twice with ACES buffer after centrifugation at 4°C for 5 min at 2,500 *g*. The spore suspension was set to $2 \cdot 10^8$ spores mL^{-1} using a Bürker-Türk haemocytometer (VWR, Amsterdam, The Netherlands) and kept on ice until further use.

Weak acid growth assay

Conidial suspension (5 μL) was inoculated in the centre of MEA plates containing 5 mM potassium sorbate, benzoic acid, or sodium propionate (all from Sigma). Medium was set at pH 4.0 using HCl, which corresponds to undissociated concentrations of 4.26, 3.07 and 4.42 mM of these acids, respectively. The absence of preservative was used as a control. Cultures were photographed after 5 days and colony surface area was measured using a manual threshold in ImageJ.

Sorbic acid resistance

Conidial suspensions of the *P. roqueforti* strains were diluted to 10^7 spores mL^{-1} and mixed in a 1:99 ratio with MEB pH 4.0 with and without 25 mM potassium sorbate. 300 μL of the resulting mixture was added in a well of a 96 wells plate (Greiner Bio-One, Cellstar 650180, www.gbo.com). Serial dilutions were made by mixing 225 μL MEB with potassium sorbate and 75 μL MEB without potassium sorbate, resulting in wells with potassium sorbate concentrations of 25, 18.75, 14.06, 10.55, 7.91, 5.93, 4.45 and 0 mM. This corresponded to undissociated sorbic acid concentrations of 21.22, 15.92, 11.94, 8.95, 6.72, 5.04, 3.78 and 0 mM, respectively. The undissociated sorbic acid concentrations were determined using the Henderson-Hasselbach equation.

The 96-wells plates were sealed with parafilm and incubated for 28 days at 25°C using biologically independent replicates. After 28 days, growth was assessed and the undissociated minimal inhibitory concentration (MIC$_u$) was determined for each strain. The MIC$_u$ was defined as the lowest undissociated concentration in which no hyphal growth was observed. An one-way ANOVA followed by a Tukey's HSD test was used to test for significant differences in MIC$_u$ ($P < 0.05$).

DNA extraction, genome sequencing, assembly and annotation

DNA extraction was performed as described [35] and Illumina NextSeq500 2x150 bp paired-end technology was used for sequencing (Utrecht Sequencing Facility, useq.nl). The sequence reads have been deposited in the Short Read Archive (SRA) under PRJNA787443, and the accession numbers of the individual strains are listed in S1 Table. The reads were trimmed on both ends when quality was lower than 15 using bbduk from the BBMap tool suite (BBmap version 37.88; https://sourceforge.net/projects/bbmap/). The trimmed reads were assembled with SPAdes v3.11.1 applying kmer lengths of 21, 33, 55, 77, 99 and 127 and the–careful setting was used to reduce the number of indels and mismatches [36]. Genes were predicted with Augustus version 3.0.3 [37] using the parameter set that was previously generated for *P. roqueforti* [35]. Functional annotation of the predicted genes was performed as described [38]. The assemblies, gene predictions and functional annotations can be accessed interactively at https://fungalgenomics.science.uu.nl. Moreover, the assemblies and gene predictions have been deposited to GenBank under BioProject ID PRJNA745079, and the accession numbers of the individual strain are listed in S1 Table. Repetitive sequences in the assembly were identified *de novo* with RepeatModeler version 2.0.3 [39], which uses RepeatScout version 1.0.6 [40], RECON version 1.08 and Tandem Repeats Finder version 4.09 [41]. RepeatMasker version 4.1.2 (http://www.repeatmasker.org) was subsequently used to annotate the identified repetitive content, using the default Dfam database provided with RepeatMasker.

Genomic phylogeny and analysis

Single-copy orthologous groups were identified and aligned using OrthoFinder v2.5.2 [42]. A maximum likelihood (ML) inference was performed using RAxML [43] under the PROT-GAMMAAUTO model. The number of bootstraps used was 200 (Average WRF = 0.43%) and *Penicillium rubens* Wisconsin 54–1255 [44] was used to root the tree. The phylogenetic tree was visualized using iTOL v5 [45].

Genome-wide association study

Two complementary approaches were taken to identify genomic regions associated with sorbic acid resistance. The first approach is based on the presence of genomic regions that are only

found in the sorbic acid resistant strains, and the second approach is based on SNPs (single nucleotide polymorphisms) that are significantly associated with sorbic acid resistance.

In the first approach, *P. roqueforti* strains were grouped into resistant (R-type) and sensitive (S-type) groups. DTO006G7 was selected from the R-group as reference for the analysis, because the assembly of this strain was the least fragmented in this group. Next, the program nucmer from the MUMmer suite (http://mummer.sourceforge.net/, version 4.0) was used to perform whole-genome alignment. Each genome was aligned to the reference and regions and genes unique for the R-type isolates were identified with the BEDtools package. The genome alignment was visualized with pyGenomeTracks [46]. A gene was considered absent when 90% or more of its sequence was not found in the genome of a strain.

In the second approach, the best practices recommended by GATK (Genome Analysis Toolkit) were used to obtain single nucleotide polymorphisms (SNPs) for each strain. In short, the sequence reads were aligned to the reference genome (DTO006G7) using Bowtie2 (version 2.2.9) and PCR duplicates were removed with Picard tools (MarkDuplicates; version 2.9.2). For variant calling, the HaplotypeCaller (GATK, version 3.7) was used with the following parameters: -stand_call_conf 30 (only retaining variants with a QUAL score > 30), -ploidy 1 and -ERC. The single-sample variant files (GVCFs) were joined into a GenomicsDB before joint genotyping. The variants were annotated using SNPeff (version 4.3) based on their pre-dicted biological effect, such as the introduction of an early stop-codon or a synonymous annotation. The number of SNPs and their putative impact on gene function ('low', 'moderate' or 'high', as defined by SNPeff) was listed for each gene per genome. The SNPs were then cor-related to sorbic acid resistance using PLINK v1.9 [21] with parameters '–maf 0.5 –allow-extra-chr'. The resulting association files were analysed and visualized using R.

Analysis of SORBUS cluster

A synteny analysis of the whole-genome alignments to the SORBUS cluster (scaffold_043 in strain DTO006G7) was performed for the R-type strains. From the MUMmer suite, nucmer was used to identify matching sequences larger than 1.5 kbp (minimal match length was set to 1.5 kbp). A Genbank file was created for each matching sequence with the SORBUS cluster and these sequences were aligned and visualized in a schematic gene overview of the SORBUS clusters using Clinker (v.0.0.23) [47].

Gene orthology was determined by reciprocal BlastP analysis (i.e., a bidirectional best blast hit) or a gene tree approach, as indicated. NCBI BLAST+ version 2.12.0 was used for the recip-rocal BlastP approach. For the gene tree approach, the protein domains indicated in the text were first identified by PFAM domain version 32 [48]. Next, the amino acid sequences corre-sponding to these domains were aligned with MAFFT version 7.310 using auto settings [49] and a gene tree was constructed using FastTree version 2.1 with default settings [50] This gene tree was visualized with DendroScope 3 [51] and the orthologs of the protein of interest were manually identified based on coclustering in the gene tree.

RNA extraction and sequencing

A genome-wide transcriptome analysis was performed on the sorbic acid sensitive *P. roqueforti* strain DTO377G3 and the sorbic acid-resistant *P. roqueforti* strain DTO006G7. Erlenmeyer flasks containing 50 mL MEB (pH 4.0) were inoculated with 100 μl ACES containing 10^7 conidia and incubated for 48 h at 25°C and 200 rpm. Mycelium was harvested using a steril-ized miracloth filter and equally divided in Erlenmeyer flasks with 50 mL MEB (pH 4.0) or 50 mL MEB containing 3 mM potassium sorbate (pH 4.0). Growth was continued for another four hours, after which the mycelium was harvested using a sterilized Miracloth filter and

frozen in liquid nitrogen. Total RNA was isolated with the RNeasy Plant Mini Kit (Qiagen) and purified by on-column DNase digestion according to the manufacturer's protocol. RNA was sequenced with Illumina NextSeq2000 2x50 bp paired-end technology (Utrecht Sequencing Facility; useq.nl). The transfer experiment and subsequent RNA-sequencing was performed in biological triplicates. The transcript lengths, counts per gene and read mapping were determined using Salmon v1.5.2 with—validateMappings [52]. The transcript abundance of reads was quantified using custom constructed indices for DTO006G7 and DTO377G3.

DESeq2 [53] was used for pairwise comparisons of the samples and the identification of differentially expressed genes. Genes with low read counts (<10) were excluded from the analysis and a gene was considered differentially expressed when the adjusted p-value was < 0.05. In addition, genes were considered up- or down-regulated when they had a $\log_2$ fold change of > 2 or < -2, respectively. A Fisher Exact test as implemented in PyRanges [54] was employed to identify over- and under-representation of functional annotation terms in sets of genes. To correct for multiple testing the False Discovery Rate method was used, with a P-value < 0.05 as cut off.

The sequence reads are available in the Short Read Archive under BioProject PRJNA796729.

Plasmid construction and generation knockout strain

A *kusA* deficient DTO013F2 strain was constructed using the protocol and plasmid pPT22.4 as described [55]. Deletion was confirmed through diagnostic PCR and nanopore sequencing (S2 Fig). CRISPR/Cas9 technology was employed to remove a 93 kbp region (from 84 to 177 kbp) of the SORBUS cluster in DTO013F2 Δ*kusA*. Two single-guide RNAs (sgRNAs) were designed to perform a simultaneous double restriction in the 93 kbp SORBUS region. Transformation procedures were performed as described, with some modifications [55]. In short, plasmid pFC332 [56] was used as a vector to express the sgRNA, *cas9* and a hygromycin selection marker (for primer sequences used in this study see S5 Table). The 5' and 3' flanking regions of sgRNA were amplified using plasmids pTLL108.1 and pTLL109.2 as template [57]. The amplified products were fused and introduced into pFC332 using Gibson assembly (NEBuilder HiFi DNA Assembly Master Mix, New England Biolabs, MA, USA). The vectors containing the sgRNA were then transformed into competent *Escherichia coli* TOP10 cells for multiplication overnight. Plasmids were recovered using Quick Plasmid Miniprep Kit (ThermoFisher, Waltham, MA, USA) and digested using SacII to verify the presence of the sgRNA. In addition, the correct integration of the sgRNA was confirmed with sequencing. To construct donor DNA, two 1 kbp homologous regions located at scaffold 43 at nucleotide position 83573 to 84629 and 177760 to 178854 were amplified and fused using a unique 23 nucleotide sequence GGAGTGGTACCAATATAAGCCGG with a PAM site for further genetic engineering.

Transformation was performed as described with adjustments [58]. In short, *P. roqueforti* conidia were incubated 48 h at 25˚C in 100 mL potato dextrose broth at 200 rpm. The mycelium was washed in SMC and incubated for 4 h at 37˚C in lysing enzymes from *Trichoderma harzianum* (Sigma) dissolved in SMC. Protoplasts were resuspended in 1 mL STC and kept on ice after centrifuging for 5 min at 3000 g. To 100 μL of this suspension, 2 μg donor DNA, 2 μg of each pFC332 vector containing sgRNA (pTF and pSdrA) and 1.025 mL of freshly made PEG solution was added (see S6 Table for the vectors used in this study). After 5 min, 2 mL STC was added and the protoplasts were mixed with 20 mL liquid MMS containing 0.3% agar and 200 μg hygromycin mL^{-1} (InvivoGen, San Diego, CA, USA). The mixture was poured on MMS containing 0.6% agar and 200 μg hygromycin mL^{-1}. Transformants were grown for 7–14 days at 25˚C and then single streaked on MM containing 100 μg mL^{-1} hygromycin until

sporulating colonies appeared. Next, the plasmid was removed by a single streak on MM plates without antibiotic. Finally, transformants were single streaked on MM, MM containing 100 μg hygromycin mL^{-1} and MEA plates to confirm that transformants lost the plasmids.

To verify the transformants, genomic DNA from DTO013F2 Δ*kusA* and two DTO013F2 Δ*kusA* Δ*SORBUS* strains was sequenced with Oxford Nanopore MinION technology (FLO-MIN106) at the Utrecht Sequencing Facility (useq.nl). The reads were assembled using Canu v2.2 using the option for raw nanopore data and guided by a genome size of 28 Mbp [59]. In addition, the nanopore reads were aligned to the DTO006G7 assembly using Minimap2 [60]. Only reads aligning once and that had a mapping quality > 60 were selected using samtools. The sequencing data generated with the MinION technology is deposited in the SRA archive under the following accession numbers: SRR17178875, SRR17178876 and SRR17178877.

Supporting information

S1 Data. Complete data set of colony growth measurements presented in Fig 1, all individual MIC values of graph presented in Fig 2, and all individual MIC values of graph presented in Fig 9.
(XLSX)

S1 Fig. Schematic clinker-style comparison of all genes present on the SORBUS cluster of the R-type strains (DTO006G1, DTO006G7, DTO012A1, DTO012A8, DTO013E5, DTO013F2). Gene identity (%) is indicated in greyscale and (partial) scaffolds corresponding to scaffold 43 of DTO006G7 are listed in order below the strain names.
(TIFF)

S2 Fig. Schematic overview of the *kusA* gene in DTO013F2. Genes are indicated with blue arrows. The nanopore sequencing read coverage of DTO013F2 Δ*kusA* is indicated in green. The orange bar indicates the targeted knock-out region, flanked by the 5' and 3' ends (in light-blue). Dotted lines and scissors indicate the loci targeted by the sgRNA.
(TIFF)

S1 Table. *Penicillium* spp. used in this study.
(XLSX)

S2 Table. Repetitive content annotations in the genome assembly of DTO006G7. The first tab provides summary statistics of repeat coverage of the full assembly as well as the SORBUS cluster. The second tab provides the individual annotations of repetitive content.
(XLSX)

S3 Table. FPKM list as determined by DESeq2 of 9993 genes of *P. roqueforti* strain DTO377G3 and DTO006G7. Strain DTO006G7 was used as reference, and the corresponding protein sequences are provided for each gene.
(XLSX)

S4 Table. Significantly differentially expressed genes between *P. roqueforti* strain DTO377G3 and DTO006G7.
(XLSX)

S5 Table. Primers used for plasmid and donor DNA construction and diagnostic PCR.
(XLSX)

S6 Table. Plasmids used in this study.
(XLSX)

References

1. Pitt JI, Hocking AD. Fungi and food spoilage. Fungi and Food Spoilage. Springer; 2009.

2. Filtenborg O, Frisvad JC, Thrane U. Moulds in food spoilage. Int J Food Microbiol. 1996; 33: 85–102. https://doi.org/10.1016/0168-1605(96)01153-1 PMID: 8913811

3. Gillot G, Jany JL, Poirier E, Maillard MB, Debaets S, Thierry A, et al. Functional diversity within the Penicillium roqueforti species. Int J Food Microbiol. 2017; 241: 141–150. https://doi.org/10.1016/j.ijfoodmicro.2016.10.001 PMID: 27771579

4. Bondi M, Messi P, Halami PM, Papadopoulou C, De Niederhausern S. Emerging microbial concerns in food safety and new control measures. Biomed Res Int. 2014; 2014: 1–3. https://doi.org/10.1155/2014/251512 PMID: 25110665

5. Dijksterhuis J. Fungal spores: Highly variable and stress-resistant vehicles for distribution and spoilage. Food Microbiol. 2019; 81: 2–11. https://doi.org/10.1016/j.fm.2018.11.006 PMID: 30910084

6. Kaczmarek M, Avery S V., Singleton I. Microbes associated with fresh produce: Sources, types and methods to reduce spoilage and contamination. Adv Appl Microbiol. 2019; 107: 29–82. https://doi.org/10.1016/bs.aambs.2019.02.001 PMID: 31128748

7. Rico-Munoz E, Samson RA, Houbraken J. Mould spoilage of foods and beverages: Using the right methodology. Food Microbiol. 2019; 81: 51–62. https://doi.org/10.1016/j.fm.2018.03.016 PMID: 30910088

8. Stratford M, Anslow PA. Evidence that sorbic acid does not inhibit yeast as a classic "weak acid preservative". Lett Appl Microbiol. 1998; 27: 203–206. https://doi.org/10.1046/j.1472-765x.1998.00424.x PMID: 9812395

9. Lubbers RJM, Dilokpimol A, Navarro J, Peng M, Wang M, Lipzen A, et al. Cinnamic Acid and Sorbic acid Conversion Are Mediated by the Same Transcriptional Regulator in Aspergillus niger. Front Bioeng Biotechnol. 2019; 7: 249. https://doi.org/10.3389/fbioe.2019.00249 PMID: 31612133

10. Plumridge A, Melin P, Stratford M, Novodvorska M, Shunburne L, Dyer PS, et al. The decarboxylation of the weak-acid preservative, sorbic acid, is encoded by linked genes in Aspergillus spp. Fungal Genet Biol. 2010; 47: 683–692. https://doi.org/10.1016/j.fgb.2010.04.011 PMID: 20452450

11. Mukai N, Masaki K, Fujii T, Kawamukai M, Iefuji H. PAD1 and FDC1 are essential for the decarboxylation of phenylacrylic acids in Saccharomyces cerevisiae. J Biosci Bioeng. 2010; 109: 564–569. https://doi.org/10.1016/j.jbiosc.2009.11.011 PMID: 20471595

12. Geoghegan IA, Stratford M, Bromley M, Archer DB, Avery S V. Weak acid resistance a (WarA), a novel transcription factor required for regulation of weak-acid resistance and spore-spore heterogeneity in Aspergillus niger. mSphere. 2020; 5: 1–39. https://doi.org/10.1101/788141

13. Stratford M, Steels H, Nebe-von-Caron G, Novodvorska M, Hayer K, Archer DB. Extreme resistance to weak-acid preservatives in the spoilage yeast Zygosaccharomyces bailii. Int J Food Microbiol. 2013; 166: 126–134. https://doi.org/10.1016/j.ijfoodmicro.2013.06.025 PMID: 23856006

14. Liewen MB, Marth EH. Viability and ATP content of conidia of sorbic acid-sensitive and-resistant strains of Penicillium roqueforti after exposure to sorbic acid. Appl Microbiol Biotechnol. 1985; 21: 113–117. https://doi.org/10.1007/BF00252372

15. Ropars J, López-Villavicencio M, Dupont J, Snirc A, Gillot G, Coton M, et al. Induction of sexual reproduction and genetic diversity in the cheese fungus Penicillium roqueforti. Evol Appl. 2014; 7: 433–441. https://doi.org/10.1111/eva.12140 PMID: 24822078

16. Ropars J, Rodríguez De La Vega RC, López-Villavicencio M, Gouzy J, Sallet E, Dumas É, et al. Adaptive horizontal gene transfers between multiple cheese-associated fungi. Curr Biol. 2015; 25: 2562–2569. https://doi.org/10.1016/j.cub.2015.08.025 PMID: 26412136

17. Cheeseman K, Ropars J, Renault P, Dupont J, Gouzy J, Branca A, et al. Multiple recent horizontal transfers of a large genomic region in cheese making fungi. Nat Commun. 2014; 5. https://doi.org/10.1038/ncomms3876 PMID: 24407037

18. Dumas E, Feurtey A, Rodríguez de la Vega RC, Le Prieur S, Snirc A, Coton M, et al. Independent domestication events in the blue-cheese fungus Penicillium roqueforti. Mol Ecol. 2020; 29: 2639–2660. https://doi.org/10.1111/mec.15359 PMID: 31960565

19. Naidu AS. Natural Food Antimicrobial Systems. Natural Food Antimicrobial Systems. CRC press; 2000.

20. Nielsen P V., Rios R. Inhibition of fungal growth on bread by volatile components from spices and herbs, and the possible application in active packaging, with special emphasis on mustard essential oil. Int J Food Microbiol. 2000; 60: 219–229. https://doi.org/10.1016/s0168-1605(00)00343-3 PMID: 11016611

21. Purcell S, Neale B, Todd-Brown K, Thomas L, Ferreira MAR, Bender D, et al. PLINK: A tool set for whole-genome association and population-based linkage analyses. Am J Hum Genet. 2007; 81: 559–575. https://doi.org/10.1086/519795 PMID: 17701901

22. Paiva S, Devaux F, Barbosa S, Jacq C, Casal M. Ady2p is essential for the acetate permease activity in the yeast Saccharomyces cerevisiae. Yeast. 2004; 21: 201–210. https://doi.org/10.1002/yea.1056 PMID: 14968426

23. Kalai S, Anzala L, Bensoussan M, Dantigny P. Modelling the effect of temperature, pH, water activity, and organic acids on the germination time of Penicillium camemberti and Penicillium roqueforti conidia. Int J Food Microbiol. 2017; 240: 124–130. https://doi.org/10.1016/j.ijfoodmicro.2016.03.024 PMID: 27090813

24. Nguyen Van Long N, Vasseur V, Couvert O, Coroller L, Burlot M, Rigalma K, et al. Modeling the effect of modified atmospheres on conidial germination of fungi from dairy foods. Front Microbiol. 2017; 8: 2109. https://doi.org/10.3389/fmicb.2017.02109 PMID: 29163403

25. Quattrini M, Liang N, Fortina MG, Xiang S, Curtis JM, Gänzle M. Exploiting synergies of sourdough and antifungal organic acids to delay fungal spoilage of bread. Int J Food Microbiol. 2019; 302: 8–14. https://doi.org/10.1016/j.ijfoodmicro.2018.09.007 PMID: 30220438

26. Suhr KI, Nielsen P V. Effect of weak acid preservatives on growth of bakery product spoilage fungi at different water activities and pH values. Int J Food Microbiol. 2004; 95: 67–78. https://doi.org/10.1016/j.ijfoodmicro.2004.02.004 PMID: 15240076

27. Blaszyk M, Blank G, Holley R, Chong J. Reduced water activity during sporogenesis in selected penicillia: Impact on spore quality. Food Res Int. 1998; 31: 503–509. https://doi.org/10.1016/S0963-9969(99)00019-8

28. Huang Y, Wilson M, Chapman B, Hocking AD. Evaluation of the efficacy of four weak acids as antifungal preservatives in low-acid intermediate moisture model food systems. Food Microbiol. 2010; 27: 33–36. https://doi.org/10.1016/j.fm.2009.07.017 PMID: 19913689

29. Razavi-Rohani SM, Griffiths MW. Antifungal effects of sorbic acid and propionic acid at different pH and NaCl conditions. J Food Saf. 1999; 19: 109–120. https://doi.org/10.1111/j.1745-4565.1999.tb00238.x

30. Bullerman LB. Effects of potassium sorbate on growth and ochratoxin production by Aspergillus ochraceus and Penicillium species. J Food Prot. 1985; 48: 162–165. https://doi.org/10.4315/0362-028X-48.2.162 PMID: 30934526

31. Lambert RJ, Stratford M. Weak-acid preservatives: Modelling microbial inhibition and response. J Appl Microbiol. 1999; 86: 157–164. https://doi.org/10.1046/j.1365-2672.1999.00646.x PMID: 10030018

32. Samson RA, Hoekstra ES, Frisvad JC. Introduction to food-and airborne fungi. Centraalbureau voor Schimmelcultures (CBS); 2004.

33. Broberg A, Jacobsson K, Ström K, Schnürer J. Metabolite profiles of lactic acid bacteria in grass silage. Appl Environ Microbiol. 2007; 73: 5547–5552. https://doi.org/10.1128/AEM.02939-06 PMID: 17616609

34. Zhang C, Meng X, Wei X, Lu L. Highly efficient CRISPR mutagenesis by microhomology-mediated end joining in Aspergillus fumigatus. Fungal Genet Biol. 2016; 86: 47–57. https://doi.org/10.1016/j.fgb.2015.12.007 PMID: 26701308

35. Punt M, van den Brule T, Teertstra WR, Dijksterhuis J, den Besten HMW, Ohm RA, et al. Impact of maturation and growth temperature on cell-size distribution, heat-resistance, compatible solute composition and transcription profiles of Penicillium roqueforti conidia. Food Res Int. 2020; 136: 109287. https://doi.org/10.1016/j.foodres.2020.109287 PMID: 32846509

36. Bankevich A, Nurk S, Antipov D, Gurevich AA, Dvorkin M, Kulikov AS, et al. SPAdes: A new genome assembly algorithm and its applications to single-cell sequencing. J Comput Biol. 2012; 19: 455–477. https://doi.org/10.1089/cmb.2012.0021 PMID: 22506599

37. Stanke M, Schöffmann O, Morgenstern B, Waack S. Gene prediction in eukaryotes with a generalized hidden Markov model that uses hints from external sources. BMC Bioinformatics. 2006; 7: 1–11. https://doi.org/10.1186/1471-2105-7-62 PMID: 16469098

38. De Bekker C, Ohm RA, Evans HC, Brachmann A, Hughes DP. Ant-infecting Ophiocordyceps genomes reveal a high diversity of potential behavioral manipulation genes and a possible major role for enterotoxins. Sci Rep. 2017; 7: 1–13. https://doi.org/10.1038/s41598-017-12863-w PMID: 28127051

39. Flynn JM, Hubley R, Goubert C, Rosen J, Clark AG, Feschotte C, et al. RepeatModeler2 for automated genomic discovery of transposable element families. Proc Natl Acad Sci U S A. 2020; 117: 9451–9457. https://doi.org/10.1073/pnas.1921046117 PMID: 32300014

40. Price AL, Jones NC, Pevzner PA. De novo identification of repeat families in large genomes. Bioinformatics. 2005; 21: i351–i358. https://doi.org/10.1093/bioinformatics/bti1018 PMID: 15961478

41. Benson G. Tandem repeats finder: A program to analyze DNA sequences. Nucleic Acids Res. 1999; 27: 573–580. https://doi.org/10.1093/nar/27.2.573 PMID: 9862982

42. Emms DM, Kelly S. OrthoFinder: Phylogenetic orthology inference for comparative genomics. Genome Biol. 2019; 20: 1–14. https://doi.org/10.1186/s13059-019-1832-y PMID: 31727128

43. Stamatakis A. RAxML version 8: A tool for phylogenetic analysis and post-analysis of large phylogenies. Bioinformatics. 2014; 30: 1312–1313. https://doi.org/10.1093/bioinformatics/btu033 PMID: 24451623

44. Van Den Berg MA, Albang R, Albermann K, Badger JH, Daran JM, Driessen AJM, et al. Genome sequencing and analysis of the filamentous fungus Penicillium chrysogenum. Nat Biotechnol. 2008; 26: 1161–1168. https://doi.org/10.1038/nbt.1498 PMID: 18820685

45. Letunic I, Bork P. Interactive Tree Of Life (iTOL) v5: an online tool for phylogenetic tree display and annotation. Nucleic Acids Res. 2021; 1–4. https://doi.org/10.1093/nar/gkab301 PMID: 33885785

46. Lopez-Delisle L, Rabbani L, Wolff J, Bhardwaj V, Backofen R, Grüning B, et al. pyGenomeTracks: reproducible plots for multivariate genomic datasets. Bioinformatics. 2021; 37: 422–423. https://doi.org/10.1093/bioinformatics/btaa692 PMID: 32745185

47. Gilchrist CLM, Chooi YH. Clinker & clustermap.js: Automatic generation of gene cluster comparison figures. Bioinformatics. 2021; 37: 2473–2475. https://doi.org/10.1093/bioinformatics/btab007 PMID: 33459763

48. Mistry J, Chuguransky S, Williams L, Qureshi M, Salazar GA, Sonnhammer ELL, et al. Pfam: The protein families database in 2021. Nucleic Acids Res. 2021; 49: D412–D419. https://doi.org/10.1093/nar/gkaa913 PMID: 33125078

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

50. Price MN, Dehal PS, Arkin AP. FastTree 2—Approximately maximum-likelihood trees for large alignments. PLoS One. 2010; 5. https://doi.org/10.1371/journal.pone.0009490 PMID: 20224823

51. Huson DH, Scornavacca C. Dendroscope 3: An interactive tool for rooted phylogenetic trees and networks. Syst Biol. 2012; 61: 1061–1067. https://doi.org/10.1093/sysbio/sys062 PMID: 22780991

52. Patro R, Duggal G, Love MI, Irizarry RA, Kingsford C. Salmon provides fast and bias-aware quantification of transcript expression. Nat Methods. 2017; 14: 417–419. https://doi.org/10.1038/nmeth.4197 PMID: 28263959

53. Love MI, Huber W, Anders S. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. Genome Biol. 2014; 15: 1–21. https://doi.org/10.1186/s13059-014-0550-8 PMID: 25516281

54. Stovner EB, Sætrom P. PyRanges: efficient comparison of genomic intervals in Python. Bioinformatics. 2020; 36: 918–919. https://doi.org/10.1093/bioinformatics/btz615 PMID: 31373614

55. Seekles SJ, Teunisse PPP, Punt M, van den Brule T, Dijksterhuis J, Houbraken J, et al. Preservation stress resistance of melanin deficient conidia from Paecilomyces variotii and Penicillium roqueforti mutants generated via CRISPR/Cas9 genome editing. Fungal Biol Biotechnol. 2021; 8: 1–13. https://doi.org/10.1186/s40694-021-00111-w PMID: 33795004

56. Nødvig CS, Nielsen JB, Kogle ME, Mortensen UH. A CRISPR-Cas9 system for genetic engineering of filamentous fungi. PLoS One. 2015/07/16. 2015; 10: 1–18. https://doi.org/10.1371/journal.pone.0133085 PMID: 26177455

57. Van Leeuwe TM, Arentshorst M, Ernst T, Alazi E, Punt PJ, Ram AFJ. Efficient marker free CRISPR/Cas9 genome editing for functional analysis of gene families in filamentous fungi. Fungal Biol Biotechnol. 2019; 6: 1–13. https://doi.org/10.1186/s40694-019-0076-7 PMID: 31559019

58. Arentshorst M, Ram AFJ, Meyer V. Using non-homologous end-joining-deficient strains for functional gene analyses in filamentous fungi. In: Bolton M, Thomma B, editors. Plant Fungal Pathogens. Humana Press; 2012. pp. 133–150. https://doi.org/10.1007/978-1-61779-501-5_9

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

60. Li H. Minimap2: Pairwise alignment for nucleotide sequences. Bioinformatics. 2018; 34: 3094–3100. https://doi.org/10.1093/bioinformatics/bty191 PMID: 29750242