Molecular Evolution of Genes Involved in Quinic Acid Utilization in Fungi

David K. Asch, Janett Ziegler, Xiang Jia Min

Department of Chemical and Biological Sciences, Youngstown State University, Youngstown, OH 44555, USA

Corresponding author email: dkasch@ysu.edu; xjm1@ysu.edu

Computational Molecular Biology, 2021, Vol.11, No.5  
doi: 10.5376/cmb.2021.11.0005

Received: 08 May, 2021  
Accepted: 22 Jun., 2021  
Published: 02 Jul., 2021

Copyright © 2021 Asch et al., This is an open access article published under the terms of the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Abstract Fungi use diverse organic compounds for their growth and development. *Neurospora crassa* can use quinic acid as its sole carbon source for its growth because of the presence of a quinic acid utilization (QUT) cluster of genes in its genome. Using bioinformatics methods we examined a total of 285 completely sequenced fungal genomes comprised of 282 unique species and found there were 117 fungal species having all 7 QUT genes in their genomes. Most species in the classes of Dothideomycetes, Eurotiomycetes, Leotiomycetes and Sordariomycetes have QUT genes, however, among 53 species in Saccharomycetes only 3 species have all 7 QUT genes. There were lineage specific losses of QUT genes, such as species in Eurotiomycetes class Ochrysosphaeraceae order lacked most of QUT utilization genes. Our survey revealed that species in Agaricomycetes, Basidiomycota, Chytridiomycetes, Exobasidiomycetes, Malasseziomycetes, Microsporidia, Schizosacharomyces, and Tremellomycetes did not have QUT utilization genes. Using concatenated protein sequences encoded by these 7 QUT genes, a robust phylogenetic tree to infer the evolution of the QUT cluster genes was constructed. In addition, we also found QUT genes from recently sequenced genome of cork oak (*Quercus suber*), however, our analysis suggests that these QUT sequences are likely from a contaminated fungal species.

Keywords Fungi; Quinic acid; Gene; Protein; Phylogeny; *Neurospora crassa*

1 Introduction

Microorganisms play many diverse roles in the environment. The kingdom Fungi consists of a very diverse group of heterotrophic, eukaryotic organisms which primarily depend on organic biomolecules made by plants and animals for food sources (Willis et al., 2018). Based on lifestyles, fungi are divided into saprobes or decomposers of the dead remains of other organisms, pathogenic, symbiotic, or parasitic fungi, living with living plants or animals (Blackwell, 2011). To live in such diverse environments, it has been advantageous for fungi to be able to utilize a wide variety of sources for basic nutrients such as carbon and nitrogen. The genes encoding enzymes involved in metabolic pathways needed to utilize these diverse compounds are sometimes organized into metabolic gene clusters (MGCs). These gene clusters are often localized in the same region on a fungal chromosome (Wisecaver et al., 2014; Wisecaver and Rokas, 2015). These MGGs are found in many fungal species (Wisecaver et al., 2014). The quinic acid utilization (QA or QUT) clusters found in *Neurospora crassa* and *Aspergillus nidulans* are among the earliest and best characterized of these gene clusters (Giles et al., 1985; Hawkins et al., 1988).

The quinic acid gene cluster (QGC) consists of seven genes that span 17.3 kb on chromosome VII in *N. crassa* (Giles et al., 1985; Galagan et al., 2003). These 7 genes include five structural genes (*qa-X*, *qa-2*, *qa-3*, *qa-4*, and *qa-Y*) and two regulatory genes (*qa-1S* and *qa-1F*). Three of the structural genes encode enzymes, *qa-2* gene encodes catabolic 3-dehydroquinase, *qa-3* encodes shikimate/quinate 5-dehydrogenase, *qa-4* encodes 3-dehydroshikimate (DHS) dehydrogenase. One structural gene *qa-Y* encodes a transporter, the quinic acid permease. The *qa-X* gene encodes a protein with unknown function. The other two genes encode proteins involved in gene regulation. The *qa-1S* gene encodes a repressor and *qa-1F* encodes an activator which stimulates expression of all the genes of the cluster (Giles et al., 1985). The QA-1F protein may activate expression of genes outside the *qa* gene cluster as well (Logan et al., 2007; Tang et al., 2011). A similar gene cluster consisting of all 7 QUT genes but having slightly different physical organization from *N. crassa* was identified in *A. nidulans* (Hawkins et al., 1988; Grant et al., 1988). The genes in the cluster were named as *qutE*, *qutB*, *qutC*, *qutG*, *qutD*, *qutA*, and *qutR*,
which were homologs of qa-2, qa-3, qa-4, qa-X, qa-Y, qa-1F, and qa-1S in N. crassa, respectively (Hawkins et al., 1988; Giles et al., 1985). For the convenience of description in this work we use the gene names of N. crassa to represent these gene homologs in fungi, and for protein sequences we use the upper case letters.

The birth, evolution and death of MGCs in fungi were recently comprehensively reviewed by Rokas et al. (2018). In this work, using computational methods the existence of QGCs was examined in completely sequenced fungal genomes and a molecular evolutionary analysis of QGCs in representative fungal species was performed. In addition, we also identified the QUT genes are spread into a plant species, which may represent a case of horizontal gene transfer from fungi to plants, if these genes are proved not to be resulted from fungal contamination.

2 Materials and Methods
2.1 Data collection
The QA gene cluster consists of seven genes that span 17.3 kb on chromosome VII in N. crassa (strain OR74A; Accession number: NC_0265.7.1). The accession numbers of the protein sequences used for downstream BLAST searches were: XP_959612.1 (QA-X), XP_959613.1 (QA-2), XP_959615.3 (QA-3), XP_959614.1 (QA-4), XP_959616.1 (QA-Y), XP_959617.2 (QA-1S), and XP_959618.1 (QA-1F).

Protein BLAST (BLASTP) was carried out with N. crassa protein sequence of each qa gene product in the QA cluster as a query to search the non-redundant fungal protein sequences which were downloaded from the RefSeq database (March 1, 2019) in the National Center for Biotechnology Information (NCBI, https://www.ncbi.nlm.nih.gov/refseq/) with a limit on fungi (taxid:4751). The E-value cutoff was set to 1e-5. In the data collection process, we found that there was only one copy for qa-2, qa-3 and qa-1F gene in N. crassa genome. However, there were two or more paralogs existed for other genes, including two paralogs for qa-X, qa-4, and qa-1S, and four paralogs of qa-Y genes in N. crassa genome. The details for these paralogs were described further in the section of results. To prevent false positives, the retrieved probable QA homologs in the RefSeq fungal database were used as queries for reciprocal BLASTP search against all protein sequences from N. crassa (strain OR74A) as a database. Only the QA protein sequences from each species having a best hit being one of the QA proteins including all QA paralogs in N. crassa in the reciprocal BLASTP were treated as real QA homologs. Thus, gene copy numbers of each QUT gene in each fungal species were estimated based on the reciprocal BALSTP of all these paralogs of N. crassa QA proteins. The list consists of 285 complete genomes from 282 species (Table 1).

2.2 Phylogenetic tree construction
The MEGA package (version 6) was used for phylogenetic analysis and tree construction (https://www.megasoftware.net/) (Tamura et al., 2013). To make the tree to be easily visualized, we selected protein sequences from 43 representative species. The selected protein sequences were aligned using MUSCLE with default parameters. We have constructed individual protein trees for all 7 QA genes. The individual protein trees did not reveal a consistent phylogenetic relationship among the species due to the long divergence time. Thus, we selected one homologous protein for each of the 7 genes in the QA cluster from each species and constructed a 7-protein tree. The 7-protein tree was constructed using the concatenated pre-aligned homologous protein sequences in each species. For species having multiple copies of QA genes, the protein sequence having the highest similarity to the protein encoded by one of the 7 QA cluster genes in N. crassa was selected for the phylogenetic tree construction. All positions with less than 95% site coverage were eliminated. That is, fewer than 5% alignment gaps, missing data, and ambiguous amino residues were allowed at any position. There were a total of 2761 positions in the final dataset. The initial trees for the heuristic search were obtained automatically by applying Neighbor-Joining and BioNJ algorithms to a matrix of pairwise distances estimated using a JTT model, and then selecting the topology as the original tree with superior log likelihood value using the Maximum Likelihood method (Jones et al., 1992). The bootstrap consensus tree that was inferred from 500 replicates was taken to represent the evolutionary history of the taxa analyzed (Felsenstein, 1985).
| Species                                                | Proteins | qa-1F | qa-1S | qa2 | qa3 | qa4 | qaX | qaY |
|--------------------------------------------------------|----------|-------|-------|-----|-----|-----|-----|-----|
| **Agaricomycetes**                                     |          |       |       |     |     |     |     |     |
| *Agaricus bisporus var. bisporus H97*                  | 10448    | 0     | 0     | 0   | 0   | 1   | 8   |
| *Agaricus bisporus var. burnetii JB137-S8*             | 11278    | 0     | 0     | 0   | 0   | 1   | 8   |
| *Coniophora puteana RW-64-598 SS2*                     | 13758    | 0     | 0     | 0   | 0   | 1   | 8   |
| *Coprinopsis cinerea okayama78130*                      | 13356    | 0     | 0     | 0   | 0   | 1   | 3   |
| *Dichomitus squalens LYAD-421 SS1*                      | 12287    | 0     | 0     | 0   | 1   | 1   | 7   |
| *Fibroporia radiculos a*                                | 9262     | 0     | 0     | 0   | 1   | 1   | 6   |
| *Fomitiporia mediterranea MF3/22*                       | 11338    | 0     | 0     | 0   | 2   | 2   | 6   |
| *Gloeophyllum trabeum ATCC 11539*                       | 11755    | 0     | 0     | 0   | 2   | 1   | 7   |
| *Heterobasidion irregular TC 32-1*                     | 13275    | 0     | 0     | 0   | 1   | 0   | 7   |
| *Laccaria bicolor S238N-H82*                            | 18215    | 0     | 0     | 0   | 0   | 1   | 3   |
| *Phanerochaete carnosus HHB-10118-sp*                   | 13925    | 0     | 0     | 0   | 1   | 1   | 7   |
| *Postia placenta MAD-698-R-SB12*                        | 12539    | 0     | 0     | 0   | 1   | 1   | 3   |
| *Punctularia strigosozonata HHB-11173 SS5*              | 11540    | 1     | 0     | 0   | 0   | 1   | 6   |
| *Schizophyllum commune H4-8*                            | 13194    | 0     | 0     | 0   | 1   | 1   | 12  |
| *Serpula lacrymans var. lacrymans S7.9*                 | 12925    | 0     | 0     | 0   | 1   | 1   | 7   |
| *Sparassis crispa*                                      | 13157    | 1     | 0     | 0   | 0   | 1   | 4   |
| *Stereum hirsutum FP-91666 SS1*                         | 14066    | 0     | 0     | 1   | 1   | 1   | 7   |
| *Trametes versicolor FP-101664 SS1*                     | 14302    | 0     | 0     | 0   | 1   | 1   | 4   |
| **Chytridiomycetes**                                   |          |       |       |     |     |     |     |     |
| *Batrachochytrium dendrobatidis JAM81*                  | 8700     | 1     | 0     | 0   | 0   | 1   | 0   |
| *Spizellomyces punctatus DAOM BR117*                    | 9422     | 1     | 0     | 0   | 0   | 1   | 0   |
| **Dothideomycetes**                                    |          |       |       |     |     |     |     |     |
| *Alternaria alternata*                                  | 13466    | 3     | 2     | 1   | 1   | 2   | 3   |
| *Aureobasidium namibiae CBS 147.97*                     | 10259    | 1     | 1     | 1   | 3   | 3   |
| *Aureobasidium subglaciale EXP-2481*                    | 10792    | 1     | 1     | 2   | 3   |
| *Baudoinia panamericana UAMH 10762*                     | 10508    | 1     | 1     | 1   | 2   | 2   |
| *Bipolaris maydis ATCC 48331*                           | 12705    | 2     | 2     | 1   | 1   |
| *Bipolaris oryzae ATCC 44560*                           | 12002    | 2     | 2     | 1   | 1   |
| *Bipolaris sorokiniana ND90Pr*                          | 12214    | 2     | 2     | 1   |
| *Bipolaris victoriae F13*                               | 12882    | 2     | 2     | 1   |
| *Bipolaris zeicola 26-R-13*                             | 12853    | 2     | 2     | 1   |
| *Cercospora beticola*                                   | 12463    | 1     | 1     | 1   |
| *Coniothyrium apollinis CBS 100218*                     | 9308     | 1     | 1     | 2   |
| *Diplodia corticola*                                    | 10839    | 1     | 1     | 1   |
| *Exserohilum turcica Et28A*                             | 11698    | 2     | 2     | 1   |
| *Leptosphaeria maculans JN3*                            | 12469    | 1     | 1     | 0   |
| *Paraphaeosphaeria sporulosa*                           | 14734    | 1     | 1     | 1   |
| *Parastagonospora nodorum SN15*                          | 15994    | 2     | 2     |
| *Pseudocercospora fijensis CIRAD86*                     | 13066    | 1     |
| *Pyrenophora tritici-repentis Pt-1C-BFP*                 | 12169    | 2     |
| *Ramularia collo-cynni*                                 | 11612    | 1     |
| *Sphaerulina musiva SO2202*                             | 10156    | 1     |
| *Verruconis gallopava*                                  | 11357    | 2     |
| *Zymoseptoria tritici IPO323*                            | 10963    | 0     | 1     |
| Species                                    | Proteins | qa-1F | qa-1S | qa2 | qa3 | qa4 | qaX | qaY |
|--------------------------------------------|----------|-------|-------|-----|-----|-----|-----|-----|
| *Eurotiomycetes*                           |          |       |       |     |     |     |     |     |
| *Aspergillus aculeatinus* CBS 121060       | 12028    | 3     | 3     | 1   | 2   | 4   | 1   | 24  |
| *Aspergillus aculeatus* ATCC 16872         | 10843    | 4     | 3     | 1   | 2   | 5   | 1   | 25  |
| *Aspergillus bombycis*                     | 12263    | 5     | 4     | 3   | 4   | 7   | 2   | 28  |
| *Aspergillus brunneoviolaceus* CBS 621.78  | 12073    | 3     | 3     | 1   | 1   | 5   | 1   | 25  |
| *Aspergillus campestris* IBT 28561         | 9756     | 2     | 2     | 1   | 1   | 3   | 1   | 11  |
| *Aspergillus candidus*                      | 9639     | 1     | 2     | 1   | 1   | 3   | 1   | 11  |
| *Aspergillus clavatus* NRRL 1              | 9121     | 2     | 2     | 1   | 1   | 4   | 2   | 14  |
| *Aspergillus costaricaensis* CBS 115574    | 11966    | 3     | 3     | 1   | 1   | 6   | 2   | 27  |
| *Aspergillus eucalyptica* CBS 122712       | 11933    | 3     | 3     | 1   | 2   | 5   | 2   | 23  |
| *Aspergillus fischeri* NRRL 181            | 10395    | 3     | 2     | 2   | 2   | 4   | 2   | 17  |
| *Aspergillus flavus* NRRL3357              | 13485    | 2     | 3     | 2   | 3   | 4   | 2   | 27  |
| *Aspergillus fumigatus* AF293              | 9630     | 3     | 3     | 2   | 2   | 5   | 2   | 15  |
| *Aspergillus glauca* CBS 516.65            | 11255    | 3     | 3     | 2   | 2   | 5   | 2   | 15  |
| *Aspergillus heteromorphus* CBS 117.55     | 11130    | 3     | 2     | 1   | 1   | 5   | 1   | 16  |
| *Aspergillus homomorphus* CBS 101889       | 11361    | 2     | 2     | 1   | 1   | 4   | 1   | 18  |
| *Aspergillus ibericus* CBS 121593          | 11680    | 2     | 2     | 1   | 2   | 4   | 1   | 24  |
| *Aspergillus japonicus* CBS 114.51         | 12022    | 3     | 3     | 0   | 2   | 5   | 1   | 25  |
| *Aspergillus lacticoffeatus* CBS 101883    | 13082    | 3     | 3     | 1   | 1   | 5   | 2   | 25  |
| *Aspergillus mulundensis*                   | 11603    | 3     | 3     | 1   | 2   | 3   | 1   | 27  |
| *Aspergillus neoerg* CBS 115656            | 11939    | 3     | 3     | 1   | 2   | 5   | 2   | 26  |
| *Aspergillus nidulans* FGSC A4              | 9556     | 2     | 2     | 1   | 1   | 3   | 1   | 22  |
| *Aspergillus niger* CBS 513.88             | 10593    | 4     | 3     | 1   | 2   | 5   | 2   | 25  |
| *Aspergillus nomius* NRRL 13137             | 11904    | 3     | 4     | 2   | 4   | 6   | 2   | 30  |
| *Aspergillus novofumigatus* IBT 16806       | 11534    | 2     | 2     | 2   | 1   | 3   | 2   | 17  |
| *Aspergillus oryzae* RIB40                  | 12074    | 2     | 3     | 1   | 3   | 5   | 2   | 27  |
| *Aspergillus piperis* CBS 112811           | 12071    | 3     | 3     | 1   | 1   | 6   | 2   | 27  |
| *Aspergillus saccharolyticus* JOP 1030-1    | 10064    | 3     | 3     | 1   | 2   | 4   | 1   | 17  |
| *Aspergillus sclerotioring* CBS 115572      | 12338    | 3     | 3     | 1   | 2   | 5   | 1   | 23  |
| *Aspergillus steynii* IBT 23096             | 13197    | 3     | 3     | 1   | 1   | 4   | 1   | 24  |
| *Aspergillus terreus* NIH2624               | 10401    | 2     | 3     | 2   | 1   | 6   | 2   | 26  |
| *Aspergillus thermomutatus*                 | 9702     | 2     | 3     | 1   | 1   | 5   | 2   | 18  |
| *Aspergillus uravum* CBS 121591             | 12014    | 4     | 3     | 1   | 2   | 5   | 1   | 23  |
| *Aspergillus vadensis* CBS 113365           | 12132    | 3     | 3     | 1   | 1   | 6   | 2   | 25  |
| *Aspergillus welwitschiae*                  | 13684    | 3     | 3     | 1   | 1   | 5   | 2   | 25  |
| *Blastomyces gilchristi* SLHI4081           | 9587     | 1     | 0     | 0   | 0   | 3   | 1   | 2   |
| *Capronia coronata* CBS 617.96              | 9231     | 1     | 1     | 1   | 1   | 2   | 2   | 5   |
| *Capronia epymycses* CBS 606.96             | 10469    | 3     | 3     | 2   | 4   | 5   | 3   | 8   |
| *Cladophialophora bantiana* CBS 173.52      | 12762    | 1     | 1     | 1   | 3   | 8   | 2   | 17  |
| *Cladophialophora carrionii* CBS 160.54     | 10373    | 1     | 1     | 1   | 2   | 4   | 2   | 12  |
| *Cladophialophora immunda*                  | 14033    | 1     | 1     | 1   | 3   | 4   | 2   | 17  |
| *Cladophialophora psammophila* CBS 110553   | 13421    | 1     | 1     | 1   | 3   | 7   | 2   | 19  |
| *Cladophialophora yegresii* CBS 114405      | 10118    | 1     | 1     | 1   | 2   | 3   | 2   | 12  |
| *Coccidioides immittis* RS                   | 9910     | 1     | 1     | 0   | 0   | 0   | 1   | 1   |
| *Coccidioides posadasii* C735 delta SOWgp   | 7226     | 1     | 1     | 0   | 0   | 0   | 1   | 1   |
| *Cyphellophora europaea* CBS 101466         | 11094    | 3     | 3     | 2   | 4   | 4   | 2   | 12  |
| *Endocarpon pusillum* Z07020                | 9238     | 1     | 1     | 1   | 1   | 3   | 1   | 0   |
| Species                      | Proteins | qa-1F | qa-1S | qa2 | qa3 | qa4 | qaX | qaY |
|------------------------------|----------|-------|-------|-----|-----|-----|-----|-----|
| Exophiala aquamarina CBS 119918 | 13118    | 2     | 2     | 2   | 3   | 7   | 3   | 23  |
| Exophiala dermatitidis NIH/UT8656 | 9578     | 0     | 0     | 0   | 0   | 1   | 1   | 5   |
| Exophiala mesophila           | 10347    | 1     | 0     | 1   | 0   | 1   | 2   | 5   |
| Exophiala oligosperma         | 13234    | 3     | 4     | 2   | 1   | 5   | 1   | 18  |
| Exophiala spinifera           | 12049    | 4     | 3     | 2   | 4   | 7   | 3   | 22  |
| Exophiala xenobiotica         | 13187    | 2     | 2     | 1   | 2   | 5   | 3   | 16  |
| Fonsecaea erecta             | 12090    | 1     | 1     | 1   | 3   | 5   | 2   | 23  |
| Fonsecaea monophora          | 11984    | 1     | 1     | 0   | 1   | 3   | 2   | 22  |
| Fonsecaea polymorpha CBS 102226 | 12369   | 4     | 4     | 1   | 5   | 8   | 3   | 21  |
| Fonsecaea rubica             | 11681    | 1     | 1     | 0   | 1   | 3   | 1   | 19  |
| Fonsecaea rubicola CBS 271.37 | 12527    | 1     | 1     | 0   | 1   | 3   | 2   | 24  |
| Histoplasma capsulatum NAm1  | 9313     | 1     | 0     | 0   | 0   | 1   | 1   | 2   |
| Microsporum canis CBS 113480 | 8765     | 0     | 0     | 0   | 0   | 0   | 1   | 2   |
| Nannizzia gypsea CBS 118893  | 8921     | 0     | 0     | 0   | 0   | 0   | 1   | 3   |
| Paracoccidioides brasiliensis Pb18 | 8390  | 0     | 0     | 0   | 0   | 1   | 1   | 2   |
| Paracoccidioides lutzii Pb01 | 8826     | 0     | 0     | 0   | 0   | 1   | 1   | 2   |
| Penicilliopsis zonata CBS 506.65 | 9870  | 2     | 2     | 1   | 1   | 3   | 2   | 13  |
| Penicillium arizense         | 12200    | 2     | 3     | 1   | 2   | 8   | 2   | 34  |
| Penicillium digitatum Pd1    | 8946     | 2     | 3     | 1   | 1   | 3   | 2   | 11  |
| Penicillium expansum         | 11060    | 2     | 2     | 1   | 1   | 6   | 2   | 19  |
| Penicillium rubens Wisconsin 54-1255 | 12791 | 2     | 2     | 1   | 1   | 6   | 2   | 23  |
| Phialophora attae            | 11848    | 1     | 4     | 1   | 3   | 6   | 2   | 23  |
| Rasamsonia emersonii CBS 393.64 | 9843   | 2     | 2     | 0   | 1   | 2   | 2   | 13  |
| Rhinocladiella mackenzie CBS 650.93 | 11382 | 2     | 2     | 1   | 6   | 6   | 2   | 18  |
| Talaromyces atroroseus       | 9523     | 2     | 2     | 1   | 2   | 3   | 1   | 16  |
| Talaromyces marneffei ATCC 18224 | 10638  | 2     | 2     | 1   | 2   | 3   | 2   | 11  |
| Talaromyces stipitatus ATCC 10500 | 13252 | 2     | 2     | 0   | 2   | 3   | 2   | 13  |
| Trichophyton benhamiae CBS 112371 | 7974  | 0     | 0     | 0   | 0   | 0   | 1   | 3   |
| Trichophyton rubrum CBS 118892 | 8706   | 0     | 0     | 0   | 0   | 0   | 1   | 3   |
| Trichophyton verrucosum HKI 0517 | 8028  | 0     | 0     | 0   | 0   | 0   | 1   | 3   |
| Uncinocarpus reesii 1704     | 7760     | 1     | 1     | 0   | 0   | 2   | 1   | 1   |

Exobasidiomycetes

Acaromyces ingoldii | 8026 | 0 | 0 | 1 | 0 | 2 | 1 | 12 |
Ceraceosorus guamensis | 7822 | 0 | 0 | 1 | 0 | 1 | 2 | 2 |
Jaminiaea rosea | 6858 | 0 | 0 | 0 | 0 | 0 | 1 | 5 |
Meira miltonrushii | 7452 | 1 | 0 | 1 | 0 | 3 | 2 | 8 |
Pseudomicrostroma glucosiphilum | 6681 | 0 | 0 | 1 | 1 | 1 | 2 | 5 |
Tilletiaria anomala UBC 951 | 6808 | 0 | 0 | 1 | 1 | 1 | 2 | 5 |
Tilletiopsis washingtonensis | 7007 | 0 | 0 | 1 | 0 | 1 | 1 | 4 |

Glomeromycetes

Rhizophagus irregularis DAOM 181602 | 26147 | 0 | 0 | 0 | 0 | 0 | 1 | 1 |

Leotiomycetes

Amorphotheca resinae ATCC 22711 | 9642 | 1 | 1 | 0 | 0 | 1 | 1 | 6 |
Botrytis cinerea B05.10 | 13703 | 1 | 1 | 1 | 1 | 3 | 2 | 15 |
### Species Proteins qa-1F qa-1S qa2 qa3 qa4 qaX qaY

| Species | Proteins | qa-1F | qa-1S | qa2 | qa3 | qa4 | qaX | qaY |
|---------|----------|-------|-------|-----|-----|-----|-----|-----|
| Glarea lozoyensis ATCC 20868 | 13083 | 1 | 1 | 1 | 1 | 3 | 2 | 9 |
| Marssonina brunnea f. sp. 'multigerntubi' MB_m1 | 10027 | 1 | 1 | 1 | 2 | 1 | 2 | 9 |
| Meliilomyces bicolor E | 18617 | 1 | 1 | 0 | 1 | 4 | 2 | 21 |
| Phialocephala scopiformis | 18567 | 1 | 1 | 1 | 3 | 4 | 2 | 20 |
| Pseudogymnoascus destructans | 9420 | 1 | 1 | 0 | 0 | 1 | 1 | 5 |
| Pseudogymnoascus verrucosus | 10573 | 2 | 2 | 1 | 1 | 3 | 2 | 24 |
| Sclerotinia sclerotiorum 1980 UF-70 | 14490 | 1 | 1 | 1 | 1 | 2 | 2 | 10 |

### Malasseziomycetes

| Species | Proteins | qa-1F | qa-1S | qa2 | qa3 | qa4 | qaX | qaY |
|---------|----------|-------|-------|-----|-----|-----|-----|-----|
| Malassezia globosa CBS 7966 | 4286 | 0 | 0 | 0 | 0 | 0 | 1 | 0 |
| Malassezia pachydermatis | 4202 | 0 | 0 | 0 | 0 | 0 | 1 | 0 |
| Malassezia restricta | 4406 | 0 | 0 | 0 | 0 | 0 | 1 | 0 |
| Malassezia sympodialis ATCC 42132 | 3318 | 0 | 0 | 0 | 0 | 0 | 1 | 0 |
| Rhodotorula graminis WP1 | 7278 | 0 | 0 | 1 | 2 | 2 | 2 | 5 |
| Rhodotorula toruloides NP11 | 8140 | 0 | 0 | 1 | 2 | 3 | 2 | 6 |

### Microsporidia

| Species | Proteins | qa-1F | qa-1S | qa2 | qa3 | qa4 | qaX | qaY |
|---------|----------|-------|-------|-----|-----|-----|-----|-----|
| Encephalitozoon cuniculi GB-M1 | 1971 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| Encephalitozoon hellem ATCC 50504 | 1847 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| Encephalitozoon intestinalis ATCC 50506 | 1938 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| Encephalitozoon romaleae SJ-2008 | 1831 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| Mitosporidium daphniae | 3330 | 0 | 0 | 0 | 0 | 0 | 1 | 0 |
| Nematocida parisi ERTm1 | 2661 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| Nosema ceranae | 3209 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| Ordospora colligata OC4 | 1820 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| Vavraia culicis subsp. floridensis | 2773 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| Vittaforma cornae ATCC 50505 | 2239 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |

### Mixiomycetes

| Species | Proteins | qa-1F | qa-1S | qa2 | qa3 | qa4 | qaX | qaY |
|---------|----------|-------|-------|-----|-----|-----|-----|-----|
| Mixia osmundae IAM 14324 | 6858 | 0 | 0 | 0 | 0 | 1 | 2 | 0 |

### Mortierellomycetes

| Species | Proteins | qa-1F | qa-1S | qa2 | qa3 | qa4 | qaX | qaY |
|---------|----------|-------|-------|-----|-----|-----|-----|-----|
| Lobosporangium transversale | 11822 | 0 | 0 | 0 | 0 | 1 | 0 | 0 |

### Mucoromycetes

| Species | Proteins | qa-1F | qa-1S | qa2 | qa3 | qa4 | qaX | qaY |
|---------|----------|-------|-------|-----|-----|-----|-----|-----|
| Phycomyces blakesleeanus NRRL 1555(+) | 16543 | 0 | 0 | 0 | 0 | 0 | 1 | 6 |
| Rhizopus microsporus ATCC 52813 | 10891 | 0 | 0 | 0 | 0 | 0 | 1 | 4 |

### Orbiliomycetes

| Species | Proteins | qa-1F | qa-1S | qa2 | qa3 | qa4 | qaX | qaY |
|---------|----------|-------|-------|-----|-----|-----|-----|-----|
| Arthrobotrys oligospora ATCC 24927 | 11479 | 0 | 0 | 0 | 0 | 2 | 6 | 0 |

### Pezizomycetes

| Species | Proteins | qa-1F | qa-1S | qa2 | qa3 | qa4 | qaX | qaY |
|---------|----------|-------|-------|-----|-----|-----|-----|-----|
| Tuber melanosporum Mel28 | 7496 | 0 | 0 | 0 | 0 | 1 | 1 | 0 |

### Pneumocystidomycetes

| Species | Proteins | qa-1F | qa-1S | qa2 | qa3 | qa4 | qaX | qaY |
|---------|----------|-------|-------|-----|-----|-----|-----|-----|
| Pneumocystis carinii B80 | 3646 | 0 | 0 | 0 | 0 | 0 | 0 | 1 |
| Pneumocystis jiroveci RU7 | 3761 | 0 | 0 | 0 | 0 | 0 | 0 | 1 |
| Pneumocystis murina B123 | 3623 | 0 | 0 | 0 | 0 | 0 | 0 | 1 |
| Species | Proteins | qa-1F | qa-1S | qa2 | qa3 | qa4 | qaX | qaY |
|---------|----------|-------|-------|-----|-----|-----|-----|-----|
| **Pucciniozymes** | | | | | | | | |
| Melampsora larici-populina 98AG31 | 16372 | 0 | 0 | 0 | 1 | 1 | 2 |
| Puccinia graminis f. sp. tritici CRL 75-36-700-3 | 15979 | 0 | 0 | 0 | 1 | 1 | 1 |
| **Saccharomyces** | | | | | | | | |
| Ascoidea rubescens DSM 1968 | 6787 | 0 | 0 | 0 | 1 | 0 | 0 |
| Babjeviella inositovora NRRL Y-12698 | 6399 | 1 | 1 | 1 | 1 | 1 | 3 |
| Candida albicans SC5314 | 6043 | 0 | 0 | 1 | 0 | 0 | 4 |
| Candida auris | 7461 | 0 | 0 | 1 | 1 | 0 | 0 |
| Candida dubliniensis CD36 | 5859 | 0 | 0 | 1 | 0 | 0 | 4 |
| Candida duobushaemulonis | 5173 | 0 | 0 | 1 | 1 | 0 | 0 |
| Candida glabrata | 5202 | 0 | 0 | 0 | 0 | 2 | 0 |
| Candida haemulonid | 5249 | 0 | 0 | 1 | 1 | 0 | 0 |
| Candida orthopsilosis Co 90-125 | 5678 | 1 | 0 | 0 | 0 | 0 | 3 |
| Candida pseudoaemulonid | 5134 | 0 | 0 | 1 | 1 | 0 | 0 |
| Candida tropicalis MYA-3404 | 6254 | 0 | 0 | 1 | 0 | 0 | 6 |
| Candida viswanathii | 10857 | 0 | 0 | 2 | 0 | 0 | 1 |
| Clavispora lusitaniae ATCC 42720 | 5936 | 0 | 0 | 1 | 0 | 0 | 3 |
| Cyblerindinae jadinii NRRL Y-1542 | 6032 | 0 | 0 | 1 | 1 | 0 | 1 |
| Debaryomyces fabryi | 6027 | 1 | 1 | 1 | 1 | 1 | 8 |
| Debaryomyces hanseni CBS767 | 6268 | 1 | 1 | 1 | 1 | 1 | 8 |
| Eremothecium cymbalariae DBVPG#7215 | 4432 | 0 | 0 | 0 | 0 | 0 | 0 |
| Eremothecium gossypii ATCC 10895 | 4776 | 0 | 0 | 0 | 0 | 0 | 0 |
| Eremothecium sinecaudum | 4536 | 0 | 0 | 0 | 0 | 0 | 0 |
| Hyphopichia burtonii NRRL Y-1933 | 5996 | 0 | 0 | 1 | 1 | 0 | 0 |
| Kazachstania africana CBS 2517 | 5375 | 0 | 0 | 0 | 0 | 2 | 0 |
| Kazachstania naganishii CBS 8797 | 5319 | 0 | 0 | 0 | 0 | 2 | 0 |
| Klyveromyces lactis | 5085 | 0 | 0 | 0 | 0 | 0 | 0 |
| Klyveromyces marxianus DMKU3-1042 | 4952 | 0 | 0 | 0 | 0 | 0 | 0 |
| Komagataella phaffii GS115 | 5040 | 0 | 0 | 0 | 0 | 1 | 1 |
| Kuraishia capsulata CBS 1993 | 5989 | 0 | 0 | 0 | 0 | 1 | 3 |
| Lachancea lantarotensis | 5056 | 0 | 0 | 0 | 0 | 2 | 1 |
| Lachancea thermotolerans CBS 6340 | 5092 | 0 | 0 | 0 | 0 | 2 | 1 |
| Lodderomyces elongisporus NRRL YB-4239 | 5799 | 0 | 0 | 0 | 0 | 0 | 2 |
| Metschnikowia bicuspidata var. bicuspidata NRRL YB-4993 | 5838 | 0 | 0 | 1 | 0 | 0 | 1 |
| Meyerozyma guilliermondii ATCC 6260 | 5920 | 0 | 0 | 1 | 0 | 0 | 3 |
| Naumovozyma castellii CBS 4309 | 5589 | 0 | 0 | 0 | 0 | 2 | 0 |
| Naumovozyma dairenensis CBS 421 | 5546 | 0 | 0 | 0 | 0 | 2 | 0 |
| Ogataea parapolymorpha DL-1 | 5325 | 0 | 0 | 0 | 1 | 0 | 2 |
| Ogataea polymorpha | 5173 | 0 | 0 | 0 | 1 | 0 | 2 |
| Pichia kudriavzevii | 5385 | 0 | 0 | 0 | 0 | 0 | 0 |
| Pichia membranifaciens NRRL Y-2026 | 5542 | 0 | 0 | 0 | 0 | 0 | 0 |
| Saccharomyces cerevisiae S288C | 6002 | 0 | 0 | 0 | 0 | 2 | 0 |
| Saccharomyces eubayanus | 5377 | 0 | 0 | 0 | 0 | 2 | 0 |
| Scheffersomyces stipitis CBS 6054 | 5818 | 1 | 1 | 1 | 1 | 0 | 6 |
| Spasthaspora passalidarum NRRL Y-27907 | 5983 | 1 | 0 | 1 | 0 | 0 | 4 |
| Sugiyamaella lignohabitans | 5135 | 2 | 1 | 0 | 0 | 1 | 3 |
| Species                        | Proteins | qa-1F | qa-1S | qa2 | qa3 | qa4 | qaX | qaY |
|-------------------------------|----------|-------|-------|-----|-----|-----|-----|-----|
| Suhomyces tanzawaensis NRRL Y-17324 | 5885     | 0     | 0     | 1   | 0   | 0   | 0   | 3   |
| Tetrapispora blattae CBS 6284  | 5388     | 0     | 0     | 0   | 0   | 0   | 2   | 0   |
| Tetrapispora phaffii CBS 4417  | 5252     | 0     | 0     | 0   | 0   | 0   | 2   | 0   |
| Torulaspora delbrueckii        | 4978     | 0     | 0     | 0   | 0   | 0   | 1   | 1   |
| Vanderwaltozyma polyspora DSM 70294 | 5367   | 0     | 0     | 0   | 0   | 0   | 2   | 0   |
| Wickerhamiella sorbophila      | 4740     | 0     | 0     | 0   | 0   | 0   | 1   | 0   |
| Wickerhamomyces anomalous NRRL Y-366-8 | 6421 | 1     | 0     | 2   | 3   | 1   | 2   | 6   |
| Wickerhamomyces ciferri        | 6702     | 0     | 0     | 1   | 1   | 0   | 1   | 5   |
| Yamadazyma tenuis ATCC 10573   | 6985     | 0     | 0     | 1   | 1   | 0   | 0   | 6   |
| Yarrowia lipolytica CLIB122     | 6448     | 0     | 0     | 0   | 0   | 0   | 1   | 4   |
| Zygosaccharomyces rouxii       | 4991     | 0     | 0     | 0   | 0   | 0   | 2   | 0   |
| Schizosaccharomycesetes        |          |       |       |     |     |     |     |     |
| Schizosaccharomyces cryophilus OY26 | 5180 | 0     | 0     | 0   | 0   | 0   | 0   | 0   |
| Schizosaccharomyces japonicus yFS275 | 4878 | 0     | 0     | 0   | 0   | 0   | 1   | 0   |
| Schizosaccharomyces octosporus yFS286 | 4986 | 0     | 0     | 0   | 0   | 0   | 0   | 0   |
| Schizosaccharomyces pombe       | 5132     | 0     | 0     | 0   | 0   | 0   | 0   | 0   |
| Sordariomycetes                |          |       |       |     |     |     |     |     |
| Beauveria bassiana ARSEF 2860  | 10364    | 0     | 0     | 0   | 1   | 3   | 1   | 8   |
| Chaetomium globosum CBS 148.51 | 11048    | 1     | 2     | 0   | 1   | 2   | 2   | 6   |
| Chaetomium thermophilum var. thermophilum DSM 1495 | 7179 | 0     | 2     | 1   | 1   | 1   | 2   | 4   |
| Colletotrichum graminicola M1.001 | 12020 | 1     | 1     | 1   | 1   | 2   | 2   | 16  |
| Colletotrichum higginsianum IMI 349063 | 14650 | 1     | 1     | 1   | 1   | 2   | 2   | 22  |
| Colletotrichum orchidophilum    | 14453    | 1     | 2     | 1   | 1   | 1   | 2   | 15  |
| Cordyceps fumosorosea ARSEF 2679 | 10061 | 1     | 1     | 1   | 1   | 1   | 1   | 10  |
| Cordyceps militaris CM01       | 9651     | 1     | 1     | 1   | 1   | 2   | 1   | 8   |
| Fusarium fujikuroi IMI 58289   | 14810    | 2     | 2     | 1   | 2   | 3   | 2   | 24  |
| Fusarium graminearum PH-1      | 13313    | 1     | 2     | 1   | 3   | 3   | 3   | 18  |
| Fusarium oxysporum f. sp. lycopersici 4287 | 27347 | 3     | 4     | 3   | 1   | 3   | 2   | 20  |
| Fusarium pseudograminearum CS3096 | 12397 | 2     | 2     | 1   | 3   | 3   | 3   | 17  |
| Fusarium venenatum             | 13932    | 2     | 2     | 1   | 3   | 3   | 2   | 21  |
| Fusarium verticillioides 7600  | 20553    | 2     | 2     | 0   | 2   | 3   | 2   | 24  |
| Gaeumannomyces tritici R3-111a-1 | 14650 | 1     | 2     | 1   | 1   | 1   | 2   | 12  |
| Grosmaniella claviger aki1407   | 8312     | 0     | 1     | 0   | 0   | 0   | 1   | 6   |
| Metarhizium acridum CQMa 102   | 9849     | 1     | 1     | 1   | 1   | 3   | 2   | 9   |
| Metarhizium brunneum ARSEF 3297 | 10689 | 1     | 1     | 1   | 1   | 2   | 2   | 7   |
| Metarhizium robertsi ARSEF 23  | 11688    | 1     | 1     | 1   | 1   | 3   | 2   | 7   |
| Nectria haematococca mpV1 77-13-4 | 15708 | 4     | 3     | 2   | 2   | 5   | 2   | 23  |
| Neurospora crassa OR74A         | 10812    | 1     | 2     | 1   | 1   | 2   | 2   | 4   |
| Neurospora tetrasperma FGSC 2508 | 10380 | 1     | 2     | 1   | 1   | 2   | 2   | 4   |
| Pestalotiopsis fici W106-1      | 15413    | 2     | 3     | 1   | 3   | 5   | 2   | 24  |
| Phaeoacremonium minimum UCRPA7 | 8834     | 0     | 1     | 0   | 0   | 0   | 1   | 18  |
| Pochonia chalmydosporia 170     | 14204    | 1     | 1     | 1   | 1   | 4   | 2   | 15  |
| Podospora anserina S mat+       | 10518    | 1     | 2     | 1   | 1   | 2   | 2   | 4   |
| Purpureocillium lilacinum       | 11763    | 1     | 1     | 1   | 1   | 1   | 2   | 14  |
| Pyricularia oryzae 70-15        | 12989    | 1     | 2     | 1   | 1   | 1   | 2   | 11  |
| Species                          | Proteins | qa-1F | qa-1S | qa2 | qa3 | qa4 | qaX | qaY |
|---------------------------------|----------|-------|-------|-----|-----|-----|-----|-----|
| Scedosporium apiospermum        | 8375     | 0     | 1     | 0   | 1   | 1   | 1   | 7   |
| Sordaria macrospora k-hell      | 9896     | 1     | 2     | 1   | 1   | 2   | 2   | 4   |
| Sporothrix schenckii 1099-18    | 10293    | 1     | 2     | 0   | 2   | 4   | 2   | 8   |
| Thermotheomyces thermophilus ATCC 42464 | 9097 | 1     | 2     | 1   | 1   | 2   | 2   | 5   |
| Thielavia terrestris NRRL 8126  | 9802     | 1     | 2     | 1   | 1   | 2   | 2   | 7   |
| Trichoderma asperellum CBS 433.97 | 12557  | 1     | 1     | 1   | 1   | 2   | 1   | 13  |
| Trichoderma atroviride IMI 206040 | 11816  | 1     | 1     | 1   | 1   | 2   | 1   | 16  |
| Trichoderma citrinoviride       | 9735     | 1     | 3     | 1   | 1   | 2   | 1   | 12  |
| Trichoderma gamsii              | 11189    | 0     | 1     | 1   | 1   | 3   | 1   | 15  |
| Trichoderma harzianum CBS 226.95 | 14065  | 1     | 1     | 1   | 1   | 3   | 1   | 15  |
| Trichoderma reeset QM6a         | 9115     | 1     | 1     | 1   | 1   | 2   | 1   | 12  |
| Trichoderma virens Gr29-8       | 12406    | 1     | 1     | 1   | 1   | 2   | 2   | 13  |
| Verticillium alfalfae VaMs.102  | 10237    | 1     | 1     | 1   | 1   | 2   | 1   | 14  |
| Verticillium dahliae VdLs.17    | 10535    | 0     | 1     | 1   | 1   | 2   | 2   | 14  |
| Taphrinomycotina                |          |       |       |     |     |     |     |     |
| Saitoella complicata NRRL Y-17804 | 7034   | 0     | 0     | 0   | 1   | 1   | 2   | 2   |
| Tremellomycetes                 |          |       |       |     |     |     |     |     |
| Cryptococcus amyloleustus CBS 6039 | 10306  | 0     | 0     | 0   | 0   | 2   | 2   | 8   |
| Cryptococcus gattii WM276        | 6561     | 0     | 0     | 0   | 0   | 5   | 2   | 14  |
| Cryptococcus neoformans var. grubii H99 | 7826 | 0     | 0     | 0   | 0   | 6   | 3   | 17  |
| Cryptococcus neoformans var. neoformans B-3501A | 6578 | 0     | 0     | 0   | 0   | 3   | 2   | 16  |
| Cryptococcus neoformans var. neoformans JEC21 | 6863 | 0     | 0     | 0   | 0   | 4   | 2   | 14  |
| Cutaneotrichosporon oleaginosum | 8320     | 0     | 0     | 1   | 0   | 1   | 2   | 6   |
| Kockovaella imperatae           | 7392     | 0     | 0     | 1   | 0   | 0   | 1   | 5   |
| Kwoniiella bestiolar CBS 10118  | 9133     | 0     | 0     | 1   | 0   | 3   | 2   | 8   |
| Kwoniiella dejecticola CBS 10117 | 8602   | 0     | 0     | 0   | 0   | 2   | 2   | 9   |
| Kwoniiella mangroviensis CBS 8507 | 8422  | 0     | 0     | 0   | 0   | 2   | 2   | 7   |
| Kwoniiella pini CBS 10737       | 7829     | 0     | 0     | 0   | 0   | 1   | 2   | 4   |
| Tremella mesenterica DSM 1558   | 8308     | 0     | 0     | 0   | 0   | 0   | 1   | 1   |
| Trichosporon asahii var. asahii CBS 2479 | 8311  | 0     | 0     | 0   | 0   | 0   | 0   | 7   |
| Tsuchiyaea wingfieldii CBS 7118 | 8094     | 0     | 0     | 0   | 0   | 2   | 2   | 4   |
| Ustilaginomycetes               |          |       |       |     |     |     |     |     |
| Anthracocystis floculosa PF-1    | 6877     | 0     | 0     | 1   | 1   | 2   | 2   | 6   |
| Kalmanomyza brasiliensis GHG001  | 5765     | 0     | 0     | 0   | 0   | 1   | 1   | 6   |
| Moesziomyces antarcticus        | 6766     | 0     | 0     | 1   | 0   | 2   | 1   | 7   |
| Pseudozyma hubeiensis SY62      | 7472     | 0     | 0     | 1   | 0   | 3   | 1   | 8   |
| Ustilago maydis 521             | 6782     | 1     | 0     | 1   | 0   | 2   | 1   | 4   |
| Wallemiomycetes                 |          |       |       |     |     |     |     |     |
| Wallemia ichthyophaga EXF-994   | 4863     | 0     | 0     | 0   | 0   | 1   | 0   | 1   |
| Wallemia mellicola CBS 633.66   | 5277     | 0     | 0     | 0   | 0   | 1   | 0   | 1   |
| Xylonomycetes                   |          |       |       |     |     |     |     |     |
| Xylona heveae TC161             | 8201     | 0     | 0     | 0   | 0   | 1   | 1   | 2   |
3 Results

3.1 A bioinformatics survey of quinic acid utilization (QUT) genes in fungi

Using the predicted protein sequences of the products of the QA cluster genes located on chromosome VII in *N. crassa* as queries to search *N. crassa* proteome, we found that qa-2, qa-3 and qa-1F genes have only a single copy, however, more than one copy of genes homologous to qa-X, qa-4, qa-Y and qa-1S exist in the genome.

The qa-X gene in *N. crassa* encodes a protein (XP_959612.1, 340 amino acids) that shares homology with inositol monophosphatase (IMPase) and other proteins with related domains (domain architecture ID 10108155). IMPase catalyzes the hydrolysis of several inositol monophosphates and the artificial substrate p-nitrophenyl-phosphate to inorganic phosphate and inositol. A gene located on chromosome IV encoding a QA-X homologous protein sequence (XP_962382.1, 305 amino acids) was identified, which was annotated as myo-inositol-1-monophosphatase, also contains IMPase domain. Strains of *N. crassa* containing a disruption of the qa-X gene are still capable of growing on quinic acid as a sole carbon source and have no detectable phenotype except production of a brown pigment during growth on quinic acid (Case et al., 1992).

The qa-3 gene encodes shikimate/quinate 5-dehydrogenase (XP_959615.3), with a length of 339 amino acids. There is only a single copy in *N. crassa* genome. However, BLASTP search revealed that XP_956000.1, encoded by *aro-1* gene, has 1563 amino acids, shares 33% identity and 49% similarity in its carboxyl terminal end of 281 residues with QA-3 protein sequences. This region (1274-1548) contains a domain of shikimate-5-dehydrogenase, fungal AROM-type (conserved domain accession: cl36977). In addition, QA-3 also shares similarities with two repressor proteins in their carboxyl residues, XP_959617.2 (918 amino acids, 27% identity and 40% similarity over 340 aligned residues) and XP_955830.2 (803 amino acids, 24% identity and 40% similarity over 330 aligned residues). XP_959617.2 is encoded by qa-1S gene, XP_955830.2 is a homolog of XP956167.2. Both of them have a Type I 3-dehydroquinase (3-dehydroquinate dehydratase or DHQase) domain and a shikimate 5-dehydrogenase domain in their carboxyl terminus. The gene (locus tag NCU04358) encoding XP_955830.2 is located on chromosome IV. These results suggest qa-3, aro-1, qa-1S, and NCU04358 might be evolutionarily related in *N. crassa*.

The qa-4 gene encodes 3-dehydroshikimate dehydratase (DHS dehydratase) (XP_959614.1) with a length of 359 amino acids. A homologous gene (locus tag: NCU00838), located on chromosome I, also encodes 3-DHS dehydratase (XP_963958.1), with a length of 340 amino acids. These two proteins share 29% identity, 48% similarity, and 15% gaps over 378 aligned residues in a global alignment.

There are four homologous QA-Y protein sequences. qa-Y gene, located on chromosome VII, encodes quinate permease (XP_959616.1) with a length of 537 amino acids. Other three homologs include XP_963898.1 (537 AAs, quinate transporter, on chromosome I), XP_960000.1 (583 AAs, quinate permease, on chromosome VII), and XP_960547.2 (565 AAs, MFS quinate transporter, on chromosome VI). These four homologs share ~30% identity and ~50% similarity in their alignment. However, deletion of the qa-Y gene prevents growth on quinic acid as a sole carbon source (Case et al., 1992).

In the fungal dataset of the RefSeq database, there were a total of 285 completely sequenced fungal genomes, consisting of 282 unique species. There were 117 fungal species having homologs for all seven QA genes, i. e. at least one homolog for each of the seven QA genes (Table 1). The retrieved protein sequences for each QUT/QA gene were available for downloading (http://proteomics.ysu.edu/publication/data/QAClusters/).

The presence or absence of QUT genes in different fungi were summarized based on the classification of Classes and Orders (Table 2). Species having QUT genes all belong to Ascomycota phylum. Among 21 species in Dothideomycetes, 19 species have genes homologous to all 7 QA genes, except *Zymoseptoria tritici* only lacked qa1F, and *Leptosphaeria maculans* lacked qa2 and qa3. In Eurotiomycetes, 40 species out of 43 species in Eurotiales order have all 7 QA utilization genes including 32 *Aspergillus* species, 4 *Penicillium* species, 2 *Talaromyces* species; 16 species out of 19 species in the order of Chaetothyriomycetidae have gene homologous to all 7 QA utilization genes. It is noted that a few fungal genomes in this Class are only lacking one of the seven
genes, such as three *Fonsecaea* species, *Aspergillus japonicas*, *Rasamsonia emersonii*, *Talaromyces stipitatus*, had all other 6 genes but lacked qa2 genes, and *Endocarpon pusillum* only lacked qaY gene. Whether the missing gene is in an un-sequenced gap of their genomes or resulting from gene losses needs to be further examined. Most of the species in the class of Leotiomycetes and Sordariomycetes have all or most of the 7 QA utilization genes (Table 1; Table 2). However, among 53 species in Saccharomycetes only 3 species have all 7 QUT genes. We have noted that there were lineage-specific QA gene losses, such as species in Onygenales order of Eurotiomycetes class lacked most of QA utilization genes, including the genes encoding metabolic enzymes, i.e. QA-2, QA-3, and QA-4, these species most likely could not utilize QA. Based on the bioinformatics survey with currently sequenced genomes, we can infer that species belonging to Schizosacharomyces, Pneumocystidomycetes, and most species in Sordariomycetes in Ascomycota, species in Basidiomycota, Microsporidia, Mucoromycota did not have QUT genes, thus, were expected not being able to utilize QA as a carbon source (Table 1; Table 2).

Table 2 Summary of distribution of quinic acid utilization genes in different class of fungal species

| Phylum          | Class               | Orders | Total Species | 7 QA genes | 6 QA genes | 5 QA genes |
|-----------------|---------------------|--------|---------------|------------|------------|------------|
| Ascomycota      | Eurotiomycetes      | 4      | 77            | 56         | 7          | 4          |
|                 | Chaetothyriomycetida| 21     | 16            | 3          | 1          |            |
|                 | Eurotiales          | 43     | 40            | 3          | 0          |            |
|                 | Onygenales          | 12     | 0             | 0          | 3          |            |
|                 | Verrucariales       | 1      | 0             | 1          | 0          |            |
|                 | Dothideomycetes     | 5      | 22            | 20         | 1          | 1          |
|                 | Leotiomycetes       | 4      | 9             | 6          | 1          | 2          |
|                 | Orbiliomycetes      | 1      | 1             | 0          | 0          |            |
|                 | Pezizomycetes       | 1      | 1             | 0          | 0          |            |
|                 | Pneumocystidomycetes| 1     | 3             | 0          | 0          |            |
|                 | Saccharomycetes     | 1      | 53            | 3          | 2          | 1          |
|                 | Schizosacharomyces  | 1      | 4             | 0          | 0          |            |
|                 | Sordariomycetes     | 8      | 42            | 32         | 6          | 0          |
|                 | Glomerellales       | 5      | 4             | 1          | 0          |            |
|                 | Hypocreales         | 22     | 19            | 2          | 0          |            |
|                 | Magnaporthales      | 2      | 2             | 0          | 0          |            |
|                 | Microascales        | 1      | 0             | 0          | 0          |            |
|                 | Ophiostomatales     | 2      | 0             | 1          | 0          |            |
|                 | Sordariales         | 8      | 6             | 2          | 0          |            |
|                 | Sordariomycetida    | 1      | 0             | 0          | 0          |            |
|                 | Xylariales          | 1      | 1             | 0          | 0          |            |
| Basidiomycota   | Taphrinomycotina    | 1      | 1             | 0          | 0          |            |
|                 | Xylonomycetes       | 1      | 1             | 0          | 0          |            |
|                 | Agaricomycetes      | 7      | 18            | 0          | 0          |            |
|                 | Exobasidiomycetes   | 5      | 7             | 0          | 0          | 2          |
|                 | Malasseziomycetes   | 1      | 4             | 0          | 0          | 0          |
|                 | Microbotryomycetes  | 1      | 2             | 0          | 0          | 2          |
|                 | Mixiomyctes         | 1      | 1             | 0          | 0          | 0          |
|                 | Pucciniomycetes     | 1      | 1             | 0          | 0          | 0          |
|                 | Tremellomycetes     | 2      | 14            | 0          | 0          | 0          |
|                 | Ustilaginomycetes   | 1      | 5             | 0          | 0          | 0          |
|                 | Wallemiomyctes      | 1      | 2             | 0          | 0          | 0          |
| Chytridiomycota | Chytridiomycetes    | 2      | 2             | 0          | 0          | 0          |
| Microsporidia   | Microsporidia       | 5      | 10            | 0          | 0          | 0          |
| Mucoromycota    | Glomeromycetes      | 1      | 1             | 0          | 0          | 0          |
|                 | Mortierellomycetes  | 1      | 1             | 0          | 0          | 0          |
|                 | Mucoromycetes       | 1      | 2             | 0          | 0          | 0          |
### 3.2 Phylogenetic analysis of protein sequences encoded by the QA utilization genes

We have constructed a phylogenetic tree using protein sequences for each QA gene from selected 43 species. However, due to long divergent time periods and relative short sequence lengths, the phylogenetic trees built with proteins encoded by individual QA utilization genes were not consistent to infer the evolution history of the QA genes. As using concatenated sequences from multiple genes or proteins proved to be a reliable method for phylogenetic analysis (Min and Hickey, 2007), thus, we concatenated the pre-aligned protein sequences encoded by the 7 QA genes in each species and constructed a 7-protein phylogenetic tree (Figure 1). The original tree and the bootstrap consensus tree were shown as Figure 1A and Figure 1B, respectively. The original tree shows the genetic distances among species of the concatenated 7-protein sequences with robust bootstrap values. All the selected species belong to Phylum Ascomycota. The overall phylogenetic tree topology was consistent with recent trees of Ascomycota trees constructed 6-genes (James et al., 2006; Schoch et al., 2009). However, it should be noted that while species in Sordariomycetes form a monophylogenetic group, species in Dothideomycetes and Eurotiomycetes have two phylogenetic groups (Figure 1A; Figure 1B).

![phylogenetic tree](image)

Figure 1 Molecular phylogenetic analysis of concatenated 7-protein sequences encoded by quinic acid utilization genes in fungi by Maximum Likelihood method

Note: (A) Original trees show the genetic distances with bootstrap values. (B) Bootstrap consensus tree to represent the inferred evolution history of the genes in selected taxa

### 3.3 QA utilization genes in a plant species – most likely a case of contamination

Using each of the 7 protein sequence encoded by QUT cluster genes in *N. crassa* as a query to search the non-redundant protein database with a limit to plant kingdom (Viridiplantae) at the National Center for Biotechnology Information (https://www.ncbi.nlm.nih.gov/), to our surprise, we identified all 7 QUT homologous proteins from recently sequenced cork oak (*Quercus suber*) plant (Ramos et al., 2018). The identified homologous proteins include 1 copy of QA2 (accession number: XP_023917533.1), 2 copies of QA3 (XP_023910457.1, ...
Molecular experimentally of species. Not.

( XP_023917535.1, 4 copies of QA4 (XP_023910453.1, XP_023917539.1, XP_023899702.1, XP_023890620.1), 2 copies of QAX (XP_023917534.1, XP_023910117.1), 3 copies of QAY (XP_023917536.1, XP_023910455.1, XP_023897749.1), QA1F (XP_023877214.1, XP_023917541.1, XP_023910451.1), and QA1S (XP_023917540.1, XP_023877217.1, XP_023910452.1), respectively. We then constructed a phylogenetic tree using the 7 concatenated protein sequences from the plant and the fungal species (Figure 2). The tree is exactly identical with the original fungal tree (Figure 1A), except with the data added from the plant species. The phylogenetic analysis showed the protein sequences of QA utilization genes in cork oak plant were clustered with protein sequences of Baudoinia panamericana, a fungal species belonging to Class Dothideomycetes.

Figure 2 Phylogentic tree constructed with 7-protein sequences encoded by quinic acid utilization genes in fungi and a plant species

Note: The tree was constructed as Figure 1A with addition of data from a plant species Quercus suber

As there is no QA utilization genes found in other plant species, we suspected that these protein sequences from cork oak plant were from a contaminated fungal species. Using the whole set of proteins from cork oak plant to search non-redundant protein databases, we found over ten thousands of proteins have a top hit with proteins from fungi. Thus, it is most likely the recently released genome sequences of cork oak plant were contaminated by a fungal species belonging to Class Dothideomycetes, although we could not determine the exact species currently. Using PCR to amplify these QA genes from uncontaminated plant tissue will be able to verify if these QA genes are present in the plant genome or not. However, if these QA genes are really part of the plant genome, it would represent a recent horizontal gene transfer from a fungus to a plant species.

4 Discussion

We performed a survey of QUT genes in 285 completely sequenced fungal genomes and found there were 117 fungal species having all 7 QUT genes in their genomes. Most species in the classes of Dothideomycetes, Eurotiomycetes, Leotiomycetes and Sordariomycetes have QUT genes, however, among 53 species in Saccharomycetes only 3 species have all 7 QUT genes. However, whether these species are able to utilize QA as a carbon source for their growth needs to be examined experimentally. Our survey revealed that species in Agaricomycetes, Basidiomycota, Chytridiomycetes, Exobasidiomycetes, Malasseziomycetes, Microsporidia, Schizosacharomycetes, and Tremellomycetes did not have QA utilization genes.
Using concatenated protein sequences encoded by 7 QUT genes, a robust phylogenetic tree to infer the evolution of the QUT cluster genes was constructed. Since there were no QA utilization genes present in Phylum Chytridiomycota, which often is used as a root for fungal phylogenetic analysis (Min and Hickey, 2007), we are not certain the exact origin of the QA utilization cluster genes in fungi. However, based on the robust bootstrap values of the concatenated 7-protein trees, it most likely reflects the evolutionary history of the QA utilization genes in fungi. In addition, we also found QUT genes from the recently sequenced genome of cork oak (Quercus suber), however, our analysis suggests that these QUT sequences are likely from a contaminated fungal species.

In summary, we identified all probable homologous protein sequences involved in utilizing QA from completely sequenced fungal genomes. These sequences can be used for further experimentally verifying those fungal species being able to utilize QA as their carbon source or not. The phylogenetic reconstruction revealed that the evolutionary history of QUT genes, which may be useful in understanding evolution of the lifestyles and metabolic gene clusters in fungi.

Authors’ contributions
DA and XM conceived the study and prepared the manuscript. JZ and XM collected the data. DA and XM performed data analysis. All authors read and approved the final manuscript.

Acknowledgements
XM is supported by the Youngstown State University Research Professorship.

References
Blackwell M., 2011, The fungi: 1, 2, 3 … 5.1 million species? Am J Bot., 98(3):426-38
https://doi.org/10.3732/ajb.1000298
PMId:21613136
Case M.E., Geever R.F., and Asch D.K., 1992, Use of gene replacement transformation to elucidate gene function in the qa gene cluster of Neurospora crassa, Genetics, 130(4): 729-736
https://doi.org/10.1093/genetics/130.4.729
PMId:1553844 PMCid:PMC1204924
Felsenstein J., 1985, Confidence limits on phylogenies: An approach using the bootstrap, Evolution, 39:783-791
https://doi.org/10.1111/j.1558-5646.1985.tb00420.x
PMId:28561359
Galagan J.E., Calvo S.E., Borkovich K.A., et al., 2003, The genome sequence of the filamentous fungus Neurospora crassa, Nature, 422(6934):859
Giles N.H., Case M.E., Baum J., Geever R., Huiet L., Patel V., and Tyler B., 1985, Gene organization and regulation in the qa (quinic acid) gene cluster of Neurospora crassa, Microbiol. Rev., 49(3):338-358
https://doi.org/10.1128/MMBR.49.3.338-358.1985
PMId:2931582 PMCid:PMC373038
Grant S., Roberts C.F., Lamb H., Stout M., and Hawkins A.R., 1988, Genetic regulation of the quinic acid utilization (QUT) gene cluster in Aspergillus nidulans, J. Gen. Microbiol., 134(2):347-358
https://doi.org/10.1099/00221287-134-2-347
PMId:3049934
Hawkins A.R., Lamb H.K., Smith M., Keyte J.W., and Roberts C.F., 1988, Molecular organisation of the quinic acid utilization (QUT) gene cluster in Aspergillus nidulans, Mol. Gen. Genet., 214(2):224-231
https://doi.org/10.1007/BF00337715
PMId:2976880
James T.Y., Kauff F., Schoch C.L., et al., 2006, Reconstructing the early evolution of Fungi using a six-gene phylogeny, Nature, 443(7113):818
Jones D.T., Taylor W.R., and Thornton J.M., 1992, The rapid generation of mutation data matrices from protein sequences, Computer Appl. Biosci., 8:275-282
https://doi.org/10.1093/bioinformatics/8.3.275
PMId:1633570
Logan, D.A., Koch A.L., Dong W., Griffith, J., Nelsen, R., Case, M.E., Schuttler, and Arnold. J., 2007, Genome-wide expression anlaysis of genetic networks in Neurospora crassa, Bioinformation, 1:390-395
https://doi.org/10.6026/97320630001390
PMId:17597928 PMCid:PMC1896053
Min X.J., and Hickey D.A., 2007, Assessing the effect of varying sequence length on DNA barcoding of fungi, Molec. Ecol. Notes, 7:365-373
https://doi.org/10.1111/j.1471-8286.2007.01698.x
PMId:18784789 PMCid:PMC1890918
