Title

The gold-standard genome of Aspergillus niger NRRL 3 enables a detailed view of the diversity of sugar catabolism in fungi

Journal

Studies in Mycology, 91(1)

ISSN

0166-0616

Authors

Aguilar-Pontes, MV

Brandl, J

McDonnell, E

et al.

Publication Date

2018-09-01

DOI

10.1016/j.simyco.2018.10.001

Peer reviewed

WESTERDIJK FUNGAL BIO DIVERSITY INSTITUTE

The gold-standard genome of *Aspergillus niger* NRRL 3 enables a detailed view of the diversity of sugar catabolism in fungi

M.V. Aguilar-Pontes[1,2], J. Brandl[3], E. McDonnell[4], K. Strasser[4], T.T.M. Nguyen[4], R. Riley[5], S. Mondo[5], A. Salamov[5], J.L. Nybo[3], T.C. Vesth[3], I.V. Grigoriev[5], M.R. Andersen[3], A. Tsang[4*], and R.P. de Vries[1,2*]

[1]*Westerdijk Fungal Biodiversity Institute, Uppsalalaan 8, 3584 CT, Utrecht, The Netherlands;* [2]*Fungal Molecular Physiology, Utrecht University, Uppsalalaan 8, 3584 CT, Utrecht, The Netherlands;* [3]*Department of Biotechnology and Biomedicine, Technical University of Denmark, Søltofts Plads 223, DK-2800, Kongens Lyngby, Denmark;* [4]*Centre for Structural and Functional Genomics, Concordia University, 7141 Sherbrooke Street West, Montreal, QC, H4B 1R6, Canada;* [5]*US Department of Energy Joint Genome Institute, 2800 Mitchell Drive, Walnut Creek, CA, 94598, USA*

Correspondence: A. Tsang, adrian.tsang@concordia.ca; R.P. de Vries, r.devries@westerdijkinstitute.nl

Abstract: The fungal kingdom is too large to be discovered exclusively by classical genetics. The access to omics data opens a new opportunity to study the diversity within the fungal kingdom and how adaptation to new environments shapes fungal metabolism. Genomes are the foundation of modern science but their quality is crucial when analysing omics data. In this study, we demonstrate how one gold-standard genome can improve functional prediction across closely related species to be able to identify key enzymes, reactions and pathways with the focus on primary carbon metabolism.

Based on this approach we identified alternative genes encoding various steps of the different sugar catabolic pathways, and as such provided leads for functional studies into this topic. We also revealed significant diversity with respect to genome content, although this did not always correlate to the ability of the species to use the corresponding sugar as a carbon source.

Key words: *Aspergillus*, Genomic diversity, Gold standard genome, Sugar catabolism.

Available online 7 October 2018; https://doi.org/10.1016/j.simyco.2018.10.001.

INTRODUCTION

The fungal kingdom is estimated to contain over 1.5 million species (Hawksworth 1991), but only a few species have been studied in depth. In recent years, several hundreds of fungal genomes have been sequenced through different initiatives to gain insights into their biology and the variation within the kingdom (MIT, Arnaud *et al.* 2012, Grigoriev *et al.* 2012). Even though sequencing technology has evolved to produce the most complete genome sequence, we are still facing hurdles with respect to gene prediction and functional annotation (Aguilar-Pontes *et al.* 2014, Watson 2018). Recognizing genes in DNA sequences remains one of the most pressing problems in genome analysis together with the functional annotation of the predicted genes. This can only be improved using a large set of -omics data, literature and human supervision, also known as manual curation. If done properly, the output is a gold-standard genome that can be used to improve the quality of other genomes, especially of related species, to study the evolutionary mechanism that allow them to adapt to their lifestyles and ecological niches. The investments required to generate gold standard genomes go beyond what is feasible in a typical genome sequencing project, as it requires the combined efforts of not only sequencing centres, but also a broad research community that covers many of the biological aspects of fungal life.

Fungi and specifically Aspergilli, can be found in almost all ecosystems. In order to survive in their niches, they have to be able to accommodate their metabolism to the energy source available. In nature, fungi need to recognize the plant biomass components in order to induce the production of the right set of degradative and metabolic enzymes that break down the complex structures forming the plant cell wall into simple molecules. Despite the complexity of the polymers forming the cell wall, the skeleton is mainly formed by simple sugars, such as D-glucose, D-fructose, D-xylose, L-arabinose, D-galactose, D-galacturonic acid and L-rhamnose (Kowalczyk *et al.* 2014). The ensemble of catabolic pathways that convert these sugars is known as primary carbon metabolism. A previous study investigated the evolution of primary carbon metabolism in *A. nidulans* and several Aspergilli based on the genome sequence available at that moment, identifying enzyme homologs and additional copies of several genes in some of the species (Flipphi *et al.* 2009). However, a recent study indicated that extra copies of an enzyme or a catabolic pathway does not necessarily affect the catabolic efficiency of the species under specific growth conditions, but correlates with the phylogenetic relationship between species (de Vries *et al.* 2017). However, a positive correlation was found between the presence of a gene predicted to encode a catabolic enzyme and their ability to grow on a specific carbon source (de Vries *et al.* 2017).

In this study, a network of reactions of primary metabolism based on the *Aspergillus niger* NRRL 3 gold-standard genome (Genozymes 2009) (manuscript in preparation) was generated and used to find orthologous genes and pathways involved in the catabolism of monosaccharides in a set of closely related fungal species (Fig. 1). The main focus was to evaluate whether there is a link between genome content and growth abilities. Our aim here is to study the genome content related to primary carbon

Peer review under responsibility of Westerdijk Fungal Biodiversity Institute.

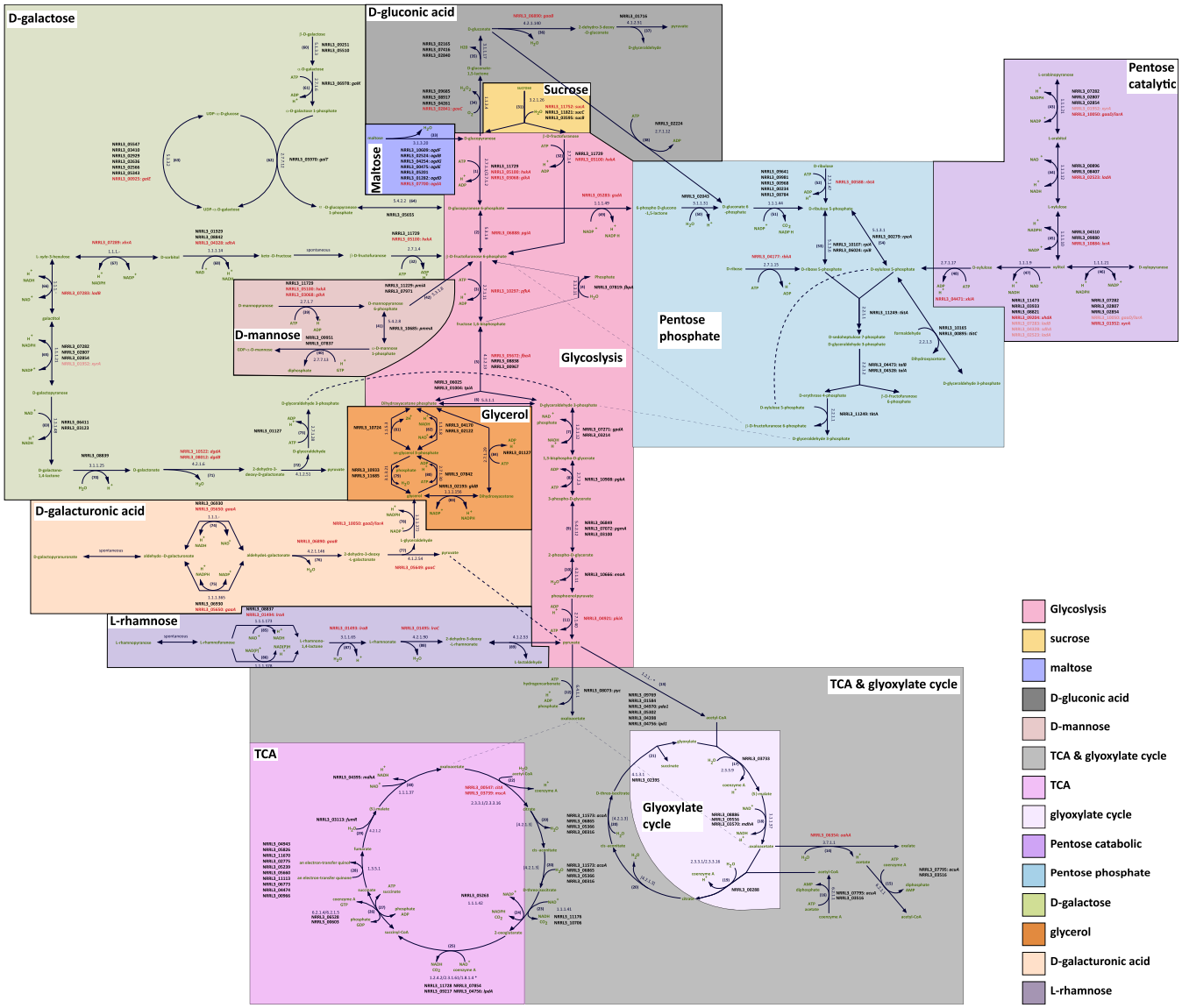

Fig. 1. Primary carbon metabolism pathways: summary overview of all pathways included in the *A. niger* NRRL 3 manually curated carbon metabolic network. Substrates are in green, reactions are depicted with an arrow, reversible reaction are indicated with double arrow. Enzyme Commission (EC) number for each reaction is indicated beside each reaction, while reactions are identified by numbers in brackets (for more information see Supplementary file 6). Proteins assigned to the reactions are indicated in black, common gene name used in *A. niger* is noted beside the protein ID where possible. Characterised enzymes are in red, enzymes involved in more than one reaction are indicated in lighter colour. Dash lines connect metabolites from different pathways. * The enzymes associated to that reaction form a complex. Each pathway is highlighted with a background shade, the legend for the shades is on the right.

metabolism to identify key enzymes or reactions that explain growth under specific conditions. The advantage of using a gold-standard genome, with telomere-to-telomere chromosomes and manually curated gene models, as a reference for this study is that it significantly reduced the risk of missing genes due to gaps in the genome sequence and errors in electronic gene calling.

MATERIALS AND METHODS

Whole genome phylogeny

Whole genome phylogeny was performed among the 28 species selected (*Saccharomyces cerevisiae* S288C, *Neurospora crassa* OR74A, *Trichoderma reesei* QM 6a, *Talaromyces marneffei* ATCC 18224, *Talaromyces stipitatus* ATCC 10500, *Penicilliopsis zonata* CBS 506.65, *Penicillium digitatum* PHI 26, *P. chrysogenum* CBS 307.48, *P. rubens* Wisconsin 54-1255, *Aspergillus glaucus* CBS 516.65, *A. wentii* CBS 141173, *A. clavatus* NRRL 1, *A. novofumigatus* CBS 117520, *A. fumigatus* Af293, *A. fischeri* NRRL 181, *A.s campestris* IBT 28561, *A.terreus* NIH 2624, *A. flavus* NRRL 3357, *A. oryzae* RIB40, *A. nidulans* FGSC A4, *A. sydowii* CBS 593.65, *A. versicolor* CBS 795.97, *A. aculeatus* ATCC 16872, *A. carbonarius* ITEM 5010, *A. brasiliensis* CBS 101740, *A. niger* NRRL 3, *A. luchuensis* CBS 106.47 and *A. tubingensis* CBS 134.48). The tree was built using 200 bidirectional best-blast hits (BBH), which were aligned using mafft V7.221 (Katoh & Standley 2013) with default parameters. Uninformative/ambiguous sites were removed from the alignment using Gblocks v0.91b (Nielsen *et al.* 1997) with parameters –t = p –e = .gb –b4 = 5. The maximum likelihood tree was built using FastTree v2.1.9 with parameters –gamma –wag.20 rate categories were used during phylogeny reconstruction. This phylogenetic tree was visualized using R language and environment v.

Table 1. List of species used in this study.

Species	Strain	Section	Species abbreviation [1]	Number of genes	Assembly length (Mb)	Reference
Saccharomyces cerevisiae	S288C		Sacce1	6 575	12	(Goffeau *et al.* 1996)
Neurospora crassa	OR74A		Neucr2	10 785	41	(Galagan *et al.* 2003)
Trichoderma reesei	QM 6a		Trire2	9 143	33	(Martinez *et al.* 2008)
Talaromyces marneffei	ATCC 18224	Talaromyces	Talma1_2	10 638	29	(Nierman *et al.* 2015)
T. stipitatus	ATCC 10500	Talaromyces	Talst1_2	13 252	36	Unpublished
Penicilliopsis zonata	CBS 506.65		Aspzo1	9 886	29	(de Vries *et al.* 2017)
Penicillium digitatum	PHI 26	Penicillium		9 118	29	(Marcet-Houben *et al.* 2012)
P. chrysogenum	CBS 307.48	Chrysogena	Pench1	11 396	31	(de Vries *et al.* 2017)
P. rubens	Wisconsin 54-1255	Chrysogena	PenchWisc1_1	13 671	32	(van den Berg *et al.* 2008)
Aspergillus glaucus	CBS 516.65	Aspergillus		11 277	28	(de Vries *et al.* 2017)
A. wentii	CBS 141173	Cremei	Aspwe1	12 442	31	(de Vries *et al.* 2017)
A. clavatus	NRRL 1	Clavati	Aspcl1	9 121	28	(Fedorova *et al.* 2008)
A. novofumigatus	CBS 117520	Fumigati		11 549	32	(Kjaerbolling *et al.* 2018)
A. fumigatus	Af293	Fumigati	Aspfu1	9 781	29	(Nierman *et al.* 2005)
A. fischeri	NRRL 181	Fumigati	Neofi1	10 406	33	(Nierman *et al.* 2005)
A. campestris	IBT 28561	Candidi	Aspcam1	9 764	28	(Kjaerbolling *et al.* 2018)
A. terreus	NIH 2624	Terrei		10 406	29	(Arnaud *et al.* 2012)
A. flavus	NRRL 3357	Flavi		12 604	37	(Payne *et al.* 2006)
A. oryzae	RIB40	Flavi		12 030	38	(Machida *et al.* 2005)
A. nidulans	FGSC A4	Nidulantes	Aspnid1	10 680	30	(Galagan *et al.* 2005)
A. sydowii	CBS 593.65	Nidulantes	Aspsy1	13 620	34	(de Vries *et al.* 2017)
A. versicolor	CBS 583.65	Nidulantes	Aspve1	13 228	33	(de Vries *et al.* 2017)
A. aculeatus	ATCC 16872	Nigri	Aspac1	10 845	35	(de Vries *et al.* 2017)
A. carbonarius	ITEM 5010	Nigri	Aspca3	11 624	36	(de Vries *et al.* 2017)
A. brasiliensis	CBS 101740	Nigri	Aspbr1	13 000	36	(de Vries *et al.* 2017)
A. niger	NRRL 3	Nigri	Aspni_NRRL3_1	11 846	35	Unpublished
A. luchuensis	CBS 106.47	Nigri	Aspfo1	13 530	37	(de Vries *et al.* 2017)
A. tubingensis	CBS 134.48	Nigri	Asptu1	12 322	35	(de Vries *et al.* 2017)

The order of the species follows the taxonomic organization in the phylogenetic tree (Fig. 2).
[1] The species abbreviations correspond to JGI acronyms used as species and genome released unique identifiers in the portal. This identifier is used as species identifier in the heatmap figures.

3.4.0 (R Core Team 2017) with the package ggtree v. 1.8.2 (Yu *et al.* 2017) from ggplot2 v. 2.2.1 (Wickham 2009).

Protein quality assessment

Protein quality assessment was performed using BUSCO v3 (Simao *et al.* 2015) with default parameters and specific lineage database accordingly (*Aspergillus, Penicillium, Penicilliopsis and Talaromyces* species against *eurotiomycetes*_odb9 database, *N. crassa* and *T. reesei* against sordariomycetes_odb9 database and *S. cerevisiae* saccharomyceta_odb9 database).

Protein functional annotation

Protein functional annotation of selected species was downloaded from the JGI Mycocosm Portal (Table 1). Exclusively InterPro and Pfam domains (Quevillon *et al.* 2005) with E-value >1e^{-15}, KEGG database information (Kanehisa *et al.* 2006) and SignalP annotation (Nielsen *et al.* 1997) was used. Additionally, *S. cerevisiae, N. crassa, T. reesei* and *A. nidulans* functional information was obtained from the SGD database (Cherry *et al.* 2012), Ensembl Fungi Biomart release 38 (Kinsella *et al.* 2011), (Benocci *et al.* 2018) and the AspGD database (Cerqueira *et al.* 2014), respectively.

Protein profiling

Gene families (clusters) were download from JGI Mycocosm Portal (Grigoriev *et al.* 2014)

https://genome.jgi.doe.gov/clm/run/fungi-2016-08.1352. Clusters were predicted by Blast (Altschul *et al.* 1990) E-value 1e^{-5} and MCL inflation parameter −2 (Enright *et al.* 2002). Clusters with only one protein assigned were removed from the original data set. In total, 112 gene clusters containing *A. niger* NRRL3 proteins of interest were selected. Amino acid sequences of selected species were downloaded from the JGI Mycocosm Portal (Table 1). Alignment of the amino acid sequences of the proteins included in the clusters was performed using mafft V7.221 (Katoh & Standley 2013) with default parameters. Alignments were manually curated and nucleotide sequences of split genes were fixed where possible, otherwise the protein sequence was removed from the alignment (Supplementary file

X1). A Neighbor joining tree was built using MEGA-CC (Kumar *et al.* 2012) with complete deletion and bootstrap 500. Phylogenetic trees were inspected manually and rooted to *S. cerevisiae* where possible. Proteins included in a monophyletic tree with an *A. niger* NRRL3 proteins will be considered as NRRL3 orthologs. Close orthologs of a NRRL3 protein will be defined as group of proteins included in a monophyletic tree after a gain or loss event where no NRRL3 protein is present. These are identified with the closest NRRL3 protein id and the suffix "like". *A. niger* protein functional information was used to assign function to ortholog groups within a cluster. Groups with functional annotation not related to monosaccharides metabolism were removed from the final set when supported by more than one protein. If no NRRL3 protein was present in the group, curated database information of other species was used (Supplementary file X2). If more than one transcript per protein was included in the tree, they were counted as one (Supplementary file X3). Membership in a group was used to determine whether a protein was conserved across the selected species. If a protein is found at least once in all members or all members of an organism subset (*e.g.* section *Nigri*), it is considered conserved. A protein was considered specific to an organism subset if it was found in at least one organism of the subset, but not in any organisms outside the subset. Heatmaps were generated using R language and environment with the package ComplexHeatmap v. 1.14.0 (Gu *et al.* 2016). Columns are ordered according to the phylogenetic tree and rows follow the order of the reactions in the pathway. Growth graphs were generated using Adobe Illustrator CC.

Growth profiling on different monosaccharides

For growth profiling, all strains were grown on MM (de Vries *et al.* 2004) with 33 different carbon sources (http:/www.fung-growth.org/). Growth was performed at 30 °C for all Aspergilli and 25 °C for the other species. Growth was continued until the largest colony of a species set almost reached the edge of the plate. Growth per species was scored from 0 when there is no visible growth to 10 when maximum growth among all carbon sources has been reach. Growth profiles containing monosaccharides and disaccharides (de Vries, *et al*.) were selected for this study (Supplementary file X4). Media with no carbon source was used as a control. If growth on a specific carbon source is the same as with no carbon source, it is considered no growth.

RESULTS

Phylogenetic relationships and proteome assessment

To provide an overview of the relationship among the selected species (Table 1) a genome-wide phylogeny was conducted (Fig. 2A). The constructed tree supports the results described before (Kocsube *et al.* 2016). Members of the section *Nigri* are in a single clade close to section *Nidulantes* that also appears as a single clade. Section *Flavi* (*A. oryzae* and *A. flavus*) are placed close to the representative of section *Terri* with section *Candidi* as an outgroup. Section *Fumigati* and section *Clavati* are in another clade with *A. wentii* from section *Cremei* as a close relative. The genus *Penicillium* sections *Chrysogena* and *Penicillium* are placed close to the related species *Penicilliopsis zonata* (previously *Aspergillus zonatus* (Kocsube *et al.* 2016)). *T. reesei* and *N. crassa* are at the base of the tree with *S. cerevisiae* at the root.

The genomes used in this study have been sequenced over the years by different consortia and technologies. *S. cerevisiae* was the first fungal genome published in 1996 (Goffeau *et al.* 1996) while *A. campestris* was published in 2018 (Kjaerbolling *et al.* 2018) sequenced with the latest technology PacBio RS. Hence the quality of assembled genomes and annotation can vary substantially. We used BUSCO (Simao *et al.* 2015) to evaluate the completeness across their predicted proteomes. The average of complete proteins is higher than 95 % in most species with the exception of *A. carbonarius*, where only 88 % of the genes are complete with more than 8 % of the core proteins missing. The proportion of the proteomes that are duplicated comprise approx. 1 % except for *S. cerevisiae, N. crassa, Talaromyces* species and *A. carbonarius* that varies between 5–10 %. With the exception of *A. carbonarius,* the genomes with the higher number of missing proteins also contain a higher number of fragmented proteins. High numbers of fragmented proteins were observed regardless the genus, technology, assembler or automatic gene calling methodology. Species with more than 150 fragmented proteins (*A. oryzae*, *A. flavus*, *A. terreus*, *P. digitatum*) were removed from the analysis (Fig. 2B).

Complete genomic sequences from diverse phylogenetic lineages reveal notable increases in genome complexity from prokaryotes to multicellular eukaryotes (Lynch & Conery 2003). However, the genome size of an organism varies from species to species and is not proportionally correlated with organismal complexity. In our dataset, the smallest genome is 12 Mb and the largest 41 Mb, corresponding to *S. cerevisiae* and *N. crassa* respectively, but the *N. crassa* genome only contains twice the number of genes of the *S. cerevisiae* genome. With the exception of *S. cerevisiae* and *N. crassa*, more than 95 % of the predicted proteins were included in the gene families downloaded from the JGI Mycocosm portal (Grigoriev *et al.* 2014) (Fig. 2C).

Genome content related to primary carbon metabolism

In contrast to the most accepted theory (Gregory 2001), the present study shows that genome size and gene content are an indication of primary carbon metabolism complexity (Supplementary Fig. 5). In general, *S. cerevisiae* lacks several of the enzymes associated with primary carbon metabolism predicted in *A. niger*. In some cases, entire pathways are missing, even though growth has been described (Oliva Neto *et al.* 2014). Therefore, in the following sections we will describe the prevalence of orthologous genes in the different species without considering *S. cerevisiae* except for special cases. Starting with glycolysis and acid production (glyoxylate, tricarboxylic and D-gluconic acid) from glycolytic products, the pathways have been ordered according to the step in which the final product enters glycolysis and to the number of carbon atoms in the main substrate: maltose, sucrose, D-mannose, D-galactose, D-galacturonic acid, L-rhamnose, pentose catabolic and pentose phosphate pathway. In each case, a brief description of the pathway with the EC numbers of each enzymatic reaction and common protein names used in *A. niger* will follow a deep analysis of the relationship between phylogeny, genome content and growth

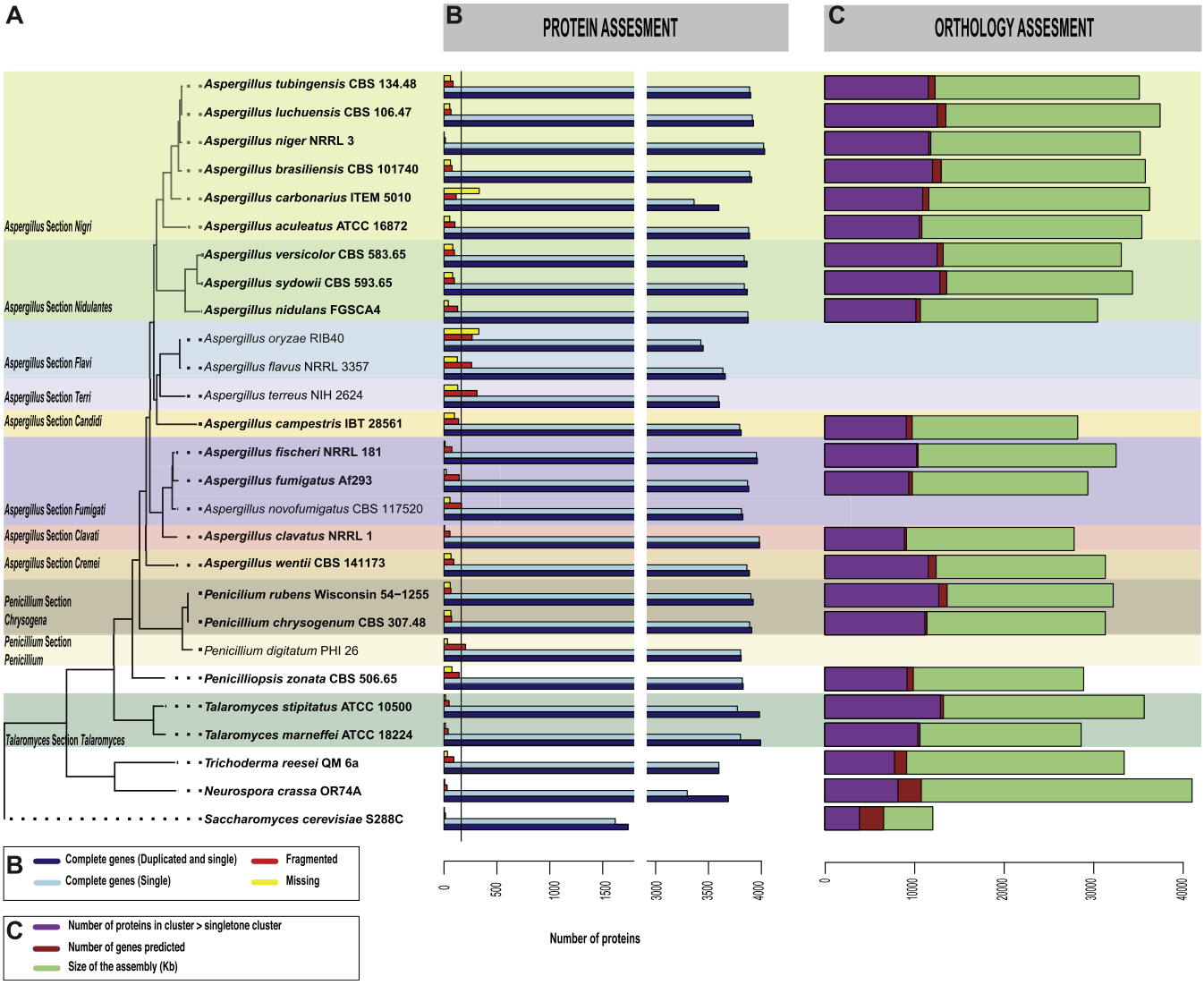

Fig. 2. Phylogenetic tree, proteome assessment and orthology assessment. A) Phylogenetic tree inferred from 200 best bidirectional hits of 28 species. Sections are identified by colours, white indicates the section is not determined. Name of the genus and sections are indicated in the left of the figure. Bold letters indicated the species selected for further analysis. **B)** Protein conservation, calculated using BUSCO v.3 indicates the degree of completeness of the proteins predicted per genome against a database of conserved proteins, bars showing number of proteins being aligned to individual species to the left-hand side. Black line indicates where the value 150 is in the graph. Species with number of fragmented proteins (red bar) higher than 150 were removed from the analysis. Dark blue: complete genes (duplicated and single), light blue: complete single genes, red: fragmented, and yellow: missing. **C)** Total number of proteins included in the gene family orthology against the total number of proteins predicted. Bars indicates the size of the genome (Kb), number of proteins in the proteome and number of proteins included in phylogeny aligned to individual species to the left-hand side. Green, size of genome (Kb); red, total proteome predicted; and purple, total number of proteins included in the gene families.

abilities. The ortholog groups will be referred to by the protein number of the *A. niger* NRRL 3 gene present in the group. If no NRRL 3 gene is present, the group will be referred to by the closest NRRL 3 paralog, followed by '-like'. Despite 'after' and 'before' not being the appropriate phylogenetic terms, we will use them to refer to moments in evolution when proteins have been gained or lost between two groups. Due to the limited number of species from each clade in this study we cannot determine when this exactly happened, *e.g.* in glycolysis, NRRL3_11729_like appears in members of the genus *Talaromyces* and is lost after section *Fumigati*. Meaning that species from clades after section *Fumigati* branched off, resulting in sections *Candidi, Nidulantes* and *Nigri* having lost this ortholog.

Glucose and fructose catabolism

D-Glucose is the preferential monomeric carbon source for most microorganisms. Although fungi rarely find free high concentration of D-glucose in their environment, it is the major component of the plant cell wall (Kowalczyk *et al.* 2014). Therefore, it is common that all species contain at least one enzyme per reaction for D-glucose catabolism (Fig. 3).

This first step in the glycolysis in the conversion of D-glucose into D-glucose 6-phosphate by two different enzymes, the hexokinase (Hxk, EC 2.7.1.1) (Panneman *et al.* 1998) and the glucokinase (Glk: EC 2.7.1.2) (Panneman *et al.* 1996). Isomerization of D-glucose 6-phosphate by glucose 6-phosphate isomerase (PfkA, EC 2.7.1.11) produces fructose 6-phosphate. This compound can be obtained from D-fructose through fructokinase reaction by hexokinase enzymes exclusively (Panneman *et al.* 1998). A kinase reaction catabolized by the 6-phosphofructose kinase (PfkA, EC 2.7.1.11) converts D-fructose 6-phosphate into D-fructose 1,6-bisphosphate (Habison *et al.* 1983). The reverse reaction, is catalysed by the D-fructose 1,6-phosphatase (FbpA, EC 3.1.3.11). Both reactions are strongly regulated by the accumulation of D-fructose 2,6-bisphosphate, which activates PfkA and inhibits FbpA (Harmsen *et al.* 1992, Ruijter & Visser 1999, Poulsen *et al.* 2005, Upadhyay & Shaw

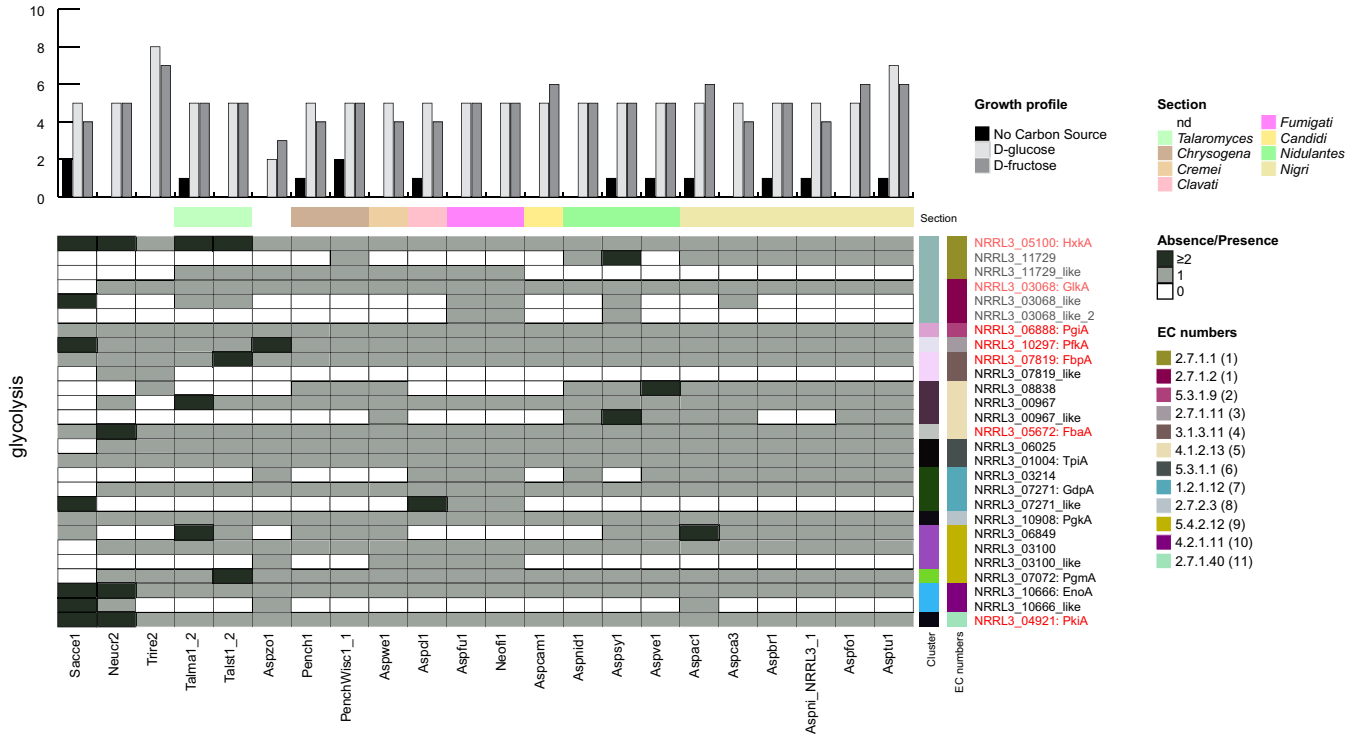

Fig. 3. Glycolysis heatmap abundance of *A. niger* orthologous proteins in the glycolytic pathway. Columns are arranged according to the phylogenetic tree and rows are arranged according to the order in which the enzymatic reaction occur in *A. niger* metabolism. **Top**: The graph on the top represents growth from 0 to 10 with no carbon source (black), D-glucose (light grey) and D-fructose (dark grey). The line below represents the section to which the species belong according to the phylogenetic tree, colours match the colours in the tree. **Heatmap row names:** Identifiers on the right of the heatmap are according to the group name (see Materials and Methods). The name after the identifier correspond to the protein common name used in *A. niger*. Red: protein has been characterised in *A. niger*. Lighter colour: the protein is involved in more than one pathway. **EC numbers column:** colours correspond to EC number associated to the function they have been assigned. EC colour legend is on the right-hand of the figure. Reaction ID number are in brackets, see Supplementary file 6 for more information. **Cluster column:** colours correspond to clusters in which proteins have been found. **Bottom legend:** species abbreviation according to Table 1. Aspzo1: *Penicilliopsis zonata,* Aspfp1: *Aspergillus luchuensis*.

2006). D-Fructose 1,6-bisphoshate is split into dihydroxyacetone phosphate and D-glyceraldehyde 3-phosphate by fructose-bisphosphate aldolase (Fba, EC 4.1.2.13) (Jagannathan *et al.* 1956). Dihydroxyacetone phosphate, also produced in glycerol metabolism, is converted into D-glyceraldehyde 3-phosphate by triphosphate isomerase (Tpi, EC 5.3.1.1) (McKnight *et al.* 1986) contributing to the equilibrium of the glycolysis. Glyceraldehyde 3-dehydogenase (Gdp, EC 1.2.1.12) (Punt *et al.* 1988) catalyses the reversible enzymatic reaction between D-glyceraldehyde 3-phosphate and 1,3-bisphosphate D-glyceraldehyde, which is further converted into 3-phosphate D-glyceraldehyde by phosphoglycerate kinase (Pgk, EC 2.7.2.3) (Clements & Roberts 1985, Streatfield *et al.* 1992, Streatfield & Roberts 1993). Phosphopyruvate hydratase, also known as enolase (EnoA, EC 4.2.1.11) (Clements & Roberts 1985, Streatfield & Roberts 1993), converts 2-phospho D-glycerate into phospho*enol*pyruvate, the substrate of the last reaction in the glycolysis, the conversion of phospho*enol*pyruvate into pyruvate by pyruvate kinase (PkiA, EC 2.7.1.40) (de Graaff *et al.* 1992).

Four out of ten enzymatic reactions in glycolysis - PgiA, PfkA, PgkA and PkiA - have a single, conserved enzyme assigned to them in all species (PgiA is present in more than one copy in *P. zonata* and PkiA in *S. cerevisiae* and *N. crassa*). A FbpA paralog was found through orthology in *N. crassa* and *T. reesei* while EnoA paralogs were only present in the genomes of *S. cerevisiae, N. crassa, P. zonata* and *A. aculeatus*.

Orthologs of HxkA and GlkA have undergone several changes in the number of proteins encoded in the genome of the tested species. Compared to a previous study (Flipphi *et al.* 2009), we found only one cluster containing a glucokinase and three hexokinases, out of which only two have catalytic function. HxkA (NRRL3_05100) is present in all species while its paralog, NRRL3_11729 has two main groups. An orthologous gene present in *P. rubens, A. nidulans, A. sydowii* and section *Nigri* genomes and a close orthologous group that appears in members of the genus *Talaromyces* and got lost after section *Fumigati*. Three different Glk groups were identified in the phylogeny. An orthologous gene of *A. niger* GlkA (NRRL3_03068) is present in all species. A second sequence (NRRL3_03068_like_1) is present in the genome of the *Talaromyces* species, section *Fumigati*, *A. sydowii* and *A. carbonarius*. A third Glk protein (NRRL3_03068_like_2) is also present in these species except for the *Talaromyces* species.

Proteins from two different clusters have been assigned to Fba activity. The ortholog of *A. nidulans* FbaA, NRRL3_05672 (FbaA) is present in all species (Roumelioti *et al.* 2010) in a single cluster. The second cluster contain two orthologous groups NRRL3_00967 and NRRL3_08838, respectively. NRRL3_00967 is present in all species except *T. reesei*, and a NRRL3_00967_like group is formed from genes of sections *Cremei, Nidulantes* and several species from section *Nigri*. A similar distribution is observed for NRRL3_08838, including members from the genus *Penicillium* and *T. reesei*. Iso-enzymes assigned to Pgm activity belong two different clusters. Both PgmA (NRRL3_07072) and the second enzyme NRRL3_03100 are present in all species. A NRRL3_03100_like orthologous gene was present in *P. zonata, A. clavatus* and section *Fumigati*.

Due to the importance of the pathway metabolizing D-glucose and D-fructose and the absence of orphan reactions, we predicted positive growth for all species on these sugars, which was confirmed by our growth data. *T. reesei* and *A. tubingensis* grew better on these substrates compared to other carbon sources

than any of the other species. Remarkably, *P. zonata* grew below average on both substrates despite the genome content not showing major differences with respect to the genes related to glycolysis compared with the rest of the species.

Glyoxylate and tricarboxylic acid cycle metabolism

The final product of glycolysis is metabolized further through the tricarboxylic acid (TCA) and the glyoxylate cycles. The glyoxylate cycle was described as a modified TCA cycle (Kornberg & Madsen 1958) with which it shares activities (Kunze *et al.* 2006). The partial parallelism between both cycles requires that identical enzymatic activities have to participate independently in different metabolic pathways, which is accomplished in most cases by paralogous proteins that are differently compartmentalized in the cell. The TCA cycle enzymes are mainly localized in the mitochondria while the glyoxylate cycle enzymes are located both inside and outside the peroxisome. The glyoxylate cycle allows cells to utilize C2 carbon sources, *e.g.* ethanol and acetate, when other carbon sources are not available and it further allows the cell to produce carbohydrates through gluconeogenesis from acetyl-CoA. The glyoxylate cycle bypasses the reactions catalysed by isocitrate dehydrogenase and 2-oxoglutarate complex. Two molecules of acetyl-CoA enter the glyoxylate cycle during each turn, while only one enters the TCA cycle. Bypassing the TCA cycle conserves carbon atoms for gluconeogenesis while simultaneously diminishing the flux of electrons into respiration (Hynes *et al.* 2007).

Pyruvate produced through glycolysis can be converted into acetyl-CoA through the pyruvate dehydrogenase complex (EC1.2.1.-*) in the mitochondria or into oxaloacetate by pyruvate carboxylase (PycA, EC 6.4.1.1) (Fig. 4). The pyruvate dehydrogenase complex involves three enzymatic reactions with

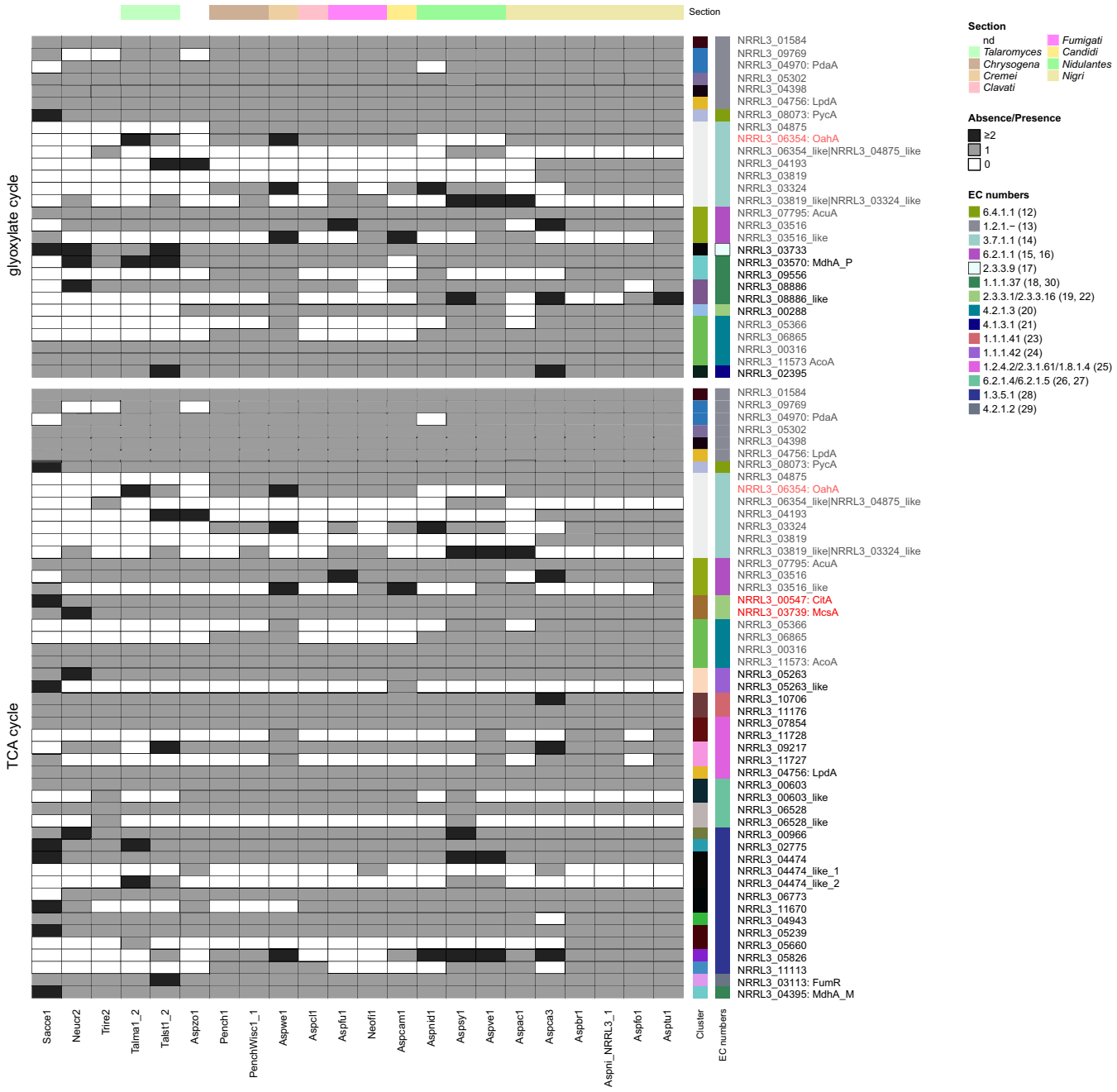

Fig. 4. Glyoxylate and TCA cycles heatmap abundance of *A. niger* orthologous proteins in the glyoxylic and TCA pathways. The legends are the same as that of Fig. 3. MdhA_P: peroxisomal MdhA and MdhA_M: mitochondrial MdhA.

independent enzymes associated to them: the pyruvate dehydrogenase alpha and beta subunits (PdaA), the dihydrolipoamide S-acetyltransferase and the lipoamide dehydrogenase (LpdA). Another source of acetyl-CoA is the reaction catalysed by the acetyl-CoA synthetase (Acu, EC 6.2.1.1) from coenzymeA and acetate. The acetate used as substrate in this reaction can be provided by oxalaceteate acetylhydrolase (Oah, EC 3.7.1.1) after converting oxaloacetate into oxalate (Ruijter *et al.* 1999, Narayanan *et al.* 2009).

The glyoxylate and TCA cycles share three enzymatic reactions, with different paralogs for malate dehydrogenase and citrate synthase, but the same enzyme for aconitase.

Malate dehydrogenase (Mdh, EC 1.1.1.37) catalyses the reversible conversion between malate and oxaloacetate (McAlister-Henn *et al.* 1995). Oxaloacetate is converted into citrate by citrate synthase (Cit, 2.3.3.1). The citrate produced in the glyoxylic cycle its transported to the mitochondria to be converted first into cis-aconitate and finally to isocitrate (D-threo-isocitrate) by aconitate hydratase (Aco, EC 4.2.1.3) that catabolizes both steps. Isocitrate relocates to the peroxisome to continue the glyoxylate cycle.

In the peroxisome, during the glyoxylate cycle, isocitrate lyase (EC 4.1.3.1) will breakdown isocitrate into glyoxylate and succinate, which then enter the TCA. Finally, glyoxylate will be used as substrate together with acetyl-CoA for malate synthase (EC 2.3.3.9) to produce malate and coenzymeA.

In the mitochondria, during the TCA cycle, isocitrate is converted into 2-oxoglutarate by two different isocitrate dehydrogenase reactions, with different cofactors (NADP, EC 1.1.1.42 and NAD, EC 1.1.1.41). The 2-oxoglutarate complex converted coenzymeA to 2-oxglutarate to produce succinylCoA. This complex requires three different enzymes, a mitochondrial 2-oxoglutarate dehydrogenase (EC 1.2.4.2), a dihydrolipoliamide succinyl-transferase (EC 2.3.1.61) and the dihydrolipoyl dehydrogenase (EC 1.8.1.4). Succinate-CoA ligases (GTP as cofactor, EC 6.2.1.4 and ATP as cofactor 6.2.1.5) catabolize the conversion of succinyl-CoA into succinate. Succinate dehydrogenase (EC 1.3.5.1) oxidizes succinate into fumarate. Finally, the last step in the TCA cycle is the conversion of fumarate into malate by the fumarate hydratase (FumR, EC 4.2.1.2).

Most of the enzymatic reactions have multiple isoenzymes assigned to them. Only PycA (NRRL3_08073), the peroxisomal malate synthase (NRRL3_03733), isocitrate lyase (NRRL3_02395) and the mitochondrial FumR (NRRL3_03113) and MdhA_M (NRRL3_04395) are present as a single enzyme in all genomes, some of them are present as more than one copy.

Within the pyruvate dehydrogenase complex, pyruvate dehydrogenase has two subunits, alpha and beta. The alpha subunits, PdaA (NRRL3_04970) and NRRL3_09769 belong to the same cluster. PdaA is present in all species except in *A. nidulans* while NRRL3_09769 is missing in *N. crassa, T. reesei* and *P. zonata*. The beta subunit NRRL3_01584, belong to a different cluster containing orthologs for all species. Enzymes from two different clusters have been assigned to dihydrolipoamide S-acetyltransferase activity, with the *A. niger* genes NRRL3_05302 and NRRL3_04398 assigned to them, and orthologs for both are present in all species. The last component of the complex, LpdA (NRRL3_04756), is present in all species. Acetyl-CoA synthase, AcuA (NRRL3_07795) is also present in all species, but its paralog (NRRL3_03516) is missing in *A. aculeatus*. A third enzyme, NRRL3_03516_like is present in *A. wentii, A. fischeri, A. campestris A. versicolor, A. aculeatus and A. tubingensis*. Even though in the network only one protein was assigned to the oxaloacetate dehydrogenase reaction, OahA (NRRL3_06354), the orthology analysis shows two additional paralogous groups (NRRL3_04875 and NRRL3_06354_like|NRRL3_04875_like). All genomes contain at least one protein assigned to this reaction. After *T. stipitatis* branched off multiple sequences are found reaching the higher number of paralogs in section *Nigri*.

In glyoxylate cycle, three isoenzymes have been assigned to the malate dehydrogenase reaction belonging to two different clusters. MdhA_P (NRRL3_03570) is present in all species, except *A. campestris*. A second enzyme in the same cluster, NRRL3_09556 is present only in section *Nigri*, *A. nidulans*, *A. wenti* and the *Penicillium* species. The third group of enzymes from a different cluster, orthologs of NRRL3_08886, is present in all species except *A. luchuensis* and *A. fischeri*. A NRRL3_08886_like group containing genes from *A. wenti,* section *Nidulantes, A. carbonarius* and *A. tubingensis* was also identified.

Initially, two citrate synthases assigned in the network were identified, NRRL3_02449 and NRRL3_11764, but according to *in house* data they do not participate during glyoxylate production. Nevertheless, A third paralog was identified through orthology. NRRL3_00288 appeared after genus *Talaromyces* split off and it is only missing in *A. aculeatus*.

In the TCA cycle, citrate synthase is assigned to two mitochondrial enzymes, CitA (NRRL3_00547) and McsA (NRRL3_03739) that belong to the same cluster and are present in all species. $NADPH^+$-dependent isocitrate dehydrogenase (NRRL3_05263) is also present in all species. This cluster also contains NRRL3_05263_like sequences from *S. cerevisiae* and *A. campestris* that group separately. $NADH^+$-dependent isocitrate dehydrogenase isoenzymes (NRRL3_10706 and NRRL3_11176) that belong to the same cluster, are present in all species.

Interestingly, one enzyme of each component of the 2-oxoglutarate dehydrogenase complex is present in almost all species, NRRL3_07854, NRRL3_09217 (missing in *T. marneffei*) and LpdA (NRRL3_04756). A second 2-oxoglutrate dehydrogenase NRRL3_11728 and a dihydrolipoamide S-succinyl transferase NRRL3_11727 are only present in *A. wentii, A. versicolor* and some members of section *Nigri*.

Succinate-CoA ligases NRRL3_06528 and NRL3_00603 belong to different clusters. Both enzymes are present in all species. In addition, a NRRL3_06528_like group contains a *T. reesei* and *A. sydowii* sequence, and these species also contain a NRRL3_00603_like sequence together with *P. zonata, Penicillium* species, *A. wenti, A. clavatus,* species from section *Fumigatus* and *A. campestris*.

Succinate dehydrogenases assigned in the network belong to several clusters. All species contain between six and eight paralogs encoded in their genome, section *Nigri* contain up to ten paralogs.

Beside production of different acids, the intermediate metabolites of both cycles are used to produce different amino acids essential for the fungus to survive in the environment. Since growth on the pyruvate precursors D-glucose and D-fructose was positive, we can predict that all species have functional TCA and glyoxylate cycles.

D-Gluconate metabolism

In the presence of D-glucose, some fungi will produce gluconate and its acid form D-gluconic acid as the primary overflow metabolite under non-limiting growth on glucose (Shindia *et al.*

2006) (Fig. 5). *A. niger* oxidizes D-glucose to D-glucono-1,5-lactone by a glucose oxidase from the carbohydrate active enzyme (CAZy) family AA2 (Lombard *et al.* 2014) (Gox, EC 1.1.3.4). D-glucono-1,5-lactone lactonodehydrolase or gluconolactonase (EC 3.1.1.17) converts D-glucono-1,5-lactone to D-gluconate, which can be dehydrated to 2-keto-3-deoxy-D-gluconate by D-galactonate dehydratase (GaaB, EC 4.2.1.146) that is also involved in D-galacturonic acid metabolism. This is further cleaved to D-glyceraldehyde and pyruvate by 2-dehydro-3-deoxy-D-gluconate D-glyceraldehyde-lyase (EC 4.1.2.51). D-gluconate can enter the pentose phosphate pathway through the conversion to D-gluconate-6-phosphate catalysed by gluconokinase (EC 2.7.1.12).

Overall, genomes from sections *Nigri, Nidulantes* and genus *Penicillium* contain several predicted proteins associated to the reactions above, explaining the higher amount of gluconic acid production reported for these fungi (de Vries *et al.* 2017). They seem to have gained additional copies of glucose oxidase (Gox), after *N. crassa* diverged from the other species. The highest number of enzymes is observed for section *Nigri* with five Gox enzymes belonging to a single cluster, including the characterised GoxC (NRRL3_02841). The high number of isoenzymes might suggest production of alternatives for lactonate, but no other enzymatic products have been described despite the broad use of GOX in industrial processes (Malherbe *et al.* 2003). Two gluconolactonases, belonging to different clusters, were predicted but the iso-enzyme NRRL3_07416 is only present in section *Nigri*. A second enzyme belonging to GaaB cluster was present in many species outside the Aspergilli, suggesting it got lost after the Aspergilli branched off the other fungi. A similar situation was detected for 2-dehydro-3-deoxy-D-gluconate D-glyceraldehyde-lyases. Only one enzyme for this function was associated to the reaction in the network (NRRL3_01716), but through orthology, we were able to identify eight different paralogous groups belonging to one of the largest clusters resulting from the orthology analysis. As mentioned above, there is a connection between D-gluconate and the pentose phosphate pathway not involving glycolysis for which NRRL3_02224, the only enzyme assigned is present in all species, including *S. cerevisiae*.

Metabolism of the disaccharides maltose and sucrose

Maltose is a disaccharide formed by two molecules of α-D-glucose with a 1,4-glucoside linkage. Through hydrolysis, α-glucosidase produces two molecules of D-glucose that enters glycolysis. α-glucosidase or maltase (EC 3.2.1.20) belongs to CAZy families GH13 and GH31 (Lombard *et al.* 2014) (Fig. 6). Seven different *A. niger* proteins have been assigned in the network. Enzymes from family GH31, AgdA (NRRL3_07700), AgdB (NRRL3_02524), AgdE (NRRL3_00475), AgdF (NRRL3_10609) and AgdG (NRRL3_04254) belong to four different clusters (Kimura *et al.* 1992, Nakamura *et al.* 1997). Enzymes from family GH13, AgdD (NRRL3_01282) and NRRL3_05201 belong to a unique cluster. All species contain more than one protein in their genome from family GH31 from different clusters. Within family GH13, NRRL3_05201 ortholog is present in all species of the Eurotiales. The NRRL3_01282_like group is mostly missing in section *Nigri*, while AgdD (NRRL3_01282) is specific for section *Nigri*.

All species in this study contain more than one α-glucosidase in their genome, which match their ability to grow on maltose as sole carbon source. *T. stipitatus, A. fischeri* and *A. niger* grew better on maltose than on D-glucose. *S. cerevisiae* is not able to grow on D-maltose which correlates with the presence of only two α-glucosidases in its genome, AgdE and AgdD orthologs. Other than this example, there is no correlation between the number of enzymes encoded and their ability to grow on maltose.

Sucrose is a disaccharide of D-glucose and D-fructose and its hydrolysis is catalysed by β-D-fructofuranoside fructohydrolase (EC 3.2.1.26) belonging to CAZy family GH32 (Lombard *et al.* 2014), also known as invertase (Fig. 7). The catalytic reaction releases two monomers, an α-D-glucose and a β-D-fructose that both enter glycolysis through the enzymatic reactions catabolized by hexokinases (HxkA and NRRL3_11729) and glucokinases (Glk), where the latter only acts on α-D-glucose.

There are three invertases (SucA-C) assigned to *A. niger* sucrose degradation (Boddy *et al.* 1993, Wallis *et al.* 1997), all of which belong to the same cluster. Invertases SucB (NRRL3_03595) and SucC (NRRL3_11821) appeared firstly in *T. stipitatus* genome. SucA (NRRL3_11752) appeared when *Penicillium* species

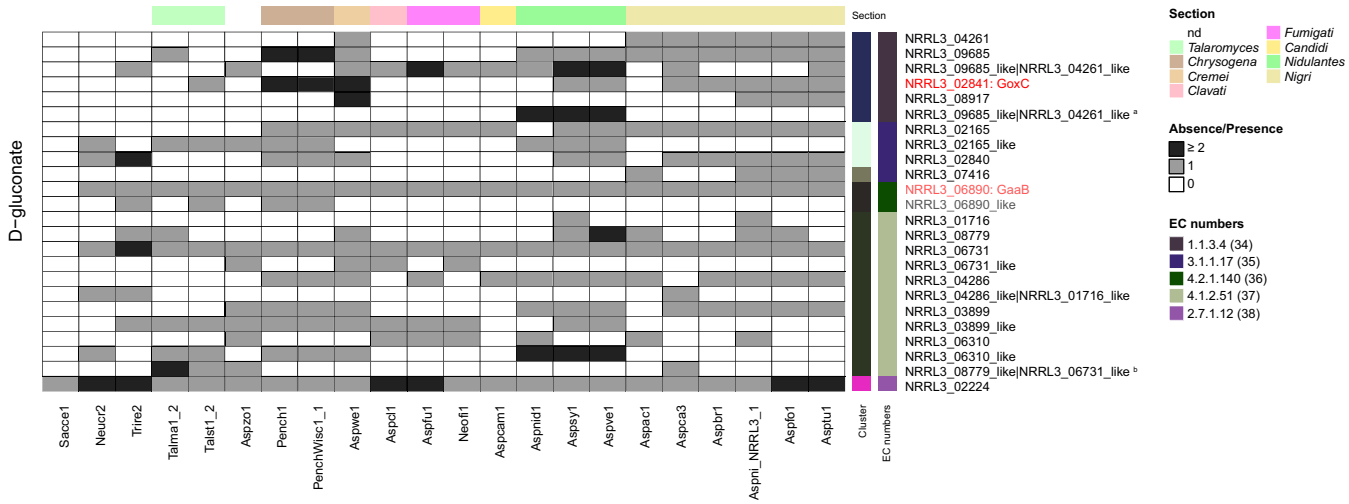

Fig. 5. D-gluconate heatmap abundance of *A. niger* orthologous proteins in the D-gluconate catabolic pathway. The legends are the same as that of Fig. 3. Full group name: [a] NRRL3_09685_like, NRRL3_04261_like, NRRL3_02841_like, NRRL3_08917_like and [b] NRRL3_08779_like, NRRL3_06731_like and NRRL3_05649_like.

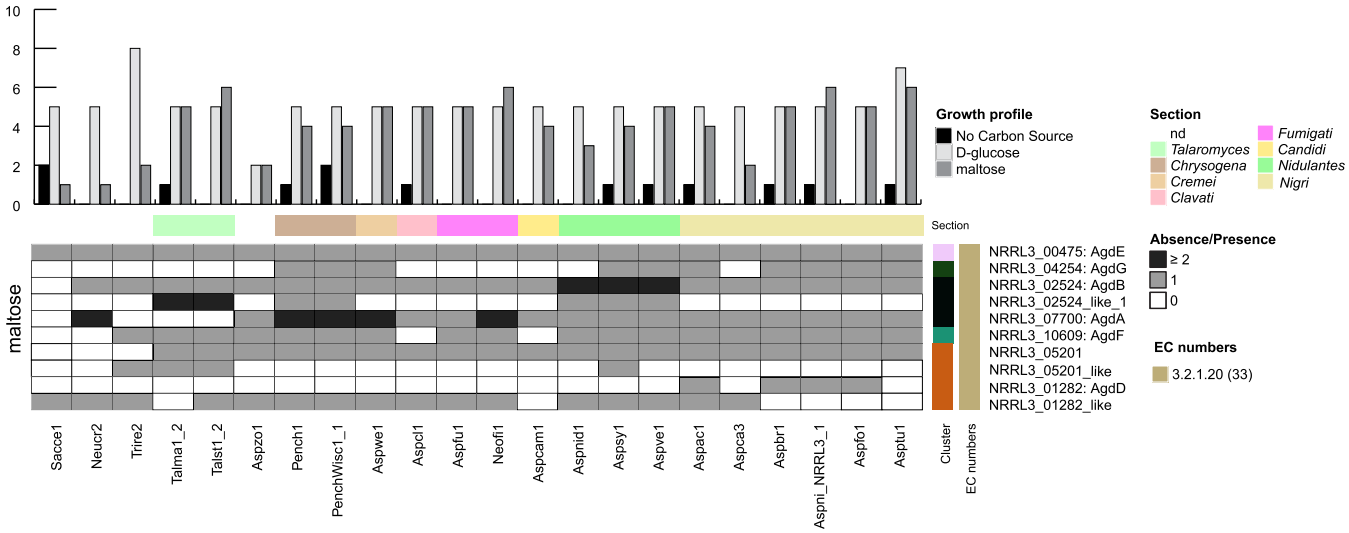

Fig. 6. Maltose heatmap abundance of *A. niger* orthologous proteins in the maltose catabolic pathway. The legends are the same as that of Fig. 3.

branched off. SucC is only present in *T. stipitatis, Penicillium* species, *A. brasiliensis* and *A. niger*, while SucB is present in all species except *P. zonata, A. clavatus* and *A. campestris*. SucA is missing in *A. clavatus, A. campestris, A. nidulans* and section *Fumigati*. In summary, *P. zonata, A. clavatus* and *A. camptestris* lack all three enzymes, section *Fumigati* only contains SucB, section *Nidulantes* and *Nigri* contain mainly SucA and SucB, while *Penicillium* species contain all three.

All species show different degrees of growth on sucrose. *T. stipitatus, A. wentii, A. brasiliensis* and *A. luchuensis* grew better on this disaccharide that on the monosacchrarides, D-glucose and D-fructose. *T. marneffei* showed reduced growth on sucrose compare to D-glucose and D-fructose. *Trichoderma reesei* is known for lacking invertases in its genome; however, it is able to grow when invertases from other species are heterologously produced (Berges *et al.* 1993), which corroborates the use of the resulting monosaccharides as substrates of glycolysis. On the other hand, *S. cerevisiae* and *N. crassa* showed growth on sucrose although they miss all three ortholog of genes identified in *A. niger* NRRL 3, which showed the sequence diversity within the CAZy family or a different mechanism to degrade sucrose.

D-Mannose metabolism

α-D-Mannose can be metabolized through glycolysis or utilized for the formation of glycoproteins through the production of GDP-α-D-mannose (Fig. 8). α-D-mannose is converted to α-D-mannose-6-phosphate by Hxk (see glycolysis metabolism section) and it is linked to glycolysis through mannose-6-phosphate isomerase (PmiA, EC 5.3.1.8) (Ruijter & Visser 1999, Upadhyay & Shaw 2006) that convers D-mannose 6-phosphate in D-fructose-6-phosphate.

To produce GDP-α-D-mannose, phosphomannomutase (PmmA, EC 5.4.2.8) converts D-mannose into α-D-mannose 1-phosphate that is further converted by the mannose-1-phosphate guanylyltransferase (EC 2.7.7.13) to GDP-α-D-mannose.

PmiA (NRRL3_11229) and its ortholog NRRL3_07971 and PmmA (NRRL3_10685) are present in all species. Two clusters encoding mannose-1-phosphate guanylyltransferase were found. All species contain enzymes from these two clusters (NRRL3_07837 and NRRL3_09951) except *A. carbonarius* that contains also a NRRL3_07837_like protein.

In general, all species show growth on D-mannose similar to D-glucose, except *T. marneffei* that showed highly reduced growth compared to D-glucose.

D-Galactose metabolism

In *A. niger* D-galactose can be metabolized through three different pathways (Fig. 9). The Leloir pathway and the oxidoreductive pathway final products of which enter glycolysis as D-fructose 6-phosphate or D-glucose 6-phosphate, and the non-

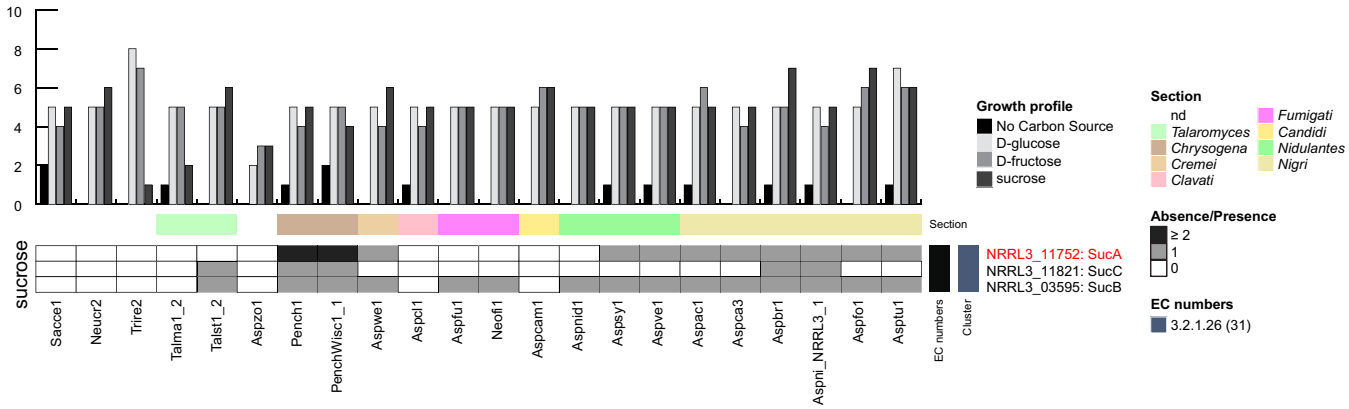

Fig. 7. Sucrose heatmap abundance of *A. niger* orthologous proteins in the sucrose catabolic pathway. The legends are the same as that of Fig. 3 .

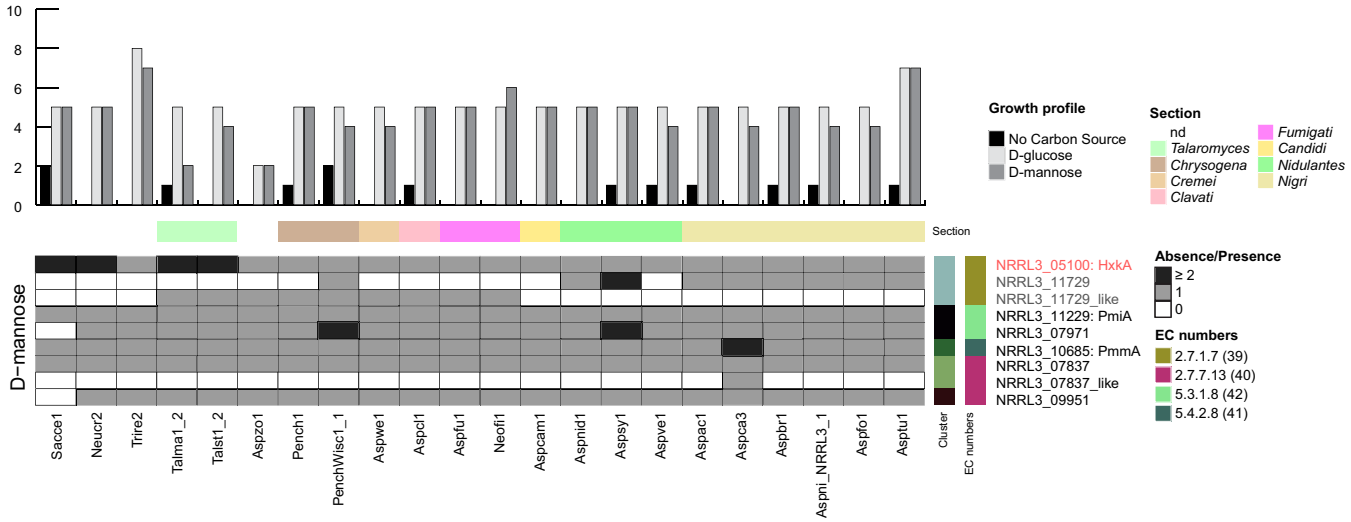

Fig. 8. D-mannose heatmap abundance of *A. niger* orthologous proteins in the D-mannose catabolic pathway. The legends are the same as that of Fig. 3.

phosphorylated DeLey-Doudoroff pathway whose final products are pyruvate and D-glyceraldehyde 3-phosphate (Elshafei & Abdel-Fatah 2001, Khosravi *et al.* 2015).

The Leloir pathway requires α-D-galactose, but the β-anomer is the most common form released during polysaccharides degradation. Aldose 1-epimerase (EC 5.1.3.3), also known as mutarotase epimerizes β-D-galactose into its α-anomer, which can occur both extracellularly and in the cytosol. α-D-galactose is phosphorylated by the galactokinase (GalK, EC 2.7.1.6) to form α-D-galactose 3-phosphate. Galactose-1-phosphate uridylyltransferase (GalT, EC 2.7.7.12), catalyses the transfer of a UMP group from UDP-glucose to galactose 1-phosphate, generating glucose 1-phosphate and UDP-galactose. To balance the reaction, UDP-galactose is converted to UDP-glucose by UDP-galactose 4-epimerase (EC 5.1.3.2). Finally, phosphoglucomutase (PgmB, EC 5.4.2.2) catalyses the conversion of glucose 1-phosphate into glucose 6-phosphate that enters glycolysis.

GalK (NRRL3_06978) and PgmB (NRRL3_05655) are present in all species while GalT (NRRL3_05970) is only missing in

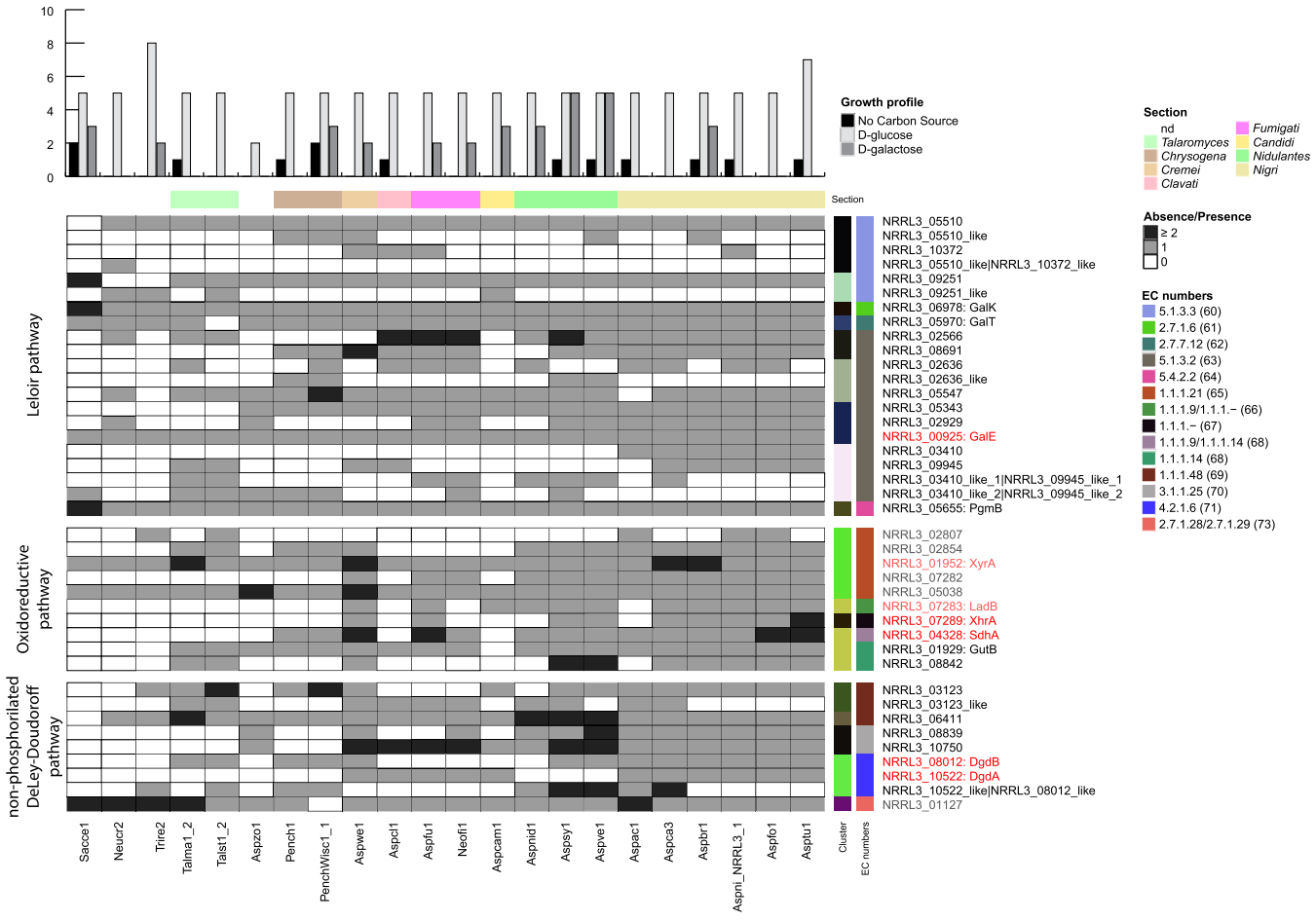

Fig. 9. D-galactose heatmap abundance of *A. niger* orthologous proteins in the D-galactose catabolic pathways. The legends are the same as that of Fig. 3.

T. stipitatus. There are two mutarotases predicted, belonging to different clusters, NRRL3_09251 (intracellular) and NRRL3_05510 (extracellular), which are both present in all species. A second extracellular mutarotase, NRRL3_10372 was found through orthology, and is present in *A. wentii, A. clavatus A. fumigatus and A. niger* exclusively. An ancestral group in this cluster contains an *N. crassa* sequence. Four clusters contain enzymes assigned to UDP-galactose 4-epimerase. GalE (NRRL3_00925) is the only ortholog present in all species. GalE and the other members of its cluster (NRRL3_05343 and NRRL3_02929) are the most conserved orthologs assigned to this activity. Phylogenetic analysis showed that almost every species contains one or more orthologs from other clusters, with the maximum number of proteins in section *Nigri*. β-D-galactose is metabolized directly by the oxidoreductive pathway. Several of the enzymes in this pathway have been shown to also be involved in L-arabinose and D-xylose metabolism (Fekete *et al.* 2004, Mojzita *et al.* 2010a,b, Mojzita *et al.* 2012a,b, Metz *et al.* 2013). The first step is the reduction of β-D-galactose to galatitol (EC 1.1.1.21). The reduction of the galactitol into L-xylo-3-hexulose is catalysed by the galactitol dehydrogenase (LadB, EC 1.1.1.-) follow by reduction to D-sorbitol catalysed by a reductase specific to D-galactose metabolism, the L-xylo-3-hexulose reductase (XhrA, EC 1.1.1.-). Sorbitol dehydrogenase (SdhA, EC 1.1.1.14) catalyses the conversion of sorbitol to keto-D-fructopyranose which through spontaneous reaction is transformed into its furanose form to enter glycolysis after Hxk transfers a phosphate group to it.

LadB (NRRL3_07283) and XhrA (NRRL3_07289) are the only single enzymes assigned in the oxidoreductive pathway, and both appear after *Penicillium* species splits off from the other species. They are missing in *A. clavatus*, *A. fischeri* and *A. aculeatus*. In contrast, there are several oxidorreductases from the same cluster assigned to the production of galactitol (EC 1.1.1.21), but XyrA (NRRL3_01952) and NRRL3_05038 are the only orthologs present in all species. The last step of the pathway, includes SdhA (NRRL3_04328) and GutB (NRRL3_01929) from two different clusters. GutB is present in all species after *T. reesei* except for *A. campestris*. While SdhA appears after *P. zonata* splits off from the other Eurotiales and is present in all species except *A. clavatus* and *A. campestris*. Neither the *A. niger* NRRL 3 genome functional annotation nor the curated network revealed an aldolase with tagatose-biphosphate affinity suggested in a previous study for the *A. nidulans* genome (Flipphi *et al.* 2009). This suggests that not all Aspergilli are able to convert D-galactose into dihydroxyacetone phosphate and D-glyceraldehyde 3-phosphate.

In the non-phosphorylated DeLey-Duodoroff pathway, the starting substrate is the β-D-galactose form which is oxidized into its lactone form, D-galactono-1,4-lactone by D-galactose dehydrogenase (EC 1.1.1.48). Gamma 1,4 lactonase (EC 3.1.1.25) converts D-galactone 1,4-lactone into D-galactonate, which is further converted by D-galactonate dehydratase (DgdA and DgdB, EC 4.2.1.6) into 2-dehydro-3-deoxy-D-galactonate. Although no aldolase has been associated to the reaction (EC 4.1.2.51), it has been shown (Elshafei & Abdel-Fatah 2001) that the last reaction produces pyruvate and D-glyceraldehyde from 2-dehydro-3-deoxy-D-galactonate. Pyruvate enters the TCA and glyoxylate cycles while D-glyceraldehyde enters glycolysis as D-glyceraldehyde-3-phosphate after a kinase (EC 2.7.1.28) transfers a phosphate group.

Two enzymes from different clusters have been assigned to D-galactose dehydrogenase. NRRL3_06411 is present in all species, while NRRL3_03123 is missing in *N. crassa*, *P. zonata, A. clavatus,* section *Fumigati* and *A. nidulans*. A second NRRL3_03123_like group that contains sequences of *Talaromyces* species, *A. wentii, A. clavati,* section *Fumigati* and some species of sections *Nidulantes* and *Nigri* was found during phylogenetic analysis. A single cluster, contains two γ-1,4-lactonase orthologs (NRRL3_08839 and NRRL3_10750) that are present in most species after *Talaromyces,* but *Penicillium* species lack both. Either DgdA or DgdB (NRRL3_08012 and NRRL3_10522) is present in all genomes after *T. reesei,* except in *P. zonata* and *A. nidulans*. An *A. nidulans* ortholog is present in a NRRL3_10522_like|NRRL3_08012_like group as well in other species before *A. campestris*. D-glyceraldehydo 3-phosphate (NRRL3_01127) is detected in all species except *P. rubens*.

For most of the species that were able to grow on D-galactose, mycelia was used as starting material. In spite of the genomes encoding the enzymes needed for D-galactose metabolism, growth from spores is impossible. Different scenarios need to be considered, transportation inside the cell or lack of induction (Fekete *et al.* 2012). Except for *A. brasiliensis* which was shown before to be the only *Nigri* able to grow from spores in D-galactose (Meijer *et al.* 2011). *A. brasiliensis* protein profile differs from the rest of the section *Nigri* in having a close ortholog NRRL3_05510_like. This group further contains *A. versicolor, A. wentii* and both *Penicillium* species orthologs. With the exception of *P. chrysosporum,* all three species show growth on D-galactose from spores. More studies need to be done to understand the different pathways involved on D-galactose metabolism.

D-galacturonic acid and L-rhamnose metabolism

D-galacturonic acid and L-rhamnose are both components from pectin and are metabolized through similar but different non-phosphorylated pathways (Alazi *et al.* 2017, Khosravi *et al.* 2017). Their recent identification explains why these pathways were not included in the previous inventory of central carbon metabolism in several Aspergilli (Flipphi *et al.* 2009).

Pectin hydrolysis released D-galacturonic acid monomers that are metabolized through the D-galacturonic pathway (Fig. 10). D-galacturonic acid reductase (GaaA, EC 1.1.1365/1.1.1.-) converts D-galacturonic acid into aldehyde-L-galactonate. (Alazi *et al.* 2017), which then is converted by GaaB (NRRL3_06890, EC 4.2.1.146), encoding a L-galactonate dehydratase, into 2-dehydro-3-deoxy-L-galactonate. 2-keto-3-deoxy-L-galactonate aldolase (GaaC, EC 4.1.2.54) catabolized the reaction splitting 2-keto-3-deoxy-L-galactonate into L-glyceraldehyde and pyruvate. Pyruvate enters TCA and glyoxylate cycles among other pathways, while L-glyceraldehyde is reduced to glycerol by aldolase reductase, (GaaD/LarA, 1.1.1.372). This enzyme has been shown to also be involved in L-arabitol, D-xylitol, D-eritritol and glycerol metabolism (de Groot *et al.* 2005, Mojzita *et al.* 2010a, Jovanovic *et al.* 2013).

All characterised enzymes, GaaA (NRRL3_05650), GaaB (NRRL3_06890), GaaC (NRRL3_05649) and GaaD/LarA (NRRL3_10050) are present in all studied species. A second enzyme (NRRL3_06930) assigned to D-galacturonic reductase activity that belongs to a different cluster than GaaA was found in all species, including *S. cerevisiae*. This enzyme could be the responsible for supporting reduced growth observed on D-galacturonic acid when *gaaA* was deleted (Alazi *et al.* 2017). Several species contain representatives of two NRRL3_06930_like groups. Another group was found in the

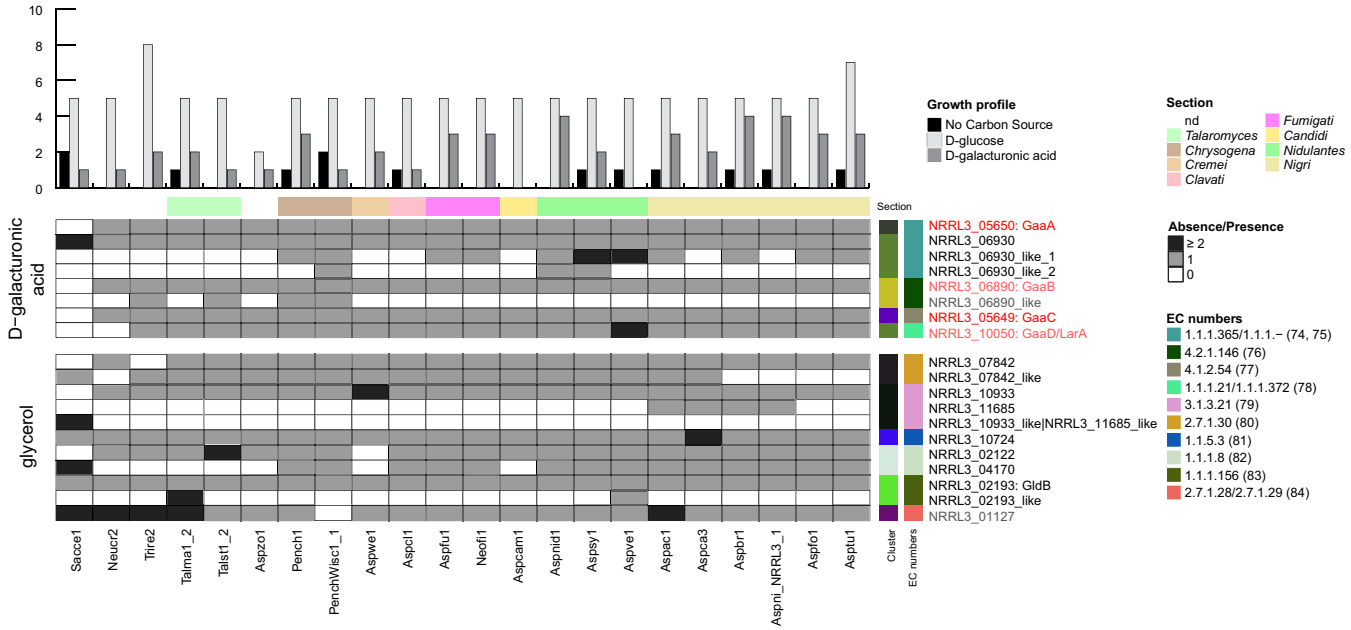

Fig. 10. D-galacturonic and glycerol heatmap abundance of *A. niger* orthologous proteins in the D-galacturonic and glycerol catabolic pathways. The legends are the same as that of Fig. 3.

GaaB cluster (NRRL3_06890_like) containing sequences from *Trichoderma reesei, Talaromyces stipitatus* and *Penicillium* species.

When evaluating growth on D-galacturonic acid, we need to take into consideration glycerol metabolism. Glycerol can be used as carbon source when present in the medium or as an intermediate of other pathways. Glycerol can be converted into *sn*-glycerol 3-phosphate by the glycerol kinase (EC 2.7.1.30). The reverse reaction is catalysed by glycerol 1-phosphatases (EC 3.1.3.21). *Sn*-glycerol 3-phosphate is converted into dihydroxyacetone phosphate that enters glycolysis, by two different reversible reactions. A flavin-dependent glycerol 3-phosphate dehydrogenase (EC 1.1.5.3) present in the mitochondrial membrane and a glycerol 3-phosphate dehydrogenases (EC 1.1.1.8). Dihydroxyacetone phosphate can be produced from dihydroxyacetone by a reversible reaction catalysed by glycerone kinase (EC 2.7.1.29). Dihydroxyacetone is produce by the reduction of glycerol catalysed by glycerol hydrogenase (GldB, EC1.1.1.156).

Glycerol 1-phosphatase NRRL3_10933, flavin-dependent glycerol 3-phosphate dehydrogenase NRRL3_10724 and GldB (NRRL3_01127) are present in all species. Glycerol kinase NRRL3_07842 is also present in all species except *S. cerevisiae* and *T. reesei*, and its ortholog group NRRL3_07842_like is only missing in section *Nigri* and *N. crassa*. Besides NRR3_10933, another phosphatase has been found in the same cluster (NRRL3_11685), which is only present in some species from section *Nigri*. The *A. wentii* genome lacks all glycerol 3-phosphatase candidates (NRRL3_02122 and NRRL3_04170), while the rest of the genomes contain at least one of the orthologs. Glycerol metabolism in *S. cerevisiae* is an important source of energy, and its genomes contains at least one copy of each enzyme.

With the exception of *P. rubens, A. clavatus, A. campestris* and *A. versicolor*, all species are able to grow on D-galacturonic acid as sole carbon source with no difference related to genome content. Therefore, we can assume, that glycerol metabolism is also functional in the rest of the species.

L-rhamnose is released from polysaccharides such as pectin or rhamnogalacturonan type I in its β-pyranose formed. Similar to D-galacturonic acid, L-rhamnose is transformed into the furanose form through a spontaneous reaction to enter the L-rhamnose degradation pathway (Fig. 11). L-rhamnose is converted into L-rhamnose-1,4-lactone by a NADH-dependent L-rhamnose-1-

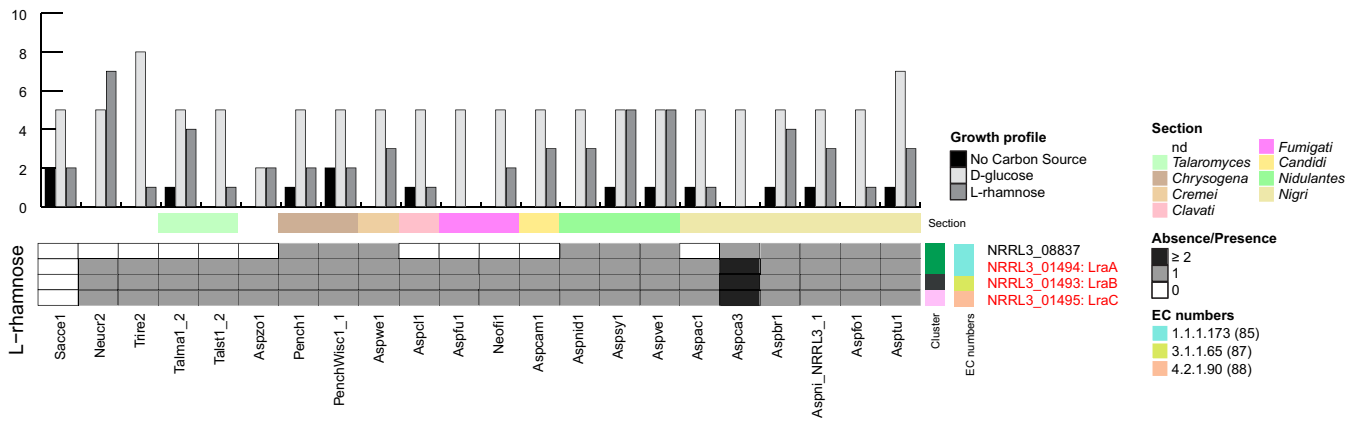

Fig. 11. L-rhamnose heatmap abundance of *A. niger* orthologous proteins in the L-rhamnose catabolic pathways. The legends are the same as that of Fig. 3.

dehydrogenase (LraA, EC 1.1.1.173) (Khosravi *et al.* 2017). The L-rhamnonic acid lactonase (LraB, EC 3.1.1.65), catalyses the hydrolysis of L-rhamnono-1,4-lactone to L-rhamnonate, while L-rhamnonate dehydratase (LraC, EC 4.1.2.53) converts this into 2-dehydro-3-deoxy-L-rhamnonate. The enzyme catalysing the last step in the pathway, an aldolase (EC 4.2.1.53) remains to be discovered.

All species contain one copy of each known Lra enzyme, LraA (NRRL3_01494), LraB (NRRL3_01493) and LraC (NRRL3_01495), except *A. carbonarius* that contain more than one copy of all three. A second paralog of LraA, NRRL3_08837 was found in the same cluster and is present in all species after *P. zonata* except *A. clavatus, A. campestris, A. aculeatus* and section *Fumigati*.

A similar phenotype was observed in species belonging to different clades. Those species that grow on L-rhamnose show either reduced or similar growth to D-glucose, except *N. crassa* which shows better growth than on D-glucose. Interestingly, *A. carbonarius* was not able to grow on L-rhamnose even though it has extra copies of all known Lra genes.

Pentose catabolism

Pentose catabolism in *A. niger* mainly involves two pathways, the pentose catabolic and pentose phosphate pathways. The pentose catabolic pathway includes L-arabinose and D-xylose catabolism to produce D-xylulose, and the pentose phosphate pathway includes D-ribulose, D-ribose and D-xylulose catabolism and the production of several glycolytic intermediates including NADPH necessary during glycolysis (Fig. 12).

In the first step in L-arabinose metabolism, L-arabinose reductase converts L-arabinose into L-arabitol (GaaD/LarA, EC 1.1.1.-, 1.1.1.21) (de Groot *et al.* 2005, Mojzita *et al.* 2010a, Jovanovic *et al.* 2013). L-arabitol oxidation to L-xylulose is catalysed by L-arabitol 4-dehydrogenase (LadA, EC 1.1.1.12) (Kim *et al.* 2010, Mojzita *et al.* 2012b). The last step is the reduction of L-xylulose to produce xylitol by xylitol dehydrogenase (LxrA, EC 1.1.1.10).

D-xylose conversion into xylitol is catabolized by D-xylose reductase (XyrA, EC 1.1.1.21). The final product of the separate L-arabinose and D-xylose pathways, xylitol, is reduced to D-xylulose by xylitol dehydrogenase (EC 1.1.1.9). which is phosphorylated by D-xylulose kinase to form D-xylulose 5-phosphate (XkiA, 2.7.1.17).

D-xylulose 5-phosphate originates from the pentose catabolic pathway and is one of the main substrates of the pentose phosphate pathway. Glycolysis connects to the pentose phosphate pathway through D-glucose 6-phosphate conversion to 6-phospho D-gluconate-1,5-lactone by the glucose-6-phosphate 1-dehydrogenase (GsdA, EC 1.1.1.49) (van den Broek *et al.* 1995). 6-phosphogluconolactonase (EC 3.1.1.31) hydrolyzes 6-phospho D-glucono-1,5-lactonate to produce D-gluconate 6-phosphate, which is further catabolized by 6-phosphogluconate dehydrogenase (EC 1.1.1.44) to D-ribulose 5-phosphate. The D-gluconate pathway also connects to the pentose phosphate pathway through D-gluconate 6-phosphate.

D-ribulose 5-phosphate can be produced from the degradation of D-ribulose catalysed by ribulokinase (RbtA, EC 2.7.1.47), and is then converted by ribulose-phosphate 3-epimerase (RpeA, EC 5.1.3.1) to D-xylulose 5-phosphate or by ribose 5-phosphate isomerase (Rpi, EC 5.3.1.6) to D-ribose 5-phosphate. Transketolase (TktA, EC 2.2.1.1) catalyses the reaction between D-ribose 5-phosphate and D-xylose 5-phosphate to produce D-sedoheptulose 7-phosphate and D-glyceraldehyde 3-phosphate. The latter can enter glycolysis or be utilized by transaldolases (Tal, EC 2.2.1.2) to produce D-erythrose 4-phosphate and β-D-fructofuranose 6-phosphate that can also enter glycolysis or other pathways. Transketolase (TktA), can also convert D-erythrose 4-phosphate and D-xylose 5-phosphate into β-D-fructose 6-phosphate and D-glyceraldehyde creating a loop in the pathway. A different transketolase (TktB, EC 2.2.1.3) catalyses the production of D-glyceraldehyde and dihydroxyacetone from D-xylulose 5-phosphate and formaldehyde.

Except for XkiA (NRRL3_04471) that is present in all species, all reactions in the pentose catabolic pathway have multiple enzymes assigned. Five enzymes have been assigned to L-arabitol and xylitol production. Through phylogenetic analysis, we found a total of eight proteins predicted to be involved in those reactions, but only two have been biochemically characterised GaaD/LarA and XyrA (Hasper *et al.* 2000, Jovanovic *et al.* 2013). *In vitro* analysis showed activity of GaaD/LarA and XyrA on L-arabinose, D-xylose and for the latter on D-galactose as well, but growth on L-arabinose or D-xylose is already reduced when LarA or XyrA are missing (Mojzita *et al.* 2010a). GaaD/LarA (NRRL3_10050) shows the highest affinity for L-arabinose and L-arabitol while XyrA (NRRL3_01952) shows higher affinity for D-xylose and xylitol compared to other enzymes. Orthologs of both genes can be found in all studied species. The NRRL3_05038 ortholog is also found in all species. Interestingly, both XyrA and NRRL3_05038 are the only enzymes for these reactions that are also present in *S. cerevisiae* and *N. crassa*. The other genomes contain different orthologs for several enzymes, up to eight in some species from section *Nigri*. LadA (NRRL3_02523) is present in all species but its ortholog NRRL3_00896 is missing in *N. crassa, T. stipitatus, P. zonata, A. clavatus* and *A. campestris*. Those species together with *Trichoderma reesei, Talaromyces marneffei, Penicillium* species*, A. campestris* and sections *Fumigati* and *Nidulantes* also miss NRRL3_08407 in their genome. LxrA (NRRL3_10884) is the only xylitol dehydrogenase characterised and it is present in all species. From the same cluster, NRRL3_04510 is missing in the genome of *A. campestris*. Species from section *Nidulantes* contain all four orthologs.

Several enzymes have been assigned to D-xylulose production but only some of them have been biochemically characterised. XdhA (NRRL3_09204) shows the highest activity with xylitol and D-xylulose, while LadA (NRRL3_02523) is also active on xylitol and D-xylulose but has higher activity on L-arabitol and L-xylulose (de Groot *et al.* 2007). XdhA, LadA and NRRL3_00319 are present in all species. From the same cluster, five additional orthologous groups were identified. *A. wentii, A. versicolor, A. brasiliensis* and *A. niger* contain representatives of all eight groups.

GsdA (NRRL3_05283), RbtA (NRRL3_00588), RpeA (NRRL3_00279), RpiA (NRRL3_10107) and TktA (NRRL3_11249) are present in all species, while PgiA (NRRL3_02043) is missing in *A. campestris*. Two different clusters contain five enzymes with 6-phosphogluconate dehydrogenase activity (EC 1.1.1.44), which were found through phylogenetic analysis. Only NRRL3_09981 is present in all species, including *S. cerevisiae*. Most of the species contain also NRRL3_09641 and NRRL3_09981 orthologs, except for *T. reesei*, but NRRL3_08784 is only present in *A. wentii, A. sydowii, A. versicolor, A. aculeatus* and *A. niger*.

Ribose 5-phosphate isomerase assigned enzymes (RpiA: NRRL3_10107 and RpiB: NRRL3_06024) belong to different clusters. RpiB is missing in *A. fumigatus*, while NRRL3_06024_like group only contains orthologs from *A. wentii*

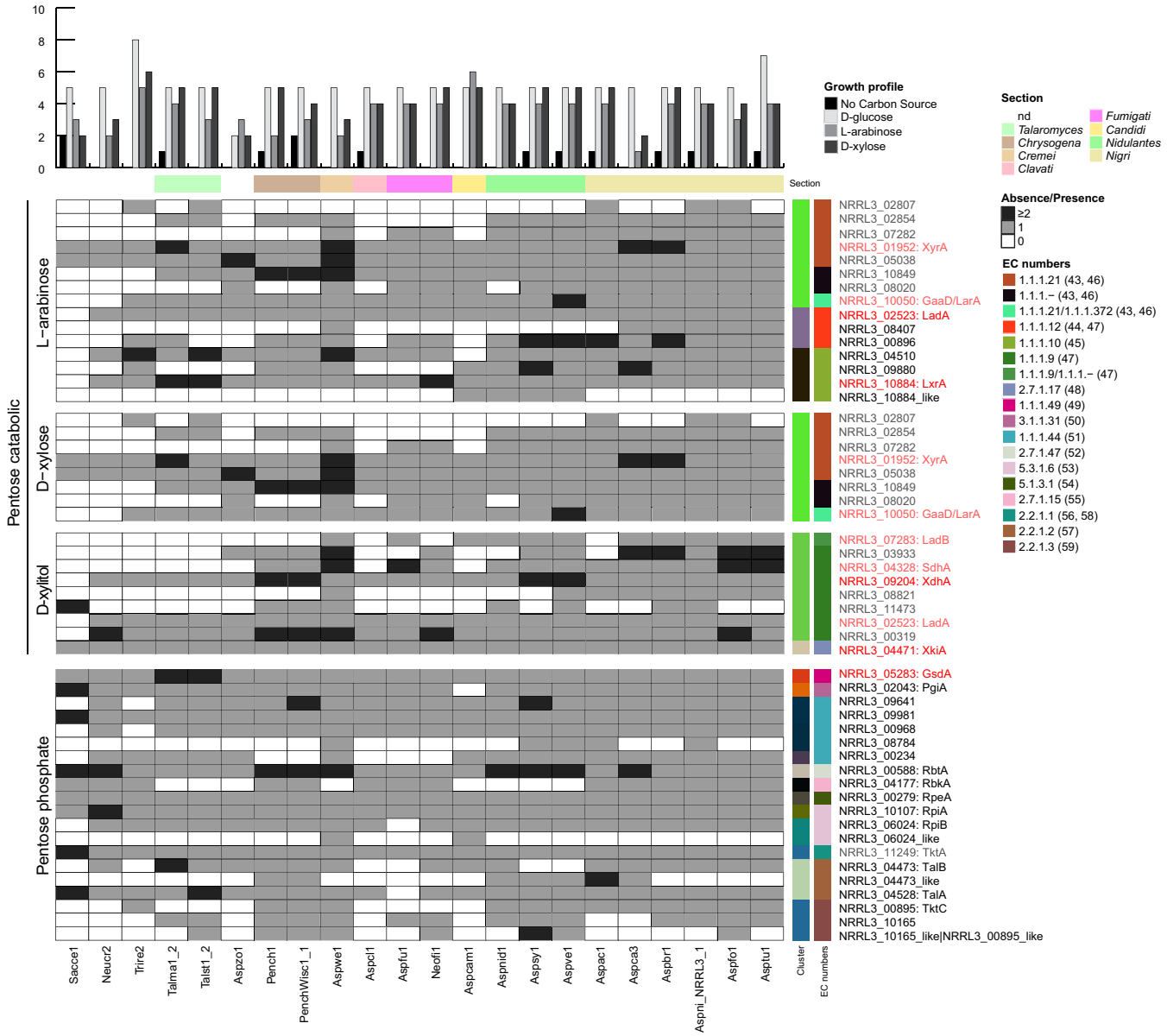

Fig. 12. Pentose metabolism heatmap abundance of *A. niger* orthologous proteins in the pentose metabolic pathways. The legends are the same as that of Fig. 3.

and *A. campestris*. There are two transaldolases predicted (TalA, NRRL3_04528 and TalB, NRRL3_04473) that belong to the same cluster. TalA is missing in *A. fumigatus* while TalB is missing in *T. reesei, A. clavatus, A. fumigatus* and some species from section *Nidulantes*. This implies that there is no transaldolase similar to these genes in *A. fumigatus*, which implies that a non-homologous gene is responsible for this function or that the gene has not been identified or assembled in the *A. fumigatus* genome. In contrast, NRRL3_04473_like orthologs are only present in *A. aculeatus, A. carbonarius,* section *Nidulantes* and *Penicillium* species. Dihydroxyacetone synthase (EC 2.2.1.3), TktC (NRRL3_00895) and NRRL3_10165 belong to the same cluster as TktA (NRRL3_11249, EC 2.2.1.1). They have been assigned a different reaction in the network. They catalysed the production of dihydroxyacetone and D-glyceraldehyde 3-phosphate from D-xylulose 5-phosphate and formaldehyde. Only *Penicillium* species, *A. versicolor* and *A. luchuensis* contain all three orthologs predicted, most species contain only two orthologs in their genomes.

All species were able to grow on L-arabinose and D-xylose to different degrees. Interestingly, even though both *Penicillium* species have the same pattern (except extra copies of NRRL3_09641) they behaved differently. *P. chrysogenum* grew on D-xylose as good as on D-glucose while *P. rubens* showed better growth on L-arabinose but always lower than on D-glucose. *A. campestris* grew better than any other species on L-arabinose compared to D-glucose, despite missing several orthologous enzymes for most of the activities. Yeasts like *S. cerevisiae* are known for being unable to grow on pentoses unless genetically modified (Jeffries 2006, Hahn-Hagerdal *et al.* 2007). As shown in the Fig. 12, they miss most of the orthologous proteins involved exclusively in pentose catabolic metabolism, but not pentose phosphate metabolism that is needed to produce NADPH for glycolysis.

DISCUSSION

Automated pipelines for genome assembly, gene prediction and functional annotation have evolved over the years to handle increasingly larger next generation sequencing output files. At the same time, human supervision has reduced compared to the initial genomes. In our protein assessment analysis, we show

how gene prediction quality varies between genomes regardless of organism, technology or consortia. Comparing the predicted proteome against a database of conserved proteins enables evaluation of the completeness of the proteome and therefore the quality of the gene annotation and genome assembly. Different options can be used to improve a new genome. The most common option is to use a similar genome from a closely related species to improve gene prediction and functional annotation, but this only provides significant improvement if that genome is of high quality. However, not many high-quality genomes are available. Gold-standard genomes are difficult to acquire, they require a high amount of omics data and manual curation by experts at every stage of the process. We showed in this study that more than 90 % of predicted proteins in genus *Penicillium* shared sequence similarity with proteins from genus *Aspergillus*. Using a gold-standard genome can improve gene prediction and protein functional annotation and, if combined with metabolic network analysis, it can be used to predict metabolic pathways with higher confidence. This analysis goes significantly beyond the use of the KEGG automatic pipeline together with orthology analysis and a partially completed genome sequence that was used in the previous study based on *A. nidulans* (Flipphi *et al.* 2009).

In this study, we also showed how orthology and protein profiling based on a manually curated metabolic network of a gold-standard genome (*A. niger* NRRL 3) can be used to predict primary carbon metabolism in a number of species and correlate the genome content with growth abilities. As expected, more distant Ascomycota, *S. cerevisiae, N. crassa* and *T. reesei*, contain lower number of orthologs of *A. niger* NRRL 3 genes compared to other members of the genus *Aspergillus*. In particular, *S. cerevisiae* lacks several proteins and even complete pathways compare to *A. niger* NRRL 3.

Glycolysis operates with remarkable efficiency if we consider the number of pathways that link to glycolysis through D-glucose 6-phosphate and D-fructose 6-phosphate. Some of the steps of the pathway have been associated with several enzymes, but we find less redundancy in key enzymatic reactions, especially in the steps leading to production of pyruvate from D-glyceraldehyde 3-phosphate. Species from section *Nigri* are known for citric acid production (Max *et al.* 2010), which correlates with the high number of orthologous genes assigned to the glyoxylate and TCA cycles. A similar pattern is observed for the genes involved in D-gluconate production from D-glucose, which also correlates with previous data showing that *Penicillium* and *Aspergillus* section *Nigri* species are high gluconate producers (de Vries *et al.* 2017).

A clear example of the presence of an alternative pathway is growth of *S. cerevisiae* on D-gluconate (Oliva Neto *et al.* 2014), while our study indicates that its genome misses all the enzymes of the pathway. Another example of alternative mechanism occurs during growth on sucrose. *S. cerevisiae, N. crassa, P. zonata, A. clavatus* and *A. campestris* genomes lack all three GH32 proteins identified in *A. niger,* even though they can grow on sucrose almost as well as on D-glucose and D-fructose. This suggests that these species contain a non-homologous alternative enzyme that can split sucrose.

In contrast, the *A. clavatus* genome contains genes that encode for all identified enzymes necessary for D-galacturonic catabolism, but it is not able to grow on D-galacturonic acid as a sole carbon source. A possible explanation for this is the absence of the galacturonic acid transporter in *A. clavatus* (Mäkelä *et al.* 2018). Tight regulation on the transporter seems to be causing the lack of growth from spores on D-galactose as sole carbon source (Fekete *et al.* 2012). Even though most species contain orthologs for the *A. niger* genes, most of them are still unable to grow from spores, similar to what is observed for *A. niger* itself. They are able to grow when small amounts of other monosaccharides are present in the media or when they are inoculated from mycelia, which supports that the problem is mainly related to uptake of D-galactose during germination.

In summary, we show that combining a gold-standard genome with a manual curated metabolic network and phylogeny can significantly improve functional annotation of less studied species. It helps with the identification of essential proteins and alternative pathways for the degradation of several compounds that otherwise could have been missed or required significant efforts through classical genetics. We observed low diversity between closely related species for monosaccharide catabolism, and even high conservation between more distant fungi. This is likely due to the important role monosaccharides have as a carbon source for fungi. We can expect a higher diversity between species for other aspects of fungal biology, such as in polysaccharide degradation, transport affinity and regulatory systems. In fact, other papers in this issue reveal high diversity with respect to plant biomass degradation (Mäkelä *et al.* 2018), stress tolerance (Emri *et al.* 2018) and sexual and asexual reproduction (Ojeda-López *et al.* 2018).

ACKNOWLEDGEMENTS

The authors would like to thank Jos Houbraken for his help with fungal taxonomy. This work was carried out on the Dutch national e-infrastructure with the support of SURF Cooperative Grant e-infra 170088, and on funding from Genome Canada and Génome Québec. M.V.A.P. was supported by a grant of the Dutch Technology Foundation STW, Applied Science division of NWO; and the Technology Program of the Dutch Ministry of Economic Affairs 016.130.609 to R.P.dV.

APPENDIX A. SUPPLEMENTARY DATA

Supplementary data to this article can be found online at https://doi.org/10.1016/j.simyco.2018.10.001.

REFERENCES

Aguilar-Pontes MV, de Vries RP, Zhou M (2014). Post-)genomics approaches in fungal research. *Briefings in Functional Genomics* **13**: 424–439.

Alazi E, Khosravi C, Homan TG, *et al.* (2017). The pathway intermediate 2-keto-3-deoxy-L-galactonate mediates the induction of genes involved in D-galacturonic acid utilization in *Aspergillus niger*. *FEBS Letters* **591**: 1408–1418.

Altschul SF, Gish W, Miller W, *et al.* (1990). Basic local alignment search tool. *Journal of Molecular Biology* **215**: 403–410.

Arnaud MB, Cerqueira GC, Inglis DO, *et al.* (2012). The *Aspergillus* Genome Database (AspGD): recent developments in comprehensive multispecies curation, comparative genomics and community resources. *Nucleic Acids Research* **40**: 653–659.

Benocci T, Aguilar-Pontes MV, Kun RS, *et al.* (2018). ARA1 regulates not only L-arabinose but also D-galactose catabolism in *Trichoderma reesei*. *FEBS Letters* **592**: 60–70.

Berges T, Barreau C, Peberdy JF, *et al.* (1993). Cloning of an *Aspergillus niger* invertase gene by expression in *Trichoderma reesei*. *Current Genetics* **24**: 53–59.

Boddy LM, Berges T, Barreau C, *et al.* (1993). Purification and characterisation of an *Aspergillus niger* invertase and its DNA sequence. *Current Genetics* **24**: 60–66.

Cerqueira GC, Arnaud MB, Inglis DO, *et al*. (2014). The Aspergillus Genome Database: multispecies curation and incorporation of RNA-Seq data to improve structural gene annotations. *Nucleic Acids Research* **42**: 705–710.

Cherry JM, Hong EL, Amundsen C, *et al*. (2012). Saccharomyces Genome Database: the genomics resource of budding yeast. *Nucleic Acid Research* **40**: 700–705.

Clements JM, Roberts CF (1985). Molecular cloning of the 3-phosphoglycerate kinase (PGK) gene from *Aspergillus nidulans*. *Current Genetics* **9**: 293–298.

de Graaff L, van den Broeck H, Visser J (1992). Isolation and characterization of the *Aspergillus niger* pyruvate kinase gene. *Current Genetics* **22**: 21–27.

de Groot MJ, Prathumpai W, Visser J, *et al*. (2005). Metabolic control analysis of *Aspergillus niger* L-arabinose catabolism. *Biotechnology Progress* **21**: 1610–1616.

de Groot MJL, van den Dool C, Wosten HAB, *et al*. (2007). Regulation of pentose cataboolic pathways genes of *Aspergillus niger*. *Food Technology and Biotechnology* **45**: 134–138.

de Vries RP, Burgers K, van de Vondervoort PJ, *et al*. (2004). A new black Aspergillus species, *A. vadensis*, is a promising host for homologous and heterologous protein production. *Applied and Environmental Microbiology* **70**: 3954–3959.

de Vries RP, Riley R, Wiebenga A, *et al*. (2017). Comparative genomics reveals high biological diversity and specific adaptations in the industrially and medically important fungal genus Aspergillus. *Genome Biology* **18**: 28.

de Vries RP, Robert V, Wiebenga A. The fung-growth database. Linking growth to genome. http://www.fung-growth.org/.

Elshafei AM, Abdel-Fatah OM (2001). Evidence for a non-phosphorylated route of galactose breackdown in cell-free extracts of *Aspergillus niger*. *Enzyme and Microbial Technology* **29**: 76–83.

Emri T, Antal K, Riley R, *et al*. (2018). Duplications and losses of genes encoding known elements of the stress defense system of the Aspergilli contribute to the evolution of these filamentous fungi but do not directly influence their environmental stress tolerance. *Studies in Mycology* **91**: 23–36.

Enright AJ, Van Dongen S, Ouzounis CA (2002). An efficient algorithm for latge-scale detection of protein families. *Nucleic Acid Research* **30**: 1575–1584.

Fedorova ND, Khaldi N, Joardar VS, *et al*. (2008). Genomic islands in the pathogenic filamentous fungus *Aspergillus fumigatus*. *PLoS Genetics* **4**: e1000046.

Fekete E, Karaffa L, Sandor E, *et al*. (2004). The alternative D-galactose degrading pathway of *Aspergillus nidulans* proceeds via L-sorbose. *Archives of Microbiology* **181**: 35–44.

Fekete E, de Vries RP, Seiboth B, *et al*. (2012). D-Galactose uptake is nonfunctional in the conidiospores of *Aspergillus niger*. *FEMS Microbiology Letters* **329**: 198–203.

Flipphi M, Sun J, Robellet X, *et al*. (2009). Biodiversity and evolution of primary carbon metabolism in *Aspergillus nidulans* and other Aspergillus spp. *Fungal Genet Biol* **46**(Suppl 1): 19–44.

Galagan JE, Calvo SE, Borkovich KA, *et al*. (2003). The genome squence of the filamentous fungus Neurospora crassa. *Nature* **422**: 859–868.

Galagan JE, Calvo SE, Cuomo C, *et al*. (2005). Sequencing of *Aspergillus nidulans* and comparative analysis with *A. fumigatus* and *A. oryzae*. *Nature* **438**: 1105–1115.

Genozymes (2009). *Genozymes for bioproducts and bioprocesses development*. http://genome.fungalgenomics.ca/.

Goffeau A, Barrell BG, Bussey H, *et al*. (1996). Life with 6000 genes. *Science* **274**(546): 563–567.

Gregory TR (2001). Coincidence, coevolution, or causation? DNA content, cell size, and the C-value enigma. *Biological Reviews* **76**: 65–101.

Grigoriev IV, Nordberg H, Shabalov I, *et al*. (2012). The genome portal of the Department of Energy Joint Genome Institute. *Nucleic Acid Research* **40**: 26–32.

Grigoriev IV, Nikitin R, Haridas S, *et al*. (2014). MycoCosm portal: gearing up for 1000 fungal genomes. *Nucleic Acids Research* **42**: 699–704.

Gu Z, Eils R, Schlesner M (2016). Complex heatmaps reveal patterns and correlation in multidimensional genomic data. *Bioinformatics* **32**: 2847–2849.

Habison A, Kubicek CP, Röhr M (1983). Partial purification and regulatory properties of phosphofructokinase from *Aspergillus niger*. *Biochemical Journal* **209**: 669–676.

Hahn-Hagerdal B, Karhumaa K, Jeppsson M, *et al*. (2007). Metabolic engineering for pentose utilization in *Saccharomyces cerevisiae*. In: *Advances in Biochemical Engineering/Biotechnology* (Olsson L, ed). Springer, Berlin, Heidelberg: 147–177.

Harmsen HJM, Kubicek-Pranz EM, Rohr M, *et al*. (1992). Regulation of 6-phosphofructo-2-kinase from the citric-acid-accumulating fungus *Aspergillus niger*. *Applied Microbiology and Biotechnology* **37**: 784–788.

Hasper AA, Visser J, de Graaf LH (2000). The *Aspergillus niger* transcriptional activator XlnR, which is involved in the degradation of the polysaccharides xylan and cellulose, also regulates D-xylose reductase gene expression. *Molecular Microbiology* **36**: 193–200.

Hawksworth DL (1991). The fungal dimension of biodiversity: magnitude, significance, and conservation. *Mycological Research* **95**: 641–665.

Hynes MJ, Szewczyk E, Murray SL, *et al*. (2007). Transcriptional control of gluconeogenesis in *Aspergillus nidulans*. *Genetics* **176**: 139–150.

Jagannathan V, Singh K, Damodaran M (1956). Carbohydrate metabolism in citric acid fermentation. 4. Purification and properties of aldolase from *Aspergillus niger*. *Biochemical Journal* **63**: 94–105.

Jeffries TW (2006). Engineering yeasts for xylose metabolism. *Current Opinion in Biotechnology* **17**: 320–326.

Jovanovic B, Mach RL, Mach-Aigner AR (2013). Characterization of erythrose reductases from filamentous fungi. *AMB Express* **3**: 43.

Kanehisa M, Goto S, Hattori M, *et al*. (2006). From genomics to chemical genomics: new developments in KEGG. *Nucleic Acids Research* **34**: 354–357.

Katoh K, Standley DM (2013). MAFFT multiple sequence alignment software version 7: improvements in performance and usability. *Molecular Biology and Evolution* **30**: 772–780.

Khosravi C, Benocci T, Battaglia E, *et al*. (2015). Sugar catabolism in Aspergillus and other fungi related to the utilization of plant biomass. *Advances in Applied Microbiology* **90**: 1–28.

Khosravi C, Kun RS, Visser J, *et al*. (2017). *In vivo* functional analysis of L-rhamnose metabolic pathway in *Aspergillus niger*: a tool to identify the potential inducer of RhaR. *BMC Microbiology* **17**: 214.

Kim B, Sullivan RP, Zhao H (2010). Cloning, characterization, and engineering of fungal L-arabinitol dehydrogenases. *Applied Microbiology Biotechnology* **87**: 1407–1414.

Kimura A, Takata M, Sakai O, *et al*. (1992). Complete amino acid sequence of crystalline alpha-glucosidase from *Aspergillus niger*. *Bioscience Biotechnology and Biochemistry* **56**: 1368–1370.

Kinsella RJ, Kahari A, Haider S, *et al*. (2011). Ensembl BioMarts: a hub for data retrieval across taxonomic space. *Database (Oxford)* **2011**: bar030.

Kjaerbolling I, Vesth TC, Frisvad JC, *et al*. (2018). Linking secondary metabolites to gene clusters through genome sequencing of six diverse Aspergillus species. *Proceedings of the National Academy of Science USA* **115**: E753–E761.

Kocsube S, Perrone G, Magista D, *et al*. (2016). *Aspergillus* is monophyletic: Evidence from multiple gene phylogenies and extrolites profiles. *Studies in Mycology* **85**: 199–213.

Kornberg HL, Madsen NB (1958). The metabolism of C2 compounds in microorganisms. 3. Synthesis of malate from acetate via the glyoxylate cycle. *Biochemical Journal* **68**: 549–557.

Kowalczyk JE, Benoit I, de Vries RP (2014). Regulation of plant biomass utilization in *Aspergillus*. *Advance in Applied Microbiology* **88**: 31–56.

Kumar S, Stecher G, Peterson D, *et al*. (2012). MEGA-CC: computing core of molecular evolutionary genetics analysis program for automated and iterative data analysis. *Bioinformatics* **28**: 2685–2686.

Kunze M, Pracharoenwattana I, Smith SM, *et al*. (2006). A central role for the peroxisomal membrane in glyoxylate cycle function. *Biochimica et Biophysica Acta (BBA) - Molecular Cell Research* **1763**: 1441–1452.

Lombard V, Golaconda Ramulu H, Drula E, *et al*. (2014). The carbohydrate-active enzymes database (CAZy) in 2013. *Nucleic Acid Research* **42**: 490–495.

Lynch M, Conery JS (2003). The origins of genome complexity. *Science* **302**: 1401–1404.

Machida M, Asai K, Sano M, *et al*. (2005). Genome sequencing and analysis of *Aspergillus oryzae*. *Nature* **438**: 1157–1161.

Mäkelä MR, DiFalco M, McDonnell E, *et al*. (2018). Genomic and proteomics diversity in plant biomass degradation approaches among Aspergilli. *Studies in Mycology* submitted.

Malherbe DF, du Toit M, Cordero Otero RR, *et al*. (2003). Expression of the *Aspergillus niger* glucose oxidase gene in Saccharomyces cerevisiae and its potential applications in wine production. *Appl Microbiol Biotechnol* **61**: 502–511.

Marcet-Houben M, Ballester AR, de la Fuente B, *et al*. (2012). Genome sequence of the necrotrophic fungus *Penicillium digitatum*, the main post-harvest pathogen of citrus. *BMC Genomics* **13**: 646.

Martinez D, Berka RM, Henrissat B, *et al*. (2008). Genome sequencing and analysis of the biomass-degrading fungus *Trichoderma reesei* (syn. *Hypocrea jecorina*). *Nature Biotechnolgy* **26**: 553–560.

Max B, Salgado JM, Rodriguez N, *et al*. (2010). Biotechnological production of citric acid. *Brazilian Journal of Microbiology* **41**: 862–875.

McAlister-Henn L, Steffan JS, Minard KI, *et al*. (1995). Expression and function of a mislocalized form of peroxisomal malate dehydrogenase (MDH3) in yeast. *Journal of Biological Chemistry* **270**: 21220–21225.

McKnight GL, O'Hara PJ, Parker ML (1986). Nucleotide sequence of the triosephosphate isomerase gene from *Aspergillus nidulans*: implications for a differential loss of introns. *Cell* **46**: 143–147.

Meijer M, Houbraken JA, Dalhuijsen S, *et al*. (2011). Growth and hydrolase profiles can be used as characteristics to distinguish *Aspergillus niger* and other black aspergilli. *Studies in Mycology* **69**: 19–30.

Metz B, Mojzita D, Herold S, *et al*. (2013). A novel L-xylulose reductase essential for L-arabinose catabolism in *Trichoderma reesei*. *Biochemistry* **52**: 2453–2460.

MIT BloHa Fungal Genome Initiative. http://www.broadinstitute.org/scientific-community/science/projects/fungal-genome-initiative/fungal-genome-initiative.

Mojzita D, Penttila M, Richard P (2010a). Identification of an L-arabinose reductase gene in *Aspergillus niger* and its role in L-arabinose catabolism. *Journal of Biological Chemistry* **285**: 23622–23628.

Mojzita D, Vuoristo K, Koivistoinen OM, *et al*. (2010b). The 'true' L-xylulose reductase of filamentous fungi identified in *Aspergillus niger*. *FEBS Letters* **584**: 3540–3544.

Mojzita D, Herold S, Metz B, *et al*. (2012a). L-xylo-3-hexulose reductase is the missing link in the oxidoreductive pathway for D-galactose catabolism in filamentous fungi. *Journal of Biological Chemistry* **287**: 26010–26018.

Mojzita D, Koivistoinen OM, Maaheimo H, *et al*. (2012b). Identification of the galactitol dehydrogenase, LadB, that is part of the oxido-reductive D-galactose catabolic pathway in *Aspergillus niger*. *Fungal Genetics and Biology* **49**: 152–159.

Nakamura A, Nishimura I, Yokoyama A, *et al*. (1997). Cloning and se*quencing of an a-glucosidase gene from Aspergillus niger and its expression in* A. nidulans. *Journal of Biotechnology* **53**: 75–84.

Narayanan B, Niu W, Joosten HJ, *et al*. (2009). Structure and function of 2,3-dimethylmalate lyase, a PEP mutase/isocitrate lyase superfamily member. *Journal of Molecular Biology* **386**: 486–503.

Nielsen H, Engelbrecht J, Brunak S, *et al*. (1997). Identification of prokaryotic and eukaryotic signal peptides and prediction of their cleavage sites. *Protein Engineering* **10**: 1–6.

Nierman WC, Pain A, Anderson MJ, *et al*. (2005). Genomic sequence of the pathogenic and allergenic filamentous fungus *Aspergillus fumigatus*. *Nature* **438**: 1151–1156.

Nierman WC, Fedorova-Abrams ND, Andrianopoulos A (2015). Genome sequence of the AIDS-associated pathogen *Penicillium marneffei* (ATCC18224) and its near taxonomic relative *Talaromyces stipitatus* (ATCC10500). *Genome Announcements* **3** e01559-01514.

Ojeda-López M, Chen W, Eagle EE, *et al*. (2018). Evolution of asexual and sexual reproduction in the Aspergilli. *Studies in Mycology* **91**: 37–59.

Oliva Neto Pd, Lima FAd, Silva KCd, *et al*. (2014). Chemical inhibition of the contaminant *Lactobacillus fermentum* from distilleries producing fuel bioethanol. *Brazilian Archives of Biology and Technology* **57**: 441–447.

Panneman H, Ruijter GJ, van den Broeck HC, *et al*. (1996). Cloning and biochemical characterisation of an *Aspergillus niger* glucokinase. Evidence for the presence of separate glucokinase and hexokinase enzymes. *European Journal of Biochemistry* **240**: 518–525.

Panneman H, Ruijter GJ, van den Broeck HC, *et al*. (1998). Cloning and biochemical characterisation of *Aspergillus niger* hexokinase. The enzyme is strongly inhibited by physiological concentrations of trehalose 6-phosphate. *European Journal of Biochemistry* **258**: 223–232.

Payne GA, Nierman WC, Wortman JR, *et al*. (2006). Whole genome comparison of *Aspergillus flavus* and *A. oryzae*. *Medical Mycology* **44**: 9–11.

Poulsen BR, Nohr J, Douthwaite S, *et al*. (2005). Increased NADPH concentration obtained by metabolic engineering of the pentose phosphate pathway in *Aspergillus niger*. *FEBS Journal* **272**: 1313–1325.

Punt PJ, Dingemanse MA, Jacobs-Meijsing BJ, *et al*. (1988). Isolation and characterization of the glyceraldehyde-3-phosphate dehydrogenase gene of *Aspergillus nidulans*. *Gene* **69**: 49–57.

Quevillon E, Silventoinen V, Pillai S, *et al*. (2005). InterProScan: protein domains identifier. *Nucleic Acids Research* **33**: 116–120.

R Core Team (2017). *R: a language and environment for statistical computing*. R Foundation for statistical computing. https://www.r-project.org/.

Roumelioti K, Vangelatos I, Sophianopoulou V (2010). A cryptic role of a glycolytic-gluconeogenic enzyme (aldolase) in amino acid transporter turnover in *Aspergillus nidulans*. *Fungal Genetics and Biology* **47**: 254–267.

Ruijter GJG, van de Vondervoort PJI, Visser J (1999). Oxalic acid production by *Aspergillus niger*: an oxalate-non-producting mutant produces citric acid at ph 5 and in the presence of manganese. *Microbiology* **145**: 2569–2576.

Ruijter GJG, Visser J (1999). Characterization of *Aspergillus niger* phosphoglucose isomerase. Use for quantitative determination of erythrose 4-phosphate. *Biochimie* **81**: 267–272.

Shindia AA, El-Sherbeny GA, El-Esawy AE, *et al*. (2006). Production of gluconic acid by some local fungi. *Mycobiology* **34**: 22–29.

Simao FA, Waterhouse RM, Ioannidis P, *et al*. (2015). BUSCO: assessing genome assembly and annotation completeness with single-copy orthologs. *Bioinformatics* **31**: 3210–3212.

Streatfield SJ, Toews S, Roberts CF (1992). Functional analysis of the expression of the 3'-phosphoglycerate kinase *pgk* gene in *Aspergillus nidulans*. *Molecular And General Genetics* **233**: 231–240.

Streatfield SJ, Roberts CF (1993). Disruption of the 3' phosphoglycrerate kinase (*pgk*) gene in *Aspergillus nidulans*. *Current Genetics* **23**: 123–128.

Upadhyay S, Shaw BD (2006). A phosphoglucose isomerase mutant in *Aspergillus nidulans* is defective in hyphal polarity and conidiation. *Fungal Genetics and Biology* **43**: 739–751.

van den Berg MA, Albang R, Albermann K, *et al*. (2008). Genome sequencing and analysis of the filamentous fungus *Penicillium chrysogenum*. *Nature Biotechnolgy* **26**: 1161–1168.

van den Broek P, Goosen T, Wennekes B, *et al*. (1995). Isolation and characterization of the glucose-6-phosphate dehydrogenase encoding gene (*gsdA*) from *Aspergillus niger*. *Molecular and General Genetics* **247**: 229–239.

Wallis GL, Hemming FW, Peberdy JF (1997). Secretion of two beta-fructofuranosidases by *Aspergillus niger* growing in sucrose. *Archives of Biochemistry and Biophysics* **345**: 214–222.

Watson M (2018). Mind the gap – ignoring errors in long read assemblies cirtically affects protein prediciton. *Biorxiv*. https://doi.org/10.1101/285049.

Wickham H (2009). *ggplot2: elegant graphics for data analysis*. Springer-Verlag, New York.

Yu G, Smith DK, Zhu H, *et al*. (2017). ggtree: an r package for visualization and annotation of phylogenetic trees with their covariates and other associated data. *Methods in Ecology and Evolution* **8**: 28–36.