Unlocking the Puzzling Biology of the Black Perigord Truffle Tuber melanosporum

ARTICLE *in* JOURNAL OF PROTEOME RESEARCH · OCTOBER 2013

Impact Factor: 4.25 · DOI: 10.1021/pr400650c · Source: PubMed

Journal of proteome research

Article

pubs.acs.org/jpr

Unlocking the Puzzling Biology of the Black Périgord Truffle *Tuber melanosporum*

Mohammad Tawhidul Islam,[†,‡,#] Abidali Mohamedali,[†,#] Gagan Garg,[†,‡,#] Javed Mohammed Khan,[†,‡] Alain-Dominique Gorse,[§] Jeremy Parsons,[§] Peter Marshall,[∥] Shoba Ranganathan,*[,†,‡,⊥] and Mark S. Baker*[,†]

[†]Department of Chemistry and Biomolecular Sciences, Macquarie University, NSW 2109, Australia

[‡]ARC Centre of Excellence in Bioinformatics, Macquarie University, NSW 2109, Australia

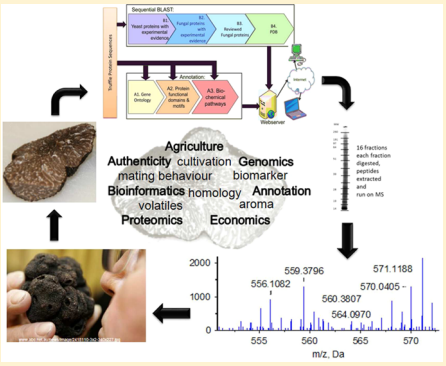
[§]QFAB, The University of Queensland, Queensland Bioscience Precinct, QLD 4072, Australia

[∥]Terra Preta Truffles, 389 Sawyers Ridge Road, Reidsdale, NSW 2622, Australia

[⊥]Department of Biochemistry, Yong Loo Lin School of Medicine, National University of Singapore, 117597 Singapore

Supporting Information

ABSTRACT: The black Périgord truffle (*Tuber melanosporum* Vittad.) is a highly prized food today, with its unique scent (i.e., perfume) and texture. Despite these attributes, it remains relatively poorly studied, lacking "omics" information to characterize its biology and biochemistry, especially changes associated with freshness and the proteins/metabolites responsible for its organoleptic properties. In this study, we have functionally annotated the truffle proteome from the 2010 *T. melanosporum* genome comprising 12 771 putative nonredundant proteins. Using sequential BLAST search strategies, we identified homologues for 2587 proteins with 2486 (96.0%) fungal homologues (available from http://biolinfo.org/protannotator/blacktruffle.php). A combined 1D PAGE and high-accuracy LC−MS/MS proteomic study was employed to validate the results of the functional annotation and identified 836 (6.5%) proteins, of which 47.5% (i.e., 397) were present in our bioinformatics studies. Our study, functionally annotating 6487 black Périgord truffle proteins and confirming 836 by proteomic experiments, is by far the most comprehensive study to date contributing significantly to the scientific community. This study has resulted in the functional characterization of novel proteins to increase our biological understanding of this organism and to uncover potential biomarkers of authenticity, freshness, and perfume maturation.



KEYWORDS: *black truffle, organoleptic, proteome, functional annotation, fungal proteomics*

INTRODUCTION

Truffles are fungi that produce subterranean fruiting bodies through the establishment of an ectomycorrhizal symbiotic relationship with the roots of host plants,[1] usually in a mutualistic fashion utilizing animals in their lifecycle to distribute spores. Among the different indigenous truffle species described, many have very pronounced organoleptic properties that are capable of attracting animals (including man) to the fruiting body. Collectively, these organoleptic properties also have accorded some truffle species their high economic importance.[2] The fruiting body of the black Périgord truffle (*Tuber melanosporum* Vittad) is one of the most prized delicacies in any gourmet food repertoire as evidenced by the exorbitant prices they fetch in world markets (≥$2,000 USD/kg).[3] This rare 'black diamond' of the kitchen has long intrigued distinguished chefs and biologists alike, due to its combination of smooth texture, pungent odor/perfume, and musty earthy flavor. In addition, its unique and often cryptic symbiotic relationship with oak and hazelnut trees has thwarted numerous efforts at routine cultivation.[4] In the past decade, the harvest of the black Périgord truffle has plummeted in Europe due to the effects of climate change, and loss of suitable arable land, encroachment of introduced species, and other factors.[5] This scarcity, coupled to increased awareness and demand for truffles has led to increasing prices and hence the inevitable replacement by similar black truffles (e.g., the Chinese black truffle *Tuber himalayensis*[6] or members of the *Tuber indicum* group[7]) into significant markets. For these reasons, a more focused study on the ecology, biology, and behavior of the black Périgord truffle has become increasingly necessary. To this end, the 2010 publication of the genome[4] of the black Périgord truffle has resulted in an understanding, on the genomic level, of its carbohydrate metabolism,[8] transcription,[9] mating behavior,[10] volatiles that produce aroma,[11] and other aspects of transcriptional and genomic control. Despite this,

Received: June 30, 2013
Published: October 22, 2013

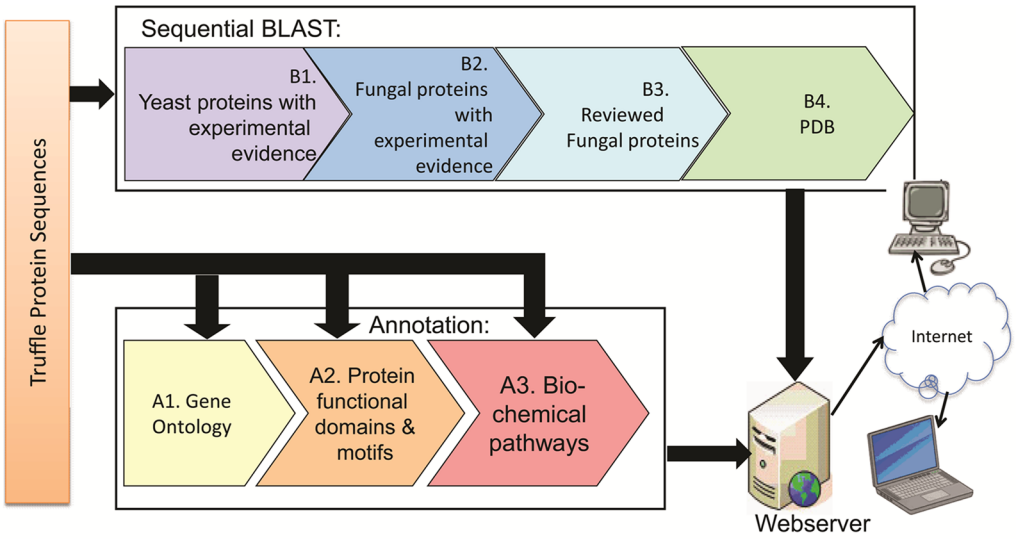


Figure 1. Summary of the pipeline used to annotate the truffle proteome. Proteins were passed through a series of databases to determine homology (sequential BLAST) as well as databases to confer annotations based on gene ontology, protein functional domains, and motifs as well as biochemical pathways.

only 14 proteins have "reviewed" annotations in UniProt[12] (v 2013_05): "reviewed" data sets contain proteins entries that are all manually annotated and reviewed in UniProtKB/Swiss-Prot database. Clearly, the full utility of the published genome in terms of the proteome remains to be realized.

To have a comprehensive understanding of the biology of this organism, it is imperative that a systems biology approach (bioinformatics and proteomics) is employed. Bioinformatics can provide compressive annotations for entire proteomes, providing valuable information regarding putative functions of proteins. Fungal proteomics over the past years has led to numerous advancements in biological understanding of the unique behaviors of filamentous fungi that have resulted in commercial gain,[13] control of pathogenic fungi,[14] and the discovery of biomarkers of freshness and authenticity[15] as well as ecology.[4] The differences between the unique biochemistries of different economically important truffle species (e.g., white truffle *Tuber magnatum* Pico, summer truffle *Tuber aestivum*, and winter truffle *Tuber brumale*) are yet to be determined comprehensively. A recent study has attempted to utilize proteomics to decipher differences in the growing conditions of the *Tuber* family with the successful identification of only 17 proteins from *T. magnatum* matched to the *T. melanosporum* gene database.[16] It is beyond doubt that the proteome of the fruiting body of the black Périgord truffle would yield a significant and useful data set for understanding this and related organisms with bioinformatics studies accelerating and guiding the complex proteomics studies.

In this study, we applied a "sequential BLAST" method previously adopted to functionally annotate the "missing" proteins of Human Chromosome 7.[17] In this approach, data similarity searching was carried out sequentially against carefully selected reference databases instead of repeated searches against the same database, as implemented in PSI-BLAST, a variant of the standard BLAST search engine,[18] to overcome the limitation of matches to predominantly "hypothetical" proteins (11 870), as reported in the Supporting Information (TuberGM_annot.xls) of the *T. melanosporum* genome.[4] This study is a combined bioinformatics and proteomics study to characterize the truffle proteome. Putative biological functions of the truffle proteome are assigned primarily by identifying homologues from well-annotated experimentally validated yeast and fungal proteins. We have also used bioinformatics analyses to ascribe functional annotations in terms of protein domains, gene ontology, and biochemical pathways. We have then attempted to validate the proteome using shotgun proteomic analysis of the fruiting body, which has provided a list of potential proteins involved in the production of the truffles' aroma profile.

MATERIALS AND METHODS

1. Bioinformatics Analysis

Data Sources. Black Périgord truffle (*T. melanosporum* Vittad.) protein sequences were extracted from the MycorWeb database [http://mycor.nancy.inra.fr/IMGC/TuberGenome/download.php?select=fast][4] in FASTA (a special file format generated by the 'FAST-All' software package) format.[4] Of the 12 826 coding DNA sequences (CDS, predicted Genoscope gene models) obtained, the truffle proteome comprised 12 771 nonredundant proteins after removing duplicate entries, although the *T. melanosporum* genome publication[4] reported only 7496 as "protein coding genes." Reviewed proteins with protein level experimental evidence are the best source of functional information, followed by reviewed proteins. Since yeast (*Saccharomyces cerevisiae*) is the most studied fungal species, our first choice for seeking truffle homologues was the set of reviewed yeast proteins with experimental evidence, followed by fungal proteins with experimental evidence and then reviewed fungal proteins. To carry out the functional characterization of the *T. melanosporum* proteins, we downloaded and set up local databases for BLAST[18] similarity searches. These databases included yeast proteins with experimental evidence (7503 sequences), fungal proteins with experimental evidence (9450 sequences), and reviewed fungal proteins (31 031 sequences). In addition, another search was conducted against all Protein Data Bank (PDB) proteins (236 604 sequences) to assign homologues from proteins with structure, as 3D structures are known to be evolutionarily conserved even under very low amino acid sequence

similarity.[17] Protein data sets for the sequential BLAST searches were extracted from the UniProt/SwissProt database (v 2013_05, release 01-05-2013).[12]

Database Similarity Searches. Database similarity searches for the truffle proteome were conducted using BLASTP.[18] A match was deemed a strong indicator of homology if the query sequence matched a database sequence with high significance (i.e., very low *E* value: $< 1 \times 10^{-5}$) and sequence identity of at least 50%. Sequential BLASTP runs were performed against the four data sets (including PDB) described above (using default parameters) for mapping a known protein sequence against a database of protein sequences. Sequences that did not have any match from the first run were passed to the next round of BLASTP to search the second data set, then the third and the fourth. As reviewed sequences with protein level experimental evidence are considered the most reliable source of homologues, these were used as the first database for BLASTP similarity searches. These proteins were also subjected to further *in silico* analyses, as described later.

Functional Annotation. Protein functional annotation in terms of protein domains, motifs, and signatures provides vital clues to biological function for experimental validation. InterProScan[19] comprises 14 programs for matching a query sequence against 13 protein domain and functional site databases and represents the most comprehensive protein functional annotation software currently available. All black Périgord truffle proteins were initially characterized through InterProScan[19] domain/motif analyses. InterProScan also provides gene ontology (GO) annotations. Pathway mapping for all of these proteins was carried out using KOBAS (KEGG Orthology-Based Annotation System, KOBAS-2.0).[20] All results from domain/motif analyses, GO annotation, and KEGG pathway mapping were used for preliminary functional annotation of these proteins.

The different bioinformatics analysis steps used for annotating the truffle proteome were integrated into a pipeline, illustrated in Figure 1.

Proteomics Studies. Proteomic studies were carried out to validate our bioinformatics approach. Freshly harvested Australian black Périgord truffles (*T. melanosporum* Vittad.) were kindly donated by Terra Preta Truffles (NSW, Australia). Truffles were stored on ice overnight for transport to the laboratory, and the best sample was selected as "representative" of the mature fruiting body. Approximately 50 mg of the inner tissue of the selected sample was freeze-crushed in liquid N_2, and the resulting powder was dissolved in 1 mL of 4× LDS buffer in the presence of both protease and phosphatase inhibitors. The sample was probe-sonicated (3 × 10 pulses, output 3) (Branson sonifier 450) until the solution was homogeneous and was centrifuged at 10 000*g* for 10 min to remove insoluble particulate matter, acetone-precipitated overnight at −80 °C, and centrifuged for 20 min at 10 000*g* . The pellet was then resuspended in 4× LDS buffer and protein-quantified using a BCA assay (Thermo-Pierce, Rockford, IL) according to the manufacturer's instructions.

1D Gel Electrophoresis and Slice-and-Dice Proteomics. Resuspended protein (100 μg) was run on precast 4−12% linear gradient SDS polyacrylamide gel (Invitrogen, USA) under reducing conditions as per the manufacturer's instructions. The gel was then fixed in 40% ethanol (v/v), 10% acetic acid (v/v) for 2 h and stained overnight with Flamingo Pink (BioRad, Hercules, USA) and imaged on the Typhoon Trio Variable Mode Laser Imager (GE Healthcare, Uppsala, Sweden) with photomultiplier tube (PMT) voltage set to 5 V below saturation of the most intense spot. The entire gel lane was divided into 16 fractions, digested using trypsin, and extracted using standard procedures described elsewhere.[21]

LC Coupled to Mass Spectrometry. The digested peptides (10 μL) were injected onto a peptide chromatography trap (Michrome peptide Captrap) on an Eksigent Ultranano LC system for preconcentration. Desalting is a standard procedure prior to LC as salt usually produces aberrant peaks on the spectrum, depending on the conductance values of the solution. Peptides were desalted using 0.1% formic acid, 2% ACN, at 5 μL/min for 10 min. The peptide trap was then switched online with an analytical column (SGE ProteCol C18, 300 Å, 3 μm, 150 μm × 10 cm). Peptides were eluted from the column using a linear solvent gradient, consisting of 0.1% formic acid as mobile phase A and 90% ACN/0.1% formic acid as mobile phase B, at 600 nL/min, starting from 2% B and going to 40% B over 140 min. After peptide elution, the column was cleaned with 80% buffer B for 19 min and then equilibrated with buffer A for 15 min before the next sample was injected. The reverse-phase nanoLC eluent was subject to positive ion nanoflow electrospray analyses in an information-dependent acquisition mode (IDA) on a Triple TOF 5600 (ABSciex, Toronto, Canada) at 15 kV acceleration voltage. MS data were collected using an ion spray voltage of 2.4 kV, curtain gas of 20 PSI, nebulizer gas of 15 PSI, and an interface heater temperature of 150 °C. In IDA mode, a TOFMS survey scan was acquired (m/z 350−1500, 0.25 s), with the 15 most intense multiply charged ions (counts >150) in the survey scan sequentially subjected to MS/MS analysis. MS/MS spectra were accumulated for 50 ms in the mass range m/z 100−1500 with the total cycle time 1.05 s, with a mass accuracy 1 ppm.

Database Searching of Proteomic Data. The experimental nanoLC−ESI−MS/MS data were submitted to Mascot after raw files were converted to .mgf format and searched against the *T. melanosporum* database (called here BT_Prot) which was derived from the 12 771 nonredundant sequences of the truffle proteome. The 16 fractions were processed individually with output files for each fraction, then merged, and a nonredundant output file was generated for protein identifications with $\log_e$ scores < 1. Search parameters included MS and MS/MS tolerance of ±100 ppm and ±0.2 Da, respectively. Carbamidomethyl was considered a fixed modification. In addition, variable modifications of methionine, threonine, and deamidation of asparagine and glutamine were also considered. Additional searching was performed against the decoy database in Mascot to evaluate false discovery rates (FDRs). Peptide FDR of a list is 2 × total numbers of peptides representing reversed protein hits in the list/total number of peptides representing all proteins in the list × 100. Protein FDR was calculated for each list of proteins using number of reversed protein hits in the list/total number of proteins in the list × 100.[22]

RESULTS AND DISCUSSION

Bioinformatics Analysis

The "sequential BLAST" approach we used involved repeated similarity searching against different select databases (Figure 1). Reviewed sequences (from UniProt) with protein level experimental evidence were used as the first database for BLASTP similarity searches. To identify the optimal sequence

identity for this study, we ran our workflow with very high sequence identity cutoff for BLAST, then reduced it by 5% on each run and compared the results against the 836 proteins that were identified by proteomics in a single preliminary shotgun approach.

At 50% sequence identity we were able to find homologues for 2486 proteins (19.5% of 12 771). Reducing the threshold to a sequence identity ≤50% yielded 7447 (58.31%) out of 12 771 proteins. The black Périgord truffle is a relatively under-studied organism with unique features and biochemistry. It was therefore expected that most of the black Périgord truffle that is homologous to other sequences will have a lower coverage (i.e., sequence identity). Although sequence identity ≤50% yielded many more protein matches, to retain high-quality results, we have only considered results with sequence identity ≥50% for this study.

Sequential-BLAST Similarity Search

In the first round of our sequential-BLAST approach, we assessed the 12 771 proteins against yeast protein sequences with experimental protein evidence. Of these 1794 (14.0%) black Périgord truffle proteins showed significant matches, with 3 hits having ≥99% sequence identity, 11 hits with 90−95% sequence identity, and 8 hits with 85−90% sequence identity. The coverage ranged from 50 to 99.2% with E values of 8.00×10^{-6} to 0. No significant matches with coverage >50% were reported for the remaining sequences (Supporting Information, Supplementary Table S1). The second BLAST search against fungal protein sequences with experimental protein evidence for the remaining 10 977 black Périgord truffle proteins showed matches for 109 sequences (0.85%) with 50−89.6% sequence coverage and E values of 2.00×10^{-6} to 0, with two hits having ≥85% sequence identity (Supporting Information, Supplementary Table S2). The third BLAST search against reviewed fungal protein sequences for the remaining 10 868 proteins yielded significant results for 583 sequences (4.6%). For these matches, the coverage ranged from 50 to 100%, with E values ranging from 9.00×10^{-6} to 0. Of these, seven had 100% sequence identity, one with 95.5% sequence identity and 10 with sequence identities between 85 and 90% (Supporting Information, Supplementary Table S3). The remainder of these proteins (10 285) were matched against solved protein structures from the PDB. 101 proteins showed matches with coverage ranging from 50 to 80% with E values of 9.00×10^{-20} to 0 (Supporting Information, Supplementary Table S4). Since structures of homologous proteins show functional conservation up to sequence identities as low as 25%,[23] the knowledge of homologous structural information for these truffle proteins provides important functional clues.

The results clearly indicate that the black truffle's unique biology in the context of its evolution shows only a distant relationship to any other commonly studied fungus (such as yeast). The proteins that matched with very high similarity (identity >70%) were proteins known to be evolutionarily conserved[24] in most eukaryotic organisms and accounted for a very small proportion of the total protein complement (552 proteins). A significant proportion of the unique biology of this organism thus lies in the 10 184 putative proteins that had very low similarity to any known proteins

Functional Annotation

InterProScan for the 12 771 *T. melanosporum* proteome provided annotations for 6487 proteins (50.8%) with 1369 unique GO annotations (Supporting Information, Supplementary Table S5), while 1309 genes were manually curated by the genome consortium but are not publicly available for comparison.[4] Our analysis on GO biological processes revealed that the majority of the proteins were involved in dUTP metabolic processes (674), oxidation−reduction process (482), metabolic process (276), translation (184), and protein phosphorylation (184). A similar analysis on GO molecular function revealed that ATP binding (763), hydrolase activity (414), and catalytic activity (335) were the most common annotations. Protein domain and family mapping provided InterPro domains for 946, family for 910, active sites for 46, conserved sites for 162, and repeats for 18 proteins compared with only protein family annotations provided for the genome.[4]

Analysis by KEGG pathways revealed a large proportion of proteins identified from metabolic pathways (961), while proteins involved in the production of secondary metabolites were also seen (239 proteins) (Supporting Information, Supplementary Table S6). The top KEGG pathways are listed in Table 1. These findings suggest that the truffle is very metabolically active, possessing a cohort of biochemical and enzymatic activity that may explain to some degree its ability to produce over 90 volatiles[25] that modulate its flavor profile as well as its complex lifecycle.

Table 1. Top 5 KEGG Pathways for Bioinformatics and Proteomics Analyses

proteins from bioinformatics analysis		proteins from MS/proteomics analysis	
description	total match	description	total match
metabolic pathways	961	metabolic pathways	158
pyrimidine metabolism	399	biosynthesis of secondary metabolites	75
biosynthesis of secondary metabolites	239	ribosome	50
cell cycle - yeast	161	glycolysis/gluconeogenesis	21
meiosis - yeast	96	protein processing in endoplasmic reticulum	19

Mass Spectrometric Evidence for *T. melanosporum* Proteins

Mass spectrometry analyses of the *T. melanosporum* proteome identified 836 proteins (Supporting Information, Supplementary Table S7) that were assigned to the BT_Prot database (Mascot version 2.3.0). Analyses of the 836 proteins showed that 91% were identified by peptides ranging between 1 and 100 per protein, with the other 9% between 101 and 1010 hits (results not shown). A protein was positively identified if it had a minimum of one unique peptide with at least 99% confidence.

Numerous mass spectra (65 260) were not assigned to any protein as the stringency of the cutoff scores for accepting a peptide match was set quite high (99%) and the peptide tolerance window for experimental results was set low (±0.2 Da); this meant that numerous spectra did not pass high stringency tests. This meant that the data used were of sufficient quality to justify accuracy of the results. The spectra obtained were matched against the predicted translations of open reading frames from the genome. Undoubtedly, there would be a degree of mismatch between experimental and predicted results. There is a possibility that the unmatched peptides may have come from cross-contamination from peptides of other organisms/species.[26] The truffle, being in a symbiotic relationship, may indeed be sharing a significant

amount of its structure with its host (hazelnut and oak trees). Additional investigations into potentially finding proteins in the unmatched peptide data were considered beyond the scope of this study, and future studies are planned to investigate this further.

The discovery of a total of 836 proteins (from a potential total of 12 771) by proteomics in a single preliminary shotgun approach in this study was higher than expected, as previous studies performed in unrelated fungal species identify a far lower proportion of nonredundant matches.[27−29] The proteomic coverage could still be increased using more sophisticated separation technologies such as PROOF fractionation,[30] which involves passing the sample through a tandem cation and anion exchange column, followed by peptide IPG-IEF prefractionation prior to MS analysis, which has been shown to significantly increase proteome coverage.[31] A more sophisticated information-independent SWATH-type analysis[32] could also be employed to ensure that any future findings can be quantitatively analyzed.

Comparison between Bioinformatics Analysis and Mass Spectrometry Evidence

Comparing the proteomics findings with our bioinformatics studies, in the first round of our BLASTP similarity search of these 836 proteins against reviewed yeast proteins with experimental evidence, 315 proteins (37.7%) were found to have significant matches (with 1 hit having ≥99% identity and 7 hits with ≥90% identity). In the second round of our BLASTP similarity search (against reviewed fungal proteins with experimental evidence), 18 proteins (2.15%) were found to have matches, with hits ≥85% identity. The third round of BLASTP similarity search (against reviewed fungal proteins) resulted in 55 proteins (6.58%; including one match with 100% identity and two matches with ≥90% identity). Nine proteins matched 3D structures in the PDB. Yeast is one of the most widely studied organisms/fungi over the years, and not surprisingly >50% of all matches found were to yeast proteins with experimental evidence. Of the *T. melanosporum* proteins identified by our proteomics study, 439 (52.5%) had no matches in our bioinformatics results with sequence identity >50%. The results have been summarized in Table 2, with details of the BLASTP matches for these 836 proteins provided as Supporting Information (Table S8).

Furthermore, we compared our findings with the 14 UniProt proteins identified in *T. melanosporum* (strain Mel28) that have been previously reviewed but not verified experimentally (Supporting Information, Supplementary Table S9). Out of these 14 reviewed proteins, three were identified by mass spectrometry with protein coverage ranging from 16 to 40% (Supporting Information, Supplementary Table S10). These proteins include a probable amino/metallo-peptidase (Gene ID: AMPP1; UniProt ID: D5GAC6), a nuclear- and cytoplasm-residing dioxygenase involved in L-methionine salvage (Gene ID: MTND; UniProt ID: D5GE59), and an integral membrane catalytic subunit of a signal peptidase complex found in the ER membrane (Gene ID: SEC11; UniProt ID: D5GNC3) that was deduced to be involved in proteolysis and signal peptide processing. In addition, of the 17 proteins reported in the recent proteomics study of *T. magnatum*,[16] we have identified *T. melanosporum* homologues to 12 proteins, of which three were found only by bioinformatics analysis and one uniquely in our proteomics data.

Overall, at least 105 of the 836 proteins could not be assigned any putative function. These proteins could be used as putative candidate markers of truffles. Alternatively, a proportion of these might be species-specific biomarkers for the black Périgord truffle (*T. melanosporum*). Further proteomics studies using single or multiple reaction monitoring experiments (SRM/MRM)[33] can quantitatively be used to study this cohort of proteins to discern markers of authenticity, a study beyond the scope of this report.

Functional annotation for the 836 proteins (Supporting Information, Supplementary Table S11) provided GO annotations for 698, InterPro domains for 768, and enzyme codes (ECs) for 225. The ECs were related to 90 corresponding KEGG biochemical pathways. Our analysis of the GO terms for the 836 proteins revealed that the majority were involved in binding (436), catalytic activity (350), localization within cells (288), and cell organelles (151) or with taking part in metabolic biological processes (416) or cellular (232) processes (Figure 2). Analysis by KEGG pathways revealed a large proportion of proteins identified from metabolic pathways (158), while proteins involved in the production of secondary metabolites (75 proteins) were also noted (Table 1). In all, 731 (87.4%) of the 836 proteins were successfully annotated with either GO, InterPro domains, or KEGG pathways. However, with the currently available biological knowledge, 105 proteins could not be annotated at all. These findings reflect those obtained from the bioinformatics studies and suggest that the proteomic assessment was representative of the currently annotated data.

Table 2. Comparative Summary of Bioinformatics Analysis and Mass Spectrometry Evidences

description	bioinformatics analysis	proteins from mass spectra (validated by bioinformatics)
total number of proteins	12 771	836
reviewed yeast protein sequences with experimental protein evidence	1794	315
reviewed fungal protein sequences with experimental protein evidence	109	18
reviewed fungal protein sequences	583	55
protein data bank (PDB)	101	9
functional annotation (GO, InterPro, KEGG)	2486	731

Proteins from Truffle That Confer Aroma

The enzymes that catalyze the production of volatiles that confer the unique aroma of the black truffle were analyzed from the pathway relating to the production of secondary metabolites (including the sulfur and methane metabolism pathways) in the KEGG database. A comprehensive analysis of the functionally annotated proteins found by proteomics and bioinformatics revealed that the proportion of proteins involved in the production of secondary metabolites in the fruiting body of the black truffle (from proteomics) (9.6%) was similar to that found *in silico* (9%). This list of proteins involved in secondary metabolism was matched against enzymes known to be involved in pathways that produce volatiles found in truffles from previous biochemical studies.[25,34−37] A total of nine proteins were identified (Table 3).

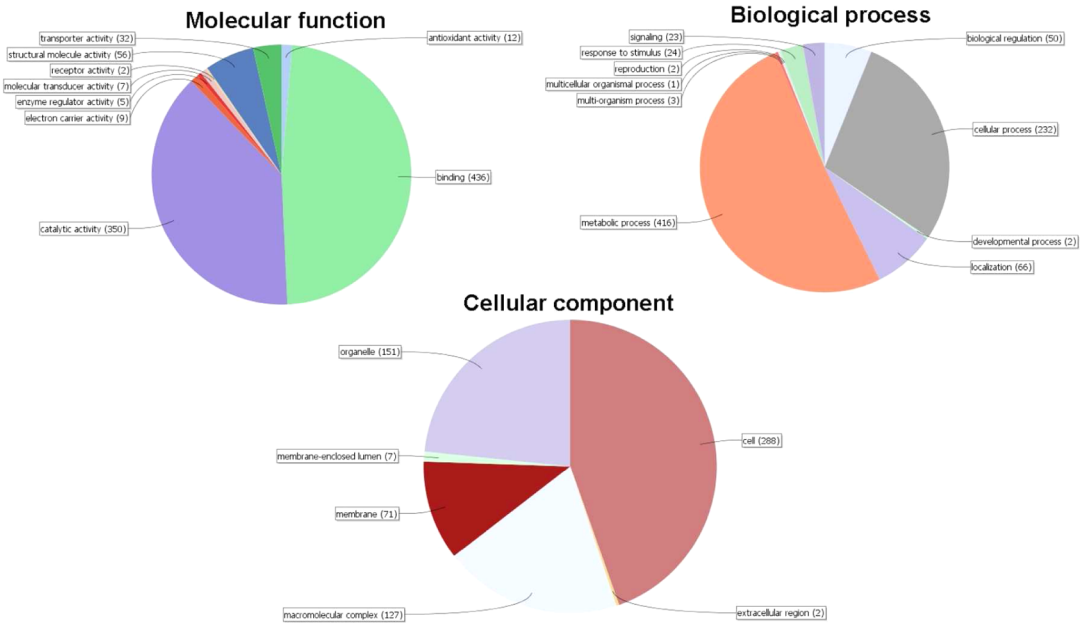


Figure 2. Pictorial representation of GO distributions for the 836 *T. melanosporum* proteins. Pie charts depicting the distributions are shown with respective molecular functions, cellular components, and biological processes labeled and number of proteins involved shown in parentheses.

Table 3. List of Previously[23,32−35] Identified Volatiles and the Enzymes Involved in Their Synthesis Matched to Proteins Obtained from the *T. melanosporum* Proteins Annotated with Bioinformatics and Substantiated with Proteomic (MS) Evidence (indicated with *)

volatile	KEGG compound code	enzymes	truffle protein	annotation
acetaldehyde	C00084	ribose-5-phosphate transaldolase, or fluorothreonine transaldolase	GSTUMT00000035001*	transaldolase
acetaldehyde	C00084	aldehyde dehydrogenase	GSTUMT00003865001*	aldehyde/histidinol dehydrogenase
acetaldehyde	C00084	alcohol dehydrogenase	GSTUMT00006862001*,[a]	alcohol dehydrogenase
anisole and methoxybenzene	C01403	phenol *O*-methyltransferase	GSTUMT00004489001*	sterol methyltransferase
2-methyl butanal	C02223	branched-chain-2-oxoacid decarboxylase	GSTUMT00001753001*, GSTUMT00000274001, GSTUMT00003265001	2-oxoacid dehydrogenase acyltransferase
propanal	C00479	propanediol dehydratase	GSTUMT00010247001	dihydroxy-acid/6-phosphogluconate dehydratase
phenylacetaldehyde	C00601	amine oxidase (pyridoxal containing)	GSTUMT00008176001	pyridoxal phosphate-dependent transferase

[a]Proteins identified as associated with volatiles in the genome publication.[4]

The organoleptic properties, particularly the unique aroma, of the truffle are arguably its most valuable asset, not only biologically in attracting animals for sporulation[38] but economically for gastronomists and food lovers. It has been previously shown from the analysis of the genome that the truffle possesses much of the machinery required for synthesizing its aroma. In this preliminary study, for the first time, we were able to potentially identify nine proteins responsible for part of the aroma profile of truffles, although more biochemical analyses need to be carried out to confirm the findings. Of these, only one (GSTUMT00006862001) has been identified as involved in secondary metabolism by the 2010 genome publication.[4]

Two compounds, DMS and 2-methylbutanal, when mixed in the right proportions, mimic the aroma of the black Périgord truffle, *T. melanosporum*.[39] The latter mixture has been used for standardly by the food industry to imitate black truffle aroma. The interesting discovery from bioinformatics with proteomics evidence of the enzyme potentially responsible for the metabolism of 2-methyl butanal validates the approaches taken.

A large proportion of the proteins annotated form secondary metabolite pathways that have proteomic as well as bioinformatics evidence did not match to known pathways of volatile synthesis. This, compounded by the fact that the enzymes involved in the production of some volatiles remain to be discovered, suggests that a more comprehensive biochemical study of the enzyme components of the *T. melanosporum* is warranted. It is hardly surprising that such low numbers of enzymes were shown to be involved in the production of over 90 volatiles considering over 70% of the proteome is yet to be annotated. The potential to discover novel enzymes that could be of economic, medicinal, or other uses remains a tantalizing possibility.

■ CONCLUSIONS

Only 14 black truffle proteins have been reviewed in UniProt. The remainder, although recently annotated, are not yet reviewed and await curation, while 1309 genes were manually curated by the truffle genome consortium.[4] We have provided high-quality bioinformatics annotations for 2587 sequences and

proteomic evidence of 836 truffle proteins. Using selected high-quality protein databases for similarity searches using BLAST sequentially, we identified homologues with experimental evidence for 14.9% of the black Périgord truffle proteome, with a further 4.6% mapping to reviewed fungal proteins and another 0.8% mapping to protein structures, totalling 2587 proteins (20.2% of the *T. melanosporum* 12 771 proteins). The acquisition of functional experimental evidence of these proteins is quite possible, as most matches were to a well-characterized fungus, *S. cerevisiae*. Additionally, using a suite of bioinformatics tools, we have assigned putative biological functions in terms of gene ontology, biochemical pathway, and domain/motif signatures for 2486 of these 2587 sequences. Using a proteomics approach, we have provided proteomic evidence of 836 proteins, none of which have been reported experimentally to date, including three of the 14 UniProt reviewed proteins that lacked proteomics evidence. Using a combination of computational strategies, we were able to identify nine proteins responsible for part of the aroma profile of truffles and for the first time suggest a potential enzymatic pathway for the production of one of the primary volatiles in black truffle. Approximately 20% of the 12 771 proteins have been assigned putative biological functionality, providing valuable clues for experimental validation and future work. We have described a generic framework that is validated by our proteomics studies and can be used to annotate the proteome of any novel organism.

■ ASSOCIATED CONTENT

Supporting Information

Table S1: BLAST hits showing high % identity against reviewed yeast proteins with experimental evidence. Hits sorted by % identity and only the hits with ≥50% sequence identities are shown. **Table S2**: BLAST hits showing high % identity against reviewed fungal proteins with experimental evidence. Hits sorted by % identity and only the hits with ≥50% sequence identities are shown. **Table S3**: BLAST hits showing high % identity against reviewed fungal proteins. Hits sorted by % identity and only the hits with ≥50% sequence identities are shown. **Table S4**: BLAST hits showing high % identity against PDB proteins. Hits sorted by % identity and only the hits with ≥50% sequence identities are shown. **Table S5**: InterPro and GO annotations of *T. melanosporum* proteins. **Table S6**: KEGG Pathways for *T. melanosporum* proteins. **Table S7**: List of proteins identified by mass spectrometry. **Table S8**: BLAST hits of proteins found by mass spectrometry. Hits sorted by % identity and only the hits with ≥50% sequence identities are shown. **Table S9**: Summary of UniProt reviewed proteins for *T. melanosporum* (strain Mel28). **Table S10**: Summary of proteins found by mass spectrometry and reviewed in UniProt (2013_05) . **Table S11**: Functional annotations of the 836 *T. melanosporum* proteins. Hits sorted by #GOs, GO, InterProScan, and KEGG Enzyme Codes. This material is available free of charge via the Internet at http://pubs.acs.org.

■ AUTHOR INFORMATION

Corresponding Authors

*Tel: +61 2 9850 8211. Fax: +61 2 9850 8313. E-mail: shoba.ranganathan@mq.edu.au.
*E-mail: mark.baker@mq.edu.au.

Author Contributions

#M.T.I., A.M., and G.G. contributed equally.

Notes

The authors declare no competing financial interest.

■ ACKNOWLEDGMENTS

We sincerely acknowledge the contributions of Dr. Alison Kan, who assisted in the proteomics analysis reported in this paper. Proteomics analysis was carried out at the Australian Proteome Analysis Facility, a Major National Research Facility of Australia. Financial support was provided by a Macquarie University Research Excellence Scholarship (MQRES) to MTI in the Department of Chemistry and Biomolecular Sciences, Macquarie University, Sydney, Australia.

■ REFERENCES

(1) Trappe, J. M.; Claridge, A. W. The hidden life of truffles. *Sci. Am.* **2010**, *302* (4), 78−82, 84.
(2) Maser, C.; Claridge, A. W.; Trappe, J. M. *Trees, Truffles, and Beasts: How Forests Function*; Rutgers University Press: New Brunswick, NJ, 2008; p xvi, 280 p, 8 p of plates.
(3) Bohannon, J. Genetics. Rooting around the truffle genome. *Science* **2009**, *323* (5917), 1006−7.
(4) Martin, F.; Kohler, A.; Murat, C.; Balestrini, R.; Coutinho, P. M.; Jaillon, O.; Montanini, B.; Morin, E.; Noel, B.; Percudani, R.; Porcel, B.; Rubini, A.; Amicucci, A.; Amselem, J.; Anthouard, V.; Arcioni, S.; Artiguenave, F.; Aury, J. M.; Ballario, P.; Bolchi, A.; Brenna, A.; Brun, A.; Buee, M.; Cantarel, B.; Chevalier, G.; Couloux, A.; Da Silva, C.; Denoeud, F.; Duplessis, S.; Ghignone, S.; Hilselberger, B.; Iotti, M.; Marcais, B.; Mello, A.; Miranda, M.; Pacioni, G.; Quesneville, H.; Riccioni, C.; Ruotolo, R.; Splivallo, R.; Stocchi, V.; Tisserant, E.; Viscomi, A. R.; Zambonelli, A.; Zampieri, E.; Henrissat, B.; Lebrun, M. H.; Paolocci, F.; Bonfante, P.; Ottonello, S.; Wincker, P. Perigord black truffle genome uncovers evolutionary origins and mechanisms of symbiosis. *Nature* **2010**, *464* (7291), 1033−8.
(5) Martin, F. Unearthing the truffle genome. *New Phytol.* **2011**, *189* (3), 645−6.
(6) Paolocci, F.; Rubini, A.; Granetti, B.; Arcioni, S. Typing *Tuber melanosporum* and Chinese black truffle species by molecular markers. *FEMS Microbiol. Lett.* **1997**, *153* (2), 255−60.
(7) Murat, C.; Martin, F. Sex and truffles: first evidence of Perigord black truffle outcrosses. *New Phytol.* **2008**, *180* (2), 260−3.
(8) Ceccaroli, P.; Buffalini, M.; Saltarelli, R.; Barbieri, E.; Polidori, E.; Ottonello, S.; Kohler, A.; Tisserant, E.; Martin, F.; Stocchi, V. Genomic profiling of carbohydrate metabolism in the ectomycorrhizal fungus *Tuber melanosporum*. *New Phytol.* **2011**, *189* (3), 751−64.
(9) Montanini, B.; Levati, E.; Bolchi, A.; Kohler, A.; Morin, E.; Tisserant, E.; Martin, F.; Ottonello, S. Genome-wide search and functional identification of transcription factors in the mycorrhizal fungus *Tuber melanosporum*. *New Phytol.* **2011**, *189* (3), 736−50.
(10) Rubini, A.; Belfiori, B.; Riccioni, C.; Arcioni, S.; Martin, F.; Paolocci, F. *Tuber melanosporum*: mating type distribution in a natural plantation and dynamics of strains of different mating types on the roots of nursery-inoculated host plants. *New Phytol.* **2011**, *189* (3), 723−35.
(11) Li, Y. Y.; Wang, G.; Li, H. M.; Zhong, J. J.; Tang, Y. J. Volatile organic compounds from a *Tuber melanosporum* fermentation system. *Food Chem.* **2012**, *135* (4), 2628−37.
(12) Reorganizing the protein space at the Universal Protein Resource (UniProt). *Nucleic Acids Res.* **2012**, *40*, (Database issue), D71-5.
(13) Doyle, S. Fungal proteomics: from identification to function. *FEMS Microbiol. Lett.* **2011**, *321* (1), 1−9.
(14) Kniemeyer, O.; Schmidt, A. D.; Vodisch, M.; Wartenberg, D.; Brakhage, A. A. Identification of virulence determinants of the human

pathogenic fungi *Aspergillus fumigatus* and *Candida albicans* by proteomics. *Int. J. Med. Microbiol.* **2011**, *301* (5), 368−77.
(15) Vincent, D.; Balesdent, M. H.; Gibon, J.; Claverol, S.; Lapaillerie, D.; Lomenech, A. M.; Blaise, F.; Rouxel, T.; Martin, F.; Bonneu, M.; Amselem, J.; Dominguez, V.; Howlett, B. J.; Wincker, P.; Joets, J.; Lebrun, M. H.; Plomion, C. Hunting down fungal secretomes using liquid-phase IEF prior to high resolution 2-DE. *Electrophoresis* **2009**, *30* (23), 4118−36.
(16) Vita, F.; Lucarotti, V.; Alpi, E.; Balestrini, R.; Mello, A.; Bachi, A.; Alessio, M.; Alpi, A. Proteins from *Tuber magnatum Pico* fruiting bodies naturally grown in different areas of Italy. *Proteome Sci.* **2013**, *11* (1), 7.
(17) Ranganathan, S.; Khan, J. M.; Garg, G.; Baker, M. S. Functional annotation of the human chromosome 7 ″missing″ proteins: a bioinformatics approach. *J. Proteome Res.* **2013**, *12* (6), 2504−10.
(18) Altschul, S. F.; Madden, T. L.; Schaffer, A. A.; Zhang, J.; Zhang, Z.; Miller, W.; Lipman, D. J. Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. *Nucleic Acids Res.* **1997**, *25* (17), 3389−402.
(19) Quevillon, E.; Silventoinen, V.; Pillai, S.; Harte, N.; Mulder, N.; Apweiler, R.; Lopez, R. InterProScan: protein domains identifier. *Nucleic Acids Res.* **2005**, *33*, W116−20.
(20) Xie, C.; Mao, X.; Huang, J.; Ding, Y.; Wu, J.; Dong, S.; Kong, L.; Gao, G.; Li, C. Y.; Wei, L. KOBAS 2.0: a web server for annotation and identification of enriched pathways and diseases. *Nucleic Acids Res.* **2011**, *39*, W316−22.
(21) Rosenfeld, J.; Capdevielle, J.; Guillemot, J. C.; Ferrara, P. In-gel digestion of proteins for internal sequence analysis after one- or two-dimensional gel electrophoresis. *Anal. Biochem.* **1992**, *203* (1), 173−9.
(22) Mirzaei, M.; Pascovici, D.; Atwell, B. J.; Haynes, P. A. Differential regulation of aquaporins, small GTPases and V-ATPases proteins in rice leaves subjected to drought stress and recovery. *Proteomics* **2012**, *12* (6), 864−877.
(23) Martin, A. C.; MacArthur, M. W.; Thornton, J. M. Assessment of comparative modeling in CASP2. *Proteins* **1997**, No. Suppl 1, 14−28.
(24) Basu, M. K.; aaCarmel, L.; Rogozin, I. B.; Koonin, E. V. Evolution of protein domain promiscuity in eukaryotes. *Genome Res.* **2008**, *18* (3), 449−61.
(25) Splivallo, R.; Ottonello, S.; Mello, A.; Karlovsky, P. Truffle volatiles: from chemical ecology to aroma biosynthesis. *New Phytol.* **2011**, *189* (3), 688−99.
(26) Ding, Q.; Xiao, L.; Xiong, S.; Jia, Y.; Que, H.; Guo, Y.; Liu, S. Unmatched masses in peptide mass fingerprints caused by cross-contamination: An updated statistical result. *Proteomics* **2003**, *3*, 1313−1317.
(27) Wang, M.; Gu, B.; Huang, J.; Jiang, S.; Chen, Y.; Yin, Y.; Pan, Y.; Yu, G.; Li, Y.; Wong, B. H.; Liang, Y.; Sun, H. Transcriptome and proteome exploration to provide a resource for the study of *Agrocybe aegerita*. *PLoS One* **2013**, *8* (2), e56686.
(28) Horie, K.; Rakwal, R.; Hirano, M.; Shibato, J.; Nam, H. W.; Kim, Y. S.; Kouzuma, Y.; Agrawal, G. K.; Masuo, Y.; Yonekura, M. Proteomics of two cultivated mushrooms *Sparassis crispa* and *Hericium erinaceum* provides insight into their numerous functional protein components and diversity. *J. Proteome Res.* **2008**, *7* (5), 1819−35.
(29) Andersson, K. M.; Meerupati, T.; Levander, F.; Friman, E.; Ahren, D.; Tunlid, A. Characterization of the proteome of the nematode-trapping cells of the fungus *Monacrosporium haptotylum*. *Appl. Environ. Microbiol.* **2013**, *79*, 4993−5004.
(30) Tan, S. H.; Mohamedali, A.; Kapur, A.; Baker, M. S. Ultradepletion of Human Plasma using Chicken Antibodies: A Proof of Concept Study. *J. Proteome Res.* **2013**, *12* (6), 2399−413.
(31) McQuade, L. R.; Schmidt, U.; Pascovici, D.; Stojanov, T.; Baker, M. S. Improved membrane proteomics coverage of human embryonic stem cells by peptide IPG-IEF. *J. Proteome Res.* **2009**, *8* (12), 5642−9.
(32) Gillet, L. C.; Navarro, P.; Tate, S.; Rost, H.; Selevsek, N.; Reiter, L.; Bonner, R.; Aebersold, R. Targeted data extraction of the MS/MS spectra generated by data-independent acquisition: a new concept for consistent and accurate proteome analysis. *Mol. Cell. Proteomics* **2012**, *11* (6), O111 016717.
(33) Boja, E. S.; Rodriguez, H. Mass spectrometry-based targeted quantitative proteomics: achieving sensitive and reproducible detection of proteins. *Proteomics* **2012**, *12* (8), 1093−110.
(34) Cullere, L.; Ferreira, V.; Venturini, M. E.; Marco, P.; Blanco, D. Potential aromatic compounds as markers to differentiate between *Tuber melanosporum* and *Tuber indicum* truffles. *Food Chem.* **2013**, *141* (1), 105−10.
(35) Davoli, P.; Bellesia, F.; Pinetti, A. Comments on Truffle aroma analysis by headspace solid phase microextraction [Is butylated hydroxytoluene (BHT) a ″natural″ volatile constituent of truffles?]. *J. Agric. Food Chem.* **2003**, *51* (15), 4483 author reply 4484..
(36) Diaz, P.; Ibanez, E.; Senorans, F. J.; Reglero, G. Truffle aroma characterization by headspace solid-phase microextraction. *J. Chromatogr., A* **2003**, *1017* (1−2), 207−14.
(37) Diaz, P.; Senorans, F. J.; Reglero, G.; Ibanez, E. Truffle aroma analysis by headspace solid phase microextraction. *J. Agric. Food Chem.* **2002**, *50* (22), 6468−72.
(38) Linde, C. C.; Selmes, H. Genetic Diversity and Mating Type Distribution of *Tuber melanosporum* and Their Significance to Truffle Cultivation in Artificially Planted Truffieres in Australia. *Appl. Environ. Microbiol.* **2012**, *78* (18), 6534−9.
(39) Cullere, L.; Ferreira, V.; Chevret, B.; Venturini, M. E.; Sanchez-Gimeno, A. C.; Blanco, D. Characterisation of aroma active compounds in black truffles (*Tuber melanosporum*) and summer truffles (*Tuber aestivum*) by gas chromatography-olfactometry. *Food Chem.* **2010**, *122* (1), 300−306.