

## Supplementary Information

### **Comparative analysis of the biosynthetic systems for fungal bicyclo[2.2.2]diazaoctane indole alkaloids: the (+)/(-)-notoamide, paraherquamide and malbrancheamide pathways**

Shengying Li,<sup>a</sup> Krithika Anand,<sup>a</sup> Hong Tran,<sup>a</sup> Fengang Yu,<sup>a</sup> Jennifer M. Finefield,<sup>b</sup> James D. Sunderhaus,<sup>b</sup> Timothy J. McAfoos,<sup>b</sup> Sachiko Tsukamoto,<sup>c</sup> Robert M. Williams,<sup>\*b,d</sup> and David H. Sherman<sup>\*a,e</sup>

<sup>a</sup>Life Sciences Institute and <sup>a</sup>Departments of Medicinal Chemistry, Microbiology & Immunology, and Chemistry, University of Michigan, Ann Arbor, Michigan 48109, USA. *Email: davidhs@umich.edu*

<sup>b</sup>Department of Chemistry, Colorado State University, Fort Collins, Colorado 80523 and <sup>d</sup>University of Colorado Cancer Center, Aurora, Colorado 80045, USA. *Email: rmw@lamar.colostate.edu*

<sup>c</sup>Graduate School of Pharmaceutical Sciences, Kumamoto University, 5-1 Oe-honmachi, Kumamoto 862-0973, Japan.

## Methods

### Preparation of genomic DNAs for Solexa sequencing

The filamentous fungal strains *Aspergillus versicolor* NRRL35600, *Penicillium fellutanum* ATCC20841, and *Malbranchea aurantiaca* RRC1813 were statically cultivated in 50 mL liquid medium (2.0% malt extract and 0.5% peptone) at 26 °C for 7 days. The fungal mats were recovered from flasks and dried with filter paper. The genomic DNAs were isolated using MasterPure Yeast DNA Purification Kit (Epicentre) by following the protocol developed for filamentous fungal mycelium. The purified genomic DNAs were further cleaned by additional RNase treatment, and sequential washing steps with phenol/chloroform and chloroform to afford high quality DNA samples for Solexa sequencing.

### Solexa genome sequencing and assembly

Three fungal genomes were 108 Cycle Paired-End sequenced using the Illumina Genome Analyzer Ix at University of Michigan DNA Sequencing Core. *De novo* assembly of genomes was performed using Velvet version 0.7.63. The summary of dataset and parameters of genome sequencing and assembly is shown in Table S1. Notably, a number of unsequenced regions in the assembled contigs containing the target biosynthetic gene clusters were complemented by sequencing the PCR products covering these unsequenced (gap) regions.

**Table S1.** Summary of genome assembly dataset and parameters

| Genome                        | Number of reads | k-mer | Coverage cutoff | N50     | Number of contigs |
|-------------------------------|-----------------|-------|-----------------|---------|-------------------|
| <i>Aspergillus versicolor</i> | 34,697,846      | 55    | 10              | 329,145 | 1243              |
| <i>Penicillium fellutanum</i> | 29,345,178      | 55    | 12              | 309,107 | 1186              |
| <i>Malbranchea aurantiaca</i> | 63,264,437      | 57    | 8               | 405,232 | 1576              |

**Table S2.** Features of the *Not*' gene products

| Protein | Size (AA) | Exons                                                                                                  | Function                  | Relative (identity/similarity [%])                                                                            | Accession number |
|---------|-----------|--------------------------------------------------------------------------------------------------------|---------------------------|---------------------------------------------------------------------------------------------------------------|------------------|
| NotA'   | 334       | 1-263, 321-659, 714-801, 926-1183                                                                      | Negative regulator        | NmrA family protein from <i>Ajellomyces capsulatus</i> G186AR (40/56)                                         | EEH03447         |
| NotB'   | 455       | 1772-1940, 1999-2591, 2649-2995, 3076-3334                                                             | FAD monooxygenase         | FAD binding domain protein from <i>Aspergillus clavatus</i> NRRL 1(47/67)                                     | XP_001268514.1   |
| NotC'   | 426       | 4233-5410, 5481-5583                                                                                   | Prenyltransferase         | FtmH from <i>Aspergillus fumigatus</i> A1163 (50/67)                                                          | BAH24002         |
| NotD'   | 612       | 6448-6730, 6784-7304, 7370-8404                                                                        | Oxidoreductase            | FAD/FMN-containing isoamyl alcohol oxidase MreA-like protein from <i>Aspergillus niger</i> CBS 513.88 (45/64) | XP_001400920.2   |
| NotE'   | 2225      | 9103-15780                                                                                             | NRPS                      | FtmA from <i>Aspergillus fumigatus</i> A1163 (48/65)                                                          | BAH23995         |
| NotF'   | 435       | 16002-16131, 16200-17377                                                                               | Prenyltransferase         | FtmPT1 from <i>Trichophyton verrucosum</i> HKI 0517 (39/58)                                                   | EFE40047         |
| NotG'   | 544       | 17901-18088, 18158-18259, 18325-18623, 18678-18799, 18863-19786                                        | P450 monooxygenase        | FtmC from <i>Aspergillus fumigatus</i> Af293 (62/75)                                                          | BAH23996         |
| NotH'   | 499       | 20439-20690, 20750-20827, 20895-20994, 21054-21122, 21180-21758, 21821-22242                           | P450 monooxygenase        | FtmG from <i>Aspergillus fumigatus</i> Af293 (47/64)                                                          | BAH24001.1       |
| NotI'   | 433       | 22791-22950, 23007-23925, 23979-24201                                                                  | FAD monooxygenase         | FAD binding domain protein from <i>Aspergillus clavatus</i> NRRL 1(43/61)                                     | EAW07088         |
| NotJ'   | 362       | 24352-25440                                                                                            | Unknown                   | Hypothetical protein from <i>Micromonospora echinospora</i> (56/67)                                           | CAI05922         |
| NotK'   | 577       | 27317-27406, 27514-27617, 27694-28612, 28668-29288                                                     | Nucleoside transporter    | Purine transporter from <i>Aspergillus nidulans</i> FGSC A4 (89/93)                                           | XP_681803        |
| NotL'   | 620       | 30047-30172, 30244-30834, 30900-31036, 31084-31257, 31310-31651, 31703-32195                           | Transcription factor      | Putative Zn(II)2Cys6 transcription factor from <i>Aspergillus nidulans</i> FGSC A4 (61/70)                    | CBF80748         |
| NotM'   | 454       | 32386-32697, 32752-32868, 32916-33236, 33330-33813, 33867-33892, 33964-34013, 34069-34123              | Unknown                   | Hypothetical protein from <i>Aspergillus nidulans</i> FGSC A4 (51/64)                                         | CBF80750         |
| NotN'   | 416       | 35562-36171, 36491-36624, 36681-37187                                                                  | Unknown                   | Hypothetical protein from <i>Aspergillus nidulans</i> FGSC A4 (61/76)                                         | EAA62570         |
| NotO'   | 462       | 37958-39346                                                                                            | Unknown                   | DUF895 domain membrane protein from <i>Aspergillus flavus</i> NRRL3357 (78/90)                                | XP_002375620     |
| NotP'   | 292       | 39545-39665, 39726-39836, 39898-40223, 40273-40593                                                     | Short chain dehydrogenase | Short-chain dehydrogenase from <i>Neosartorya fischeri</i> NRRL 181 (79/87)                                   | XP_001258601.1   |
| NotQ'   | 506       | 40679-40821, 40996-41134, 41190-41256, 41343-41858, 41915-42123, 42175-42437, 42490-42644, 42701-42729 | Transcription factor      | Putative C6 transcription factor, from <i>Aspergillus flavus</i> NRRL3357 (24/39)                             | XP_002372282     |
| NotR'   | 172       | 43297-43815                                                                                            | Unknown                   | Hypothetical protein NFIA_060620 from <i>Neosartorya fischeri</i> NRRL 181 (80/88)                            | XP_001258603     |

**Table S3.** Features of the *Phq* gene products

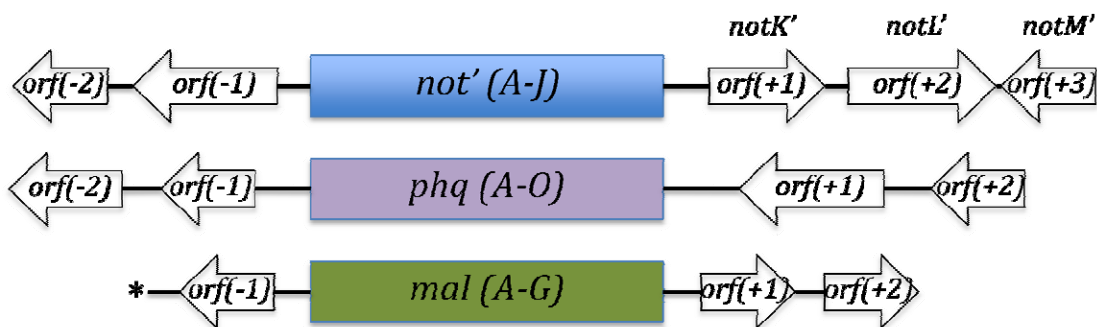
| Protein | Size (AA) | Exons                                                                                                                            | Function                         | Relative (identity/similarity [%])                                                                            | Accession number |
|---------|-----------|----------------------------------------------------------------------------------------------------------------------------------|----------------------------------|---------------------------------------------------------------------------------------------------------------|------------------|
| PhqA    | 405       | 1-1118, 1182-1281                                                                                                                | Prenyltransferase                | <i>Cyclo-L-Trp-L-Trp</i> prenyltransferase from <i>Aspergillus oryzae</i> (31/53)                             | ADI60056.1       |
| PhqB    | 2449      | 1606-8955                                                                                                                        | NRPS                             | Putative nonribosomal peptide synthase, putative from <i>Neosartorya fischeri</i> NRRL 181 (32/48)            | XP_001262287.1   |
| PhqC    | 353       | 9530-9739, 9796-9975, 10025-10696                                                                                                | 2OG-Fe(II)-oxygenase             | gibberellin 2-oxidase from <i>Aspergillus oryzae</i> RIB40 (43/59)                                            | XP_001818578.2   |
| PhqD    | 322       | 10913-11829, 11925-11976                                                                                                         | Pyroline-5-carboxylate reductase | Pyroline-5-carboxylate reductase from <i>Talaromyces stipitatus</i> ATCC 10500 (34/49)                        | XP_002485467.1   |
| PhqE    | 265       | 12804-13601                                                                                                                      | Short chain dehydrogenase        | Short chain dehydrogenase from <i>Pyrenophora triticirepentis</i> Pt-1C-BFP (45/66)                           | XP_001930671.1   |
| PhqF    | 411       | 14289-14438, 14486-14556, 14748-14831, 14927-15359, 15406-15493, 15656-16065                                                     | Efflux pump                      | TqAJ from <i>Penicillium aethiopicum</i> (40/59)                                                              | ADY16699.1       |
| PhqG    | 338       | 17781-17826, 17882-18099, 18174-18327, 18381-18713, 18799-19064                                                                  | Negative regulator               | NmrA family protein from <i>Trichophyton equinum</i> CBS 127.97 (42/63)                                       | EGE06939.1       |
| PhqH    | 602       | 21682-21767, 21856-22555, 22628-23650                                                                                            | Oxidoreductase                   | FAD/FMN-containing isoamyl alcohol oxidase MreA-like protein from <i>Aspergillus niger</i> CBS 513.88 (39/57) | XP_001400920.2   |
| PhqI    | 462       | 25396-26784                                                                                                                      | Prenyltransferase                | Putative tryptophan dimethylallyltransferase from <i>Coccidioides posadasii</i> C735 delta SOWgp (46/64)      | XP_003066904.1   |
| PhqJ    | 406       | 29106-30053, 30105-30377                                                                                                         | Prenyltransferase                | Reverse prenyltransferase NotF from <i>Aspergillus</i> sp. MF297-2 (37/54)                                    | ADM34132.1       |
| PhqK    | 459       | 31443-32081, 32133-32737, 32799-32934                                                                                            | FAD monooxygenase                | TqaH from <i>Penicillium aethiopicum</i> (39/56)                                                              | ADY16696.1       |
| PhqL    | 563       | 34979-35567, 35630-35973, 36029-36147, 36201-36840                                                                               | P450 monooxygenase               | benzoate 4-monooxygenase cytochrome P450 from <i>Verticillium albo-atrum</i> VaMs.102 (37/58)                 | XP_003006919.1   |
| PhqM    | 536       | 37357-37564, 37639-37814, 37857-37967, 38043-38363, 38413-38504, 38556-38670, 38741-39002, 39056-39119, 39180-39300, 39498-39638 | P450 monooxygenase               | Elymoclavine monooxygenase from <i>Claviceps paspali</i> (42/62)                                              | AET79203.1       |
| PhqN    | 326       | 43471-44349, 44542-44643                                                                                                         | Methyltransferase                | Sterol 24-C-methyltransferase from <i>Arthroderma otae</i> CBS 113480 (45/66)                                 | XP_002849812.1   |
| PhqO    | 451       | 46048-46275, 46372-46452, 46684-46740, 46863-47385, 47463-47884                                                                  | P450 monooxygenase               | Cytochrome P450 from <i>Aspergillus niger</i> CBS 513.88 (36/54)                                              | XP_001392409.2   |

**Table S4.** Features of the *Mal* gene products

| Protein | Size (AA) | Exons                                                 | Function                    | Relative (identity/similarity [%])                                                     | Accession number |
|---------|-----------|-------------------------------------------------------|-----------------------------|----------------------------------------------------------------------------------------|------------------|
| MalA    | 667       | 1-2004                                                | Flavin-dependent halogenase | Conserved hypothetical protein from <i>Aspergillus clavatus</i> NRRL 1 (67/82)         | XP_001273636.1   |
| MalB    | 369       | 3054-4163                                             | Prenyltransferase           | <i>Cyclo-L-Trp-L-Trp</i> prenyltransferase from <i>Aspergillus oryzae</i> (34/57)      | ADI60056.1       |
| MalC    | 264       | 4886-5680                                             | Short chain dehydrogenase   | Short chain dehydrogenase from <i>Pyrenophora tritircirepentis</i> Pt-1C-BFP (43/64)   | XP_001930671.1   |
| MalD    | 336       | 5937-6202, 6299-6628, 6707-6860, 6932-7149, 7228-7270 | Negative regulator          | NmrA-like family protein from <i>Aspergillus flavus</i> NRRL3357 (44/63)               | XP_002384043.1   |
| MalE    | 438       | 8047-9200, 9297-9459                                  | Prenyltransferase           | Reverse prenyltransferase NotF from <i>Aspergillus</i> sp. MF297-2 (42/65)             | ADM34132.1       |
| MalF    | 590       | 10457-11437, 11521-12047, 12104-12368                 | Oxidoreductase              | NotD from <i>Aspergillus</i> sp. MF297-2 (41/60)                                       | ADM34137.1       |
| MalG    | 2345      | 13142-20179                                           | NRPS                        | Non-ribosomal peptide synthetase from <i>Trichoderma atroviride</i> IMI 206040 (32/52) | EHK45804.1       |

**Table S5.** Features of the 5' and 3' boundary of *not'*, *phq* and *mal*

| Gene cluster | ORF                                 | Size (AA) | Putative function                 | Relative (identity/similarity [%])                                                               | Accession number |
|--------------|-------------------------------------|-----------|-----------------------------------|--------------------------------------------------------------------------------------------------|------------------|
| <i>not'</i>  | <i>orf(-1)</i>                      | 558       | Polysaccharide synthase           | Capsule polysaccharide biosynthesis protein from <i>Aspergillus fumigatus</i> Af293 (32/50)      | XP_748327        |
|              | <i>orf(-2)</i>                      | 335       | Transcription factor              | Zn(II)2Cys6 transcription factor from <i>Aspergillus oryzae</i> RIB40 (65/79)                    | XP_003189787.1   |
|              | <i>orf(+1)</i> (i.e. <i>notK'</i> ) | 577       | Nucleoside transporter            | Purine transporter from <i>Aspergillus nidulans</i> FGSC A4 (89/93)                              | XP_681803        |
|              | <i>orf(+2)</i> (i.e. <i>notL'</i> ) | 620       | Transcription factor              | Putative Zn(II)2Cys6 transcription factor from <i>Aspergillus nidulans</i> FGSC A4 (61/70)       | CBF80748         |
|              | <i>orf(+3)</i> (i.e. <i>notM'</i> ) | 454       | Unknown                           | Hypothetical protein from <i>Aspergillus nidulans</i> FGSC A4 (51/64)                            | CBF80750         |
| <i>phq</i>   | <i>orf(-1)</i>                      | 463       | Beta-glucosidase precursor        | Beta-glucosidase precursor <i>Talaromyces stipitatus</i> ATCC 10500 (48/67)                      | XP_002483357.1   |
|              | <i>orf(-2)</i>                      | 540       | D-lactate dehydrogenase           | D-lactate dehydrogenase from <i>Aspergillus oryzae</i> RIB40 (56/73)                             | XP_001823427.2   |
|              | <i>orf(+1)</i>                      | 769       | Response regulator                | Putative response regulator from <i>Penicillium chrysogenum</i> Wisconsin 54-1255 (72/79)        | XP_002565550.1   |
|              | <i>orf(+2)</i>                      | 279       | PQ loop repeat protein            | PQ loop repeat protein from <i>Neosartorya fischeri</i> NRRL 181 (74/85)                         | XP_001259845.1   |
| <i>mal</i>   | <i>orf(-1)</i>                      | 252       | MIP transporter                   | Putative MIP transporter from <i>Penicillium chrysogenum</i> Wisconsin 54-1255 (60/67)           | XP_002562918.1   |
|              | <i>orf(+1)</i>                      | 420       | Ankyrin repeat-containing protein | Putative ankyrin repeat-containing protein from <i>Penicillium marneffeii</i> ATCC 18224 (35/52) | XP_002148863.1   |
|              | <i>orf(+2)</i>                      | 356       | Unknown                           | Hypothetical protein TERG_02024 from <i>Trichophyton rubrum</i> CBS 118892 (45/61)               | XP_003237302.1   |



**Figure S1.** Features of the 5' and 3' boundary of *not'*, *phq* and *mal* (The asterisk denotes the boundary of the assembled contig)