Aspergillus collagen-like (*acl*) genes: identification, sequence polymorphism and assessment for PCR-based pathogen detection

Kiril Tuntevski[1], Brandon C. Durney[2], Anna K. Snyder[3], P. Rocco LaSala[4,5], Ajay P. Nayak[6], Brett J. Green[6], Donald H. Beezhold[6], Rita V. M. Rio[3], Lisa A. Holland[2], and Slawomir Lukomski[1#]

[1]Department of Microbiology, Immunology, and Cell Biology; [2]Department of Chemistry; [3]Department of Biology, [4]Department of Pathology, West Virginia University, and [5]Clinical Laboratory, West Virginia University Healthcare, Morgantown, West Virginia, 26506; [6]Allergy and Clinical Immunology Branch, Health Effects Laboratory Division, National Institute for Occupational Safety and Health (NIOSH), Centers for Disease Control and Prevention, Morgantown, WV 26505, USA.

G1X2T6_ARTOA [*Arthrobotrys oligospora* (strain ATCC 24927 / CBS 115.81 / DSM 1491) (Nematode-trapping fungus) (*Didymozoophaga oligospora*)] Putative uncharacterized protein (75 residues)

B8LU57_TALSN [*Talaromyces stipitatus* (strain ATCC 10500 / CBS 375.48 / QM 6759 / NRRL 1006) (*Penicillium stipitatum*)] Cuticle collagen, putative (193 residues)

G7XZR3_ASPKW [*Aspergillus kawachii* (strain NBRC 4308) (White koji mold) (*Aspergillus awamori* var. *kawachi*)] Putative uncharacterized protein (239 residues)

Q0CUF0_ASPTN [*Aspergillus terreus* (strain NIH 2624 / FGSC A1156)] Predicted protein (265 residues)

Q0CXL1_ASPTN [*Aspergillus terreus* (strain NIH 2624 / FGSC A1156)] Predicted protein (358 residues)

B8NLK1_ASPFN [*Aspergillus flavus* (strain ATCC 200026 / FGSC A1120 / NRRL 3357 / JCM 12722 / SRRC 167)] Putative uncharacterized protein (554 residues)

G3YG25_ASPNA [*Aspergillus niger* (strain ATCC 1015 / CBS 113.46 / FGSC A1144 / LSHB Ac4 / NCTC 3858a / NRRL 328 / USDA 3528.7)] Putative uncharacterized protein (556 residues)

Q2TZV0_ASPOR [*Aspergillus oryzae* (strain ATCC 42149 / RIB 40) (Yellow koji mold)] Predicted protein (114 residues)

Q4WW98_ASPFU [*Neosartorya fumigata* (strain ATCC MYA-4609 / Af293 / CBS 101355 / FGSC A1100) (*Aspergillus fumigatus*)] Collagen triple helix repeat protein (143 residues)

Q4WBU6_ASPFU [*Neosartorya fumigata* (strain ATCC MYA-4609 / Af293 / CBS 101355 / FGSC A1100) (*Aspergillus fumigatus*)] Extracellular proline rich protein (359 residues)

B0Y9H1_ASPFC [*Neosartorya fumigata* (strain CEA10 / CBS 144.89 / FGSC A1163) (*Aspergillus fumigatus*)] Extracellular proline rich protein (343 residues)

B0Y8Y8_ASPFC [*Neosartorya fumigata* (strain CEA10 / CBS 144.89 / FGSC A1163) (*Aspergillus fumigatus*)] Collagen triple helix repeat protein (125 residues)

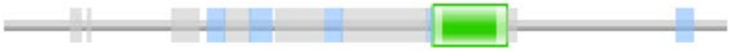
Q5B8U8_EMENI [*Emericella nidulans* (strain FGSC A4 / ATCC 38163 / CBS 112.46 / NRRL 194 / M139) (*Aspergillus nidulans*)] Putative GPI-anchored protein (Eurofung) (432 residues)

C1GFJ3_PARBD [*Paracoccidioides brasiliensis* (strain Pb18)] Transcription initiation factor TFIID subunit 12 (779 residues)

C1H5C3_PARBA [*Paracoccidioides brasiliensis* (strain ATCC MYA-826 / Pb01)] Transcription initiation factor TFIID subunit 12 (735 residues)

F2PPM5_TRIEC [*Trichophyton equinum* (strain ATCC MYA-4606 / CBS 127.97) (Horse ringworm fungus)] Putative uncharacterized protein (822 residues)

F2SFI2_TRIRC [*Trichophyton rubrum* (strain ATCC MYA-4607 / CBS 118892) (Athlete's foot fungus)] Putative uncharacterized protein (612 residues)



F2SMI7_TRIRC [*Trichophyton rubrum* (strain ATCC MYA-4607 / CBS 118892) (Athlete's foot fungus)] Putative uncharacterized protein (1009 residues)

F2RUU2_TRIT1 [*Trichophyton tonsurans* (strain CBS 112818) (Scalp ringworm fungus)] Extracellular serine-threonine rich protein (760 residues)

F2RUB4_TRIT1 [*Trichophyton tonsurans* (strain CBS 112818) (Scalp ringworm fungus)] Putative uncharacterized protein (Fragment) (513 residues)

H0EK56_GLAL7 [*Glarea lozoyensis* (strain ATCC 74030 / MF5533)] Putative Complement C1q tumor necrosis factor-related protein 9 (62 residues)

H0EFU5_GLAL7 [*Glarea lozoyensis* (strain ATCC 74030 / MF5533)] Putative Collagen alpha-1(XXVII) chain (80 residues)

H0EUR9_GLAL7 [*Glarea lozoyensis* (strain ATCC 74030 / MF5533)] Putative Collagen-like protein 7 (326 residues)

H0EFU6_GLAL7 [*Glarea lozoyensis* (strain ATCC 74030 / MF5533)] Putative Collagen alpha-1(IX) chain (395 residues)

H0EK55_GLAL7 [*Glarea lozoyensis* (strain ATCC 74030 / MF5533)] Putative Collagen alpha-1(XVI) chain (76 residues)

G2YLV9_BOTF4 [*Botryotinia fuckeliana* (strain T4) (Noble rot fungus) (*Botrytis cinerea*)] Putative uncharacterized protein BofuT4_uP000540.1 (59 residues)

G2XV05_BOTF4 [*Botryotinia fuckeliana* (strain T4) (Noble rot fungus) (*Botrytis cinerea*)] Putative uncharacterized protein BofuT4_P059050.1 (480 residues)

A7E9G8_SCLS1 [*Sclerotinia sclerotiorum* (strain ATCC 18683 / 1980 / Ss-1) (White mold) (*Whetzelinia sclerotiorum*)] Predicted protein (84 residues)

A7F644_SCLS1 [*Sclerotinia sclerotiorum* (strain ATCC 18683 / 1980 / Ss-1) (White mold) (*Whetzelinia sclerotiorum*)] Predicted protein (39 residues)

A7EJU0_SCLS1 [*Sclerotinia sclerotiorum* (strain ATCC 18683 / 1980 / Ss-1) (White mold) (*Whetzelinia sclerotiorum*)] Predicted protein (91 residues)

A7EMQ4_SCLS1 [*Sclerotinia sclerotiorum* (strain ATCC 18683 / 1980 / Ss-1) (White mold) (*Whetzelinia sclerotiorum*)] Predicted protein (80 residues)

Q0UAZ0_PHANO [*Phaeosphaeria nodorum* (strain SN15 / ATCC MYA-4574 / FGSC 10173) (Glume blotch fungus) (*Septoria nodorum*)] Putative uncharacterized protein (488 residues)

B2WP93_PYRTR [*Pyrenophora tritici-repentis* (strain Pt-1C-BFP) (Wheat tan spot fungus) (*Drechslera tritici-repentis*)] Putative uncharacterized protein (76 residues)

E4ZY10_LEPMJ [*Leptosphaeria maculans* (strain JN3 / isolate v23.1.3 / race Av1-4-5-6-7-8) (Blackleg fungus) (*Phoma lingam*)] Putative uncharacterized protein (1270 residues)

E4ZN07_LEPMJ [*Leptosphaeria maculans* (strain JN3 / isolate v23.1.3 / race Av1-4-5-6-7-8) (Blackleg fungus) (*Phoma lingam*)] Predicted protein (69 residues)

I1SA91_GIBZE [*Gibberella zeae* (strain PH-1 / ATCC MYA-4620 / FGSC 9075 / NRRL 31084) (Wheat head blight fungus) (*Fusarium graminearum*)] Uncharacterized protein (217 residues)

G3JPL4_CORMM [*Cordyceps militaris* (strain CM01) (Caterpillar fungus)] Putative uncharacterized protein (264 residues)

G9NZ61_HYPAI [*Hypocrea atroviridis* (strain ATCC 20476 / IMI 206040) (*Trichoderma atroviride*)] Putative uncharacterized protein (986 residues)

G9P0I6_HYPAI [*Hypocrea atroviridis* (strain ATCC 20476 / IMI 206040) (*Trichoderma atroviride*)] Putative uncharacterized protein (744 residues)

G0RQ54_HYPJQ [*Hypocrea jecorina* (strain QM6a) (*Trichoderma reesei*)] Predicted protein (643 residues)

F9FLM8_FUSOF [*Fusarium oxysporum* (strain Fo5176) (Panama disease fungus)] Putative uncharacterized protein (493 residues)

E9FAT3_METAR [*Metarhizium robertsii* (strain ARSEF 23 / ATCC MYA-3075) (*Metarhizium anisopliae*)] Putative uncharacterized protein (233 residues)

E9DQY7_METAQ [*Metarhizium acridum* (strain CQMa 102)] Putative uncharacterized protein (223 residues)

G2XEJ8_VERDV [*Verticillium dahliae* (strain VdLs.17 / ATCC MYA-4575 / FGSC 10137)] Putative uncharacterized protein (447 residues)

G4NGY8_MAGO7 [*Magnaporthe oryzae* (strain 70-15 / ATCC MYA-4617 / FGSC 8958) (Rice blast fungus) (*Pyricularia oryzae*)] Putative uncharacterized protein (222 residues)

G4MLC0_MAGO7 [*Magnaporthe oryzae* (strain 70-15 / ATCC MYA-4617 / FGSC 8958) (Rice blast fungus) (*Pyricularia oryzae*)] Putative uncharacterized protein (830 residues)

F0X784_GROCL [*Grosmannia clavigera* (strain kw1407 / UAMH 11150) (Blue stain fungus) (*Graphiocladiella clavigera*)] Putative uncharacterized protein (170 residues)

F0XDL9_GROCL [*Grosmannia clavigera* (strain kw1407 / UAMH 11150) (Blue stain fungus) (*Graphiocladiella clavigera*)] Putative uncharacterized protein (844 residues)

B2AD52_PODAN [*Podospora anserina* (strain S / ATCC MYA-4624 / DSM 980 / FGSC 10383) (*Pleurage anserina*)] Predicted CDS Pa_4_105 (284 residues)

C4XYJ5_CLAL4 [*Clavispora lusitaniae* (strain ATCC 42720) (Yeast) (*Candida lusitaniae*)] Predicted protein (953 residues)

A5E4W8_LODEL [*Lodderomyces elongisporus* (strain ATCC 11503 / CBS 2605 / JCM 1781 / NBRC 1676 / NRRL YB-4239) (Yeast) (*Saccharomyces elongisporus*)] Predicted protein (1069 residues)

Q6BP25_DEBHA [*Debaryomyces hansenii* (strain ATCC 36239 / CBS 767 / JCM 1990 / NBRC 0083 / IGC 2968) (Yeast) (*Torulaspora hansenii*)] DEHA2E17094p (234 residues)

A5DGD0_PICGU [*Meyerozyma guilliermondii* (strain ATCC 6260 / CBS 566 / DSM 6381 / JCM 1539 / NBRC 10279 / NRRL Y-324) (Yeast) (*Candida guilliermondii*)] Putative uncharacterized protein (2198 residues)

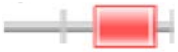
C5MGH0_CANTT [*Candida tropicalis* (strain ATCC MYA-3404 / T1) (Yeast)] Predicted protein (531 residues)

Q6FY13_CANGA [*Candida glabrata* (strain ATCC 2001 / CBS 138 / JCM 3761 / NBRC 0622 / NRRL Y-65) (Yeast) (*Torulopsis glabrata*)] Similarity (794 residues)

G0VGD0_NAUCC [*Naumovozyma castellii* (strain ATCC 76901 / CBS 4309 / NBRC 1992 / NRRL Y-12630) (Yeast) (*Saccharomyces castellii*)] Putative uncharacterized protein NCAS0F00660 (504 residues)

G0V741_NAUCC [*Naumovozyma castellii* (strain ATCC 76901 / CBS 4309 / NBRC 1992 / NRRL Y-12630) (Yeast) (*Saccharomyces castellii*)] Putative uncharacterized protein NCAS0A07310 (677 residues)

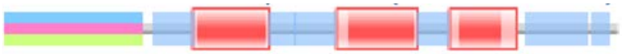
Q4P6H7_USTMA [*Ustilago maydis* (strain 521 / FGSC 9021) (Smut fungus)] Putative uncharacterized protein (144 residues)



E7A1G3_SPORE [*Sporisorium reilianum* (strain SRZ2) (Maize head smut fungus)] Putative uncharacterized protein (1335 residues)

F4RHV5_MELLP [*Melampsora larici-populina* (strain 98AG31 / pathotype 3-4-7) (Poplar leaf rust fungus)] Putative uncharacterized protein (480 residues)

E3KIN2_PUCGT [*Puccinia graminis* f. sp. *tritici* (strain CRL 75-36-700-3 / race SCCL) (Black stem rust fungus)] Putative uncharacterized protein (519 residues)



A8NU51_COPC7 [*Coprinopsis cinerea* (strain Okayama-7 / 130 / ATCC MYA-4618 / FGSC 9003) (Inky cap fungus) (*Hormographiella aspergillata*)] Putative uncharacterized protein (641 residues)

FIG. S1. Schematic representation of 62 fungal collagen-like proteins (CLPs). Several known domains are predicted to be present in fungal CLPs, such as: Ribonuclease (PF00545), TFIID_20kDa (PF03847), CFEM (PF05730), Aminotransferase I & II (PF00155), Peptidase C2 (PF00648), WSC domain (PF01822), Flo11 (PF10182), and DUF3246 (PF11596).

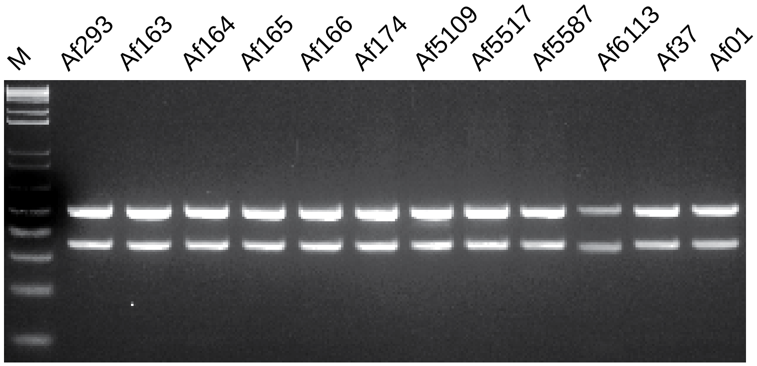


FIG. S2**.** Testing two-plex PCR targeting the *aclF1-5'* and *aclF1-3'* regions. Amplification with genomic DNA extracts from all 12 *A. fumigatus* strains is shown. PCR amplification was resolved by 2% agarose gel electrophoresis. M, 1kb Plus TrackIt DNA Ladder.

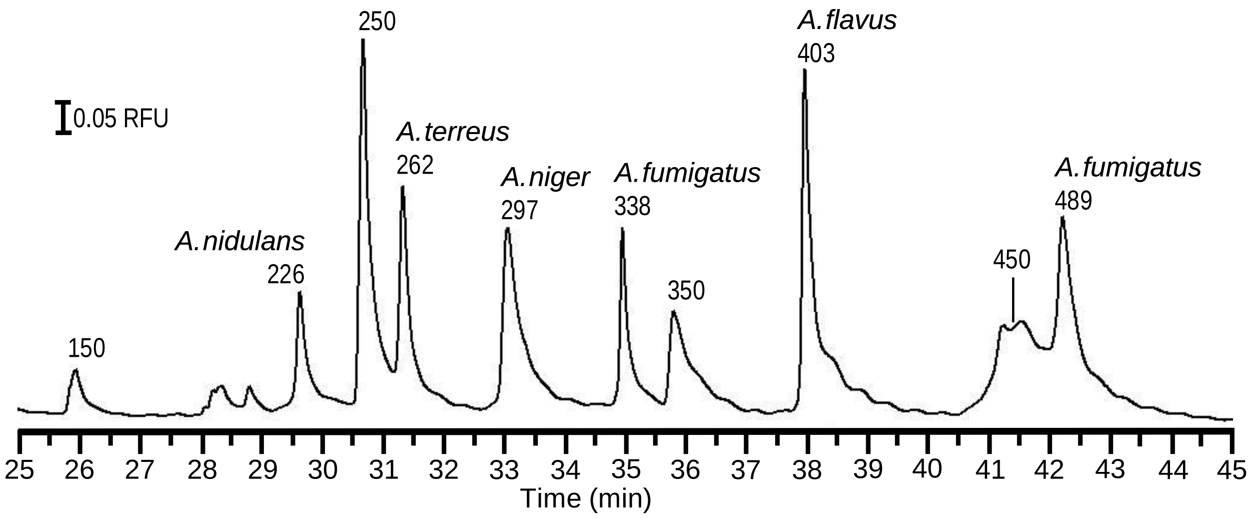


FIG. S3. Capillary electrophoresis separation of 5 marker mix and internal standards. Nanogel electrophoretic separation of five PCR amplified markers of *A. nidulans*, *A. terreus*, *A. niger*, *A. fumigatus*, and *A. flavus* and three internal size standards (150 bp, 350 bp, and 450 bp) detected with the intercalating dye SYBR Green 1. The separation is accomplished using a 25 µm inner diameter capillary with an effective length of 40.2 cm, E = 100 V/cm, 30 °C, 10% nanogel with [DMPC]/[DHPC] = 2.5, 6kV for 2 sec injection.