

真菌基因组数据库概况

周建芬¹ 曾昭清² 高元钢² 余知和^{2*}

(1. 长江大学图书馆 荆州 434023)

(2. 长江大学生命科学学院 荆州 434025)

摘要: 阐明不同生物基因组 DNA 序列信息及破译相关遗传学背景的基因组学是生物学和医学研究的核心学科。最近十年来真菌基因组学研究发生根本性的变化, 真菌已成为真核生物基因组研究的最佳模式生物。至 2008 年 6 月, 近 80 种隶属于真菌、微孢子虫和卵菌的全基因组序列公布, 代表着最广泛的真核生物, 它们的基因组大小从 2.5 Mb~81.5 Mb。本文整理了这些数据的相关信息。

关键词: 基因组, 真菌, 微孢子虫, 卵菌, 网络资源

The Summary of the Database of Fungal Genomes Publicly Available

ZHOU Jian-Fen¹ ZENG Zhao-Qing² GAO Yuan-Gang² YU Zhi-He^{2*}

(1. The Library, Yangtze University, Jingzhou 434023)

(2. College of Life Sciences, Yangtze University, Jingzhou 434025)

Abstract: Genomics that indicates DNA sequence of genomes of different organisms and their genetics background is a core of discipline closely related with biology and medical research. The last decade has witnessed a revolution in the genomics of the fungal kingdom and fungi have become the best model organisms of the eukaryotes. By June of 2008, there were less than 80 fungal genomes that were completely or almost completely sequenced and publicly accessible representing the widest sampling of genomes from any eukaryotic kingdom. These genomes belonging to true fungi, Microsporidia and Oomyces vary in size from approximately 2.5 Mb to 81.5 Mb. This review here provides an overview of available fungal genomes.

Keywords: Genomes, Fungi, Microsporidia, Oomyces, Online sources

1985 年美国科学家率先提出人类基因组计划, 其主要目标是阐明人类基因组 30 亿个碱基对的序列, 发现所有人类基因并认识其在染色体上的位置, 破译人类全部遗传信息。1986 年美国科学家 Thomas Roderick 提出基因组学概念。1990 年 10 月国际人类基因组计划在美国启动, 美、英、日、德、法和

中国相继参加这一宏伟计划, 2001 年 6 月 26 日人类基因组工作框架图如期完成。人类基因组计划带动了模式生物和其他重要生物的基因组学研究, 多个物种全基因组测序先后完成。阐明不同生物基因组 DNA 序列信息及破译相关遗传学背景的基因组学已经成为生物学和医学研究的核心学科。由欧洲、

美国、加拿大和日本等近百个实验室 600 多位科学家通力协作,第一个真核生物酿酒酵母*Saccharomyces cerevisiae*的基因组测序于 1996 年完成,这对于酵母菌类群来说是一个革命性的里程碑并且激起了真核生物基因功能和表达的全球性研究^[1]。随后粟酒裂殖酵母*Schizosaccharomyces pombe*^[2]和粗糙脉孢菌*Neurospora crassa*^[3]基因组的完成,显露出酿酒酵母作为所有其他真菌模式的局限性。尽管如此,真菌基因组测序的进展最初是缓慢的。为加快真菌基因组研究的步伐,2000 年由美国 Broad 研究所与真菌学研究团体发起真菌基因组行动(fungal genome initiative, FGI),目的是促进在医药、农业和工业上具有重要作用的真菌代表性物种的基因组测序。2002 年 2 月 FGI 发表了第一份关于测定 15 种真菌基因组计划的白皮书。2003 年 6 月,FGI 发表了第二份白皮书,列出了 44 种真菌作为测序的目标,强调对其中的 10 属即青霉*Penicillium*、曲霉*Aspergillus*、组织胞浆菌*Histoplasma*、球孢子菌*Coccidioides*、镰刀菌*Fusarium*、脉孢菌*Neurospora*、假丝酵母*Candida*、裂殖酵母*Schizosaccharomyces*、隐球酵母*Cryptococcus*和锈病菌*Puccinia*的物种优先进行测序。之后,经过 FGI、法国基因组学研究项目联盟(Génolevures Consortium)、美国能源部联合基因组研究所(The DOE Joint Genome Institute, JGI)、基因组研究院(The Institute for Genomic Research, TIGR)、英国 The Wellcome Trust Sanger Institute 和华盛顿大学基因组测序中心等共同努力,得到包括美国国家人类染色体研究所、美国国家科学基金会、美国农业部和能源部等以及来自产业集团如 Monsanto、Syngenta、Biozentrum、Bayer Crop Science AG 和 Exelixis 等的资助,在最近的几年里,真菌基因组学研究取得重大突破。至 2008 年 6 月 1 日,国际上约有 3806 种生物的全基因组序列测定工作已经完成或正在进行^[4],公开发表的 813 个完整的基因组中,约 75 种广义真菌或菌物的基因组测序工作已经组装完成或正在组装,包括子囊菌(Ascomycete) 59 种、担子菌(Basidiomycete) 8 种、接合菌(Zygomycete) 2 种、壶菌(Chytridiomycete) 1 种、微孢子虫(Microsporidia) 1 种、卵菌(Oomycete) 4 种(见表 1)。此外,还有*Ajellomyces dermatitidis* 和*Antonospora locustae*等 20 余种真菌基因组序列正在

测定中。这些真菌都是重要的人类病原菌、植物病原菌和腐生菌或者模式菌株,基因组大小为 2.5 Mb~81.5 Mb,包含酵母和丝状真菌,或者具有二型性(或多型性)生长,拥有与动物和植物细胞一样的细胞生理学和遗传学特征,包括多细胞性、细胞骨架结构、生长发育、有性生殖、细胞周期、细胞间信号传递、生理规律、DNA 甲基化和遗传修饰等,充分体现具有 9 亿年进化史真菌的生物学多样性。

1 真菌基因组学研究特色

较其他真核生物的基因组而言,真菌基因组相对简单,容易被测序和注解,而且便于进行遗传操作和基因修饰,因此,真菌成为真核生物基因组研究的最佳模式生物,真菌比较基因组学正处在真核生物比较基因组学的研究前沿,它能为其他很多重要的研究包括药物靶基因研究及药物开发、系统发生学、基因功能研究以及物种进化提供依据^[5,6]。最近有关真菌基因组的研究焦点集中在真菌与食品应用工程、真菌致病性(农业、医学)、抗真菌药物的发现、揭示人类基因与真菌基因的同源性以及酵母比较基因组学等^[7]。一方面,真菌基因组测序的种类涉及到真菌的各个类群(子囊菌、担子菌、接合菌、壶菌),重点选择子囊菌门位于同一演化分支(clade)的多个近缘种进行测序,如*Candida* clade、*Coccidioides* clade、*Fusarium* clade、*Histoplasma* clade、*Penicillium/Aspergillus* clade、*Neurospora* clade、*Schizosaccharomyces* clade 等,也包括担子菌门的*Cryptococcus* clade、*Puccinia* clade,或对同一个种(如*Saccharomyces cerevisiae*)的不同菌株进行测序,以进行进化和比较基因组学研究。另一方面,真菌基因组学研究的重要目的就是探讨真菌演化过程中的一些关键事件的遗传基础,如由法国 10 多个研究团体组织的 Génolevures Consortium,其目标旨在大规模地比较半子囊菌纲内*Saccharomyces cerevisiae* 与其他不同分支的酵母代表的基因组,并试图利用酵母相对小的酵母基因组来诠释真菌物种和类群的特异基因、功能基因簇中基因的分布、基因编码蛋白质的氨基酸分歧速率及染色体 shuffling 等分子进化方面的基本问题,从而为探索真核生物的基因组进化提供可靠信息。

表 1 已发表真菌及卵菌基因组的相关信息
Table 1 Complete fungal and oomycetes genomes

物种及菌株 Species and strain	分类 Taxonomy	测序中心 Sequencing center (s)	基因组大小 Size (Mb)	在线数据 Online database
<i>*Ajellomyces (Histoplasma) capsulatus</i> NAM1	Ascomycete	Broad Institute & Washington University Genome Sequencing Center	32.99	GenBank AAJI01000000
<i>Ajellomyces (Blastomyces) dermatitidis</i> ATCC 26199	Ascomycete	Washington University Genome Sequencing Center	28	http://genome.wustl.edu/genome.cgi?GENOME=Blastomyces%20dermatitidis&SECTION=assemblies
<i>Alternaria brassicicola</i> ATCC 96866	Ascomycete	Washington University Genome Sequencing Center, Virginia Polytechnic Institute, State University	30	http://genome.wustl.edu/genome_group.cgi?GROUP=5&SUB_GROUP=6
<i>Ascosphaera apis</i> USDA-ARSEF 7405	Ascomycete	Baylor College of Medicine	24	GenBank AARE01000000
<i>Ashbya (Eremothecium) gossypii</i> ATCC 10895	Ascomycete	Syngenta AG & Basel University	9.2	GenBank NC_005782 to 89
<i>Aspergillus clavatus</i> NRRL1	Ascomycete	The Institute for Genomic Research	27.86	GenBank AAKD03000000
<i>Aspergillus flavus</i> NRRL3357	Ascomycete	The Institute for Genomic Research & North Carolina State University	36.79	GenBank AAIH00000000
<i>*Aspergillus fumigatus</i> AF293	Ascomycete	The Institute for Genomic Research, Sanger Centre, Institute Pasteur	29.38	GenBank AAHF00000000
<i>Aspergillus nidulans</i> FGSC-A4	Ascomycete	Broad Institute	30.07	GenBank AACD01000000
<i>*Aspergillus niger</i> CBS 513.88	Ascomycete	DSM, The Netherlands	37.2	http://www.dsm.com/en_US/html/dfs/genomics_aniger.htm
<i>Aspergillus oryzae</i> RIB40	Ascomycete	National Institute of Technology and Evaluation, AIST and A. oryzae Genome Analysis Consortium	37.12	http://www.bio.nite.go.jp/ngac/e/rib40-e.html
<i>*Aspergillus terreus</i> NIH2624	Ascomycete	Broad Institute	29.33	GenBank AAJN01000000
<i>*Batrachomyces dendrobatidis</i> JEL423	Chytridiomycete	Broad Institute	23.7	GenBank AATT01000000
<i>Botryotinia fuckeliana (Botrytis cinerea)</i> B05.10	Ascomycete	Syngenta AG & Broad Institute	42.66	GenBank AAID01000000
<i>*Candida albicans</i> SC5314	Ascomycete	Stanford University	14.42	www.candidagenome.org
<i>Candida dubliniensis</i> CD36	Ascomycete	Sanger Institute, University of Aberdeen, Trinity College	14.6	http://www.sanger.ac.uk/sequencing/Candida/dubliniensis/
<i>Candida glabrata</i> CBS138	Ascomycete	Institut Pasteur & Genoscope	12.28	GenBank NC_005967 to NC_006036
<i>Candida parapsilosis</i> CCTCC M203011	Ascomycete	Sanger Institute, University of Aberdeen, Trinity College	26	http://www.sanger.ac.uk/sequencing/Candida/parapsilosis/
<i>Candida tropicalis</i> MYA-3404	Ascomycete	Broad Institute	14.58	GenBank AAFN01000000
<i>Chaetomium globosum</i> CBS148.51	Ascomycete	Broad Institute	34.89	GenBank AAFU01000000
<i>Clavispora (Candida) lusitanae</i> ATCC42720	Ascomycete	Broad Institute	12.12	GenBank AAFT01000000
<i>*Coccidioides immitis</i> RS	Ascomycete	Broad Institute	28.89	GenBank AAEC02000000
<i>*Coccidioides posadasii</i> RMSCC3488	Ascomycete	Broad Institute	28.15	GenBank ABBB01000000
<i>Coprinopsis (Coprinus) cinerea</i> okayama7#130	Basidiomycete	Broad Institute	36.26	GenBank AACS00000000
<i>Cryptococcus neoformans</i> serotype A, H99	Basidiomycete	Broad Institute	19.46	http://www.broad.mit.edu/annotation/genome/cryptococcus_neoformans/Home.html
<i>*Cryptococcus neoformans</i> serotype B, R265	Basidiomycete	Broad Institute	17.5	GenBank AAFF01000000

续表

物种及菌株 Species and strain	分类 Taxonomy	测序中心 Sequencing center (s)	基因组大小 Size (Mb)	在线数据 Online database
* <i>Cryptococcus neoformans</i> serotype D, JEC21	Basidiomycete	The Institute for Genomic Research & Stanford University	19.05	GenBank NC_006670 to 94
<i>Debaryomyces hansenii</i> CBS767	Ascomycete	Inland Northwest Research Alliance & Genoscope	12.2	GenBank NC_006043 to 49
<i>Encephalitozoon cuniculi</i> GB-M1	Microsporidia	NCBI Genome Project, Genoscope & University Blaise Pascal	2.5	GenBank NC_003229 to 42
<i>Epichloe festucae</i> ATCC56430	Ascomycete	University of Kentucky	30	ftp://ftp.genome.ou.edu/pub/e_festucae
<i>Fusarium oxysporum</i> f. sp. <i>lycopersici</i> 4286	Ascomycete	Broad Institute	61.34	GenBank AAXH000000000
<i>Gibberella moniliformis</i> (<i>Fusarium verticillioides</i>) 7600	Ascomycete	Broad Institute & Syngenta	41.77	GenBank AAIM020000000
<i>Gibberella zeae</i> (<i>Fusarium graminearum</i>) PH-1	Ascomycete	Broad Institute	36.43	GenBank AACM020000000
* <i>Hypocrea jecorina</i> (<i>Trichoderma reesei</i>) QM6a	Ascomycete	DOE Joint Genome Institute & Los Alamos National Laboratory	33	GenBank AAIL010000000
<i>Kluyveromyces lactis</i> NRRL Y-1140	Ascomycete	University Claude Bernard & Genoscope	10.69	GenBank NC_006037 to 42
<i>Kluyveromyces waltii</i> NCYC2644	Ascomycete	Broad Institute	11	GenBank AADM000000000
<i>Laccaria bicolor</i> S238N-H82	Basidiomycete	DOE Joint Genome Institute	66	GenBank ABFE010000000
<i>Lodderomyces elongisporus</i> NRRL YB-4239	Ascomycete	Broad Institute	15.4	GenBank AAP000000000
<i>Magnaporthe grisea</i> 70-15	Ascomycete	Broad Institute & North Carolina State University	41.62	GenBank AACU020000000
<i>Malassezia globosa</i> CBS7966	Basidiomycete	Procter & Gamble Company	9	GenBank AAYY010000000
<i>Malassezia restricta</i> CBS7877	Basidiomycete	Procter & Gamble Company	9	GenBank AAXK010000000
<i>Mycosphaerella fijiensis</i> CIRAD 86	Ascomycete	DOE Joint Genome Institute, United States Department of Agriculture, Purdue University	73.4	http://genome.jgi-psf.org/Mycf11/Mycf11.download.ftp.html
<i>Mycosphaerella graminicola</i> (<i>Septoria tritici</i>) IPO323	Ascomycete	DOE Joint Genome Institute, United States Department of Agriculture, Purdue University	41.2	http://genome.jgi-psf.org/Mycgr1/Mycgr1.download.ftp.html
<i>Nectria haematococca</i> MPVI	Ascomycete	DOE Joint Genome Institute	51.15	GenBank AF178385
<i>Neosartorya</i> (<i>Aspergillus</i>) <i>fischeri</i> NRRL181	Ascomycete	The Institute for Genomic Research	32.55	GenBank AAKE000000000
<i>Neurospora crassa</i> OR74A	Ascomycete	Broad Institute, OGI School of Science & Engineering, University of Kentucky, University of Kansas Washington University Genome Sequencing Center	39	GenBank AABX000000000
<i>Peronospora</i> (<i>Hyaloperonospora</i>) <i>parasitica</i> Emoy2	Oomycete		79.5	http://genome.wustl.edu/pub/organism/Fungi/Hyaloperonospora_parasitica/assembly/Hyaloperonospora_parasitica-2.0/
<i>Phaeosphaeria</i> (<i>Stagonospora</i>) <i>nodorum</i> SN15	Ascomycete	Broad Institute	37.24	GenBank AAGI000000000
<i>Phanerochaete chrysosporium</i> RP-78	Basidiomycete	DOE Joine Genome Institute	29.6	GenBank AADS000000000
<i>Phycomyces blakesleeanus</i> NRRL1555	Zygomycete	DOE Joint Genome Institute	55.9	http://genome.jgi-psf.org/Phyb11/Phyb11.download.ftp.html
<i>Phytophthora infestans</i> T30-4	Oomycete	Broad Institute	225.6	http://www.broad.mit.edu/annotation/genome/phytophthora_infestans/Home.html
<i>Phytophthora ramorum</i> UCD Pr4	Oomycete	DOE Joine Genome Institute, Virginia Polytechnic Institute	65	http://genome.jgi-psf.org/Phyra1_1/Phyra1_1.info.html

续表

物种及菌株 Species and strain	分类 Taxonomy	测序中心 Sequencing center (s)	基因组大小 Size (Mb)	在线数据 Online database
<i>Phytophthora sojae</i> P6497	Oomycete	DOE Joine Genome Institute, Virginia Polytechnic Institute	95	GenBank AAQY000000000
* <i>Pichia (Hansenula) angusta (polymorpha)</i> RB11	Ascomycete	Génolevures	9.5	GenBank AF204738
<i>Pichia guilliermondii</i> ATCC6260	Ascomycete	Broad Institute	29.33	GenBank AAFM010000000
<i>Pichia stipitis</i> CBS6054	Ascomycete	DOE Joint Genome Institute	15.4	GenBank AAVQ010000000
<i>Pneumocystis carinii</i>	Ascomycete	Broad Institute & University of Cincinnati	7.7	http://pgp.cchmc.org/
<i>Podospira anserina</i> S Mat+	Ascomycete	The French National Center for Scientific Research & Genoscope	34	http://podospira.igmors.u-psud.fr/index.html
<i>Puccinia graminis</i> f. sp. <i>tritici</i> CRL75-36-700-3	Basidiomycete	Broad Institute	81.52	GenBank AAWC010000000
<i>Pyrenophora tritici-repentis</i> Pt-1C-BFP	Ascomycete	Broad Institute	37.21	GenBank AAXI010000000
<i>Rhizopus oryzae</i> RA 99-880	Zygomycete	Broad Institute	46.09	GenBank AACW020000000
* <i>Saccharomyces bayanus</i> MCYC 623	Ascomycete	Washington University Genome Sequencing Center & Broad Institute	11.54	GenBank AACA000000000
<i>Saccharomyces castellii</i> NRRL Y-12630	Ascomycete	Washington University Genome Sequencing Center	12	GenBank AACF000000000
* <i>Saccharomyces cerevisiae</i> S288c	Ascomycete	Sanger Institute	11.74	www.yeastgenome.org
<i>Saccharomyces kluyveri</i> NRRL Y-12651	Ascomycete	Washington University School of Medicine, NCBI	12.6	GenBank AACE000000000
<i>Saccharomyces kudriavzevii</i> IFO1802	Ascomycete	Washington University Genome Sequencing Center	10.6	GenBank AACI000000000
<i>Saccharomyces mikatae</i> IFO1815	Ascomycete	Washington University Genome Sequencing Center & Broad Institute	12.12	GenBank AACH010000000
<i>Saccharomyces paradoxus</i> NRRL Y-17217	Ascomycete	Broad Institute	11.75	GenBank AABY000000000
<i>Schizosaccharomyces japonicus</i> yFS275	Ascomycete	Broad Institute	11.3	GenBank AATM010000000
<i>Schizosaccharomyces octosporus</i> yFS286	Ascomycete	Broad Institute	10.84	http://www.broad.mit.edu/annotation/genome/schizosaccharomyces_group/MultiHome.html
<i>Schizosaccharomyces pombe</i> 972h	Ascomycete	Sanger Institute & Cold Spring Harbor Laboratory	12.53	GenBank NC_003421 to 24 but not 22
<i>Sclerotinia sclerotiorum</i> 1980	Ascomycete	Broad Institute	38.33	GenBank AAGT010000000
<i>Uncinocarpus reesii</i> 1704	Ascomycete	Broad Institute	22.33	GenBank AAIW010000000
<i>Ustilago maydis</i> 521	Basidiomycete	Broad Institute	19.68	GenBank AAP010000000
<i>Vanderwaltozyma polyspora</i> DSM70294	Ascomycete	Trinity College	14.7	GenBank AAZN000000000
* <i>Yarrowia lipolytica</i> CLIB122	Ascomycete	Génolevures	20.5	GenBank NC_006067 to 72
<i>Zygosaccharomyces rouxii</i> CBS732	Ascomycete	Génolevures	12.8	GenBank AL392203 to AL397138

注：*多个菌株被测序

Note: More than one strain sequenced

2 部分完成全基因组序列的真菌种类(包括卵菌和微孢子虫)简介

2.1 子囊菌

子囊菌是真菌中最大的类群,也是已完成基因组序列最多的类群,共有 59 种子囊菌基因组序列已公布,主要包括酵母菌类群和青霉菌类及重要的人类和植物病原菌。其中曲霉属 7 个种,包括棒曲霉 *Aspergillus clavatus* 一种罕见的病原真菌,产生可能使人类和动物致病的真菌毒素——棒曲霉素(patulin);费希尔曲霉 *A. fischeri* 也称为 *Neosartorya fischeri*,一种有性的曲霉菌,引起角膜炎和移植病人肺部曲霉病;黄曲霉 *Aspergillus flavus* 感染谷物,威胁人类健康,产生高致癌的黄曲霉素(aflatoxin);烟曲霉 *A. fumigatus* 感染玉米和豆粕等谷物,是临床上重要的条件致病真菌,过敏体质者、器官移植患者及免疫缺陷患者等更易感染;米曲霉 *A. oryzae* 产生蛋白酶、淀粉酶、糖化酶、纤维素酶和植酸酶等复合酶,是生产在大肠杆菌中不能表达的真核生物活性蛋白的理想载体;构巢曲霉 *A. nidulans* 能诱导产生异核体或二倍体、有性和无性孢子,是遗传学和生物化学研究的模式生物之一;土曲霉 *A. terreus* 产生临床有关的次生代谢产物——他汀(lovastatin)被广泛应用于预防心脏病。此外,也能产生对人类和其他动物有害的真菌毒素——棒曲霉素和橘霉素(citrinin)。这些曲霉近缘种的基因组大约为 30 Mb~37 Mb,分布于 5 或 8 条染色体,通过比较基因组学研究,将有助于了解生物学、工业和医学等领域的重大问题^[8]。

白假丝酵母 *Candida albicans* 一种重要的条件致病菌,可在人的多个系统或器官与宿主共栖生存,最常见的是人口腔和阴道。白假丝酵母基因组最令人感兴趣的特点是染色体长度多态性现象,即假丝酵母染色体发生数值和结构性的重排、相互易位、缺失和个别染色体的三倍体化,这些染色体改变导致适应环境的显型改变。光滑假丝酵母 *C. glabrata* 是假丝酵母属中除白假丝酵母外第二个常见的人体条件病原菌,与酿酒酵母的亲缘关系密切。热带假丝酵母 *C. tropicalis* 不仅是造成器官深部感染的致病性假丝酵母,也是应用于工业生产制备聚酯、聚酰胺、香水和木糖醇,研究过氧化物酶体生成和过氧化物酶蛋白表达的重要生物体。葡萄牙假丝酵母 *C. lusitaniae* 也称为 *Clavispora lusitaniae*,引起的感染

约占假丝酵母病的 1%,并且引起假丝酵母菌血症(candidemia),部分菌株对抗菌药物有抗性。在致病假丝酵母中,白假丝酵母和热带假丝酵母是二倍体,而光滑假丝酵母和葡萄牙假丝酵母是单倍体,葡萄牙假丝酵母与白假丝酵母亲缘关系最远。洛德酵母 *Lodderomyces elongisporus* 是白假丝酵母 clade 中已知的唯一产生子囊孢子的种类,通常产生单个子囊孢子,无致病性,与 *C. parapsilosis*、*C. albicans* 和 *C. tropicalis* 的亲缘关系密切^[9]。

粗球孢子菌 *Coccidioides immitis* 与 *Coccidioides posadasii* 形态相同,是一种二型的土壤真菌,但遗传性质明显不同。两种真菌均引起球孢子菌病(Coccidioidomycosis),又称裂谷热,该病在美国的西北部、墨西哥的某些地方以及中南美洲是地方性流行病,但在非疫区仅有零星病例报道。粗球孢子菌是研究致病真菌进化生物学的模型,同时已被列为潜在的生物恐怖因子。两者的基因组大小相当,但前者含 14 条染色体而后者为 4 条染色体^[10]。*Uncinocarpus reesii* 是一种环境腐生型的非致病性真菌,单倍体,具有性阶段,与上述两种致病球粗球孢子菌关系近缘,22 个染色体,据估计与粗球孢子菌的进化距离约为 2000~3000 万年。

荚膜组织胞浆菌 *Ajellomyces capsulatus* 也是一种二型真菌,无性型为 *Histoplasma capsulatum*,在富氮环境中,荚膜组织胞浆菌形成菌丝体和分生孢子,由于物理扰动,分生孢子成雾状散开,如果人体被动吸入孢子,就会发生肺部感染,有时还会扩散到其他器官。荚膜组织胞浆菌基因组分布于 7 条染色体。

灰葡萄孢霉 *Botryotinia fuckeliana* 是一种植物致病菌,感染超过 200 种植物,包括大部分的蔬菜、水果、多种乔木、灌木、花卉和杂草,全世界每年因灰葡萄孢霉造成的损失估计有 10~100 亿欧元。单倍体,分布于 30 条染色体,可能有 10000~12000 基因。

球毛壳菌 *Chaetomium globosum* 是一种丝状真菌,可感染健康人的皮肤和指甲,引起免疫功能低下患者致命的全身性感染。球毛壳菌可产生毒枝菌素(mycotoxins),在农业生产上被广泛应用于果树、花卉、蔬菜和农作物等植物病害的生物防治。

串珠状赤霉 *Gibberella moniliformis* 又称为轮枝镰刀菌 *Fusarium verticillioide*,全球性地感染玉米,导致农作物重大的经济损失。单倍体,分布于 12 条染色体。

玉蜀黍赤霉 *Gibberella zeae* 又称禾谷镰刀菌 *Fusarium graminearum*, 引起小麦和大麦赤霉病的病原体, 还产生对人类及动物健康危害严重的呕吐毒素(vomitoxin)。

稻瘟病菌 *Magnaporthe grisea* 是一种世界性水稻病害的病原体, 无性型为 *Pyricularia grisea*, 已成为研究植物病原真菌和宿主相互作用的模式生物。

粗糙脉孢菌是 20 世纪现代遗传学和分子生物学研究的模式物种, 随着近年来对该物种研究的深入, 在基因的甲基化、生命节律、防御机制、基因沉默、DNA 修复、基因进化以及细胞信号传导等方面取得了许多重要的进展。基因组约 39 Mb, 有 7 条染色体, 可能含有 10082 个蛋白质编码基因, 已注释 1100 个基因。

核盘菌 *Sclerotinia sclerotiorum* 是已知具有最广泛宿主的真菌病原体, 引起菜豆、向日葵、油菜和大豆等菌核病, 是研究寄主-病原物相互作用的模型。单倍体, 有 3 条染色体。

瑞氏木霉 *Trichoderma reesei* 是多细胞的真核微生物, 其菌株可产生纤维素酶、半纤维素酶、蛋白酶和淀粉酶等酶类。单倍体, 7 条染色体。

由于酵母类真菌与人类的生产活动密切相关, 广义的酵母类群就理所当然地成为真菌基因组研究的一个重要目标。酿酒酵母作为一种模式生物在实验系统研究方面具有许多优势, 基因组约 12 Mb, 16 对染色体, 共有 6275 个开放阅读框, 预计功能基因 5800 个, 据估计约有 23% 与人类同源^[1]。其亲缘关系相近的其他芽殖酵母 *Saccharomyces bayanus*、*S. castellii*、*S. kluyveri*、*S. kudriavzevii*、*S. mikatae* 和 *S. paradoxus* 也都完成基因组测序, 它们的分化时间为 500~2000 万年前, 染色体数为 9、8 或 16 对。基因组大约都是 12 Mb, 与酿酒酵母基因组同线性高度保守。

与芽殖酵母不同的是, 粟酒裂殖酵母的繁殖方式为裂殖, 其遗传背景清晰, 是研究细胞周期调控、有丝分裂、减数分裂、DNA 修复与重组和基因组稳定性的模式生物, 也是针对细胞周期紊乱的抗肿瘤药物的理想细胞初筛模型。基因组大小为 13.8 Mb, 分布在 3 条染色体上, 预测含有 4997 个开放阅读框。而日本裂殖酵母 *Schizosaccharomyces japonicus* 是一个二型裂殖酵母, 能够形成真菌丝, 其侵染力很强, 可能成为侵袭性真菌生长的模式。

此外, 汉氏德巴利氏酵母 *Debaryomyces han-*

senii、*Eremothecium gossypii*、乳酸克鲁维酵母 *Kluyveromyces lactis*、瓦尔特克鲁维酵母 *K. waltii*、季氏毕赤酵母 *Pichia guilliermondii*、树干毕赤酵母 *P. stipitis* 和解脂耶氏酵母 *Yarrowia lipolytica* 等也相继完成或正在测序中, 旨在揭示酵母类真菌的系统演化关系^[11]。

2.2 担子菌

灰盖鬼伞 *Coprinopsis cinerea* 又称 *Coprinus cinereus*, 是一个多细胞、具有成形典型伞状结构并具备完整性周期的担子菌, 是研究多细胞有机体发展和调节的模型。

新型隐球酵母 *Cryptococcus neoformans* 是一种二型的异宗配合真菌, 在土壤中广泛存在, 是土壤和瓜果的腐生菌, 另外经呼吸道和消化道等感染免疫缺陷的人群, 危害极大。新型隐球酵母为单倍体, 有明确的有性期, 14 条染色体, 编码约 6574 个基因。

黄孢原毛平革菌 *Phanerochaete chrysosporium* 是研究最为深入的能够分解木质素的白腐菌, 也是第一个被测序的担子菌类真菌, 理解其基因组结构和组成可促进该菌在生物技术中的应用。

禾柄锈菌 *Puccinia graminis* 引起谷类作物秸秆锈病, 该菌具有复杂的生命周期, 包括 5 个孢子阶段和两个截然不同的宿主, 即以杂草为初宿主, 通常以小藜属植物为转换寄主。禾柄锈菌为单倍体, 分布于 18 条染色体。

玉米瘤黑粉菌 *Ustilago maydis* 是一种通常以一种丝状菌丝体存在的异宗配合担子菌, 其生活史有两种不同形态的细胞, 即单倍体细胞(担孢子)和双核菌丝体, 单倍体细胞没有致病性, 在特定培养基上芽殖产生“酵母”状菌落。不同遗传型的单倍体细胞融合形成的双核菌丝能在寄主植物体内迅速发育, 刺激寄主组织形成肿瘤, 并继而经过细胞核融合, 产生双倍体的冬孢子。玉米瘤黑粉菌是一个极好的研究寄主-病原物相互作用的模型系统, 尤其是分析响应寄主信号而引起感染菌株的形态改变, 有助于抗真菌药物分子靶标的研究。

2.3 接合菌和壶菌

布拉克须霉 *Phycomyces blakesleeanus* 因其对光线的敏感尤其是巨大孢囊梗的趋光性和对重力、风等环境反应的适应性而成为重要研究对象, 其基因组序列将有助于从分子水平上研究这些信号传递调控途径, 并为真菌的基因组进化提供关键信息。同时, 布拉克须霉也被作为真菌中 β -胡萝卜素生物合成与

调控的模式菌株,其基因组约有 14792 个基因。

米根霉 *Rhizopus oryzae* 是一种常见的引起毛霉病的病原,同时又是条件致病菌,存在于土壤、腐烂的水果和蔬菜及动物粪便,在亚洲常用来发酵食品和饮料,但是该菌产生对人畜有毒的麦角碱。

壶菌是真菌中唯一的产生游动孢子、既可腐生又可寄生的两栖群体,在自然界中具有重要的生态作用。*Batrachochytrium dendrobatidis* 无有性阶段,不产生休眠孢子,引起两栖动物的病害,是目前壶菌中唯一寄生脊椎动物的种类,也是第一个完成基因组序列的壶菌代表。因为壶菌处在真菌进化树中最基部,与动物的亲缘关系密切,因此 *Batrachochytrium dendrobatidis* 的基因组序列有利于生物的进化研究。

2.4 其他

微孢子虫是一类专营细胞内寄生、无线粒体和过氧化物酶体的单细胞真核生物,是昆虫、鱼类及哺乳动物的病原微生物,长期被归类为原生生物,最新研究表明可能源自专性寄生生活的真菌^[12]。兔脑炎原虫 *Encephalitozoon cuniculi* 是最早从兔中发现的感染哺乳动物的微孢子虫,宿主广泛,包括无脊椎动物(昆虫)、啮齿类动物、草、肉食性动物、禽类和灵长类动物等。感染人体,可影响神经系统、呼吸和消化器官。近十几年来有大量有关兔脑炎原虫引起艾滋病患者致死性感染的临床报道,其基因组是已知最小的真核生物基因组之一,仅 2.5 Mb,分布于 11 条染色体,约有 2000 个蛋白质编码基因。

卵菌与真菌趋同进化,有着真菌的许多特点,长期以来被作为真菌的类群而研究,现在归到与硅藻和海藻等一类的 Stramenopiles 界。卵菌是一类重要的植物病原菌,寄生范围很广,引起高等植物严重病害如腐霉菌、疫霉菌、霜霉菌和白锈菌等。其中,重要的病原代表疫霉属 *Phytophthora* 有 59 种,而且致病疫霉 *P. infestans*、*P. ramorum* 和 *P. sojae* 等因具有良好的遗传背景而作为研究该属的模式菌株,基因组相继被测序,预测三者基因组分别有 22658、15743 和 19027 个基因^[13]。

3 结束语

真菌种类庞大而多样,分布广泛,甚至在某些海洋环境和极端环境下也有真菌存在。应用比较基因组学方法对真菌基因组学的了解和应用有助于人们理解不同生境下真菌的生理特性、形态差异、进

化和代谢多样性。重要的人类病原菌、植物病原菌和腐生菌基因组序列的测定,不仅推动繁殖方式的选择、顶端生长、致病机制和宿主-真菌相互作用等基础研究,还能更好地寻找抗菌药物靶点,促进疫苗和抗菌药物的开发研制。相信随着迅速发展的高通量方法应用于真菌基因组学研究,将会大大地推动此领域的研究进程。

致谢:加拿大 Guelph 大学 Tom Hsiang 博士提供相关文献和部分基因组数据库信息,特此感谢。

参考文献

- [1] Goffeau A, Barrell BG, Bussey H, et al. Life with 6000 genes. *Science*, 1996, **274**: 546-567.
- [2] Wood VR, Gwilliam MA, Rajandream M, et al. The genome sequence of *Schizosaccharomyces pombe*. *Nature*, 2002, **415**: 871-880.
- [3] Galagan JE, Calvo SE, Borkovich KA, et al. The genome sequence of the filamentous fungus *Neurospora crassa*. *Nature*, 2003, **422**: 859-868.
- [4] Bernal A, Ear U, Kyripides N, et al. Genomes OnLine Database (GOLD): a monitor of genome projects world-wide. *Nucleic Acids Res*, 2001, **29**: 126-127.
- [5] Hsiang T, Baillie DL. Issues in comparative fungal genomics. *Applied Mycology and Biotechnology*, 2006, **6**: 100-122.
- [6] Xu JR, Peng YL, Dickman MB, et al. The dawn of fungal pathogen genomics. *Annu Rev Phytopathol*, 2006, **44**: 337-366.
- [7] Galagan JE, Henn MR, Ma LJ, et al. Genomics of the fungal Kingdom: insights into eukaryotic biology. *Genome Res*, 2005, **15**: 1620-1631.
- [8] Galagan JE, Calvo SE, Cuomo C, et al. Sequencing of *Aspergillus nidulans* and comparative analysis with *A. fumigatus* and *A. oryzae*. *Nature*, 2005, **438**: 1105-1115.
- [9] Kurtzman CP, Robnett CJ. Identification and phylogeny of ascomycetous yeasts from analysis of nuclear large subunit (26S) ribosomal DNA partial sequences. *Antonie van Leeuwenhoek*, 1998, **73**: 331-371.
- [10] Fisher MC, Koenig CL, White TJ, et al. Molecular and phenotypic description of *Coccidioides posadasii* sp. nov., previously recognized as the non-California population of *Coccidioides immitis*. *Mycologia*, 2002, **94**: 73-84.
- [11] Rokas A, Willaims BL, King N, et al. Genome-scale approaches to resolving incongruence in molecular phylogenies. *Nature*, 2003, **425**: 798-804.
- [12] Baldauf SL, Roger AJ, Wenk-Siefert I, et al. A kingdom-level phylogeny of eukaryotes based on combined protein data. *Science*, 2000, **290**: 972-976.
- [13] Randall TA, Dwyer RA, Huitema E, et al. Large-scale gene discovery in the oomycete *Phytophthora infestans* reveals likely components of phytopathogenicity shared with true fungi. *Mol Plant-Microbe Interact*, 2005, **18**: 229-243.