Infection, Genetics and Evolution 14 (2013) 302–312

Contents lists available at SciVerse ScienceDirect

Infection, Genetics and Evolution

journal homepage: www.elsevier.com/locate/meegid

Intraspecific comparative genomics of *Candida albicans* mitochondria reveals non-coding regions under neutral evolution

Thais F. Bartelli [a], Renata C. Ferreira [a,b], Arnaldo L. Colombo [b], Marcelo R.S. Briones [a,*]

[a] Laboratório de Genômica Evolutiva e Biocomplexidade, Departamento de Microbiologia, Imunologia e Parasitologia, Disciplina de Microbiologia, Universidade Federal de São Paulo, Rua Pedro de Toledo, 669, 4° andar, fundos, São Paulo, SP, CEP 04039-032, Brazil
[b] Laboratório Especial de Micologia, Disciplina de Infectologia, Universidade Federal de São Paulo, Rua Pedro de Toledo, 669, 5° andar, São Paulo, SP, CEP 04039-032, Brazil

ARTICLE INFO

Article history:
Available online 29 December 2012

Keywords:
Candida albicans
Mitochondrial genome
Molecular evolution

ABSTRACT

The opportunistic fungal pathogen *Candida albicans* causes serious hematogenic hospital acquired candidiasis with worldwide impact on public health. Because of its importance as a nosocomial etiologic agent, *C. albicans* genome has been largely studied to identify intraspecific variation and several typing methods have been developed to distinguish closely related strains. Mitochondrial DNA can be useful for this purpose because, as compared to nuclear DNA, its higher mutational load and evolutionary rate readily reveals microvariants. Accordingly, we sequenced and assembled, with 8-fold coverage, the mitochondrial genomes of two *C. albicans* clinical isolates (L296 and L757) and compared these sequences with the genome sequence of reference strain SC5314. The genome alignment of 33,928 positions revealed 372 polymorphic sites being 230 in coding and 142 in non-coding regions. Three intergenic regions located between genes tRNAGly/*COX1*, *NAD3*/*COB* and ssurRNA/*NAD4L*, named IG1, IG2 and IG3, respectively, which showed high number of neutral substitutions, were amplified and sequenced from 18 clinical isolates from different locations in Latin America and 2 ATCC standard *C. albicans* strains. High variability of sequence and size were observed, ranging up to 56 bp size difference and phylogenies based on IG1, IG2 and IG3 revealed three groups. Insertions of up to 49 bp were observed exclusively in Argentinean strains relative to the other sequences which could suggest clustering by geographical polymorphism. Because of neutral evolution, high variability, easy isolation by PCR and full length sequencing these mitochondrial intergenic regions can contribute with a novel perspective in molecular studies of *C. albicans* isolates, complementing well established multilocus sequence typing methods.

1. Introduction

Candida spp. are important opportunistic fungal pathogens and one of the major leading causes of superficial and life-threatening bloodstream infections, especially in hospitalized immunocompromised hosts (Koh et al., 2008; Lim et al., 2012; Pfaller,1996). In Brazil, the overall incidence reported in a surveillance study showed 2.49 cases per 1000 hospital admissions which is 2–15 times greater than in countries in the Northern Hemisphere, such as the United States (Colombo et al. 2006). The primary source of most of these infections is endogenous, though there is severe risk of acquisition of *Candida* spp. from the hospital environment by contaminated plastic devices and staff skin (Dorko et al., 1999; Fanello et al., 2001; Pfaller, 1996).

The genome of *C. albicans* has been extensively studied to identify intraspecific variability and several typing methods were developed to effectively elucidate the epidemiology of *C. albicans* and to discriminate clinical isolates to help identify the source of contamination (Cliff et al., 2008; Fanello et al., 2001). DNA fingerprinting methods such as restriction fragment length polymorphism (RFLP), randomly amplified polymorphic DNA (RAPD) and pulsed field gel electrophoresis (PFGE), have been widely used for *C. albicans* typing (Fanello et al., 2001; Heo et al., 2011; Noumi et al., 2009; Ruiz-Diez et al., 1997). However, these techniques are prone to ambiguity and subjective interpretations because of variations in electrophoretic patterns such as band size and intensity. Moreover, these techniques are not indicated for estimating genetic distances and phylogenetic inference, because they underestimate the real number of evolutionary events, are subject to systematic errors and cannot be readily assessed in terms of probability models (Mello et al., 1998; Soll, 2000). More reliable molecular studying methods based on sequencing, such as the gold standard multilocus sequence typing (MLST), relies on the analysis of at least six nuclear housekeeping genes (Robles et al., 2004) and though several authors have used *C. albicans* mtDNA in molecular analysis (Anderson et al., 2005; Aranishi, 2006; Jacobsen et al.,

* Corresponding author. Tel.: +55 11 5576 4537; fax: +55 11 5572 4711.
E-mail addresses: briones.marcelo.rs@gmail.com, marcelo.briones@unifesp.br (M.R.S. Briones).

1567-1348
http://dx.doi.org/10.1016/j.meegid.2012.12.012

2008; Sanson and Briones, 2000; Watanabe et al., 2005), more studies are needed to investigate fully its intraspecific nucleotide diversity in *C. albicans*.

Mitochondrial DNA (mtDNA) is more susceptible to damage and mutations than nuclear DNA, mainly because of the presence of reactive oxygen species generated during ATP synthesis and less efficient repair system of gamma DNA polymerase (Kang and Hamasaki, 2002; Kaguni, 2004). The high mutation number and the faster evolutionary rate, from 5 to 10 times higher than nuclear DNA (Brown et al., 1979), makes mtDNA suitable for discrimination of closely related organisms and recent evolutionary events. Furthermore, because it is haploid and present in multiple copies in cells, greater efforts and high technology are not usually required for the amplification and sequencing of specific PCR products. Despite the high variability of mitochondrial genes, their use can be limited in genetic analysis of closely related populations because of low intraspecific variability, probably constrained by negative selection on functional domains (Aranishi, 2006; Sanson and Briones, 2000; Watanabe et al., 2005). Non-coding regions (*e.g.* introns, pseudogenes, intergenic) evolve neutrally or are at least significantly less susceptible to natural selection and fitness interference than coding regions. Therefore, these genomic segments are expected to have a higher number of polymorphic sites and to evolve faster, making them interesting sequences to explore intraspecific mitochondrial nucleotide variability (Aranishi, 2006; Watanabe et al., 2005).

In this study, we have sequenced the complete mitochondrial genomes of two *C. albicans* clinical isolates and compared them with the genome sequence of the reference strain SC5314, to identify intraspecific hypervariable sites. We demonstrated that intergenic regions evolve under neutrality and are the most variable segments in the mtDNA, interesting features that could bring light into the usefulness of these sequences in molecular studies of *C. albicans* microvariability.

2. Materials and methods

2.1. Strains and mtDNA isolation

C. albicans clinical isolates were obtained from the collection of the "Laboratório Especial de Micologia (LEMI), Disciplina de Doenças Infecciosas e Parasitárias (DIPA), Departamento de Medicina, Universidade Federal de São Paulo". 18 isolates were collected from patients with hematogenic infection by *C. albicans* from 1997 to 2010 in different locations in Latin America. Two standard *C. albicans* ATCC (American Type Culture Collection) strains were also used in the analysis (Table 1). Cultures were grown in YPD medium (1% yeast extract, 2% peptone, 2% dextrose) at 30 °C before experiments. Mitochondrial DNA for whole genome sequencing or PCR amplifications was isolated by the method described previously (Defontaine et al., 1990).

2.2. Yeast nuclei purification and DNA extraction

Yeast nuclei purification was performed according to the method described previously (Hahn, 2006). Nuclear DNA was extracted by adding 200 µl of Solution B (100 mM NaCl, 10 mM EDTA, 1% Sarkosyl, 50 mM Tris–HCl pH 7.8) and incubated for 30 min at room temperature, followed by purification with phenol–chloroform, washed in 70% Ethanol, Ethanol precipitated and resuspended in TE buffer.

2.3. Whole mitochondrial genome sequencing and assembly

The complete mitochondrial genome sequences of two *C. albicans* clinical isolates (L296 and L757) were obtained using the whole genome shotgun method (Fleischmann et al., 1995). For mitochondrial genomic library construction, mtDNA was randomly sheared by sonication (Sambrook and Russel, 2001) and fragments of size from 1 to 2 kb were blunt cloned into pBluescript IISK (Stratagene) prior to sequencing. mtDNA sequences were determined by dideoxynucleotide chain termination method of Sanger et al., (1977) using fluorescent BigDye terminator cycle sequencing kit (version 3.1; Applied Biosystems) in an ABI Prism 3100 automated sequencer (Applied Biosystems) according to the manufacturer's instructions. Assembly of finished sequences from chromatograms was generated using Phred (Ewing and Green, 1998; Ewing et al., 1998a), Phrap and Consed (Gordon et al., 1998). Sequences were considered finished when Phred scores were above 40, which corresponds to less than one estimated error per 10 kb assembled.

Table 1
C. albicans clinical isolates and accession number of nucleotide sequences used in this study. IG1 = tRNA-Gly/*COX1*, IG2 = *NAD3*/*COB* and IG3 = ssurRNA/*NAD4L*. Brazilian states RJ (Rio de Janeiro); SP (São Paulo); PR (Paraná); BA (Bahia). Strains in bold indicate that the complete mitochondrial genome sequence was used as source for nucleotide sequence. *COB* = Cytochrome *b*, ITS1 and ITS2 = rDNA ITS excluding 5.8S rDNA.

Strain	Source		Year	GenBank Accession No.					
	Clinical	Geographic		IG1	IG2	IG3	COB	ITS1	ITS2
SC5314	Blood	USA		NC002653	NC002653	NC002653	NC002653	NC002653	NC002653
ATCC 24433	Nail	USA		JQ814087	JQ814119	JQ814140	–	JX494812	JX494813
ATCC 90029	Blood	USA		JQ814086	JQ814120	JQ814141	–	JX494814	JX494815
34 ptc	Catheter	?	?	JQ814102	JQ814105	JQ814125	–	JX494790	JX494791
L296	Blood	Brazil/RJ	1997	JQ864234	JQ864234	JQ864234	JQ864234	JQ814076	JQ814082
L757	Blood	Brazil/SP	2001	JQ864233	JQ864233	JQ864233	JQ864233	JQ814077	JQ814083
6965	Blood	Brazil/SP	2010	JQ814098	JQ814109	JQ814129	–	JX494798	JX494799
6944A	Blood	Brazil/SP	2010	JQ814100	JQ814106	JQ814127	–	JX494794	JX494795
7060A	Blood	Brazil/SP	2010	JQ814097	JQ814123	JQ814130	–	JX494800	JX494800
6945	Blood	Brazil/SP	2010	JQ814099	JQ814108	JQ814128	–	JX494796	JX494797
6921	Blood	Brazil/PR	2010	JQ814101	JQ814107	JQ814126	–	JX494792	JX494792
7252A	Blood	Brazil/PR	2010	JQ814094	JQ814112	JQ814133	–	JX494804	JX494805
7251	Blood	Brazil/PR	2010	JQ814095	JQ814111	JQ814132	JQ814068	JQ814072	JQ814078
7082	Blood	Brazil/PR	2010	JQ814096	JQ814110	JQ814131	–	JX494802	JX494803
6924	Blood	Brazil/BA	2010	JQ814103	JQ814104	JQ814124	–	JX494788	JX494789
5147	Blood	Ecuador	2009	JQ814089	JQ814117	JQ814138	–	JQ814074	JQ814080
6592	Blood	Ecuador	2009	JQ814091	JQ814115	JQ814136	–	JX494808	JX494809
5982	Blood	Argentina	2009	JQ814088	JQ814118	JQ814139	JQ814067	JQ814075	JQ814081
6779	Blood	Argentina	2009	JQ814090	JQ814116	JQ814137	JQ814069	JX494810	JX494811
6185	Blood	Venezuela	2009	JQ814093	JQ814113	JQ814134	–	JX494806	JX494807
6461	Blood	Colombia	2009	JQ814092	JQ814114	JQ814135	–	JQ814073	JQ814079