UNIVERSIDADE DE LISBOA

FACULDADE DE CIÊNCIAS

DEPARTAMENTO DE BIOLOGIA VEGETAL

Biogeography of *Xanthophyllomyces dendrorhous*, a yeast with biotechnological potential – population mapping and host associations

Márcia Cristina David Palma

MESTRADO EM MICROBIOLOGIA APLICADA

2011

UNIVERSIDADE DE LISBOA

FACULDADE DE CIÊNCIAS

DEPARTAMENTO DE BIOLOGIA VEGETAL

Biogeography of *Xanthophyllomyces dendrorhous*, a yeast with biotechnological potential – population mapping and host associations

Dissertação orientada por Prof. Dr. José Paulo Sampaio (CREM, FCT-UNL)

e Profª. Drª. Margarida Barata (FCUL)

Márcia Cristina David Palma

MESTRADO EM MICROBIOLOGIA APLICADA

2011

FACULDADE DE CIÊNCIAS E TECNOLOGIA
UNIVERSIDADE NOVA DE LISBOA

CREM
CENTRO DE RECURSOS MICROBIOLÓGICOS

Biogeography of *Xanthophyllomyces dendrorhous*, a yeast with biotechnological potential – population mapping and host associations

Márcia Cristina David Palma

MASTER THESIS

2011

This thesis was fully performed at CREM (Centro de Recursos Microbiológicos), Departamento de Ciências da Vida of Universidade Nova de Lisboa under the direct supervision of Prof. Dr. José Paulo Sampaio.

Prof. Dr. Margarida Barata was the internal designated supervisor in the scope of the *Master in Applied Microbiology* of the Faculty of Sciences of the University of Lisbon.

To my father, the best of teachers

Acknowledgments

I wish to express my gratitude to my supervisor, Prof. Dr. José Paulo Sampaio first, for allowing me to integrate his team as a research fellow and for giving me the opportunity to learn and grow as a scientist. Secondly, I would like to thank him for accepting me as a master student, for giving me time to write this thesis and for all the scientific knowledge passed on to me during this period.

I would like to convey my acknowledgement to the host laboratory CREM (Centro de Recursos Microbiológicos) and FCT-UNL, for giving me the necessary conditions in order to perform my work. I also thank FCT (Fundação para a Ciência e Tecnologia) for providing financial support to the project in which this work was included (PTDC/BIA-BDE/73566/2006).

A special thanks to Dr. Diego Libkind for all the strains that he conveyed for this study, for all the information regarding the genome of *X. dendrorhous* and for all his collaboration in so many aspects of this work.

I would like to thank Prof. Dr. Rogério Tenreiro and Prof. Dr. Margarida Barata for their help, answering promptly to all my questions during the time of preparation of this thesis.

I would like to thank all my lab colleagues whose knowledge, encouragement and cheer kept me up when science got me down, especially Marco Coelho, Dr. Elisabete Valério, Carla Gonçalves and Jorge Santos.

To my best friend, my love and my life, thank you for being all this and more.

My parents deserve special mention for their limitless support and for all the things that they taught me. Words fail me to express my never ending gratitude and love for my mother whose support has allowed me to pursue my dream of becoming a scientist. No words are necessary to thank my father for wherever he may be, he knows that my inquisitive spirit was born from his own and my unwearyingly willpower and work ethics is but a fragment of his.

Resumo

Xanthophyllomyces dendrorhous é uma levedura basidiomiceta cuja grande relevância biotecnológica advém do facto de ser a única espécie cujo principal pigmento produzido é astaxantina. Este carotenoide possui várias atividades biológicas relevantes e inúmeras aplicações industriais, sendo o segundo carotenoide mais rentável do mercado. Apesar do interesse nesta levedura devido às suas aplicações biotecnológicas, ainda são limitados os conhecimentos sobre a sua biologia e biogeografia.

X. dendrorhous foi isolada pela primeira vez durante a década de 60 do século passado a partir de exsudados de árvores de folha caduca em diversas localizações no Hemisfério Norte, tendo-se considerado durante muito tempo restrita a estes habitats e latitudes. Inicialmente descrita como uma levedura assexuada designada *Phaffia rhodozyma*, em 1995 o seu estado sexuado, *Xanthophyllomyces dendrorhous,* foi observado e descrito pela primeira vez. O seu ciclo sexuado é caracterizado pela conjugação entre a célula-mãe e a gémula, após a qual ocorre a formação de um holobasídio aéreo que subsequentemente dá origem a basidiósporos na sua extremidade apical. Este estado sexuado é despoletado pela presença de polióis no meio de cultura. Taxonomicamente esta levedura encontra-se classificada na ordem Cystofilobasidiales, classe Tremellomycetes e subfilo Agaricomycotina. A descoberta de um novo habitat de *X. dendrorhous* em 2007 na Argentina (Patagónia), veio mostrar que esta levedura possui uma distribuição geográfica mais diversificada do que aquela considerada inicialmente. Nas florestas da Patagónia, *X. dendrorhous* encontra-se associada aos corpos frutíferos de um fungo ascomiceta, parasita obrigatório de árvores *Nothofagus,* denominado *Cyttaria*. À semelhança dos exsudados de árvores, os corpos frutíferos de *Cyttaria* são ricos em açúcares simples, estando presentes nos troncos de *Nothofagus* uma vez por ano. Após a descoberta de novas populações de *X. dendrorhous* no Hemisfério Sul foi realizado um estudo filogenético da região ITS (*internal transcribed spacer*) do rDNA das estirpes de *X. dendrorhous* e das diferentes árvores hospedeiras tendo-se verificado concordância entre ambas as filogenias. Este facto levou à formulação da hipótese de que a variabilidade intraespecífica observada em *X. dendrorhous*, poderia dever-se ao facto de cada linhagem se encontrar associada a um hospedeiro diferente. Após a descoberta de *X. dendrorhous* no Hemisfério Sul, a exploração de outras regiões geográficas onde, tanto *Nothofagus* como *Cyttaria* estão presentes, foi levada a cabo em 2009 por elementos do laboratório de acolhimento. As amostras biológicas recolhidas na Austrália e Nova Zelândia resultaram na obtenção de novos isolados de *Xanthophyllomyces*.

Face ao número crescente de isolados de *Xanthophyllomyces* existentes, o presente trabalho teve como principal objetivo o estudo das relações filogenéticas entre as diferentes

linhagens existentes assim como, a exploração de possíveis relações entre a variabilidade genética encontrada em *Xanthophyllomyces* e os hospedeiros aos quais esta levedura se encontra associada. De forma a atingir este objetivo, duas abordagens distintas foram exploradas: (i) amplificação, sequenciação e estudo filogenético de sete fragmentos de genes de *Xanthophyllomyces* e (ii) a cariotipagem de isolados representativos das diferentes linhagens encontradas. Os dados obtidos foram analisados sob o ponto de vista da associação aos hospedeiros e localização geográfica dos isolados. Os resultados mostraram a existência de quatro populações distintas em *X. dendrorhous*. A estrutura populacional encontrada sugere que as populações naturais se propagam preferencialmente de forma clonal. Cada uma das diferentes populações apresenta-se associada a diferentes plantas hospedeiras, sendo esse o fator que mais parece influenciar a estrutura populacional. No entanto verificou-se também alguma diferenciação geográfica apesar da sua influência parecer menos relevante. Na filogenia concatenada dos sete fragmentos de genes estudados foi observado que cada grupo filogenético principal agrupa estirpes isoladas a partir de um mesmo hospedeiro, sendo igualmente possível verificar a existência de subdivisões dentro destes mesmos grupos, que estão por sua vez associadas à origem geográfica dos isolados. Este facto foi corroborado pelas networks filogenéticas construídas no *software* Splistree4 e pela análise da estrutura populacional usando uma abordagem bayesiana implementada no *software* Structure. A determinação dos cariótipos de estirpes representativas de cada uma das populações encontradas em *X. dendrorhous* revelou padrões únicos de cromossomas para cada estirpe. O número de cromossomas das onze estirpes estudadas varia entre seis e 11, sendo o seu tamanho entre 0,8 Mb e 3,5 Mb. Não foi observada uma relação entre a estrutura populacional de *X. dendrorhous* e os cariótipos observados para as diferentes estirpes.

Verificou-se que um dos isolados em estudo, possui três genes heterozigóticos. Os alelos de cada gene foram separados por clonagem e posteriormente sequenciados, tendo-se verificado a existência de dois alelos distintos para cada gene heterozigótico. A análise filogenética permitiu observar que, para os três genes heterozigóticos, os alelos encontrados pertencem a duas populações distintas originárias respectivamente da Australásia (em associação com *Nothofagus*) e Japão (em associação com *Betula*). Este facto sugere a existência de contacto entre as duas populações.

Diversos estudos biogeográficos têm sido realizados sobre os géneros *Nothofagus*, *Betula* e *Cornus*. Enquanto que a biogeografia de *Nothofagus* parece ter sido moldada por fenómenos como a fragmentação do super continente Gondwana e eventos de dispersão a longas distâncias, as histórias biogeográficas de *Betula* e *Cornus* parecem ter sido influenciadas por migrações intercontinentais através das pontes terrestres de Bering e do Oceano Atlântico durante o período geológico Terciário. A análise da atual distribuição de *X.*

dendrorhous levando em consideração as histórias biogeográfias dos seus hospedeiros parece sugerir uma correlação entre ambas. É possível que a distribuição de *X. dendrorhous* no Hemisfério Sul esteja relacionada com a biogeografia de *Nothofagus*, assim como a sua distribuição no hemisfério Norte parece ser coerente com as histórias biogeográficas de *Betula* e *Cornus*. A coexistência de Nothofagaceae, Fagaceae e Betulaceae na região do arquipélago indo-australiano poderá ter permitido um salto de hospedeiro por parte de *X. dendrorhous* que assim colonizou outras áreas no Hemisfério Norte.

Entre os isolados recolhidos na Australásia foram encontradas duas putativas novas espécies de *Xanthophyllomyces*. É possível observar que a divergência destas novas espécies relativamente a *X. dendrorhous* se verifica ao nível molecular tanto na região ITS como nas sete regiões genómicas estudadas. As duas espécies, *Xanthophyllomyces* sp. I e *Xanthophyllomyces* sp. II apresentam o mesmo estado sexuado que *X. dendrorhous*, assim como produção de astaxantina. O presente estudo constitui o primeiro registo da existência de mais do que uma espécie no género *Xanthophyllomyces.* Com este estudo iniciou-se a exploração da recém descoberta biodiversidade existente na Australásia, identificando-a como um possível *hotspot* de diversidade e possivelmente como ponto de origem do género *Xanthophyllomyces.*

Palavras chave: *Xanthophyllomyces; Biogeografia; Associação a hospedeiros; Espécies novas; Diversidade.*

Abstract

The basidiomycetous yeast *Xanthophyllomyces dendrorhous* has biotechnological relevance due to the production of astaxanthin, a carotenoid with remarkable biological properties. First isolated during the 1960s from exudates of deciduous trees throughout the Northern Hemisphere, this yeast was for long considered to be restricted to these habitats and latitudes. The discovery of a new *Xanthophyllomyces* habitat in the Southern Hemisphere in 2007 brought a better understanding of the geographic distribution and ecology of this yeast. In Patagonian forests, *X. dendrorhous* was discovered to be associated with the fruiting bodies of *Cyttaria*, an ascomycetous parasite of *Nothofagus* trees. Given that the *Cyttaria-Nothofagus* host system is also present in Oceania, sampling of selected localities in Australasia was carried out in 2009 yielding a considerable number of new isolates of *Xanthophyllomyces*.

Benefiting from the growing collection of *Xanthophyllomyces* strains available, that encompasses now novel Australasian lineages, the major intend of this work was to elucidate the phylogenetic relationships between the different lineages, and to further explore the genetic variation within *Xanthophyllomyces*. To achieve this goal a representative group of isolates was studied using multilocus sequence analysis and karyotyping assays within an ecological and geographic framework.

The results showed that *X. dendrorhous* is composed by four well structured populations, which suggests that natural populations of this yeast are mostly clonal. The existence of an isolate bearing alleles from two distinct populations suggests contact between populations from Australasia and Japan. *Xanthophyllomyces dendrorhous* populations have different ecological associations and appear to have diverged first according to host tree and secondly by geography. Different karyotypes were presented by each of the tested strains and no correlation with the revealed population structure was observed. The Australasian population harbors two putative new species that appear to be endemic to the area. This study represents the first record of the existence of more than one species in the *Xanthophyllomyces* genus. The extant distribution of the *X. dendrorhous* in the Southern Hemisphere may be related with the biogeographic history of *Nothofagus* trees, while its Northern Hemisphere distribution appears to be coherent with the biogeographic histories of *Betula* and *Cornus* trees.

Keywords: *Xanthophyllomyces; Biogeography; Host association; Novel species; Diversity.*

Table of contents

Figure Index

Table Index

Symbol and abbreviations list

AIC – Akaike information criterion

ATCC - American Type Culture Collection

bp – Base pairs

CBS – Centraalbureau voor Schimmelcultures, Netherlands

CHEF – Contour-clamped homogeneous electric field

CLP – Chromosome length polymorphism

CN – Chromosome number

CRUB - Centro Regional Universitario Bariloche

Cy. – *Cystofilobasidium*

E. coli – *Escherichia coli*

ITS – Internal transcribed spacer

k – Population

MCMC – Markov Chain Monte Carlo

ML – Maximum likelihood

MLSA – Multilocus sequence analysis

MP – Maximum parsimony

MYA – Million years ago

PALM – Phylogenetic reconstruction by automated likelihood model selector

PCR – Polymerase chain reaction

PFGE – Pulse field gel electrophoreses

rDNA – Ribosomal deoxyribonucleic acid

RN – Reticulate network

SDS – Sodium dodecyl sulfate

SN – Splits network

Tris – 2-amino-2-hydroxymethyl-propane-1,3-diol

w/v – Weight / Volume

Δk – Delta k (determined by the rate of change in the log probability of data between successive k values)

1. Introduction

Xanthophyllomyces dendrorhous (asexual state *Phaffia rhodozyma*) is the only basidiomycetous yeast that produces astaxanthin, a high-value carotenoid, as its main pigment (Johnson 2003). Astaxanthin has remarkable antioxidant, neuroprotective, anti-inflammatory and antitumor properties (Naguib 2000, Kurihara *et al.* 2002, Frengova and Beshkova 2009, Lee *et al.* 2010) and is the second most important carotenoid in the market (Schmidt *et al.* 2010). The ability to ferment sugars, a rare attribute for basidiomycetous yeasts, makes *X. dendrorhous* a promising candidate for a cost-competitive biotechnological production of astaxanthin (Rodriguez-Saiz *et al.* 2010).

This yeast was first isolated by Herman Phaff in 1967 from exudates of deciduous trees in Japan and Alaska (Miller *et al.* 1976, Golubev 1995) and was described as an asexual species, *Phaffia rhodozyma,* in 1976 (Miller *et al.* 1976). On the following years, other isolates were obtained also from exudates of deciduous trees in Finland, Germany and Russia (Golubev 1995, Weber *et al.* 2006). Through phylogenetic analysis using rDNA sequences, it was observed that the species of genus *Cystofilobasidium* were the closest relatives of *Phaffia rhodozyma.* The molecular phylogenies placed this yeast in the order Cystofilobasidiales, class Tremellomycetes and subphylum Agaricomycotina (Scorzetti *et al.* 2002). In 1995 the sexual state was observed for the first time after inoculation of strains in an agar medium containing polyols. The sexual state was described as *Xanthophyllomyces dendrorhous* and is characterized by the conjugation between cell and bud (pedogamic sexual process), after which a slender cylindrical holobasidium forms. Basidiospores, usually four, are produced in the terminal apex of the basidium (Golubev 1995) (Figure 1). This sexual state is not yet fully understood since the real ploidy of the species remains uncertain. Several studies have provided evidence that the vegetative stage is diploid but suggested also the existence of different levels of ploidy among the few tested strains, including possible aneuploids (Kucsera *et al.* 1998, Medwid 1998, Hermosilla *et al.* 2003). Another interesting aspect of *Xanthophyllomyces* is the existence of chromosome length polymorphisms (CLP) between strains. Despite the fact that the number of strains used was small no two strains presented the same chromosomal pattern in the karyotyping studies performed to date (Nagy *et al.* 1994, Adrio *et al.* 1995, Cifuentes *et al.* 1997, Nagy *et al.* 1997). Besides polyploidy and CLP, determination of genetic diversity using different regions of the rDNA repeat, revealed that *Xanthophyllomyces* has a high intra-specific variability (Fell *et al.* 2007). Because so little was known regarding the ecology and biogeography of this yeast, there were no factors that could be directly related to such differences.

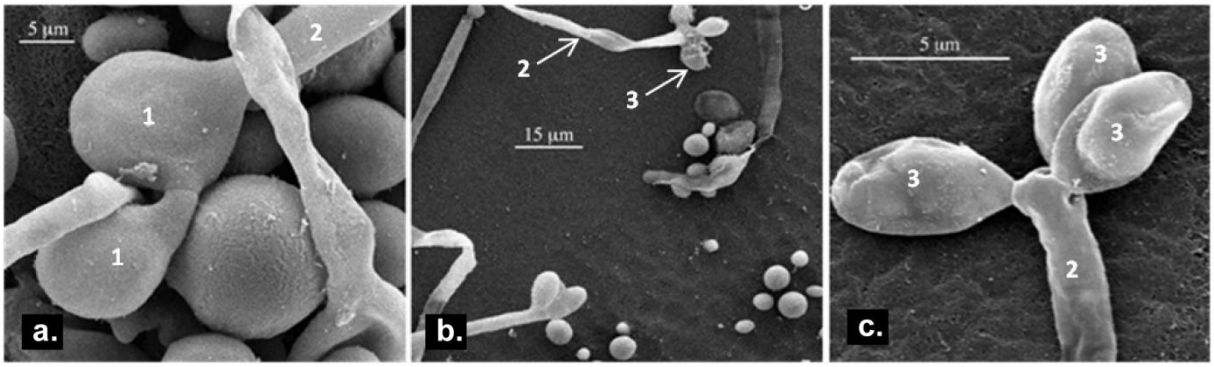

Figure 1 – *Xanthophyllomyces dendrorhous* sexual state. (a) Scanning electron microscopy photos showing two conjugated cells [1] originating a non-septate basidium [2]; **(b)** aerial basidium [2] with apical basidiospores [3]; **(c)** detail of the basidium apex [2] and basidiospores [3]. Image adapted from the website of the American Society for Microbiology, *http://202.195.144.50/ASM/106-Introduce.htm*.

Since all *Xanthophyllomyces* isolates were obtained from tree exudates belonging to genera *Betula*, *Alnus*, *Fagus* and *Cornus* it was assumed that the distribution of *X. dendrorhous* was limited to these host trees. Moreover, considering the biogeography of the plants (Chen *et al.* 1999, Chen and Li 2004, Manchester *et al.* 2007), it was assumed that *Xanthophyllomyces* was restricted to the Northern Hemisphere (Golubev 1995, Fell and Blatt 1999). However, the discovery of a new *Xanthophyllomyces* habitat in the Southern Hemisphere brought novel insights regarding the geographic distribution and ecology of this yeast. In Patagonia (Argentina), *X. dendrorhous* was found to be associated with the fruiting bodies of *Cyttaria*, an obligate biotrophic ascomycete fungus restricted to *Nothofagus* trees (Rawlings 1956) (Figure 2). The mycelium of *Cyttaria* develops inside the tree and persistent tumors are formed in the tree branches (Rawlings 1956). On these tumors, fruiting bodies rich in simple sugars are annually produced, and from them *X. dendrorhous* is consistently isolated (Libkind *et al.* 2007). In the same study a collection of strains from diverse geographical origins and also from different host systems comprising four different plant families, Betulaceae, Cornaceae, Fagaceae, and Nothofagaceae, was compared using the internal transcribed spacer (ITS) sequences of both the yeast and the host trees. Since a general concordance was observed between the two ITS phylogenies it was hypothesized that the genetic polymorphisms that segregated the strains into different lineages could represent intraspecific variation resulting from host-specific association. It was also hypothesized that the association of *X. dendrorhous* with *Cyttaria* derived from an earlier association of the yeast with *Nothofagus* trees. The genus *Nothofagus* forms a monophyletic group, sister to the rest of the Fagales that has been instrumental in understanding the biogeography of the Southern Hemisphere (Cook and Crisp 2005, Peterson *et al.* 2010).

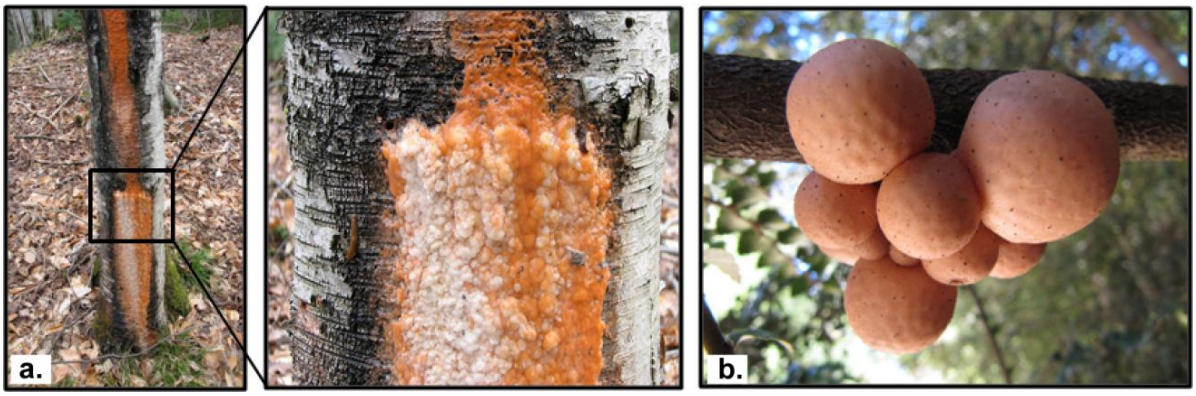

Figure 2 – Presently known habitats for *Xanthophyllomyces.* (a) *Betula* tree trunk with an orange colored exudate in North America (enlarged detail on the right); **(b)** branch of *Nothofagus cunninghamii* with *Cyttaria gunnii* fruiting bodies (Tasmania).

The extensive fossil record of *Nothofagus* pollen found throughout what was once the Gondwana supercontinent, has allowed tracking major shifts in geology and ecology that have taken place in the last 80 million years (Cook and Crisp 2005).

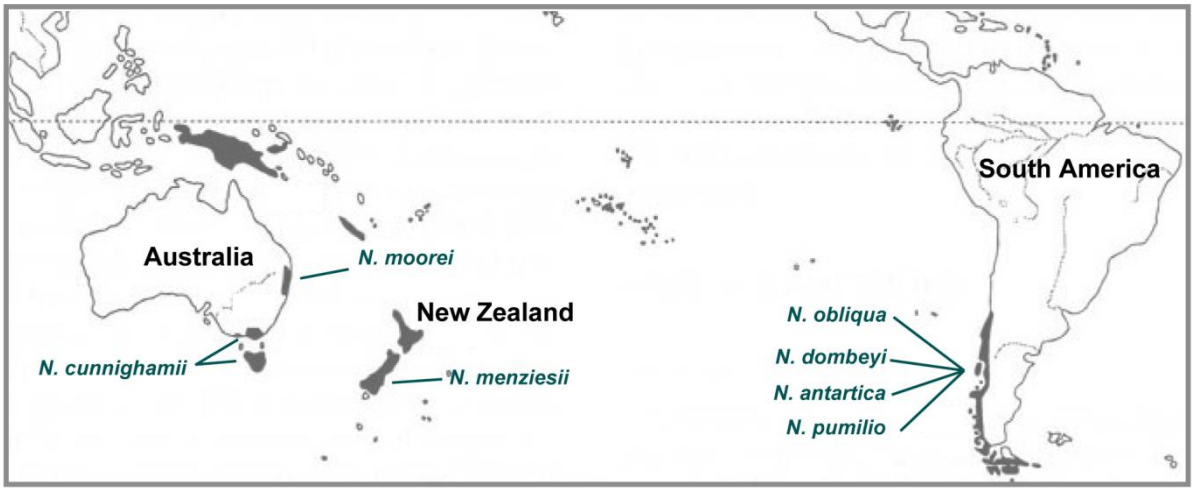

Figure 3 – *Nothofagus* world distribution. Map depicting in dark gray the present geographic distribution of *Nothofagus.* The distribution of several species is also indicated. Adapted from Knapp et al. (2005) and Swenson et al. (2001).

The extant disjunctive distribution of *Nothofagus* in the Southern Hemisphere, both in South America and Australasia has been correlated with the Gondwana break up and the continental drift theory, as well as more recent (in geological terms 44.6 – 28.5 million years ago) long-distance dispersal events within Australasia (Swenson *et al.* 2001, Mathiasen and Premoli 2010, Peterson *et al.* 2010). After the discovery of the association of *X. dendrorhous* with *Nothofagus* in Patagonia, the existence of unknown populations of this yeast in other regions colonized by *Nothofagus* had to be considered. In 2009 a collecting trip to *Nothofagus* forests in New Zealand and Australia by members of the host laboratory, yielded

a considerable number of new *Xanthophyllomyces* isolates confirming the existence of additional populations associated with *Nothofagus* trees. As the number of *Xanthophyllomyces* strains increased, also did the possibilities of better understanding the biology, ecology and world distribution of this microorganism.

With the increase of sequence information available, a broad application of phylogenetic methods has taken hold of fungal taxonomy and of biogeography studies. The multilocus sequence analysis (MLSA) or complete genome comparisons are being widely applied in detriment of the single genetic marker approach, given the extensively grater information yielded (Taylor *et al.* 1999). These techniques allow not only the recognition of species (Taylor *et al.* 2000) but also the study of distinct populations among a species (Taylor and Fisher, 2003). The statistical phylogeographic approach connects phylogeography methodologies and classical population genetics methods, providing insights on the evolutionary history of organisms (Knowles, 2009). Bioinformatics tools available today allow the inference of parameters such as recombination, population structure and to assign individuals to a population based on their ancestry composition from MLSA datasets (Crandall *et al.* 2000, Posada and Crandall 2001, Excoffier and Heckel 2006, Huson and Bryant 2006, Falush *et al.* 2007, Perez-Losada *et al.* 2007, Chen *et al.* 2009)

Taking advantage of the growing collection of *Xanthophyllomyces* strains that encompasses novel Australasian lineages, the major intend of this work was to shed light on the phylogenetic relationships between the different lineages, and to explore in more detail the genetic variation previously observed. To achieve this goal a representative group of isolates was investigated by MLSA and karyotype analysis within an ecological and geographic framework.

2. Material and methods

2.1. Studied strains

Forty strains from nine different geographical localizations and two distinct host systems were studied (Table 1).

Table 1 – List of strains used in the study and relevant information regarding them.

Species	Strain	Locality	Substrate
X. dendrorhous	CBS 5905	Kyoto, Japan	Exudate of *Fagus crenata*
	ATCC 24230	Japan	Exudate of *Betula tauschii*
	ATCC 24229	Hiroshima, Japan	Exudate of *Cornus brachypoda*
	CBS 5908	Kiso, Japan	Exudate of *Alnus japonica*
	ATCC 24201	Hiroshima, Japan	Exudate of *Cornus brachypoda*
	ATCC 24261	Yamagata, Japan	Exudate of *Betula maximowicziana*
	GY13L04	Taiwan	Soil sample near *Stauntonia purpurea*
	ATCC 24228	Kenai Lake, Alaska, United States of America	Exudate of *Betula papyritosa*
	NRRL Y-17432	Wisconsin, United States of America	Exudate of *Betula populifolia*
	NRRL Y-17433	Illinois, United States of America	Exudate of *Betula populifolia*
	NRRL Y-17434	Illinois, United States of America	Exudate of *Betula populifolia*
	NRRL Y-27348	Illinois, United States of America	Exudate of *Betula populifolia*
	CBS 7918T	Moscow, Russia	Exudate of *Betula verrucosa*
	CBS 6938	Finland	Exudate of *Betula* sp.
	KBP 2604	Novgorod, Russia	Exudate of *Betula* sp.
	KBP 2607	Komi Republic, Russia	Exudate of *Betula* sp.
	CRUB 0853	Gutierrez Lake coast, Patagonia, Argentina	*C. hariotii* on *N.dombeyi*
	CRUB 1149	Ilón Lake, Patagonia, Argentina	Water sample near *C. hariotii* on *N. pumilio*
	CRUB 1490	Gutierrez Lake coast, Patagonia, Argentina	*C.hariotii* on *N.dombeyi*
	CRUB 1613	Suisse colony, Patagonia, Argentina	*C.hariotii* on *N.dombeyi*
	CRUB 1614	Llao-llao Forest, Patagonia, Argentina	*C. hariotii* on *N.dombeyi*
	CRUB 1617	Tronador area, Patagonia, Argentina	*C. hariotii* on *N.antartica*
	CRUB 1620	Chile border, Patagonia, Argentina	*C. hariotii* on *N.dombeyi*
	CRUB 1621	Tronador area, Patagonia, Argentina	*C. hariotii* on *N. antarctica*
	CRUB 1151	Ilón Lake, Patagonia, Argentina	Water sample near *C. hariotii* on *N. pumilio*
	CRUB 1492	Gutierrez Lake coast, Patagonia, Argentina	Stroma of *C. hariotii* on *N. dombeyi*
	CRUB 1917	Hua-Hum, Patagonia, Argentina	*C.espinosae* on *N. obliqua*
	CRUB 1896	South America	Leaves of *N. pumilio*
	ZP 869	(Australia) Tasmania	*C.gunnii* on *N.cunninghamii*
	ZP 874	(Australia) Tasmania	*C. gunnii* on *N.cunninghamii*
	ZP 922	Haast Pass, New Zealand	*C.nigra* on *N.menziesii*
	ZP 888	Lewis Pass, New Zealand	*C. gunnii* on *N.menziesii*
***Xanthophyllomyces* sp. I**	ZP 955	Lamington Park, Queensland, Australia	Leaves of *N. mooreii*
	ZP 928	Lamington Park, Queensland, Australia	Leaves of *N. mooreii*
	ZP 938	Lamington Park, Queensland, Australia	Leaves of *N. mooreii*
	ZP 909	Haast Pass, New Zealand	*C.nigra* on *N. menziesii*
	ZP 884	Lewis Pass, New Zealand	*C.nigra* on *N.menziesii*
***Xanthophyllomyces* sp. II**	ZP 863	Mount Field National Park, Tasmania, Australia	*C.gunnii* on *N. cunninghamii*
	ZP 878	Mount Field National Park, Tasmania, Australia	*C. gunnii* on *N.cunninghamii*
	ZP 875	Mount Field National Park, Tasmania, Australia	*C.gunnii* on *N. cunninghamii*

All the strains from Australasia (referred to as ZP) were isolated (Libkind *et al.* 2007) during the research project in which the present study is included. Reference cultures were obtained from the following culture collections: Centraalbureau voor Schimmelcultures, Netherlands (CBS); Department of Soil Biology, Moscow State University (KBP); American Type Culture Collection (ATCC), Agricultural Research Service Culture Collection (also referred to as NRRL Collection) and from Centro Regional Universitario Bariloche (CRUB) in Argentina. Strain GY13L04 was provided by Dr. Ching-Fu Lee of the National Hsinchu University of Education in Taiwan.

2.2. DNA extraction, MLSA primers and PCR conditions

Genomic DNA of all strains studied was obtained as previously described (Gonçalves *et al.* 2011) and used to amplify different genomic regions. Primers for the MLSA were designed using sequences publicly available on GenBank (Table S1). The ITS region of the rDNA comprising the ITS1, *5.8S* gene and the ITS2 region was amplified and sequenced using universal primers (Sampaio *et al.* 2001) for all *Xanthophyllomyces* isolates from Australasia and Taiwan (Table S2). For MLSA a total of seven nuclear genes with distinct biological functions were selected and fragments were amplified by PCR using the conditions presented in Table S2. The evaluation of the chemical properties of each of the pair of primers was performed with software Primer Premier 5.0 (PREMIER Biosoft International, Palo Alto, California) and also with AutoDimer v1 software (Vallone and Butler 2004). After the PCR reaction, all products were purified either with NucleoSpin® kit from Macherey-Nagel Company or the GFX^{TM} PCR DNA and Gel Band Purification Kit from GE Healthcare Company according to the manufacturers protocols. All PCR products were sequenced by STAB VIDA. All sequences presented in this study will became available on Genbank once the results of this thesis are published in a scientific paper.

2.3. Cloning assays

The detection of heterozygous sequences of *CRTS*, *IDI* and *EPH1* genes in CBS 5905, made necessary the separation of the different haplotypes through cloning. The genes fragments were amplified and the products of the PCR reaction were purified as previously mentioned. Those products were then inserted into the vector pJET1/blunt and the cloning protocol of the CloneJET™ PCR Cloning Kit from Fermentas® was carried out. Competent DH5-α *E. coli* cells were prepared according to previously described method (Sambrook J. 1989) and then used for cloning assays. Due to specific characteristics of the cloning vector used, only transformed cells are able to grow and no additional screening method was

necessary. Colony PCR was performed using gene-specific primers (Table S2) in the obtained clones and the resulting products were sequenced as previously mentioned.

2.4. Phylogenetic analysis

All sequences produced in this study were aligned using ClustalW Multiple alignment tool (Thompson *et al.* 1994), present in the BioEdit software (Hall 1999). The MLSA phylogenetic analysis was performed using two different inference methods: Maximum Parsimony (MP) and Maximum Likelihood (ML). All MP phylogenies were performed in MEGA5 software (Tamura *et al.* 2011) using 1000 replicates (bootstrap analysis) for each phylogeny. The ML phylogenetic trees were performed in PALM server (phylogenetic reconstruction by automated likelihood model selector), which performs automated evolution model selection (Chen *et al.* 2009). The best fitting model was determined by Akaike Information Criterion (AIC) (Akaike 1987) for all alignments (Table S3). The nucleotide alignments were submitted in PHYLIP format (Felsenstein 1981), the number of substitution rate category was four (Chen *et al.* 2009). Individual phylogenetic trees were constructed for each of the seven partially sequenced genes (*CRTS; CRTI; IDI; L41; INV; EPH1* and *GDHA*) using both MP and ML. The concatenated alignment was composed by the seven genes fragments and 41 sequences (CBS 5905 is represented by two sequences consisting of the two different haplotypes resolved by cloning). The multilocus alignment was constituted by a total of 3187 positions (1287 informative positions). ML phylogenies are presented as supplementary data.

The phylogenetic analysis of *Xanthophyllomyces* ITS sequences was performed using sequences produced in this study together with sequences already available in GenBank. Moreover, 17 representative sequences of *Cystofilobasidium* available in GenBank were included in the ITS phylogenetic analysis, having the resulting alignment a total of 45 sequences and 369 positions. The phylogenetic analysis of the ITS region of several *X. dendrorhous* host trees was also performed. For both ITS phylogenies, the MP method was employed as detailed before and in these two phylogenies the complete deletion option was used.

2.5. Phylogenetic Networks analysis

The multilocus alignment was also used to construct phylogenetic networks. This method allows a more accurate representation of reticulate evolutionary events and also to evaluate the strength of the phylogenetic signal of a dataset. Two different types of networks were constructed: (i) Splits Networks (SN) using the Neighbor-net model, and (ii) Reticulate networks (RN) using the Recombination Network model (Huson and Bryant 2006). In a SN

the length of an edge is proportional to the weight of the associated split analogously to the length of a branch in a phylogenetic tree. A set of incompatible splits is represented in a network using bands of parallel edges. In a RN the internal nodes represent ancestral species and nodes with more than one parental group correspond to reticulated events. Additionally the Phi-test was also performed. This test uses the pairwise homoplasy index, Phi statistic, to detect refined incompatibility indicating recombination (Bruen *et al.* 2006). Both the construction of networks and the Phi-test were conducted in SplitsTree4 software (Huson and Bryant 2006).

2.6. Analysis of *X. dendrorhous* population structure by a quantitative clustering method

In order to investigate the population structure of *X. dendrorhous*, the multi-locus sequence dataset was analyzed in Structure software (Falush *et al.* 2007, Hubisz *et al.* 2009). This model-based clustering method (Pritchard *et al.* 2000) consists on a Bayesian approach using Markov Chain Monte Carlo (MCMC) (Hastings 1970). It detects the underlying genetic population structure among a set of individuals genotyped at multiple markers and also computes the proportion of the genetic information of an individual originating from each inferred population using a quantitative clustering method. The main characteristics of the computation process were a burning length of 100.000 and a MCMC after burning of 1.000.000. The ancestry model used was admixture and the number of populations (K) tested was from 1 up to 10 with 20 runs per each K value. The assumed ploidy of the genome was diploid. The results obtained by Structure were treated according to a mathematical method developed in order to perform a more formal analysis of the output (Evanno *et al.* 2005). This method uses an *ad hoc* statistic (ΔK) based on the rate of change in the log probability of data between successive K values thus allowing accurately detecting the uppermost hierarchical level of population structure in a dataset.

2.7. Karyotyping assays

With the aim of characterizing the chromosomal profile of strains representative of each of the *X. dendrorhous* lineages, an optimized protocol for PFGE-CHEF based on previous works was developed (Nagy *et al.* 1994, Adrio *et al.* 1995, Santopietro *et al.* 1995). Briefly, a pre-inoculum of each strain was prepared in a complete medium [Yeast extract, 0.3% (w/v); peptone, 0.5% (w/v); malt extract, 0.3% (w/v); glucose, 1.0% (w/v)], and incubated at 20ºC and 150 rpm for 24 h. A portion of that culture was then used to inoculate ($DO_{600\ nm}$= 0.2) 100 ml of new liquid culture medium. After 24 h, cells were collected by centrifugation, washed with sterile water and quantified on the hemocytometer. A pellet containing $4x10^8$ cells was

resuspended in 8 mL of a pre-treatment solution (0.1 M β-mercaptoethanol, 0.01 M EDTA, 0.1 M Tris-HCl, pH 7.4) and incubated for 30 minutes at 25°C, followed by a washing step and gentle centrifugation. The pellet was subsequently resuspended in an enzyme solution containing 10 mg/ml of lysing enzymes from *Trichoderma harzianum* from Sigma prepared with 1.2 M sorbitol in sodium citrate buffer. Incubation was done in a 50 ml Erlenmeyer at 25°C and 30 rpm. Because the incubation time is strain-specific, periodic microscopic observations were done and the cells were only collected when at least 90% had formed protoplasts. The protoplasts were washed with the osmotic stabilizer solution (1.2 M sorbitol) and then added to 1 ml of low melting point agarose [1.3% (w/v)] also prepared with the osmotic stabilizer solution. Using Bio-Rad plug molds, protoplast plugs are prepared and subsequently treated with Proteinase K enzyme solution (1 mg/ml Proteinase K from Sigma; 0.01 M Tris-HCl; 1% SDS; 0.5 M EDTA) for 24 h at 50°C. The plugs were then washed with 0.05 M EDTA and stored at 4ºC until used. For strains that did not form protoplasts conveniently an extra step was introduced, consisting in the addition of $MnCl_2$ to the liquid culture medium 12 h after inoculation (to a final concentration of 4.0 mM), which resulted in a weakening of the cell wall (Pera *et al.* 1999). The preparation of protoplasts in sufficient quantity to obtain distinguishable bands without degraded genetic material represents one limiting step of this technique. *Saccharomyces cerevisiae* CBS 8803 was used as a standard for the PFGE-CHEF assays, and its plugs were prepared with the same protocol but the incubation time with the enzyme solution was only 20 minutes. Two different sets of conditions were used for the PFGE-CHEF assays. First a shorter preliminary assay, in order to access the absence/presence of degraded genetic material. If good quality plugs were obtained, another PFGE run would be done under adequate conditions for the separation of the *X. dendrorhous* chromosomes (Table 2). All PFGE-CHEF assays were performed in a CHEF-DR III system from Bio-Rad.

Table 2 – Summary of the PFGE running parameters used.

Parameters	Preliminary assay	Final assay
Agarose	1% Mega Base Agarose prepared in 0.5X TBE (45 mM Tris-borate: 1 mM EDTA, pH 8.0)	0.9% Mega Base Agarose prepared in 0.5X TBE (45 mM Tris-borate: 1 mM EDTA, pH 8.0)
Buffer	0.5X TBE	0.5X TBE
Temperature (ºC)	14	14
Voltage (V/cm)	6	2.7
Pulse (s)	50-90	200-1100
Run time (h)	24	72
Angle(º)	120	120

The gel was stained with an ethidium bromide solution (0.5 µg/ml) for 30 min and then unstained for 30 min in sterile water. Images of the gel were collected with a GelDoc (Bio-Rad) image acquisition system and analyzed in Gelcompar v4.1 from Applied Maths.

3. Results and discussion

3.1. *Xanthophyllomyces* multilocus phylogeny.

With the aim of assessing the phylogenetic relationships between the different *Xanthophyllomyces* lineages a MLSA approach was devised using seven different genomic regions encoding proteins involved in several metabolic processes (Table S2). The biotechnological interest in the production of astaxanthin has driven the sequencing and characterization of the genes associated with its biosynthesis. Three of those genes were chosen for MLSA due to the fact that each of them had been sequenced for more than one strain, enabling the selection of regions comprising a larger number of single nucleotide polymorphisms (SNP). Unfortunately for genes not related to astaxanthin biosynthesis only one sequence was usually available therefore, the four other genes were chosen for the multi-locus scheme based on the information of one single sequence each. The expanded collection of isolates studied made it necessary to construct several sets of primers for four of the seven genes (Table S2). Previously, it was determined that *CRTI* and *CRTS* genes are not in the same chromosome (Niklitschek *et al.* 2008) but regarding the rest of the 5 genes used for the MLSA no such information is available. However the genes chosen codify proteins involved in very distinct roles in the yeast metabolism, lowering the probability of being genetically linked. Moreover, unpublished data regarding the genome draft of *X. dendrorhous* type strain (CBS 7918^T) indicated that all seven genes are present in separate contigs, being the smallest ~3.0 kb long (personal communication by Dr. Diego Libkind). Although this does not prove that they are genetically unlinked, it points in favor of this hypothesis.

Each of the seven genes was amplified for the collection of 40 strains and the resulting alignments were used to construct MP (Figure 4c) and ML (Figure S1) based phylogenies. The individual phylogenies yielded similar results in both methods and were found to be generally concordant. The multilocus alignment of the seven gene fragments was used to construct a phylogeny using MP (Figure 4b) and ML (Figure S2). The concatenated dataset increased the resolution of the analysis in comparison to the individual ones. On every individual phylogeny a total of six major clades were identified whatever the inference method used and the same number of clades was resolved in the multilocus phylogeny (Figure 4b). Before this study, a phylogenetic analysis using the ITS region resolved *Xanthophyllomyces* into 3 major clades/lineages (Libkind *et al.* 2007) that in the multilocus phylogeny (Figure 4b) would partially correspond to clade A, C and D. The phylogeny also had a more divergent branch corresponding to strain CBS 5905 that was phylogenetically isolated from the other lineages (Libkind *et al.* 2007).

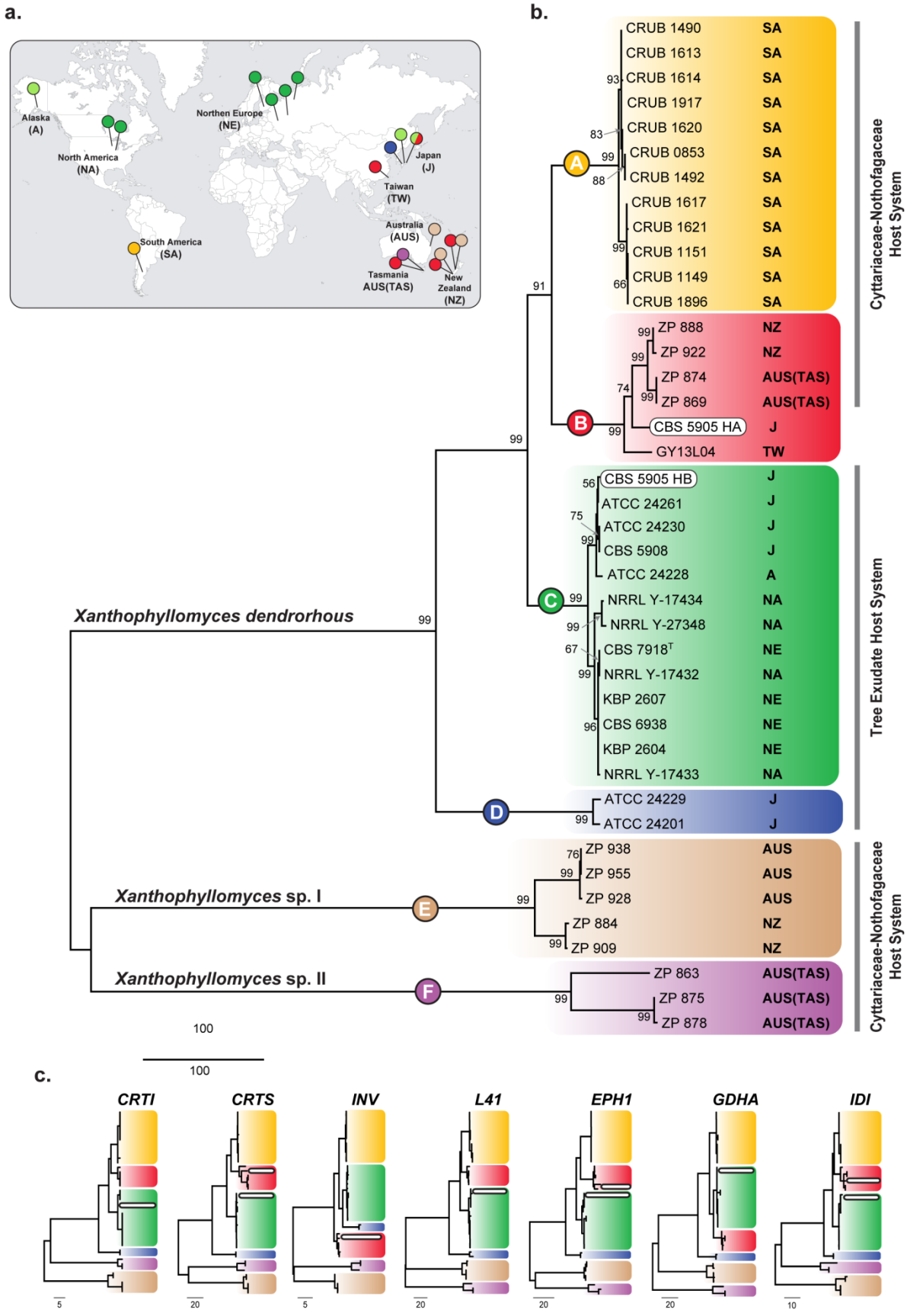

Figure 4 - *Xanthophyllomyces* multilocus phylogeny and distribution.

Figure 4 - *Xanthophyllomyces* multilocus phylogeny and distribution. (a) World map pinpointing the geographical origins of the strains belonging to each of the main phylogenetic clades presented in the multilocus phylogeny. Colored pins correspond to the phylogenetic clade to which the isolate belongs. **(b)** MP multilocus phylogeny identifying six major clades shown in distinct colors and represented with the letters A - F. Next to each strain the geographical origin is indicated and also the host system from which it was isolated. Bootstraps values from 1000 replicates are shown next to the branches. There were a total of 3187 positions in the final dataset (1287 informative positions). The scale is in the units of the number of changes over the whole sequence. Strain CBS 5905 presents two different haplotypes (HA and HB) that are highlighted by two white rounded boxes. **(c)** Individual MP phylogenies of the seven genes used in MLSA are depicted. Clades are presented in the same colors has shown above. Each phylogeny is identified by the abbreviated gene names. Full information regarding the genes is present in Table S2.

In the multilocus phylogeny the clades A, C and D include not only the strains used in the study by Libkind *et al*. (Libkind *et al.* 2007), but also additional isolates comprising a more diverse sampling both in geography and host trees. From the six resolved clades, three represent novel *Xanthophyllomyces* clades (B, E and F), comprising strains from Australasia and also including one strain (GY13L04) from Taiwan. Within the cluster formed by clades A, B, C and D small differences in the topology exist between the individual gene phylogenies, coherent with what is expected under the Phylogenetic Species Recognition model proposed by Taylor and co-authors (Taylor *et al.* 2000). In this case, the species cluster of *X. dendrorhous* encompasses four different lineages (Figure 4b) where clade B represents a novel clade. Inside each of the four clades of *X. dendrorhous*, subgroups of strains are also resolved and well supported by high bootstrap values. For example, in clade A two subgroups exist and in clade C, whereas all the strains from Japan group together with the strain from Alaska, the rest of the strains cluster separately (strains from NA and NE) (Figure 4a and b).The clades E and F always appear in a most divergent position in relation to the other 4 clades. The fact that the concatenated phylogeny, which has a greater resolution power than any of the individual ones, is not able to present one of the referred clades as being the most divergent is probably related to the fact that an appropriate outgroup does not exist for this multilocus phylogeny.

Regarding strain CBS 5905 it was found that three of the seven gene fragments sequenced were heterozygotic (Figure 4c; genes *CRTS*, *EPH1* and *IDI*). The haplotypes were separated by cloning and then sequenced. It was found that each heterozygotic gene presented two different haplotypes that always grouped in clades B and C in the respective gene phylogeny. The other 4 genes were found to be homozygotic for strain CBS 5905 and presented haplotypes that grouped either in clade B (*INV* gene) or in clade C (*CRTI*, *L41* and *GDHA* genes) (Figure 4c), which further confirms the results obtained for the heterozygotic genes. These results seem to indicate that this strain is a result of a cross between members of two lineages represented by clades B and C, which originate respectively from

Australasia/Taiwan and Japan. This suggests that a contact area between lineages B and C might have existed allowing gene flow between them.

Concerning the host system, it is possible to observe that the strains from the *Cyttaria* - *Nothofagus* host system are dispersed along the phylogeny in opposition with the isolates collected from tree exudates that are confined to clades C and D (Figure 4b).

3.2. *Xanthophyllomyces* novel species.

The recently found Australasian populations of *Xanthophyllomyces* encompass the largest genetic diversity ever observed in this genus. The Australasian genotypes are divided into three distinct lineages, one (clade B) somewhat similar to *X. dendrorhous* found in Southern America and two others (clades E and F) that are distinct from any other population ever reported (Figure 4b). Although the isolates from the latter lineages present the same life cycle as *X. dendrorhous* (Figure S3) and also produce astaxanthin (personal communication by Dr. Diego Libkind), they have a considerable genetic divergence in comparison to all known populations of *X. dendrorhous* (Figure 4b and Figure 5). The relevant divergence existing between clade E and F and the remaining clades suggests that these novel Australasian clades might in fact represent two new distinct species within the genus *Xanthophyllomyces* (Figure 4b). The separation of *Xanthophyllomyces* into three species is not only supported by the multilocus phylogeny but also by the ITS phylogeny (Figure 5). The ITS phylogeny presented in Figure 5 includes the genera *Xanthophyllomyces* and *Cystofilobasidium* and allowed comparing the genetic divergence of the different species within both genera. The degree of genetic divergence observed in the ITS sequences within *X. dendrorhous* species is somewhat similar to that observed within *Cy. capitatum* or *Cy. macerans* (Figure 5). However, the number of different genotypes included in *X. dendrorhous* appears to be higher than that found in each *Cystofilobasidium* species, which seem to be much more homogenous. The two putative novel species of *Xanthophyllomyces* also exhibit genetic diversity. Within clades, E and F, two different lineages appear to exist both in ITS (Figure 5) and in the multilocus phylogeny (Figure 4b). The study of more strains from these two new species might elucidate the biogeographic reasons that are accountable for this apparent lineage differentiation within each species. On clade E, strains ZP 938 and ZP 928 were collected from leafs of *N. mooreii* in Australia, while strains ZP 909 and ZP 884 were collected from fruiting bodies of *C. nigra* on *N. menziesii* in New Zealand (Table 1). Therefore geography and host are different and might justify the observed genetic diversity of *Xanthophyllomyces* sp. I.

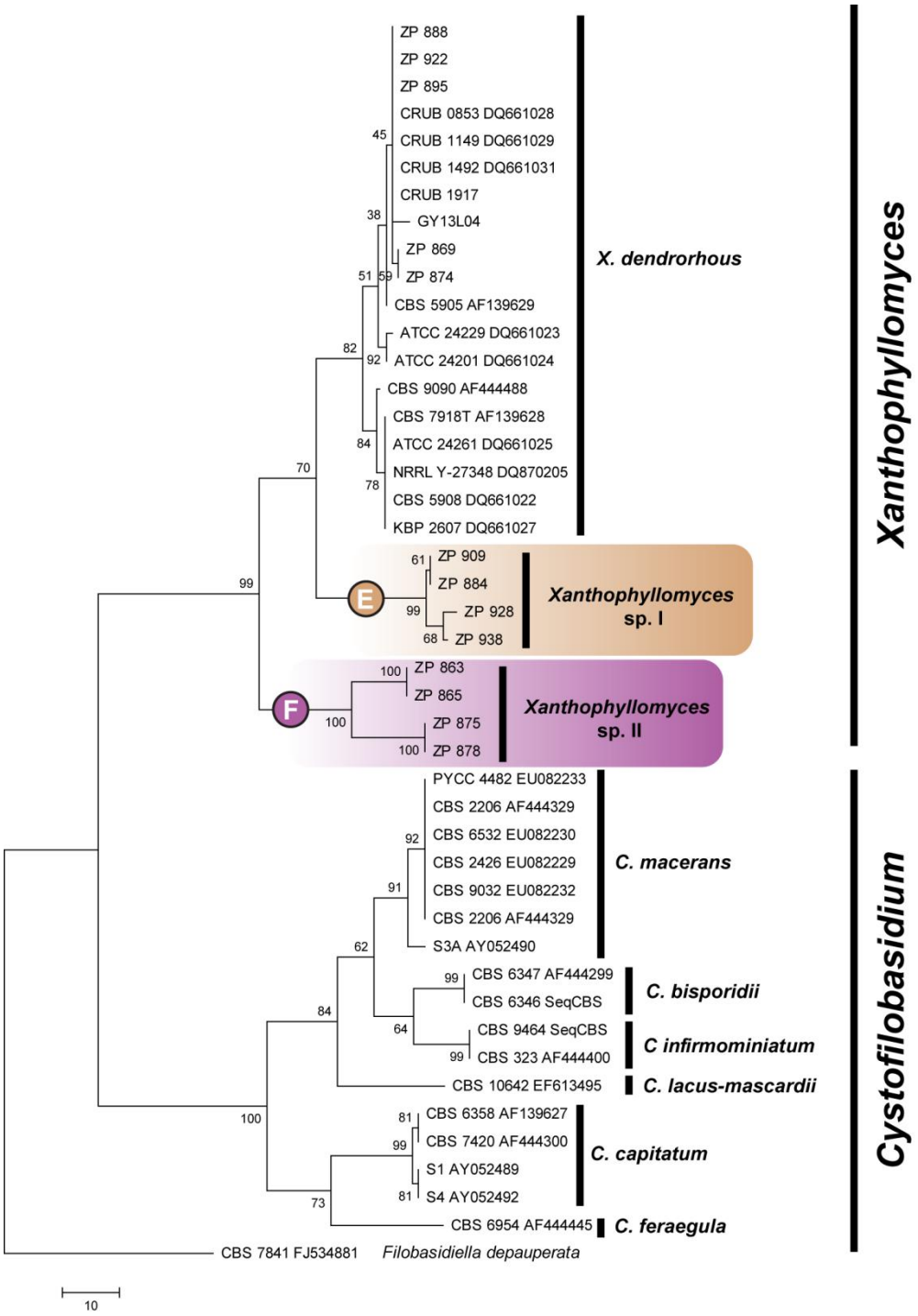

Figure 5 - ITS phylogeny of *Xanthophyllomyces* and *Cystofilobasidium*. The phylogeny depicts the relationships of the two new species of *Xanthophyllomyces* with *X. dendrorhous* and *Cystofilobasidium* species. A total of 45 sequences of representative strains of different species from both genera *Xanthophyllomyces* and *Cystofilobasidium* were included. The phylogeny was constructed using the MP method and the bootstrap values, from 1000 replicates, are presented next to the branches. All positions containing gaps and missing data were eliminated. There were a total of 369 positions in the final dataset. Strains that have ITS sequences already available in GenBank have the corresponding access number in front of the strain name. Clades E and F are identified in the phylogeny and correspond to the homonymous clades in the multilocus phylogeny. *Filobasidiella depauperata* was used as outgroup.

On the other hand in clade F all *Xanthophyllomyces* sp. II isolates share the same geographical origin and host system (*C. gunnii* on *N. cunninghamii* of Tasmania) but present clearly distinguishable genotypes. Although studied strains belonging to this new species are few, it seems that the factors influencing their population structure may differ from those influencing the population structure of *X. dendrorhous*. While *X. dendrorhous* has a vast geographic distribution (Figure 4c, orange, green, red and blue pins), these new putative species appear to be endemic to Australasia. Whereas *X. dendrorhous* apparently adapted to different host trees as new geographic areas where colonized and therefore formed different lineages, the Australasian species remained in the same host (clade E and F) and in the same geographical area (Clade F). Both, sympatric speciation or a phenomenon of secondary contact may be related to the formation or present existence of the different species encountered in Tasmanian samples (*X. dendrorhous* and *Xanthophyllomyces* sp. II). A more extensive study of isolates of the new species and of the Australasian habitats is needed in order to draw further insights on their biogeography and evolutionary history.

3.3. Exploring the population structure of *X. dendrorhous*.

The multilocus sequence analysis provided new information regarding the phylogenetic relationships of the different lineages that had previously been identified in *X. dendrorhous* by sequencing the ITS region. In light of the data provided by the MLSA, *X. dendrorhous* seems to be composed by four different lineages represented in the multilocus phylogeny by four distinct clades (A, B, C and D in Figure 4b). The assessment of evolutionary relationships using the phylogenetic tree model is done under the assumption that the evolution process is dominated by two types of events, mutation and speciation (Huson and Scornavacca 2011). Under more complex models of evolution that might involve events as complex as horizontal gene transfer, hybridization and recombination, phylogenetic trees may not provide an adequate representation of the phylogenetic history of the organism. Phylogenetic networks comprise more realistic models which are able to clearly represent events such as recombination (Huson 1998, Huson and Bryant 2006). The utilization of phylogenetic networks also allows determining if the dataset under analysis has a strong phylogenetic signal or not. In order to assess the strength of the phylogenetic signal (Figure 6) and if reticulate events as recombination played a relevant role in the evolution of this species (Figure S4), phylogenetic networks were constructed using the multilocus dataset. Two types of networks were tested: (i) Splits Networks (SN), which constitutes a combinatorial generalization of phylogenetic trees therefore enabling the visualization of incompatibilities within and between datasets (Huson and Scornavacca 2011) and (ii) Reticulate networks (RN) that provide a explicit representation of the evolutionary history in

the presence of reticulate events such as recombination (Huson and Bryant 2006, Huson and Scornavacca 2011).

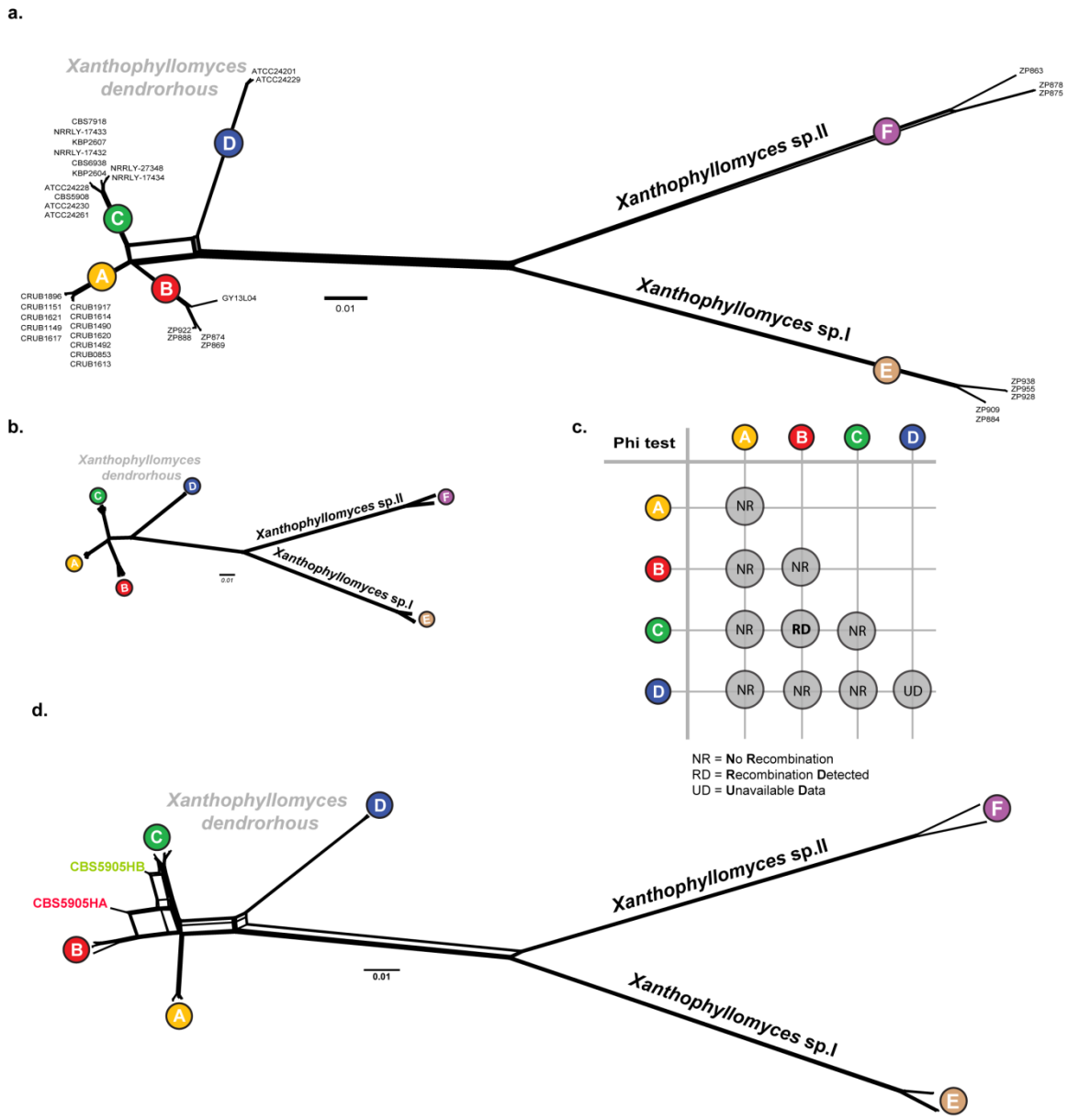

Figure 6 - Splits network of *Xanthophyllomyces*. (a) Neighbor-net network constructed with the multilocus dataset, depicting all *Xanthophyllomyces* strains except CBS 5905 (39 strains). The main edges of the network correspond to the clade of the multilocus phylogeny of Figure 4b. **(b)** Neighbor joining tree constructed with the multilocus dataset, depicting all *Xanthophyllomyces* strains except CBS 5905 (39 strains)**. (c)** Phi-test results between and within populations. Analysis was performed using the multilocus dataset with all *X. dendrorhous* strains **(d)** Neighbor-net network depicting all 40 strains of *Xanthophyllomyces.*

In order to determine if the phylogenetic signals obtained from the seven different genomic regions were strong and coherent, individual alignments of each genetic marker were introduce and concatenated in SplitsTree4 to build Neighbor-net networks (a category of SN) that implicitly represents such differences. Three different multilocus sub datasets were

tested in network analysis: (i) one comprising all *Xanthophyllomyces* species (40 strains) (Figure 6d); (ii) another comprising all *Xanthophyllomyces* strains except CBS 5905 (39 strains) (Figure 6a); (iii) and finally one compose only by the strains belonging to *X. dendrorhous* but without CBS 5905 (31 strains) (Figure S4). The Neighbor-net network on Figure 6a presents a group of incompatible splits regarding the relation of clade D with the other 3 clades that compose *X. dendrorhous*. This, and the small topological differences observed when comparing the individual phylogenetic trees (Figures 4c and 5) suggests more clearly that the limit of the species *X. dendrorhous* lies in the transition from concordance among branches to incongruity among branches observed within clades A to D (Taylor *et al.* 2000). On Figure 6d the network differs slightly from the tree-like form (Figure 6b) because it includes the two haplotypes of the CBS 5905 heterozygotic genes. The network clearly shows that alleles in CBS 5905 most likely derived from representatives of clade B and from a subgroup of Japanese strains in clade C. The remaining strains present a position in the network similar to that in the phylogenetic tree. A RN was constructed for the *X. dendrorhous* multilocus dataset but did not yield any additional information regarding the occurrence of reticulate events since the obtained topology overlapped that of the SN (Figure S4). Because all SN have generally a tree-like form (Figure 6a, 6b and 6d), it is possible to assess that reticulate events were not dominant in the evolution of *X. dendrorhous* and also that the phylogenetic signal in the multilocus dataset is strong (Wagele and Mayer 2007).

The pairwise homoplasy index (Phi) test (Bruen *et al.* 2006) was applied in order to determine if signs of recombination within and between the four distinct populations in *X. dendrorhous* could be detected. This test detects refined incompatibility in sequence data (Bruen *et al.* 2006) assuming the infinite sites model of evolution (Nei 1973) in which the detection of incompatibility for a pair of sites indicates recombination. The test was performed in a two by two comparison as represented in Figure 6c using the multilocus dataset with all the strains from *X. dendrorhous*. Since at least three sequences are necessary to perform the test, it was not possible to perform Phi-test within clade D. Recombination was detected only between clades B and C. When the same test was performed using the multilocus dataset composed by all *X. dendrorhous* strains with the exception of CBS 5905 strain, no statistically significant values were retrieved from the analysis, therefore no recombination was detected. Recombination was already demonstrated in the laboratory between strains CBS 6932 and CBS 5908 (both belonging to clade C) by crossing auxotrophic mutants and analyzing of the progeny (Kucsera *et al.* 1998). However no recombination was detected by Phi-test within that clade. This might imply that although the strains are able to cross and undergo meiosis in the laboratory, in natural populations of *X. dendrorhous* inbreeding appears to be dominant. The association to

very specific ecological niches and the peculiar mating process between parental cell and bud might also be contributing to the well structured populations detected in *X. dendrorhous*.

The analysis of the multilocus alignment of *X. dendrorhous* was also performed using the Structure software (Figure 7). Given a number of populations (K) and assuming Hardy–Weinberg and linkage equilibrium within populations, Structure estimates allele frequencies in each cluster and population memberships for every individual strain. It is able to detect the underlying genetic population structure among a set of individuals genotyped at multiple markers and also of computing the proportion of the genetic information of an individual originating from each inferred population (quantitative clustering method). Therefore, in order to determine the ancestry of the *X. dendrorhous* strains the respective multilocus dataset was tested (Figure 7). The graphic output of Structure clearly shows the existence of four distinct populations that coincide with the clades defined by the phylogenetic analysis of the same dataset (Figure 7). Strain CBS 5905 shows a mixed ancestry with more than half belonging to the population identified in green and corresponding to clade C.

Figure 7 - Ancestry of *X. dendrorhous* estimated with Structure software. Each column represents a different strain and the different colors represent distinct populations inferred by Structure. The vertical axis represents the amount of ancestry that each strains has from each of the inferred populations. The phylogenetic clades where each strain appears in the phylogenetic tree are identified in the upper part. The graph corresponds to the highest probability run for the number of populations inferred by Structure.

The complex ancestry of CBS 5905 detected in Structure is congruent with the two disjoint haplotypes depicted in the phylogenetic tree (Figure 4) in clade B (haplotype HA) and in clade C (haplotype HB) and on the splits network (Figure 6d). Therefore this confirms that this strain is indeed the result of a cross between members of populations originating from Australasia and Japan.

3.4. *X. dendrorhous* karyotype.

The use of Pulsed-field gel electrophoresis (PFGE) coupled with contour-clamped homogeneous electric fields (CHEF) apparatus allows the determination of karyotypes as well as chromosome-length polymorphisms (CLP), thus facilitating evolutionary and population studies (Lin *et al.* 2007). Aiming to elucidate the chromosomal composition of strains from different lineages of *X. dendrorhous* and to assess if the population structure observed by phylogenetic methods had a correlation with the chromosomal profile of the strains, the PFGE-CHEF technique was employed. Approximately 25 strains were tested by this technique although only 11 yielded satisfactory results due to the fact that each strain presented a unique response to the experimental conditions used (Figure 8). The bands observed in the PFGE gels are considered to represent chromosomes. In this study all the chromosomes were identified by visual scrutiny of the gels and by the analysis of the correspondent intensity graphics. On Figure 8 the image corresponding to the lane of each strain is presented with indication of the observed bands. All the strains were tested in duplicate and the images presented in Figure 8 were all normalized with GelCompar software using the *S. cerevisiae* standards as reference for the normalization. Observing the first column of Figure 8, were the *S. cerevisiae* standards are represented, it is possible to see that there exists reproducibility among all performed assays (referred to as gel number in Figure 8). Observing the PFGE results obtained for the *X. dendrorhous* strains it can be observed that no two strains present the same chromosomal profile. Nevertheless some chromosomal bands of similar size were observed in strains CRUB 1620, CRUB 1614 and CRUB 1149 (Figure 8). It is possible that some of the differences observed in the chromosome number (CN) of some strains might be due to the limited ability that PFGE technique has for the resolution of co-migrating chromosomes with equal size, a fact that may lead to the underestimation of CN. The obtained karyotypes are in the size range of 0.8 Mb to 3.5 Mb, which corroborates the results from previous studies [i.e. chromosomes sizes ranging from 0.83 Mb to 3.1 Mb (Nagy *et al.* 1994, Adrio *et al.* 1995, Cifuentes *et al.* 1997, Nagy *et al.* 1997)]. Furthermore, in these studies the CN of *X. dendrorhous* was determined to vary from seven to 17 but, because the different studies used different experimental conditions and different assumptions regarding the possibility of co-migration of the bands, no direct comparison of the CN is possible (Nagy *et al.* 1994, Adrio *et al.* 1995, Cifuentes *et al.* 1997, Nagy *et al.* 1997). In the current study, CN varies from six to 11, but most of the strains have between six to nine chromosomal bands, being CBS 5905 the only strain with 11 chromosomes.

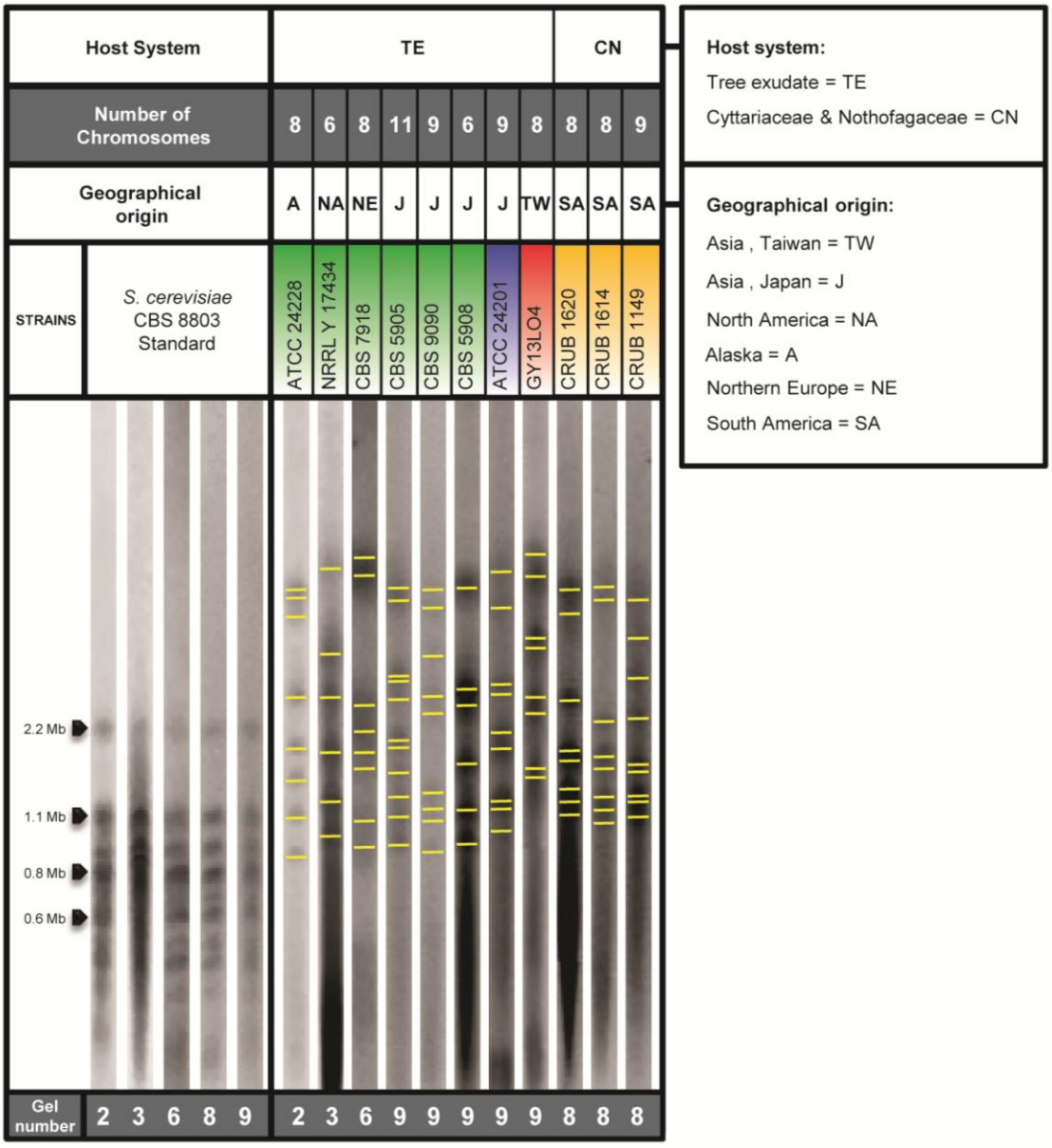

Figure 8 - *X. dendrorhous* karyotypes. Chromosomal patterns of eleven strains and information regarding host system, geographic localization, and CN are also presented. The color associated to the strain number identifies the population (A, B, C or D) to which each strain was assigned.

Among the group of 11 strains analyzed by PFGE, a strain that is a copy of CBS 5905 (named CBS 9090) was included. It is presently known that these two strains have differences in the ITS sequence (Figure 5). It is assumed that the repeated transfers over decades, a procedure commonly used in culture collections, might have influenced the accumulation of these differences (Libkind *et al.* 2007). The chromosomal pattern of these two strains was also observed to be different in both the number of chromosomes and in their sizes (Figure 8). From the preliminary results presented in Figure 8, there seems to be

no direct relation between the chromosomal profiles and the population structure of *X. dendrorhous*.

Although *X. dendrorhous* karyotypes show a considerable variability in chromosomal patterns, this species does not represent a unique situation among Fungi, since many other species present CLP (Zolan 1995, Pérez-Ortín *et al.* 2002, Giraud *et al.* 2008).The existence of intraspecific CLP can arise mainly by rearrangements of linked groups of genes or gains and losses of non coding sequences (Zolan 1995, Poláková *et al.* 2009). It was shown for the basidiomycetous *Coprinopsis cinerea* (*Coprinus cinereus)* that two strains differing in the length of most of their chromosomes are interfertile, and that new CLP appear after meiosis (Zolan 1995). This means that the progeny of such cross presents a variety of karyotypes, indicating that the changes in length observed in the chromosomes represent gains and losses of nonessential sequences (Zolan 1995). A similar situation has been observed when crossing strains of *X. dendrorhous.* The progeny showed not only composed karyotypes from both parental strains, but also novel CLP what could indicate that the genomic regions altered in the karyotypes of the progeny might not be essential (Kucsera *et al.* 1998). Nevertheless in order to assess the importance of the CLP observed in *X. dendrorhous*, an understanding of how the genome is arranged is necessary (Zolan 1995).

3.5. *X. dendrorhous* population structure, host system and geographical distribution.

In the present study *X. dendrorhous* was found to have a well defined population structure with four distinct populations. Considering both host and geographic origin, it seems that the populations that colonize the same host tree are more similar among themselves than those sharing the same geographical origin but a different host tree (Figure 4). The ITS phylogenies of yeast and host trees were compared by Libkind *et al.* in 2007 and a correlation between both was proposed. In light of the broader sampling of *Xanthophyllomyces* isolates presently available, this analysis was updated with greater detail. In Figure 9a the ITS phylogeny of the host trees shows that all the strains isolated from *Nothofagus* are present either in clade A or clade B of the multilocus phylogeny (Figure 9b). Moreover, strains isolated from *Betula* grouped in clade C and those isolated from *Cornus* grouped in clade D (of the multilocus phylogeny) (Figure 9). Two strains, CBS 5905 and GY13L04, present a unique situation regarding the yeast-host association. Strain CBS 5905 was isolated from an exudate of *Fagus crenata* and due to its mixed ancestry presents two different haplotypes, one in clade B and another in clade C of the multilocus phylogeny (Figure 9b). Strain GY13L04 was collected in Taiwan from *Stauntonia purpurea* (Figure 9a). Unfortunately the lack of more isolates from the same host tree or same geographic location

does not allow detailed ecological inferences. Excluding these two mentioned strains, if the multilocus phylogeny of *X. dendrorhous* is analyzed in light of the phylogeny of the host trees, a division between clades according to host becomes clear (Figure 9b). The strains isolated from trees belonging to the Fagales, comprising *Nothofagus* and *Betula* are phylogenetically closer between themselves than with the strains collected from Cornales (a different order). Therefore the population structure of *X. dendrorhous* shows coherence with the phylogenetic split existing between Fagales and Cornales, and also with the subsequent phylogenetic split in the Fagales between the two genera *Nothofagus* and *Betula* (Figure 9).

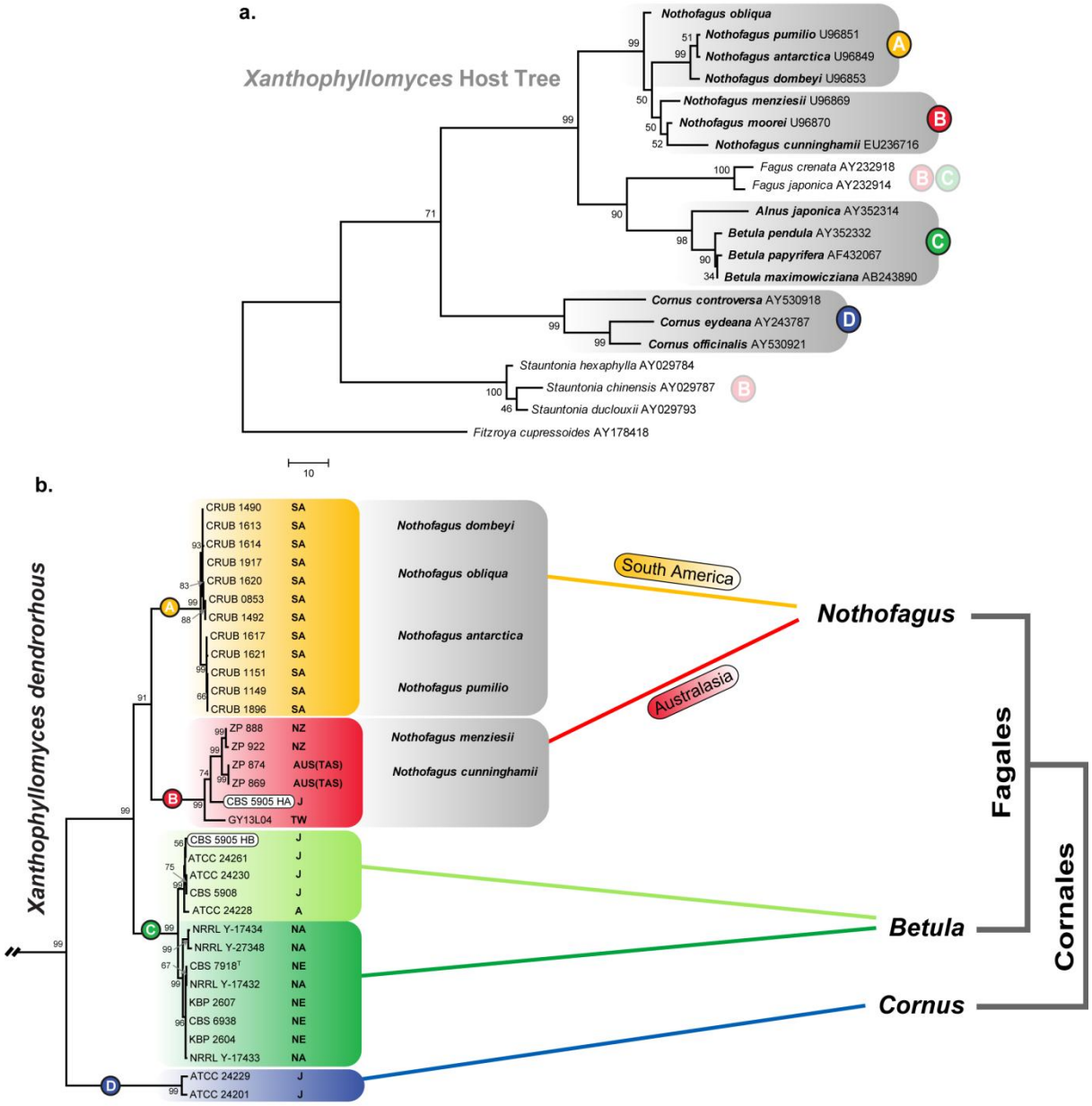

Figure 9 - *X. dendrorhous* population structure and host tree phylogeny. (a) ITS phylogeny (performed by MP method) of representatives of the host trees. GenBank accession numbers are indicated next to each species name and the bootstrap values are presented next to the branches. The clades of the multilocus phylogeny of the yeast are also depicted. **(b)** Combined representation of the yeast multilocus phylogeny and the host phylogeny.

Isolates from Japan, sharing same geography but distinct hosts, grouped in distinct clades and isolates from North America and Northern Europe, with distinct geography but same host end up grouping on the same clade (Figure 9b).

On the other hand, additional evolutionary factors, such as geographic origin, appear to shape *X. dendrorhous* population structure. Within the strains isolated from *Nothofagus*, those originating from South America grouped together (clade A) and are separated from those originating from Australasia (clade B) (Figure 9b). In clade C, an existing subdivision also appears to be related with geographic origin. The isolates from Japan always grouped together with the isolate from Alaska, while those from NA and NE clustered separately but in the same clade, as can be observed in clade C represented by two different shades of green (Figure 9b).

X. dendrorhous was found to be present not only in the Northern Hemisphere but also in the Southern one, namely Patagonia and now more recently in New Zealand and Tasmania (Figure 9). The existence of *X. dendrorhous* in both Australasia and South America might suggest that just as Nothofagus trees were disseminated though the southern part of Gondwana so was *X. dendrorhous*. This might therefore present a possible explanation for the presence of *X. dendrorhous* in both geographic locations (grey arrow on Figure 10). Nevertheless the fact that populations from clades A and B are phylogenetic closer among themselves seems contradicting to the vicariance scenario implied by the fragmentation of Gondwana. The biogeographic history of *Nothofagus* and of its obligate parasite, *Cyttaria* are considered to be a reflection of the break up of Gondwana but also of transoceanic long dispersal events (Peterson *et al.* 2010). It is therefore possible that the extant distribution of *X. dendrorhous* in the Southern Hemisphere may be related with the same factors affecting *Nothofagus* biogeography. In order to test the different hypotheses regarding the events that might have led to the present *X. dendrorhous* distribution in the Southern Hemisphere (vicariance or dispersal), assessment of parameters such as the divergence time between clades A and B and comparison of those dates with the break up process of Gondwana would be necessary (Crisp *et al.* 2011). Unfortunately such analyses in yeasts are challenging because of the absence of fossil records for the correct calibration of phylogenies, and thus permitting only indirect methods of calibration (Crisp *et al.* 2011).

The fact that similar genotypes of *X. dendrorhous* are present in Australia, New Zealand and Asia (Taiwan and Japan) and the mixed ancestry of CBS 5905 may suggest that an area of contact might have existed between Australasia and Asia (Figure 10, arrow 1). The biogeographic history of Southeast Asia and the West Pacific (Turner *et al.* 2001) is considered to comprise two general patterns: (i) a pattern of Southeast Asian elements, perhaps of Laurasian origin, expanding into Australian areas, and (ii) a reverse pattern for Australian elements, perhaps of Gondwanan origin. Both patterns would make the mentioned

area a possible contact zone and hypothetical route for the colonization of Asia by *X. dendrorhous*. The existence in New Guinea of Fagaceae and Nothofagaceae and the fact that Australia and New Guinea were periodically connected by the Torres Strait land-bridge (Voris 2000) are facts favoring this possible route of migration towards the Northern Hemisphere (Figure 10, arrow 1). The overlapping area of distribution of Fagaceae and Nothofagaceae could have allowed the occurrence of host shift and the association of *X. dendrorhous* with other trees like *Fagus*, *Betula* and *Cornus*. Once associated with these new hosts, the pattern of geographical distribution of the yeast might have been driven by the distribution of these new hosts (Figure 10).

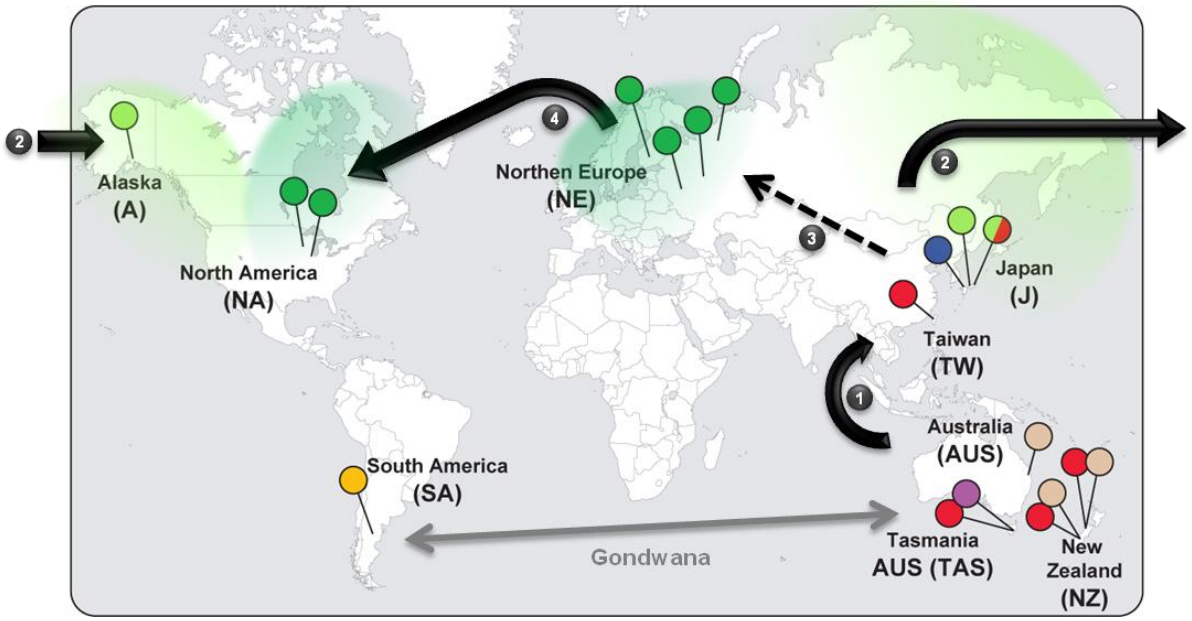

Figure 10 - *Xanthophyllomyces* extant distribution and hypothetical routes for the colonization of the Northern Hemisphere. World map pinpointing the geographical origins of the strains belonging to each of the main phylogenetic clades (color codes are the same as in figure 9). Black arrows illustrate the hypothetical routes of colonization of the Northern Hemisphere. Areas in two different shades of green represent two distinct biogeographic areas. Grey arrow represents the possible relationship between the breakage of Gondwana and the distribution of *Xanthophyllomyces* in the Southern Hemisphere.

A phylogeographic study done in the Betulaceae (Chen *et al.* 1999) considered that during the Cretaceous and early Tertiary, migration between Eurasia and North America was possible via the North Atlantic (Figure 10, arrow 4) and Bering land-bridge (Figure 10, arrows 2) (Milne 2006) and today the results of such intercontinental migration can be observed among several genera of Betulaceae (Chen *et al.* 1999). The colonization of the different regions of the Northern Hemisphere by X. dendrorhous might be related with the events affecting the host trees biogeography. Pollen records show that the Northern Hemisphere was divided into two biogeographic regions: (i) Western North America plus most of Asia

(Figure 10, light green area) and (ii) Eastern North America plus Europe (Figure 10, dark green area) (Milne 2006). Possible evidence of these two biogeographic areas appear to exist within clade C, were a separation between strains of Asia from the Occidental ones is noticeable (Figure 9, clade C, division represented by two shades of green). The knowledge of the historical evolution of the tree hosts enables the formulation of hypotheses regarding the history of colonization of the Northern Hemisphere by *X. dendrorhous*. Nevertheless, one must consider that a set of results can be consistent with a set of alternative explanations (Crisp *et al.* 2011). Therefore in order to fully explore the present hypothesis, a large sampling would be necessary from the vast area between Australasia and Asia and further study would be necessary in order to test the different scenarios that could be responsible for the present distribution of *X. dendrorhous*.

4. Final considerations and future perspectives

X. dendrorhous is composed by four different populations, presenting a well defined population structure that suggests that natural populations of *X. dendrorhous* are mostly clonal and in nature inbreeding supersedes outcrossing. Strain CBS 5905 shows a mixed ancestry and may represent an event of contact between two distinct populations of *X. dendrorhous*. The different populations of *X. dendrorhous* have different ecological associations that appear to have diverged first according to host tree and secondly by geographic origin. The association between yeast and the host system *Cyttaria–Nothofagus* appears to be older than the association of the yeast with the tree exudates host system. Furthermore, the association of *X. dendrorhous* with *Cyttaria* appears to have derived from an earlier association of the yeast with *Nothofagus* trees, due to the consistent isolation of strains from *Nothofagus* leaves. Different karyotypes were presented by each of the tested *X. dendrorhous* strains and no correlation with the revealed population structure was observed. The existence of *X. dendrorhous* in both Australasia and South America may indicate that the extant distribution of the species in the Southern Hemisphere is a relic of the Gondwana break up process. The results obtained in this work suggest that Southeast Asia and Australasia are a *Xanthophyllomyces* diversity hotspot from which different lineages have radiated. Moreover, Australasia harbors two putative new species that appear to be endemic to the area, representing this study the first record of the existence of more than one species in the *Xanthophyllomyces* genus. The outstanding diversity found in Australasia, with three distinct Xanthophyllomyces species and the fact that the Northern Hemisphere isolates show greater genetic homogeneity suggests that Australasia might represent the place of origin of the *Xanthophyllomyces* genus. The extant distribution of the *X. dendrorhous* appears to be coherent with the biogeographic histories of the trees colonized by the yeast. Although in need of testing, a hypothetical route of colonization of the Northern Hemisphere by *X. dendrorhous* was drawn from the biogeographic processes undergone by the trees colonized and by the extant genetic distribution observed in the population structure of the yeast.

The sampling of areas where known hosts exist, like New Guinea, could provide valuable information regarding the existence (or not) of more strains with mixed ancestry that could prove contact and cross between elements of different populations and also test the hypothetical route of colonization of the Northern Hemisphere by *X. dendrorhous*. The availability of the complete genome sequence of *X. dendrorhous* would certainly be a valuable tool in understanding the CLP observed among strains and also for the biotechnological research concerning astaxanthin production.

5. References

Adrio JL, López M, Casqueiro J, Fernández C and Veiga M (1995) Electrophoretic karyotype of the astaxanthin-producing yeast *Phaffia rhodozyma*. *Cur. Genet.* **27**, 447-450.

Akaike H (1987) Factor analysis and AIC. *Psychometrika* **52**, 317-332.

Bruen TC, Philippe H and Bryant D (2006) A Simple and Robust Statistical Test for Detecting the Presence of Recombination. *Genetics* **172**, 2665-2681.

Chen S-H, Su S-Y, Lo C-Z, Chen K-H, Huang T-J, Kuo B-H and Lin C-Y (2009) PALM: A Paralleled and Integrated Framework for Phylogenetic Inference with Automatic Likelihood Model Selectors. *PLoS ONE* **4**, e8116.

Chen Z-D, Manchester SR and Sun H-Y (1999) Phylogeny and evolution of the Betulaceae as inferred from DNA sequences, morphology, and paleobotany. *Am. J. of Bot.* **86**, 1168-1181.

Chen ZD and Li J (2004) Phylogenetics and biogeography of *Alnus* (Betulaceae) inferred from sequences of nuclear ribosomal DNA ITS region. *Int. J. Plant Sci.* **165**, 325-335.

Cifuentes V, Hermosilla G, Martinez C, Leon R, Pincheira G and Jimenez A (1997) Genetics and electrophoretic karyotyping of wild-type and astaxanthin mutant strains of *Phaffia rhodozyma A. Van Leeuw.* **72**, 111-117.

Cook LG and Crisp MD (2005) Not so ancient: the extant crown group of Nothofagus represents a post-Gondwanan radiation. *Proc. Biol. Sci.* **272**, 2535-2544.

Crandall K, Fetzner J, Jara C and Buckup L (2000) On the phylogenetic positioning of the South American freshwater crayfish genera (Decapoda: Parastacidae). *J. Crust. Biol.* **20**, 530 - 540.

Crisp MD, Trewick SA and Cook LG (2011) Hypothesis testing in biogeography. *Trends Ecol. & Evol.* **26**, 66-72.

Evanno G, Regnaut S and Goudet J (2005) Detecting the number of clusters of individuals using the software STRUCTURE: a simulation study. *Mol. Ecol.* **14**, 2611-20.

Excoffier L and Heckel G (2006) Computer programs for population genetics data analysis: a survival guide. *Nat. Rev. Genet.* **7**, 745-758.

Falush D, Stephens M and Pritchard JK (2007) Inference of population structure using multilocus genotype data: dominant markers and null alleles. *Mol. Ecol. Notes* **7**, 574-578.

Fell JW and Blatt GM (1999) Separation of strains of the yeasts *Xanthophyllomyces dendrorhous* and *Phaffia rhodozyma* based on rDNA IGS and ITS sequence analysis. *J. Ind. Microbiol. Biotechnol.* **23**, 677-681.

Fell JW, Scorzetti G, Statzell-Tallman A and Boundy-Mills K (2007) Molecular diversity and intragenomic variability in the yeast genus *Xanthophyllomyces* : the origin of *Phaffia rhodozyma* ? *FEMS Yeast Res.* **7**, 1399-1408.

Felsenstein J (1981) Evolutionary trees from DNA sequences: a maximum likelihood approach. *J. Mol. Evol*. **17**, 368-76.

Frengova GI and Beshkova DM (2009) Carotenoids from *Rhodotorula* and *Phaffia* : yeasts of biotechnological importance. *J. Ind. Microbiol. Biotechnol.* **36**, 163-180.

Giraud T, Refregier G, Le GM, de Vienne DM and Hood ME (2008) Speciation in fungi. *Fungal.Genet. Biol.* **45**, 791-802.

Golubev WI (1995) Perfect state of *Rhodomyces dendrorhous* (*Phaffia rhodozyma*). *Yeast* **11**, 101-110.

Gonçalves P, Valério E, Correia C, de Almeida JMGCF and Sampaio JP (2011) Evidence for Divergent Evolution of Growth Temperature Preference in Sympatric *Saccharomyces* Species. *PLoS ONE* **6**, e20739.

Hall T (1999) BioEdit: a user-friendly biological sequence alignment editor and analysis program for Windows 95/98/NT. *Nucl. Acids. Symp. Ser* **41**, 95 - 98.

Hastings WK (1970) Monte Carlo sampling methods using Markov chains and their applications. *Biometrika* **57**, 97-109.

Hermosilla G, Martinez C, Retamales P, Leon R and Cifuentes V (2003) Genetic determination of ploidy level in *Xanthophyllomyces dendrorhous. A. Van Leeuw.* **84**, 279-287.

Hubisz MJ, Falush D, Stephens M and Pritchard JK (2009) Inferring weak population structure with the assistance of sample group information. *Mol. Ecol. Resour.* **9**, 1322-32.

Huson DH (1998) SplitsTree: analyzing and visualizing evolutionary data. *Bioinformatics* **14**, 68-73.

Huson DH and Bryant D (2006) Application of phylogenetic networks in evolutionary studies. *Mol. Biol. Evol.* **23**, 254-267.

Huson DH and Scornavacca C (2011) A Survey of Combinatorial Methods for Phylogenetic Networks. *Genome Biol. Evol.* **3**, 23-35.

Johnson EA (2003) *Phaffia rhodozyma*: colorful odyssey. *Int. Microbiol.* **6**, 169-174.

Kucsera J, Pfeiffer I and Ferenczy L (1998) Homothallic life cycle in the diploid red yeast *Xanthophyllomyces dendrorhous* (*Phaffia rhodozyma*). *A. Van Leeuw.* **73**, 163-168.

Kurihara H, Koda H, Asami S, Kiso Y and Tanaka T (2002) Contribution of the antioxidative property of astaxanthin to its protective effect on the promotion of cancer metastasis in mice treated with restraint stress. *Life Sci.* **Vol. 70**.

Lee DH, Kim CS and Lee YJ (2010) Astaxanthin protects against MPTP/MPP+-induced mitochondrial dysfunction and ROS production in vivo and in vitro. *Food Chem.Toxicol.***49**,271- 80

Libkind D, Ruffini A, van BM, Alves L and Sampaio JP (2007) Biogeography, host specificity, and molecular phylogeny of the basidiomycetous yeast *Phaffia rhodozyma* and its sexual form, *Xanthophyllomyces dendrorhous Appl. Environ. Microbiol.* **73**, 1120-1125.

Lin CY, Chen YC, Lo HJ, Chen KW and Li SY (2007) Assessment of *Candida glabrata* strain relatedness by pulsed-field gel electrophoresis and multilocus sequence typing. *J. Clin. Microbiol.* **45**, 2452-9.

Manchester SR, Xiang Q-Y and Xiang Q-P (2007) *Curtisia* (Cornales) from the Eocene of Europe and its phytogeographical significance. *Bot. J. Linn. Soc.* **155**, 127-134.

Mathiasen P and Premoli AC (2010) Out in the cold: genetic variation of *Nothofagus pumilio* (Nothofagaceae) provides evidence for latitudinally distinct evolutionary histories in austral South America. *Mol. Ecol.* **19**, 371-385.

Medwid RD (1998) *Phaffia rhodozyma* is polyploid. *J. Ind. Microbiol. Biot.* **21**, 228-232.

Miller MW, Yoneyama MINO and Soneda MASA (1976) *Phaffia*, a New Yeast Genus in the Deuteromycotina (Blastomycetes). *Int. J. Syst. Bacteriol.* **26**, 286-291.

Milne R (2006) Northern Hemisphere Plant Disjunctions: A Window on Tertiary Land Bridges and Climate Change? *Ann. Bot.* **98**, 465-472.

Naguib YM (2000) Antioxidant activities of astaxanthin and related carotenoids. *J. Agric. Food Chem.* **48**, 1150-1154.

Nagy Á, Garamszegi N, Vágvölgyi C and Ferenczy L (1994) Electrophoretic karyotypes of *Phaffia rhodozyma* strains. *FEMS Microbiol. Lett.* **123**, 315-318.

Nagy Á, Palágyi Z, Ferenczy L and Vágvölgyi C (1997) Radiation-induced chromosomal rearrangement as an aid to analysis of the genetic constitution of *Phaffia rhodozyma*. *FEMS Microbiol. Lett.* **152**, 249-254.

Nei M (1973) Analysis of gene diversity in subdivided populations. *Proc. Natl. Acad. Sci. USA* **70**, 3321-3.

Niklitschek M, Alcaino J, Barahona S, Sepulveda D, Lozano C, Carmona M, Marcoleta A, Martinez C, Lodato P, Baeza M*, et al.* (2008) Genomic organization of the structural genes controlling the astaxanthin biosynthesis pathway of *Xanthophyllomyces dendrorhous*. *Biol. Res.* **41**, 93-108.

Pera LM, Rubinstein L, Baigorí MD, de Figueroa LaIC and Callieri DA (1999) Influence of manganese on cell morphology, protoplasts formation and β-d-glucosidase activity in *Phaffia rhodozyma*. *FEMS Microbiol. Lett.* **171**, 155-160.

Perez-Losada M, Porter ML, Tazi L and Crandall KA (2007) New methods for inferring population dynamics from microbial sequences. *Infect. Genet. Evol.* **7**, 24-43.

Pérez-Ortín JE, Querol A, Puig S and Barrio E (2002) Molecular Characterization of a Chromosomal Rearrangement Involved in the Adaptive Evolution of Yeast Strains. *Genome Res.* **12**, 1533-1539.

Peterson KR, Pfister DH and Bell CD (2010) Cophylogeny and biogeography of the fungal parasite *Cyttaria* and its host *Nothofagus*, southern beech. *Mycologia* **102**, 1417-25.

Poláková S, Blume C, Zárate JÁ, Mentel M, Jørck-Ramberg D, Stenderup J and Piškur J (2009) Formation of new chromosomes as a virulence mechanism in yeast *Candida glabrata*. *Proc. Nat. Acad. .Sci.* **106**, 2688-2693.

Posada D and Crandall KA (2001) Intraspecific gene genealogies: trees grafting into networks. *Trends Ecol. Evol.* **16**, 37-45.

Pritchard JK, Stephens M and Donnelly P (2000) Inference of population structure using multilocus genotype data. *Genetics* **155**, 945-59.

Rawlings GB (1956) *Australasian Cyttariaceae*. *Trans. & Proc. Roy. Soc. New Zealand.* **84**, 19-28.

Rodriguez-Saiz M, de la Fuente JL and Barredo JL (2010) *Xanthophyllomyces dendrorhous* for the industrial production of astaxanthin. *Appl. Microbiol. Biotechnol.* **88**, 645-658.

Sambrook J. FEF, Maniatis T. 1989. *Molecular cloning: a Laboratory Manual*. New York: Cold Spring Harbor Laboratory Press.

Sampaio JP, Gadanho M, Santos S, Duarte FL, Pais C, Fonseca A and Fell JW (2001) Polyphasic taxonomy of the basidiomycetous yeast genus *Rhodosporidium*: *Rhodosporidium kratochvilovae* and related anamorphic species. *Int. J. Syst. Evol. Microbiol.* **51**, 687-97.

Santopietro LMD, Spencer JFT, Spencer DM and Siñeriz F (1995) Formation of protoplasts from *Phaffia rhodozyma*. *Biot. Tech.* **9**, 339-344.

Schmidt I, Schewe H, Gassel S, Jin C, Buckingham J, Humbelin M, Sandmann G and Schrader J (2010) Biotechnological production of astaxanthin with *Phaffia rhodozyma* / *Xanthophyllomyces dendrorhous. Appl.Microbiol.Biotechnol.* **89** (3), 555-571

Scorzetti G, Fell JW, Fonseca A and Statzell-Tallman A (2002) Systematics of basidiomycetous yeasts: a comparison of large subunit D1/D2 and internal transcribed spacer rDNA regions. *FEMS Yeast Res.* **2**, 495-517.

Swenson U, Backlund A, McLoughlin S and Hill RS (2001) *Nothofagus* Biogeography Revisited with Special Emphasis on the Enigmatic Distribution of Subgenus Brassospora in New Caledonia. *Cladistics* **17**, 28-47.

Tamura K, Peterson D, Peterson N, Stecher G, Nei M and Kumar S (2011) MEGA5: Molecular Evolutionary Genetics Analysis using maximum likelihood, evolutionary distance, and maximum parsimony methods. *Mol. Biol. Evol.,*[Epub ahead of print].

Taylor J, Jacobson D and Fisher M (1999) The evolution of THE EVOLUTION OF ASEXUAL FUNGI: Reproduction, Speciation and Classification. *Annu. Rev. Phytopathol.* **37**, 197-246.

Taylor JW, Jacobson DJ, Kroken S, Kasuga T, Geiser DM, Hibbett DS and Fisher MC (2000) Phylogenetic species recognition and species concepts in fungi. *Fungal. Genet. Biol.* **31**, 21-32.

Thompson J, Higgins D and Gibson T (1994) CLUSTAL W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice. *Nucl. Acids Res.* **22**, 4673 - 4680.

Turner, Hovenkamp and Van W (2001) Biogeography of Southeast Asia and the West Pacific. *J. Biogeo.* **28**, 217-230.

Vallone PM and Butler JM (2004) AutoDimer: a screening tool for primer-dimer and hairpin structures. *Biotechniques* **37**, 226-31.

Voris HK (2000) Maps of Pleistocene sea levels in Southeast Asia: shorelines, river systems and time durations. *J. Biogeo.* **27**, 1153-1167.

Wagele J and Mayer C (2007) Visualizing differences in phylogenetic information content of alignments and distinction of three classes of long-branch effects. *BMC Evol. Biol.* **7**, 147.

Weber RWS, Davoli P and Anke H (2006) A microbial consortium involving the astaxanthin producer *Xanthophyllomyces dendrorhous* on freshly cut birch stumps in Germany. *Mycologist* **20**, 57-61.

Zolan ME (1995) Chromosome-length polymorphism in fungi. *Microbiol Rev* **59**, 686-98.

6. Supplementary data

Table S1 – Sequences used for the construction of the primers for the multilocus sequence analysis.

Gene	Species	Strains	GenBank Accession Number
CRTI	*X. dendrohous*	ATCC 24230	DQ028748
	X. dendrohous	CBS 6938	Y15007
	Neurospora crassa	OR74A	XM_959620
	Pyrenophora tritici-repentis	Pt-1C-BFP	XM_001937660
	Cercospora nicotianae	ATCC 18366	U03903
CRTS	*X. dendrohous*	ATCC 96594	AX034666
	X. dendrohous	CBS 6938	AX034665
	X. dendrohous	ATCC 24203	DQ202402
	X. dendrohous	ATCC 96815mut	AY946023
	X. dendrohous	ATX5	EU713462
IDI	*X. dendrohous*	ATCC 24230	DQ235686
	X. dendrohous	CBS 6938	Y15811
EPH1	*X. dendrohous*	CBS 6938	AF166258
GDHA	*X. dendrohous*	VKPM 2410	EU791456
L41	*X. dendrohous*	ATCC 24230	AF004672
INV	*X. dendrohous*	ATCC MYA-131	FJ539193

Table S2 – List genes and genetic regions used in this study and relevant information regarding them.

Gene	Biological function	Primers (5'- 3')	Amplified fragment (bp)	General PCR conditions
CRTI	Phytoene desaturase	CRTI_F: GATGCTATCGTCATTCTTG CRTI_R: GCAAGAACACCAACGGAT	641	95ºC for 3 min; (95ºC for 30s, 48-51ºC for 1min, 72ºC for 1min) 35 cycles; 72ºC for 7 min
		CRTI_F01: AGTTTGTGATCCAMACGCTTTC CRTI_R01: TCCAAATCCATGAACTTCAGCT	409	95ºC for 3 min; (95ºC for 30s, 55-55.2ºC for 1min, 72ºC for 1min) 35 cycles; 72ºC for 7 min
		CRTI_F02: GGRATCMTTYGACACAATCTTCGAG CRTI_R02: GCWCCAGCCAAKACRATGGG	532	95ºC for 3 min; (95ºC for 30s, 57-59ºC for 1min, 72ºC for 1min) 35 cycles; 72ºC for 7 min
CRTS	Hidroxylase-ketolase	CRTS_F: TGGACGTCATGGCTCTTG CRTS_R: CTAGGAGACTTAGGATATC	567	95ºC for 3 min; (95ºC for 30s, 48-50ºC for 1min, 72ºC for 1min) 35 cycles; 72ºC for 7 min
		CRTS_F01: CGACTCGCTCCAGAACAAGACC CRTS_R01: GATCGGAAGCTGARCCRAGCAC	391	95ºC for 3 min; (95ºC for 30s, 60ºC for 1min, 72ºC for 1min) 35 cycles; 72ºC for 7 min
		CRTS_F02: GAAGCCTAYGAYTWYCCGAAACC CRTS_R02: GTSTTSTGAGAGTCGGTRAAACATCCA	1382	95ºC for 3 min; (95ºC for 30s, 59ºC for 1min, 72ºC for 1min) 35 cycles; 72ºC for 7 min
IDI	Isopentenyl diphosphte isomerase	IDI_F: GTGTTGCAGTCATCCTTTG IDI_R: GAGAAGTTGATCCCACCA	631	95ºC for 3 min; (95ºC for 30s, 50-52ºC for 1min, 72ºC for 1min) 35 cycles; 72ºC for 7 min
		IDI_F01: AGCTGCGTCCCGAAAGTTGGAG IDI_R01: GCYTGRAGCTCRGGCTTGGT	341	95ºC for 3 min; (95ºC for 30s, 60ºC for 1min, 72ºC for 1min) 35 cycles; 72ºC for 7 min
EPH1	Epoxide hydrolase	EPH1_F: GGGCCAACGAGTGGAGCTG EPH1_R: GCGCAACGACAGCGAACTC	794	95ºC for 3 min; (95ºC for 30s, 59-60ºC for 1min, 72ºC for 1min) 35 cycles; 72ºC for 7 min
		EPH1_F01: CTGTCTCATGGATGGCCTTCATCG EPH1_R01: CCCWGCTCCGCAGGYATATG	407	95ºC for 3 min; (95ºC for 30s, 59-60ºC for 1min, 72ºC for 1min) 35 cycles; 72ºC for 7 min
L41	Ribosomal protein L41	L41_F: GGTCAACGTTCCCAAGACTC L41_R: CCCTTCTTGTACTGGGTCAC	640	95ºC for 3 min; (95ºC for 30s, 55-56ºC for 1min, 72ºC for 1min) 35 cycles; 72ºC for 7 min

Table S2 (cont.) – List genes and genetic regions used in this study and relevant information regarding them.

Gene	Biological function	Primers (5'- 3')	Amplified fragment (bp)	General PCR conditions
INV	Invertase	INV_F: CCACGCCGGGTATCAATC INV_R: CCTTGATCTCGACGGATCCC	557	95ºC for 3 min; (95ºC for 30s, 54-56ºC for 1min, 72ºC for 1min) 35 cycles; 72ºC for 7 min
GDHA	Glutamate dehydrogenase	GDHA_F: GAGCCCGAGTTCGAGCAG GDHA_R: TCCCAAAGCGGAGTTGAA	610	95ºC for 3 min; (95ºC for 30s, 52-53.5ºC for 1min, 72ºC for 1min) 35 cycles; 72ºC for 7 min
ITS	*ITS1, 5.8S, ITS2* region of rDNA	*ITS5*: GGAAGTAAAAGTCGTAACAAGG *ITS4*: TCCTCCGCTTATTGATATGC	Aprox.600	95ºC for 5 min; (95ºC for 30s, 52ºC for 1min, 72ºC for 1min) 40 cycles; 72ºC for 7 min

Table S3 – Composition of the multilocus sequence alignment.

Gene	Aligned Fragment Length (bp)	Parsimony significant sites	ML Selected model
CRTI	359	99	TrNef+G
CRTS	523	223	K81uf
IDI	290	107	TVM+G
EPH1	365	160	HKY
L41	543	273	TrN
INV	506	225	TVM
GDHA	568	226	TIM
Multilocus	3187	1287	GTR+G

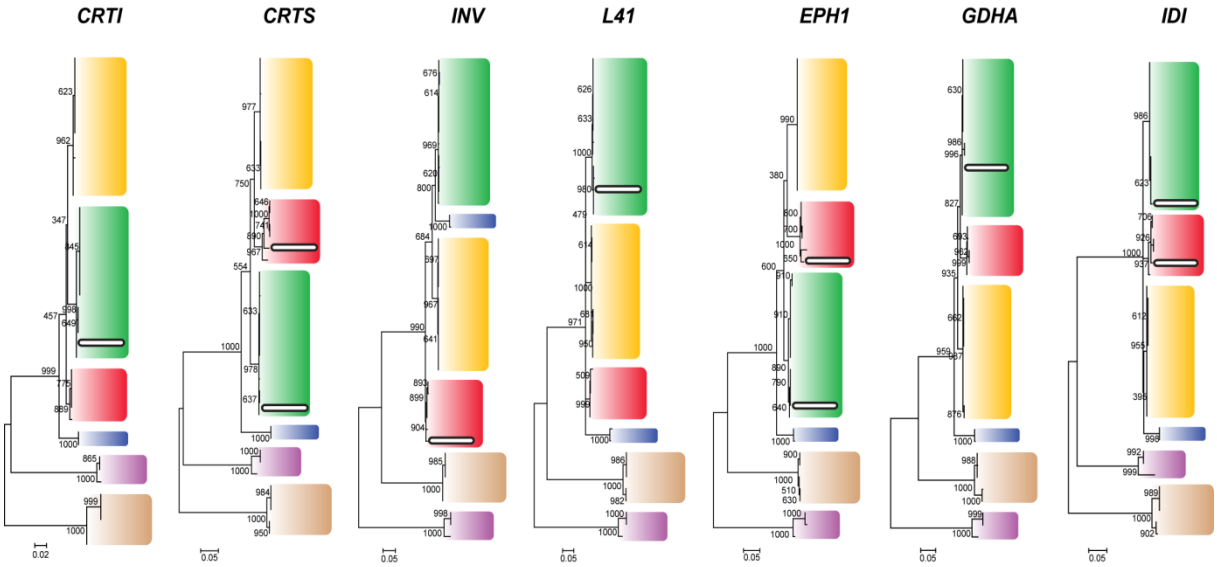

Figure S1 - Individual phylogenies of *Xanthophyllomyces* by ML method. Individual ML phylogenies of the 7 gene fragments used in the MLSA. Bootstraps values from 1000 replicates are shown next to the branches. Clades are presented in the same colors as shown in all other figures. Strain CBS 5905 presents two different haplotypes (HA and HB) that are highlighted by a white rounded box. Phylogenetic relevant information for each of the 7 trees is presented in table S3. Each phylogeny is identified by the abbreviated gene names. Full information regarding the genes is present in Table S2.

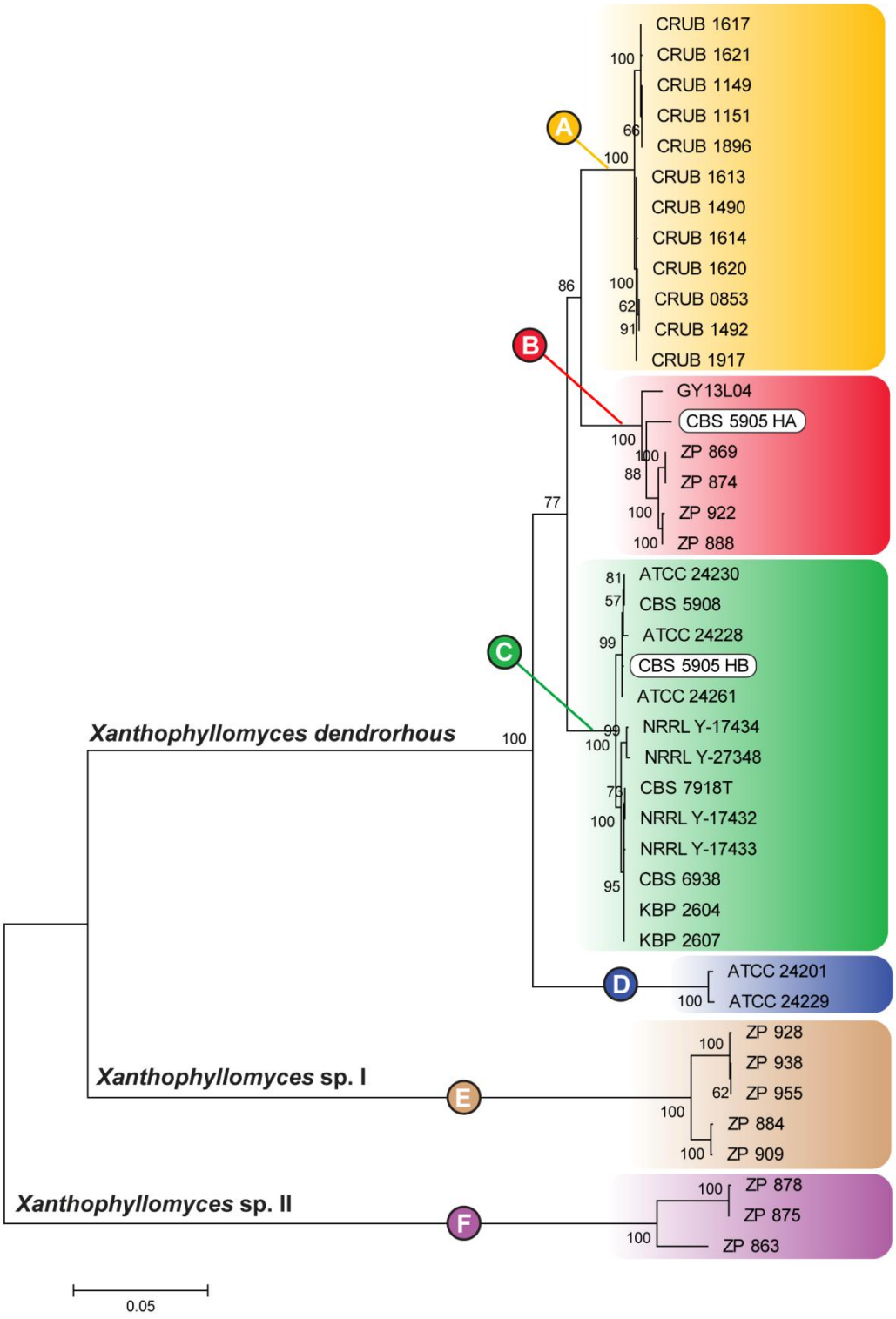

Figure S2 - Multilocus phylogeny of *Xanthophyllomyces* by ML method. Maximum likelihood analysis of the multilocus alignment was performed using the PALM server. On the phylogeny colored circles are shown identifying 6 major clades also depicted in distinct colors and represented with the letters A - F. Strain CBS 5905 presents two different haplotypes (HA and HB) that are highlighted by a white rounded box. The alignment used has a total of 3187 nucleotides and 41 DNA sequences corresponding to 40 strains. The best model was determined by AIC to be GTR+G. Bootstrap values (1000 replicates) are showed next to the respective branches. The gamma distribution shape parameter (G) was 0.0075.

Figure S3 - Sexual state of *Xanthophyllomyces* strains from Australasia. Photographs depicting the morphological structures associated with *Xanthophyllomyces* sexual state presented by different strains isolated in Australasia when inoculated in 0.5% ribitol water/agar culture medium and incubated at 18 ºC for one week.

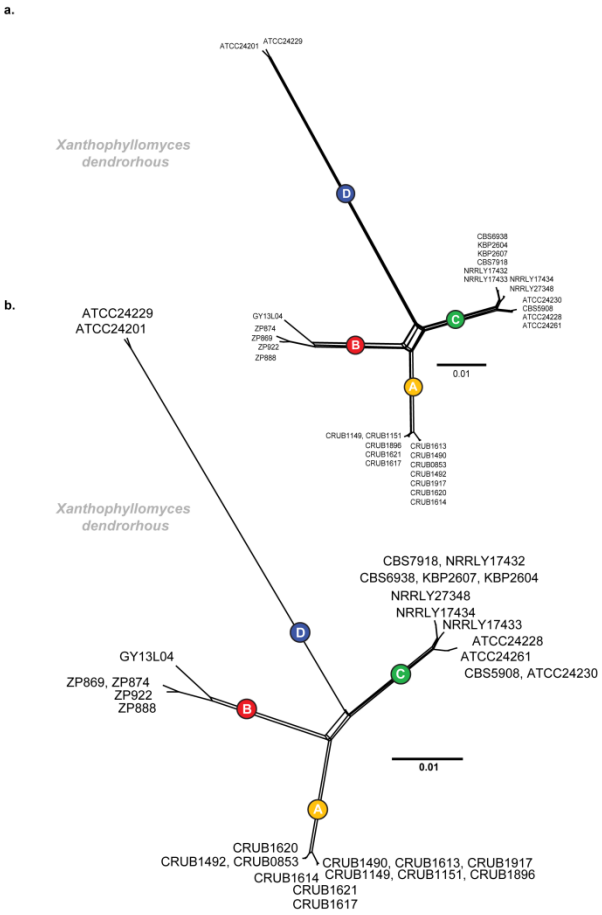

Figure S4 - *X. dendrorhous* SN and RN. **(a)** SN built with the multilocus dataset for all 31 *X. dendrorhous* strains (not including CBS 5905). **(b)** RN built with the multilocus dataset for all 31 *X. dendrorhous* strains (not including CBS 5905). Analysis was performed in SplitsTree4 software. Colored circles are shown identifying 4 existing clades in *X. dendrorhous*.