Fungal Genetics and Biology 170 (2024) 103861

Contents lists available at ScienceDirect

Fungal Genetics and Biology

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

Multi-locus sequence typing and phylogenetics of *Cryptococcus neoformans* AD hybrids

M. Cogliati [a,*], P.E. Chidebelu [b], M. Hitchcock [c], M. Chen [d], V. Rickerts [e], S. Ackermann [e], M. Desnos Ollivier [f], J. Inácio [g], U. Nawrot [h], M. Florek [i], K.J. Kwon-Chung [j], D.-H. Yang [j], C. Firacative [k], C.A. Puime [l], P. Escandon [m], S. Bertout [n], F. Roger [n], J. Xu [c]

[a] *Lab. Medical Mycology, Dept. Biomedical Sciences for Health, Università degli Studi di Milano, Milano, Italy*
[b] *Department of Microbiology, University of Nigeria, Nsukka, Enugu State, Nigeria*
[c] *Department of Biology, McMaster University, Hamilton, Ontario, Canada*
[d] *Shanghai Key Laboratory of Molecular Medical Mycology, Department of Dermatology, Chanzheng Hospital, Second Military Medical University, Shanghai, China*
[e] *Robert Koch Institute, Berlin, Germany*
[f] *Institut Pasteur, Université de Paris, CNRS UMR2000, Molecular Mycology Unit, National Reference Center for Invasive Mycoses and Antifungals, Paris, France*
[g] *School of Applied Sciences, University of Brighton, Brighton, UK*
[h] *Department of Pharmaceutical Microbiology and Parasitology, Wroclaw Medical University, Wroclaw, Poland*
[i] *Department of Pathology, Faculty of Veterinary Medicine, Wroclaw University of Environmental and Life Sciences, Wroclaw, Poland*
[j] *Molecular Microbiology Section, Laboratory of Clinical Immunology and Microbiology, National Institute of Allergy and Infectious Diseases NIH, Bethesda, USA*
[k] *Studies in Translational Microbiology and Emerging Diseases (MICROS) Research Group, School of Medicine and Health Sciences, Universidad de Rosario, Bogotá, Colombia*
[l] *Unidad de Parasitología y Micología, Departamento de Laboratorios de Salud Pública, Ministerio de Salud Pública, Montevideo, Uruguay*
[m] *Grupo de Microbiología, Instituto Nacional de Salud, Bogotá, Colombia*
[n] *Laboratoire de Parasitologie et Mycologie Médicale, UMI 233 TransVIHMI, University of Montpellier, IRD, INSERM, Montpellier, France*

ARTICLE INFO

Keywords:
Cryptococcus neoformans
AD hybrids
MLST
Phylogenetics
Loss of heterozygosity
Multiple origin

ABSTRACT

Hybrid AD strains of the human pathogenic *Cryptococcus neoformans* species complex have been reported from many parts of the world. However, their origin, diversity, and evolution are incompletely understood. In this study, we analyzed 102 AD hybrid strains representing 21 countries on five continents. For each strain, we obtained its mating type and its allelic sequences at each of the seven loci that have been used for genotyping haploid serotypes A and D strains of the species complex by the *Cryptococcus* research community. Our results showed that most AD hybrids exhibited loss of heterozygosity at one or more of the seven analyzed loci. Phylogenetic and population genetic analyses of the allelic sequences revealed multiple origins of the hybrids within each continent, dating back to one million years ago in Africa and up to the present in other continents. We found evidence for clonal reproduction and long-distance dispersal of these hybrids in nature. Comparisons with the global haploid serotypes A and D strains identified new alleles and new haploid multi-locus genotypes in AD hybrids, consistent with the presence of yet-to-be discovered genetic diversity in haploid populations of this species complex in nature. Together, our results indicate that AD hybrids can be effectively genotyped using the same multi-locus sequencing type approach as that established for serotypes A and D strains. Our comparisons of the AD hybrids among each other as well as with the global haploid serotypes A and D strains revealed novel genetic diversity as well as evidence for multiple origins and dynamic evolution of these hybrids in nature.

1. Introduction

Strains in the *Cryptococcus neoformans* species complex and *Cryptococcus gattii* species complex are opportunistic fungal pathogens responsible for causing cryptococcosis, a potentially life-threatening infection primarily affecting immunocompromised and more rarely immunocompetent individuals (D'Souza et al., 2011; Pappas, 2013; Rajasingham et al., 2022; Williamson et al., 2017; Yang et al., 2021). These fungi are present in the environment and have been isolated from different sources such as bird excreta, trees, soil, mammal feces, and

* Corresponding author at: Lab. Medical Mycology, Dept. Biomedical Sciences for Health, Università degli Studi di Milano, Via Pascal 36, 20133 Milano, Italy.
E-mail address: massimo.cogliati@unimi.it (M. Cogliati).

https://doi.org/10.1016/j.fgb.2023.103861
Received 19 September 2023; Received in revised form 6 December 2023; Accepted 18 December 2023
Available online 19 December 2023

water (Cogliati, 2013). The propagules able to infect susceptible humans or animals have not yet been directly seen in the environment but it is likely that desiccated yeast cells, basidiospores and small-size blastoconidia could reach pulmonary alveoli playing a key role in the onset of the infection (Velagapudi et al., 2009). Most of the cryptococcosis cases are caused by strains of *C. neoformans* species complex that are distributed worldwide, whereas cases due to *C. gattii* species complex are more prevalent in Australia as well as in other endemic geographical areas both in subtropical (California, Colombia, Brazil, Northwestern Argentina, Mexico, India, Southern Africa) and temperate (Southwestern Canada, Northwestern USA, Europe) regions (Francisco et al., 2021).

Both *C. neoformans* species complex and *C. gattii* species complex contain differentiated evolutionary lineages and, in a recent taxonomic revision, some of them were elevated to species level (Hagen et al., 2015). Specifically, *C. gattii* species complex have been classified in five molecular types/species: VGI (serotype B, *C. gattii sensu stricto*), VGII (serotype B, *C. deuterogattii*), VGIII (serotypes B or C, *C. bacillisporus*), VGIV (serotype C, *C. tetragattii*) and VGVI (serotype C, *C. decagattii*) (Monroy-Nieto et al., 2022). In addition, a new lineage, molecular type VGV, was recently identified as a potential new species (Farrer et al., 2019). On the other hand, *C. neoformans* species complex are commonly distinguished in molecular types VNI, VNII, and VNB (serotype A, *C. neoformans* var. *grubii*, *C. neoformans sensu stricto*), VNIV (serotype D, *C. neoformans* var. *neoformans*, *C. deneoformans*) and their hybrids VNIII (AD hybrids). Although hybrids between the two species complexes have been rarely identified (Bovers et al., 2006), diploid or aneuploid AD hybrids, resulting from the fusion of one serotype A and one serotype D strain, are frequently isolated (Cogliati, 2013), challenging the reproductive isolation and species status of these two lineages (Hagen et al., 2017; Kwon-Chung et al., 2017). At present, based on mating type combinations, three types of AD hybrids have been identified: αAD**a** and **a**ADα, originated by the fusion of opposite mating types (mating type α × mating type **a**), and αADα originated by same-sex mating (mating type α × mating type α) (Cogliati et al., 2011; Samarasinghe and Xu, 2018).

Based on serotyping and multi-locus sequence typing (MLST), the prevalence of AD hybrids has shown to vary among geographical areas. Among the continents, Europe was found to have the highest prevalence (20–30 %), followed by North and South America (10 %), but AD hybrids have also been isolated less frequently from Africa and Asia (2 %) (Cogliati, 2013). Most of the AD hybrids reported were from clinical cases and only a low percentage of them were recovered from environmental sources suggesting that probably their real environmental niches have been under-explored. Although AD hybrids are largely sterile and unable to mate, due to a 10–15 % nucleotide divergence between parental strains, they can propagate asexually and generate diverse genotypes by non-disjunction, aberrant meiosis, mitotic recombination, and gene conversion that often produce aneuploids and that, under stress conditions, can increase the fitness of the derived progeny (Michelotti et al., 2022; Samarasinghe et al., 2020a; Samarasinghe et al., 2020b; You et al., 2021). Elucidating where, when, and how AD hybrids originated represents a crucial step to understand the evolution of *C. neoformans* species complex, however, such studies have been hindered by the absence of efficient protocols for sequence typing of this group of hybrid yeasts. Only recently new primers, specific for serotype A and D alleles of the seven loci chosen for standard multi-locus sequence typing (MLST) of the human pathogenic *Cryptococcus* have been developed, allowing the direct typing of AD hybrids and their comparison with the global *C. neoformans* species complex MLST profiles of haploid isolates (Cogliati et al., 2020). The present study analyzes, for the first time, a large number of AD hybrids by MLST in order to understand their diversity and relationships with each other and with the global *C. neoformans* serotype A and D haplotypes.

2. Materials and methods

2.1. Isolates

A total of 102 *C. neoformans* AD-hybrid isolates were investigated (Supplementary Table 1). Fifty-eight were from Europe (Belgium, Croatia, France, Germany, Greece, Hungary, Italy, Poland, Portugal, Spain, Sweden, and Turkey), 18 from Africa (Ivory Coast and Rwanda), 15 from North America (USA), six from South America (Brazil, Chile, Colombia, Uruguay), and five from Asia (China and Japan). Forty-four isolates were mating type **a**ADα, 42 were αAD**a**, eight were αADα, and eight were -ADα for which the mating type A allele was lacking. Most of the hybrids were isolated from clinical sources (n = 89) and only 13 were from environmental sources.

2.2. Multi-locus sequence typing (MLST)

Genomic DNA of isolates was extracted as previously described (Viviani et al., 1997) and an MLST protocol specific for *C. neoformans* AD hybrids was applied to amplify the seven standard loci: *CAP59*, *GDP1*, IGS1, *LAC1*, *PLB1*, *SOD1*, and *URA5*. The protocol was previously described (Cogliati et al., 2020) and was modified only for *CAP59* locus (Supplementary Table S2, https://sites.google.com/view/screenprojectcryptococcus/screen-project-studies/cryptococcus-hybrids-sequencing-cryhs-study/primers-and-thermal-cycling-protocol). Allele type (AT) numbers and the sequence type (ST) numbers for haplotypes were assigned according to global MLST database curators (https://www.mycologylab.org). If one or more alleles were lacking they were indicated with 0 in the allelic combination, and ST of this combinations were codified using the number of the complete ST minus the lacking allele (example: ST637 lacking *PLB1* allele = ST637-P). If the complete ST was not present in the global MLST database a new number was assigned and used only in the hybrid database (example: new combination lacking *PLB1* = ST900-P). Hybrids allele combination and sequence concatenation used the following scheme: *CAP59A/CAP59D-GDP1A/GDP1D*-IGS1A/IGS1D-*LAC1A/LAC1D-PLB1A/PLB1D-SOD1A/SOD1D-URA5A/URA5D*. Hybrid sequence type (HST) codes were progressively assigned and deposited in a new database, including only hybrids, located at SCrEEN Project website (https://sites.google.com/view/screenprojectcryptococcus). New allele sequences have been deposited at GenBank (https://www.ncbi.nlm.nih.gov/genbank) with the following accession numbers: OR402885-OR402889.

2.3. Minimum spanning analysis

Minimum spanning network analysis with GoeBurst algorithm was performed using Phyloviz software v2.0 (https://www.phyloviz.net). Analysis was performed separately using the allelic combinations of hybrids, the allelic combinations of the A-portion of hybrids, and the allelic combinations of the D-portion of hybrids. A clonal complex (CC) was defined when two or more genotypes differed in one or two loci. The lowest common ancestor (LCA) of each CC determined also the code of the CC (example: if LCA = HST1 then CC = CC1). For each analysis, three types of data were labeled: geographical origin, mating type, and source. The main hybrids CCs were used to determine the A and D parental CCs that originated them, as well as to determine the mating type, geographical origin, and source of potential parents of hybrids.

To understand the relationship of hybrid parents with the genotypic clusters of global haploid isolates, the allelic combinations of the A portion of hybrids was compared with all VNI, VNII, and VNB MLST profiles present in the global MLST database. Similarly, the allelic combinations of D portion of hybrids were compared with all the VNIV MLST profiles.

2.4. Phylogenetic analysis

Phylogenetic analysis was performed using maximum likelihood method implemented in MEGA software v11.0.13 (https://www.megasoftware.net) and confirmed by a Bayesian method using Beast software v2.7.4 package, including BEAUty, Beast, and DensiTree (https://www.beast2.org). Concatenated sequences of hybrids and concatenated sequences of A and D portion of hybrids were input to run three separate analyses. Only complete sequences, without missing alleles, of the serotype A and/or D portion of hybrids were considered in this analysis to avoid the potentially biased inferences of sequence types and age estimates. The maximum likelihood phylogenetic trees were inferred by adopting the Tamura-Nei substitution model and a bootstrap statistical analysis with 1000 replicates. An ultrametric tree was generated using the maximum likelihood tree as a starting point and fixing the age of the most recent common ancestor involved in the *C. neoformans-C. gattii* split at 34 million years. This age was derived using the neutral mutation rate of 2E-9 per nucleotide per year for protein-coding genes (D'Souza et al., 2011). *Cryptococcus gattii* ST51, *C. neoformans* var. *grubii* ST31, and *C. neoformans* var. *neoformans* ST121 were chosen as outgroup genotypes. *Cryptococcus gattii* ST51 was artificially duplicated to fit as outgroup for hybrids analysis.

Bayesian analysis first required to format the sequence alignment to. xml format by BEAUty software, then was run by Beast, and the resulting analysis was visualized by DensiTree.

The phylogenetic trees of A and D portions of hybrids were compared with each other and the parental clusters that generated the main hybrid CCs were connected to infer how and when they originated. Finally, the concatenated sequences of A and D portions of hybrids were aligned with the main global serotype A STs and serotype D STs, respectively, to infer the phylogenetic relationships among them.

2.5. Population genetic analysis

Intra- (A-portion, D-portion, and whole hybrids) and inter-population genetic variabilities (A-portion versus D-portion of hybrids) were estimated by several parameters: number of haplotypes (h), number of segregating sites (S), haplotype diversity (Hd), nucleotide diversity (Pi), and average number of nucleotide differences (k). Minimum number of recombination events for each locus in each population was also estimated. In addition, the number of fixed mutations, the number of shared mutations, and the number of monomorphic mutations in one population and polymorphic in the other, were calculated to estimate genetic divergence between populations. Inter-population comparison of nucleotide diversity was also performed. All parameters and tests were calculated using DNAsp v5.10.1 software (https://www.ub.edu/dnasp). Linkage disequilibrium test between pairs of loci was also performed by Arlequin software v3.0 (cmpg.unibe.ch/software/arlequin3) to evaluate the level of linkage between loci in the hybrid population. Phylogenetic compatibility and index of association were calculated by MultiLocus V1.3 (Agapow and Burt, 2001).

3. RESULTS

3.1. Genotypes of AD hybrids

Among the 102 strains, a total of 52 HSTs were identified by MLST analysis (Supplementary Table 1). The most common HST among the investigated hybrids was HST1, including 21 isolates (20.6 %), followed by HST19 (7.8 %), HST27 (6.9 %), HST13 (4.9 %), HST37 (3.9 %), and the other 47 HSTs included 3 or fewer isolates each (<3 %). Most of the hybrid MLST profiles lacked one or more alleles in either the A or the D portions or both the A and D portions but for alleles at different loci (35/52, 67.3 %). Amplification of missing alleles was repeated several times using different annealing temperatures and/or alternative primers to confirm their missing status. For samples with missing alleles, additional amplifications, using the non-serotype-specific primers of the standard MLST scheme, were performed to obtain an amplicon that, when sequenced, produced a sequence identical to the originally obtained, clean allele sequence, further confirming that the investigated allele was missing.

Interestingly, only five new allele types (ATs) were identified in these hybrids that have not been reported in the global serotypes A and D populations, respectively.

Separate analysis of the A portion of hybrids identified 35 different STs, among which the most prevalent was ST632 present in 28.4 % (29/102) of hybrid isolates, followed by ST69 (12.7 %, 13/102), and the others including less than 10 % of isolates each. The percentage of A-portion STs in which one or more alleles were lacking, was 51.4 % (18/35).

Separate analysis of D portion of hybrids identified 32 different STs, among which the most prevalent was ST636 present in 22.5 % (23/102) of hybrid isolates, followed by ST637 (18.6 %, 19/102), and the others including less than 10 % of isolates each. The percentage of D-portion STs in which one or more alleles were lacking, was 46.8 % (15/32).

3.2. AD-hybrid clonal complexes

Analysis of hybrid HSTs by GoeBurst algorithm generated a minimum spanning tree that identified six main CCs. The main CC was represented by CC19 which included 15 HSTs, followed by CC1 (13 HSTs), CC13 (7 HSTs), CC37 (3 HSTs), CC17 and CC35 (2 HSTs each) and ten singleton HSTs (Fig. 1A).

Mating type **a**ADα hybrids were grouped in clusters CC1, CC13 (including also -ADα isolates) and two singleton isolates. Mating type αAD**a** hybrids were distributed in two different clusters (CC19, and CC37) and eight singleton genotypes. All same-sex hybrids αADα were grouped in clusters CC35 and CC17 (Fig. 1B).

The hybrids belonging to CC19 were distributed among all continents except for Asia, whereas those belonging to CC1 were from all continents except for Africa. Clusters CC17 and CC37 included isolates from Africa and Europe, respectively, and CC13 contained isolates from both Africa and Europe. Finally, CC35 isolates were from Europe and South America (Fig. 1C).

Isolates from both clinical and environmental sources were present in four of the six main hybrid CCs. Only CC13 and CC17 found exclusively among clinical isolates (Fig. 1D).

3.3. Clonal complexes of the A-portion of hybrids

Minimum spanning network analysis of the STs of the A portion of hybrids identified five main clonal clusters (Fig. 2A). These clonal clusters seemed associated with mating types. Specifically, three CCs (CC69, CC174, and CC634) and two singleton genotypes grouped all hybrid isolates containing the mating type α (51 isolates), with CC69 including 60 % (30/51) of them. Two CCs (CC632 and CC633) grouped 52 isolates containing the **a**A mating type and those lacking mating type A allele (-A), with CC632 including most of them (73 %, 38/52).

3.4. Clonal complexes of the D-portion of hybrids

When minimum spanning network analysis was performed for the D portion of hybrids, six main clonal clusters were identified (Fig. 2B). Similar to that of the A portion, these clonal clusters seemed associated with mating types. Specifically, three CCs (CC636, CC635, and CC116) and one singleton genotype grouped all hybrid isolates containing the mating type αD (60 isolates), with CC636 including 63.3 % (38/60) of them. In addition, hybrids containing the **a**D mating type were grouped in three CCs (CC637, CC682 and CC514) and four singleton genotypes (42 isolates), with CC637 including 69 % (29/42) of them.

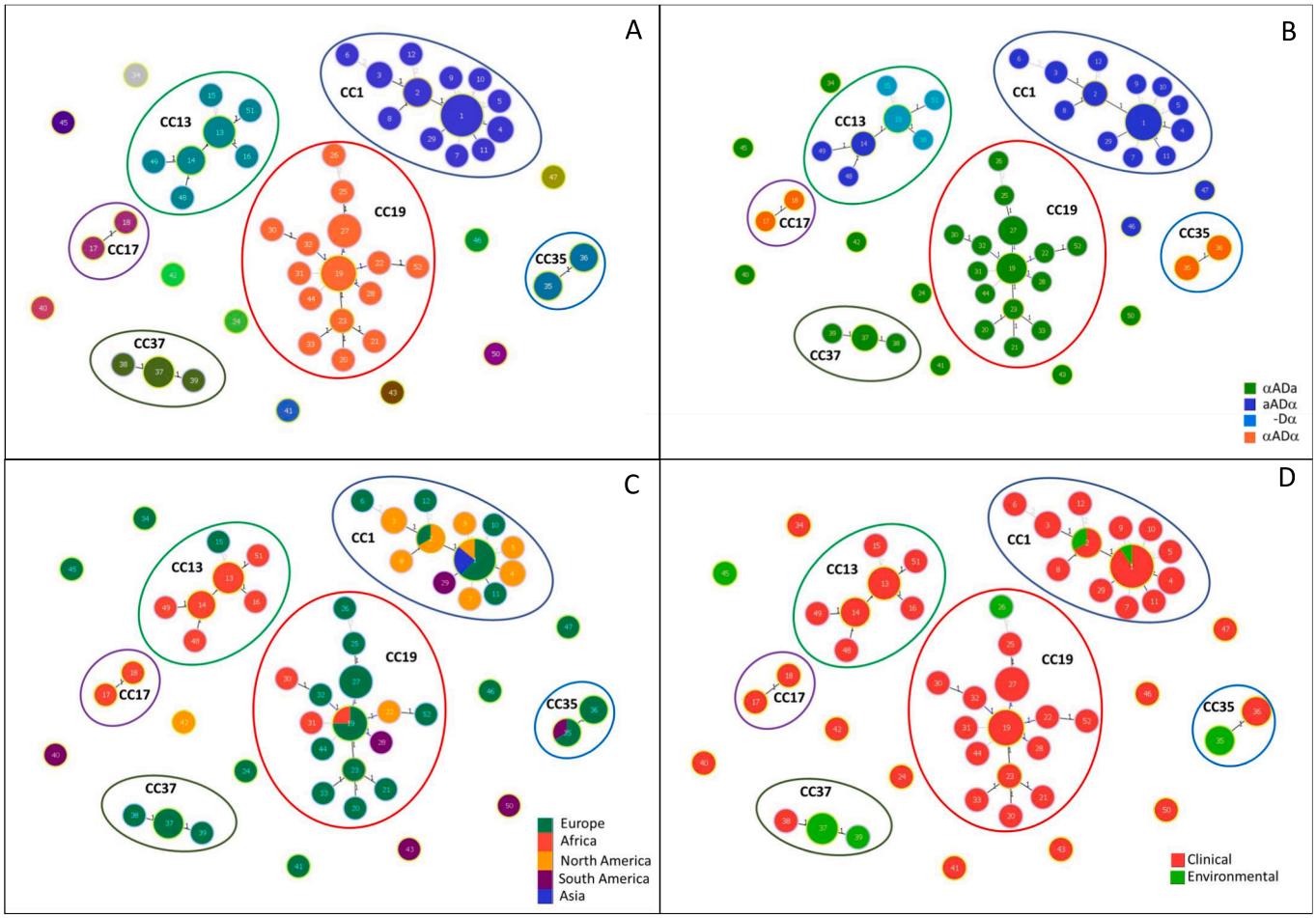

Fig. 1. (A) Minimum spanning tree inferred by GoeBurst algorithm showing the main clonal complexes (CC, grouped inside closed lines) among AD hybrid genotypes. In the same tree were reported separately (B) mating types, (C) geographical origin, and (D) source. The number inside each circle indicates the hybrid sequence type (HST). The number beside each linkage line represents the number of different loci. The size of each circle is proportional to the number of isolates with the same HST.

3.5. AD-hybrid combinations

Mating type αAD**a** hybrids mainly originated from two different combinations between A and D parents (Fig. 3). The hybrids belonging to CC19 originated from the mating of parental A isolates included in CC69 and parental D isolates from CC637. The geographical origin of both A and D parental isolates was from all continents except Asia. Clinical and environmental isolates were represented in both parental CCs. On the other hand, CC37 hybrids were originated from the combination of A parents belonging to CC634 and D parents belonging to CC514. The A parents were from Europe, Africa and South America but the D parents were all from Europe. Clinical and environmental isolates were represented in both parental CCs.

Two main combinations were identified among mating type **a**ADα hybrids (Fig. 4). The CC1 hybrid cluster appear to be generated by the combination of CC632 A parents and CC636 D parents which included isolates from different geographical origin except from Africa. Environmental isolates were present among both A and D parents. However, the hybrids belonging to CC13 originated from CC633 (A parents) and CC635 (D parents), which were mainly from Africa, with only one isolate from Europe, and did not include environmental isolates.

Same-sex hybrids αADα shared all an A parent belonging to CC634 including isolates from Europe, Africa, and South America and mainly from environmental sources (Fig. 5). For hybrids belonging to CC35, the D parents were from CC116 cluster and were from Europe and South America, whereas for hybrids belonging to CC17, the D parents were grouped in CC635 cluster including only isolates from Africa and Europe.

3.6. Comparison with global haploid isolates

Only four STs (ST69, ST174, ST58, and ST23) among the 35 haplotypes of the A portion of hybrids matched with global haplotypes of VNI, VNII, and VNB deposited in the MLST global database, none of the other 31 haplotypes were present in the database. Twenty haplotypes (57.1 %) belonged to the main VNI cluster, whereas the others (42.9 %), including CC632, CC633 and two singleton genotypes, were not located in this cluster (Fig. 6).

Comparison of haplotypes of hybrids D-portion with VNIV global haplotypes showed that only two STs were already present in the database (ST116 and ST514). The remaining 30 were all new to the database. Most of the serotype D haplotypes (20/32, 62.5 %) in the hybrids belonged to the main VNIV cluster. The remaining 12 non-correlated haplotypes belonged to CC635, CC682, CC578 (including ST674 and ST677), and two singleton genotypes (Fig. 7).

3.7. Phylogenetic analysis of hybrids

Maximum likelihood phylogenetic tree using the hybrids sequences, concatenated as described above, showed that the most ancient CCs of hybrids are CC13 (**a**ADα) and CC17 (αADα) that likely originated from Africa about 1 million years ago. The genetic divergence of these two

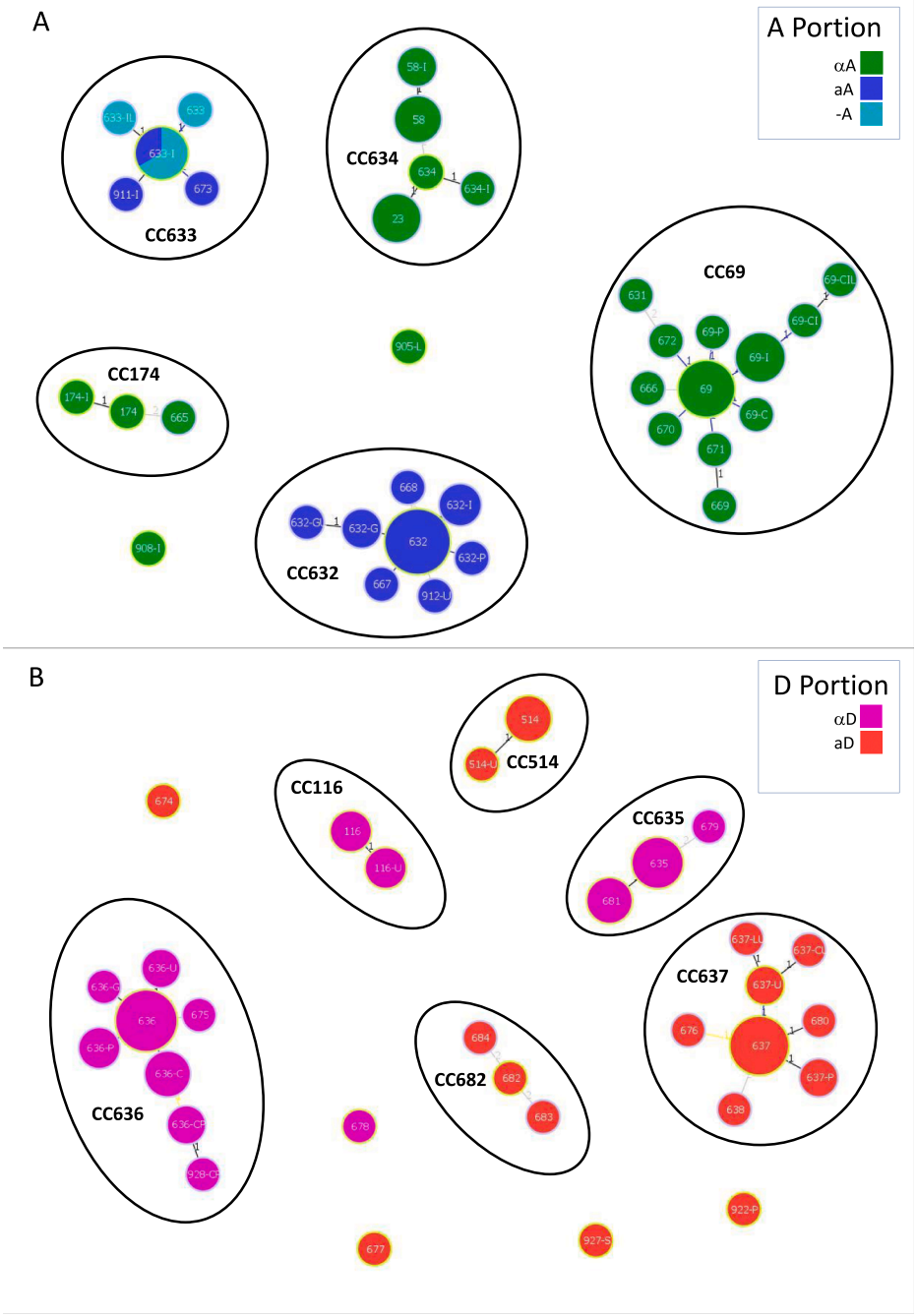

Fig. 2. Minimum spanning tree inferred by GoeBurst algorithm showing the main clonal complexes (CC, grouped inside closed lines) among the genotypes of the A-portion (A) and the D-portion (B) of AD hybrids. Mating types are also reported. The number inside each circle indicates the sequence type (ST). The number beside each linkage line represents the number of different loci. The size of each circle is proportional to the number of isolates with the same ST.

groups of hybrids from the other hybrids was supported by 100 % of trees generated by bootstrap analysis. On the other hand, the hybrid cluster CC37 (αAD**a**) originated from Europe about 600,000 years ago, although supported by a weaker statistical analysis (73 %), whereas the remaining three hybrid CCs (CC1, CC19, and CC35) showed a more recent origin (Fig. 8). When the same set of sequences was used to infer a phylogenetic tree with Bayesian method the resulting tree topology was identical confirming the above results (Supplementary Figure S1).

3.8. Phylogenetics of hybrid combinations

Comparison of maximum likelihood phylogenetic trees inferred separately for A-portion and D-portion concatenated sequences of hybrids revealed that the hybrid cluster CC13, originated from CC633 (**a**A parents) and CC635 (αD parents) was the most ancient combination that occurred about 700,000–1,000,000 years ago. The hybrid cluster CC1 originated from the combination of ancient **a**A parents (CC632) and recent αD parents (CC636). On the contrary, the CC17 hybrid cluster originated from the same-sex combination of recent αA parents (CC634) and ancient αD parents (CC635). The other three combinations originating the remaining three main hybrid CCs (CC37, CC35, and CC19) have a more recent origin (Fig. 9).

3.9. Phylogenetic comparison with MLST global database

Fig. 10 shows the phylogenetic relationships between the haplotypes

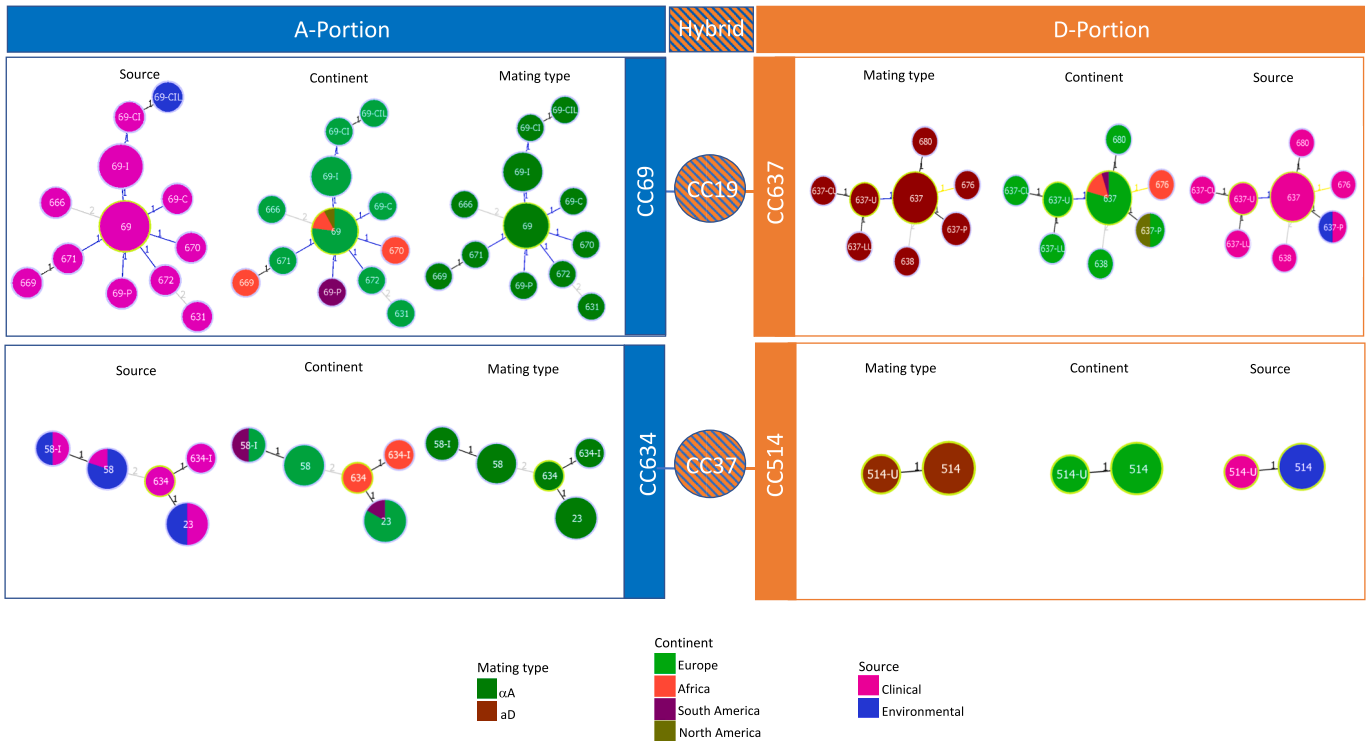

Fig. 3. Parental clonal complexes (CC) generating the main αAD**a** hybrid combinations. Left and right panels report the parental sequence types (ST) belonging to each CC and their correlation with mating types, geographical origin, and source.

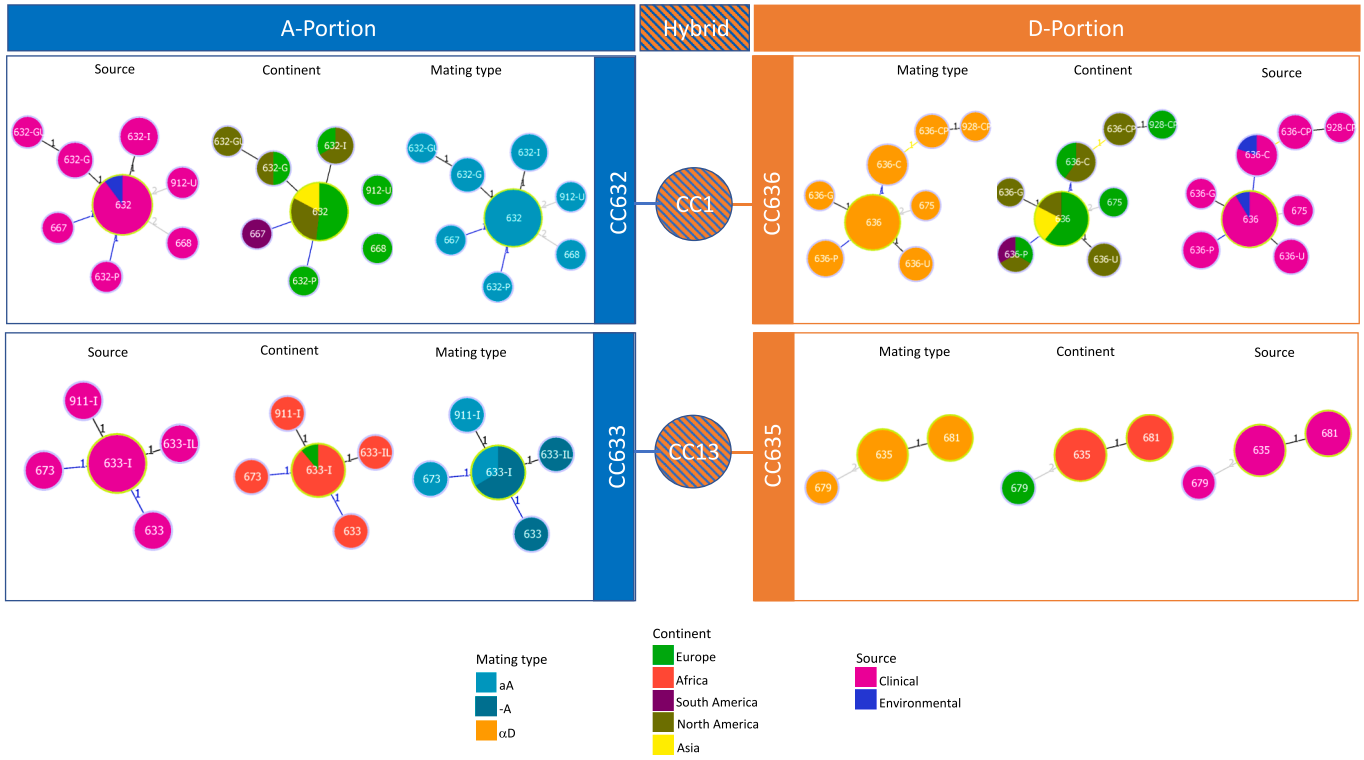

Fig. 4. Parental clonal complexes (CC) generating the main **a**ADα hybrid combinations. Left and right panels report the parental sequence types (ST) belonging to each CC and their correlation with mating types, geographical origin, and source.

of hybrids A-portion and the most representative VNI STs deposited in the MLST global database, as well as ST40 and ST14 used as VNII and VNB reference ST, respectively. Results confirmed that STs of CC633 (**a**A) from A portion of hybrids were the most ancient STs identified so far, originating about 1 million years ago and belonging to VNB molecular type. In contrast, CC632 which grouped all the other **a**A isolates belonged to molecular type VNI and originated more recently as well as the other VNI STs. None of the A-portion haplotypes of hybrids clustered

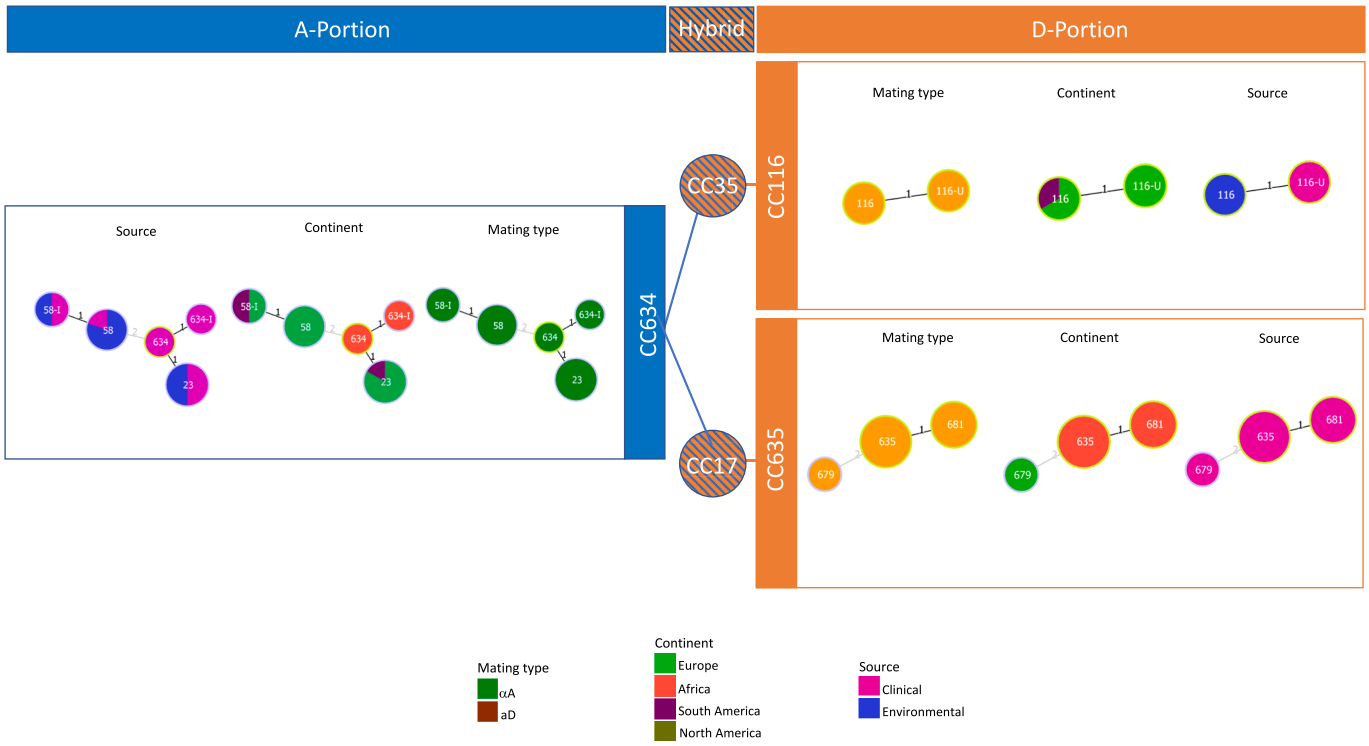

Fig. 5. Parental clonal complexes (CC) generating the main αADα hybrid combinations. Left and right panels report the parental sequence types (ST) belonging to each CC and their correlation with mating types, geographical origin, and source.

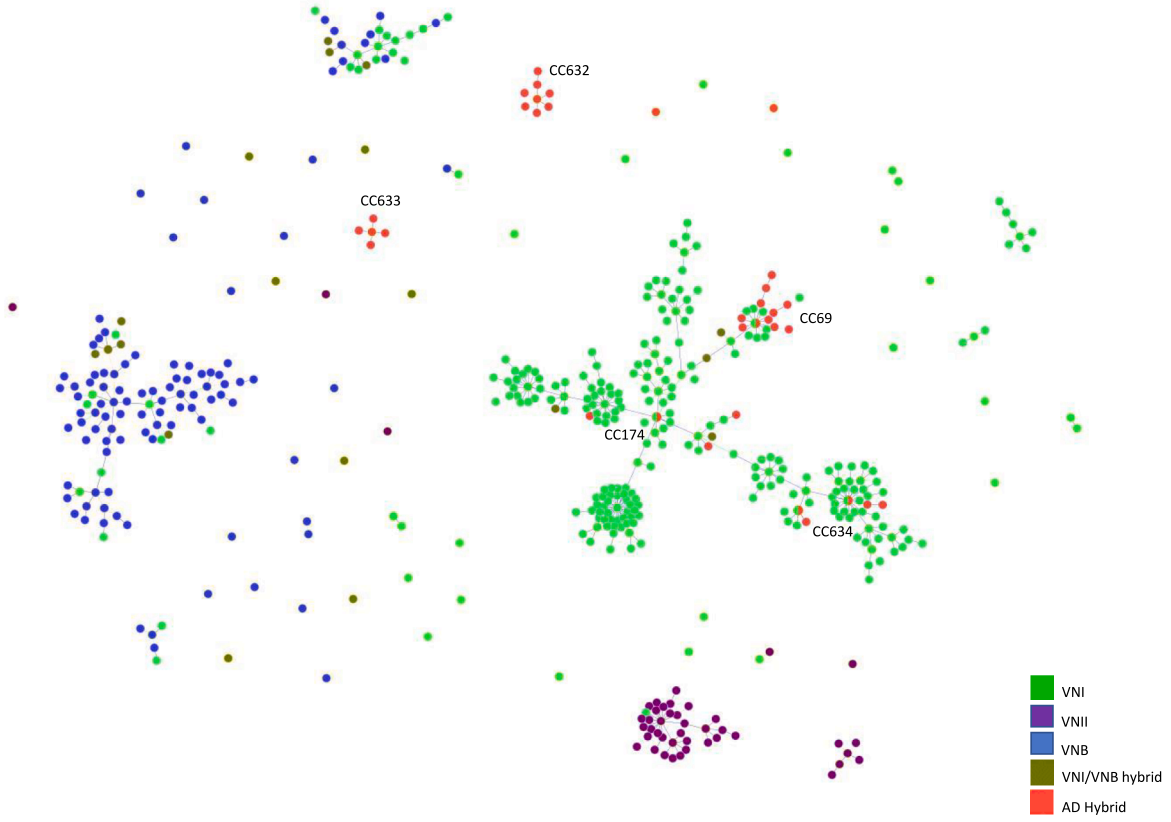

Fig. 6. Minimum spanning tree inferred by GoeBurst algorithm showing the relationship between the A-parent sequence types (ST) of hybrids and the 540 haploid STs of VNI, VNII and VNB molecular types present in the global MLST database. The location of five main clonal complexes of A-portion of hybrids is indicated.

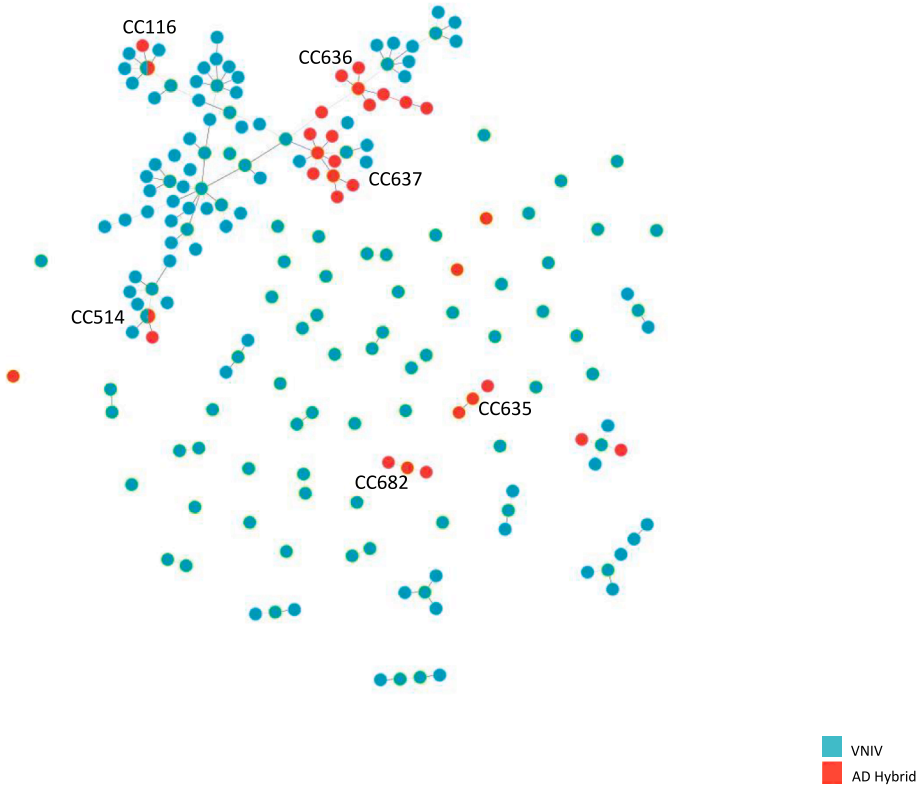

Fig. 7. Minimum spanning tree inferred by GoeBurst algorithm showing the relationship between the D-parent sequence types (ST) of hybrids and the 157 haploid STs of VNIV molecular type present in the global MLST database. The location of the six main clonal complexes of D-portion of hybrids is indicated.

with VNII reference ST. Comparison of haplotypes of D-portion of hybrids with the main VNIV STs included in the global MLST database showed that STs belonging to CC635 (ST635, ST679, and ST681) represented the most ancient cluster among VNIV STs and originated during a range of time from 1 million (ST679) to 150,000 years ago (ST635, ST681). All the other STs were more recent and originated within the last 30,000 years (Fig. 11).

3.10. Genetic analysis of hybrids population

Both A-portion and D-portion in hybrids have similar and moderate values of Hd, k, and Pi confirming that they are grouped in few clusters. In contrast, D-portion has a higher number of polymorphic sites and recombination events than A-portion indicating that it is a more dynamic population (Supplementary Figure S2. A).

Comparison between the A-portion and D-portion showed that they are genetically divergent with a high k value (345.295), extremely low gene flow (Fst = 0.95086), few recombinant events (observed only in *LAC1* and *URA5* loci), and no shared mutations (Supplementary Figure S2. A).

When the A and D portions where compared with the global population, both showed a statistically significant higher value of phylogenetic compatibility (33–38 % vs. 0 %) and association index (0.92–1.29 vs. 0.69–0.3). This means that both the putative A and D parental populations for the hybrids were not random samples of the global A and D populations but were from a limited group of genotypes (Supplementary Table S3).

In the investigated hybrid population, the number of A/D alleles at each locus were relatively few, from five in *SOD1* to 11 in IGS1, much lower than those in the global A and D populations. Interestingly, the five combinations identified in the *SOD1* locus were correlated to different hybrid CCs (Supplementary Figure S2. B-C). Linkage disequilibrium between loci ($p < 0.001$) revealed significant departure from random mating for the serotypes A and D populations, confirming that only specific combinations of alleles at different loci from the natural haploid populations are present in our population of hybrids.

4. Discussion

Genotyping by MLST of 102 AD hybrids revealed the presence of 52 HSTs which were not equally represented among isolates, with 40 % of them belonging to five HSTs. These results indicate a low genotypic variability among AD hybrids, due potentially to a high clonal reproduction and dispersal ability of a few hybrids (Lengeler et al., 2001; Michelotti et al., 2022; Samarasinghe et al., 2020a). It is likely that only few hybrid genotypes are able to survive to the environmental stresses and reproduce by asexual cycle for long time. In addition, our results showed that most of the investigated hybrids (62 %) presented loss of heterozygosity (LOH). This mechanism was previously observed by the genetic analysis of progenies from experimental crossing of A and D strains, and by genotyping of few natural isolates (Dong et al., 2019; Cogliati et al., 2012; Lengeler et al., 2001; Li et al., 2012; Litvinseva et al., 2007; Michelotti et al., 2022). Therefore, this is the first report on high prevalence of LOH among a large number of natural AD hybrids by genotype analysis. The separate analysis of the portion A and D of hybrids showed that missing alleles were observed in both portions in a similar rate (about 50 %) suggesting that LOH was relatively balanced and random, and neither of the two parental genomes was preferentially conserved. Previous studies reported that, in experimental crossings, AD-hybrid progeny tends to keep alleles from the mitochondrial-donor parent (mating type **a** parent) (Samarasinghe et al., 2020a). Therefore, the similar percentage of LOH observed in our study in the two parental genomes likely reflect the equal proportion in the investigated populations of **a**ADα and αAD**a** isolates. Interestingly, MLST identified only

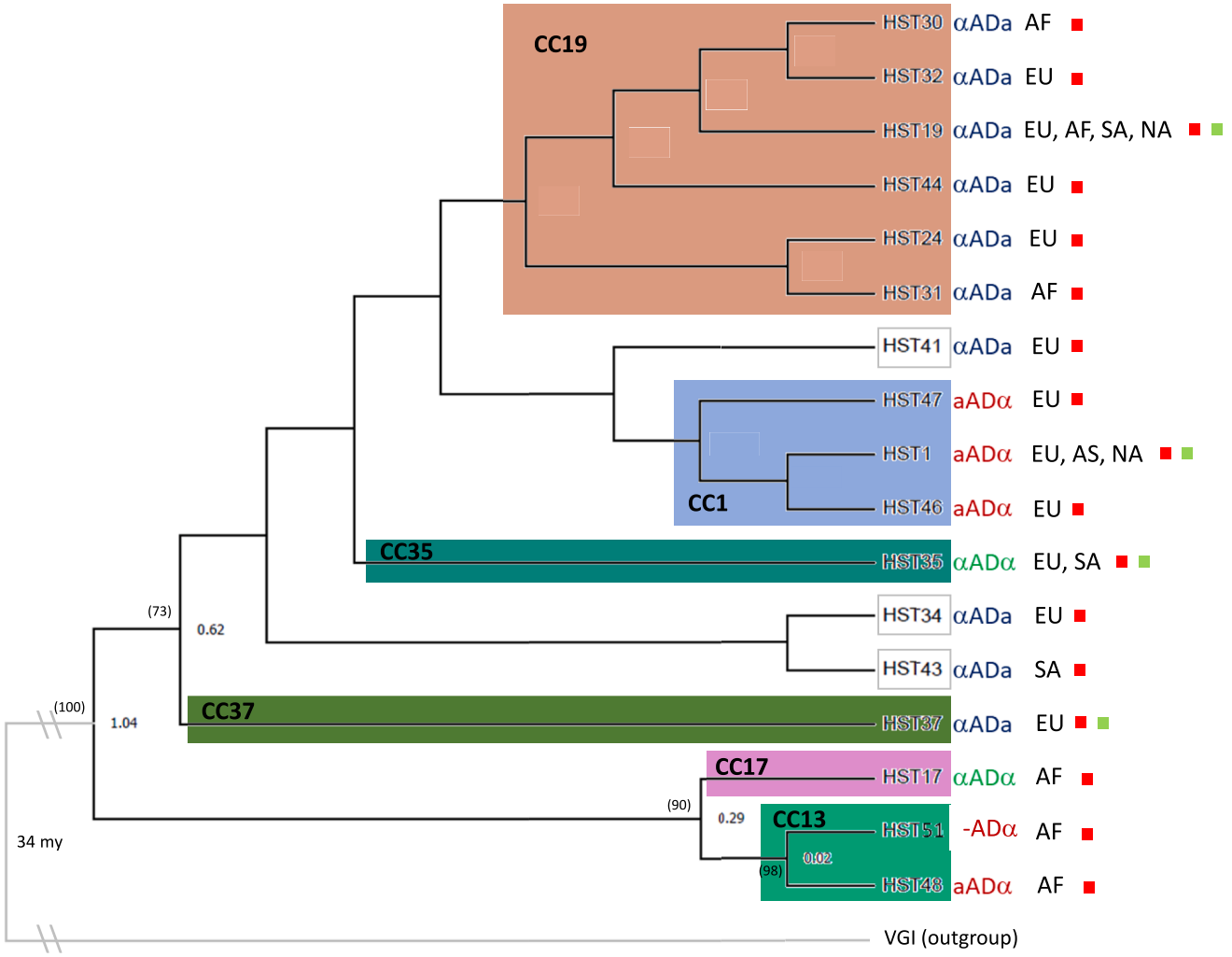

Fig. 8. Maximum likelihood phylogenetic tree inferred by aligning the seven-loci concatenated sequences of the AD hybrids. ST51 belonging to VGI molecular type was artificially duplicated and used as outgroup. Divergence time (million years) from present (number at rightside of the node) and bootstrap values (numbers in the brackets), obtained after 1000 replicates, are reported only for nodes with a bootstrap value above 70 %. The different hybrid clonal complexes (CC) are highlighted with colored boxes. White boxes indicate singleton hybrid sequence types (HST). The small red squares on the right side indicate that the HST included clinical isolates, whereas the green ones environmental isolates. EU = Europe; AF = Africa; AS = Asia; SA = South America; NA = North America. The length of the branches is not proportional to the divergence time to better represent the genetic relationships between the different HSTs. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

five new ATs while the majority of the ATs were already represented in the global haploid MLST database, indicating that rare mutations don't seem to play a primary role in the evolution of hybrids. However, LOH could be over-estimated in our analysis due to possible false negative PCRs and, therefore, future investigations on AD hybrids could benefit by the application of more sensitive techniques such as whole genome analysis and comparative genome hybridization.

Cluster analysis identified different CCs which were associated with specific mating types and geographical regions. Particularly, CC1, grouping most of the **a**ADα isolates, did not include any African isolate although they represented 18 % of the investigated population. On the contrary, most of them were clustered in CC13 suggesting that this latter cluster could be endemic in Africa. Similarly, the CC19 cluster, grouping most of the αAD**a** isolates, did not include any Asian isolate, but in this case the result could be biased by the low number of Asian isolates investigated. Two distinct geographical origins could also be suggested for αADα isolates presenting two clusters, one for Europe (CC35) and one for Africa (CC17). Environmental isolates were present in all hybrid clusters except CC13 and CC17 which, however, included mostly isolates of Africa from which none environmental isolate was incorporated in the study. The close relationship between clinical and environmental isolates, although these latter were underrepresented in our population, in the main hybrid clusters corroborates the fact that hybrids could spread between the two ecological niches. When parental clusters and their combinations to form hybrids were analyzed, a specific cluster in the portion A of hybrids (CC634, αA) was identified to have been originated both bisexual and unisexual hybrid combinations with parental D isolates mostly from Europe or Africa. Interestingly, the **a**D parents of bisexual combinations were all from Europe (CC514) whereas αD parents of unisexual combinations were from Africa (CC635) or Europe and South America (CC116) suggesting that **a**D isolates belonging to CC514 could be endemic in Europe.

Comparison with global haploid isolates showed limited overlap between hybrid parental haplotypes and those in the MLST database. Only six haplotypes (four for portion A and two for portion D) matched haplotypes previously identified. Interestingly, all these haplotypes were found in Europe and Mediterranean area during an extensive population genetic analysis suggesting once again that Europe

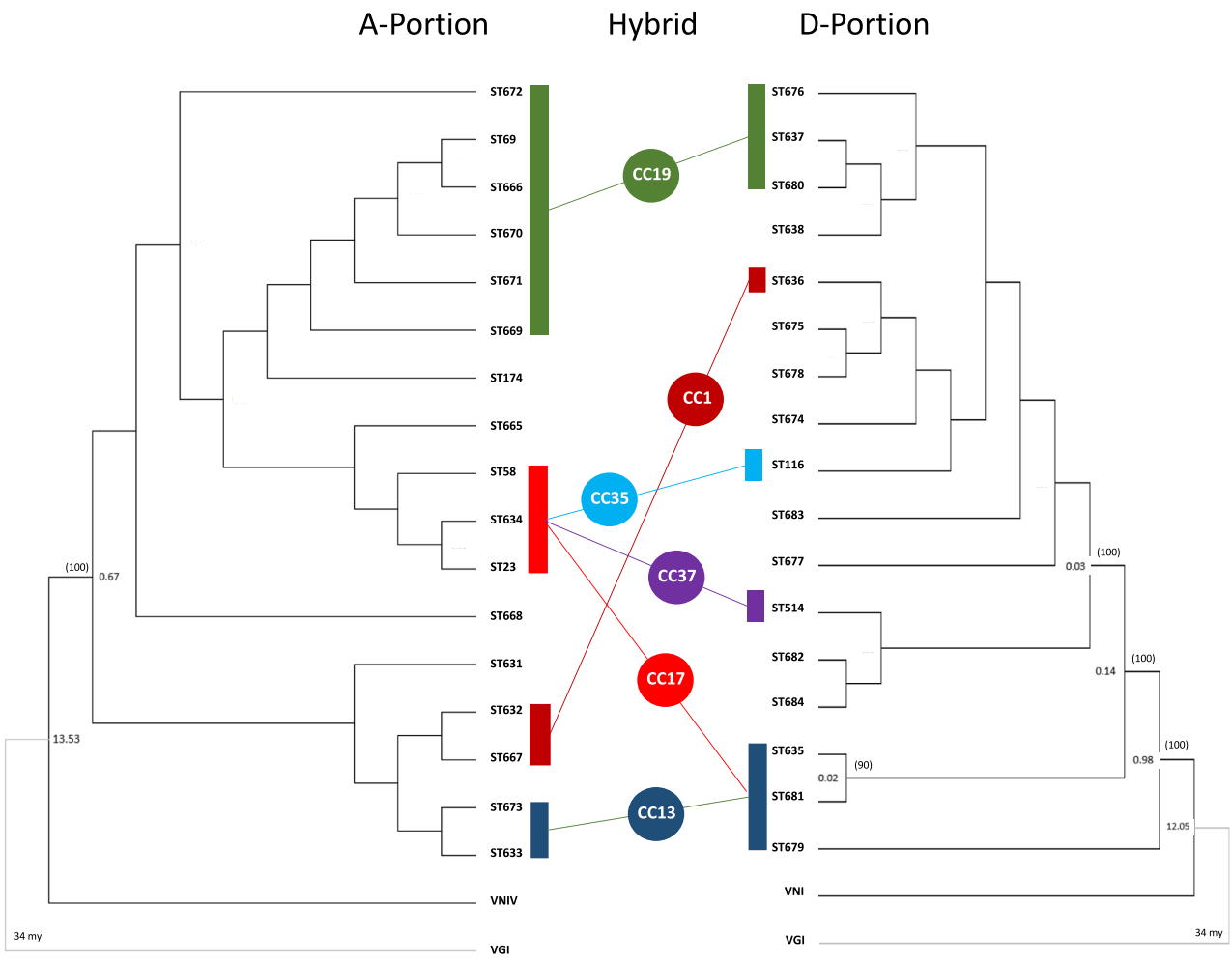

Fig. 9. Maximum likelihood phylogenetic tree inferred by aligning the seven-loci concatenated sequences of the A-portion (left panel) and D-portion (right panel) of AD hybrids. ST51 belonging to VGI molecular type, ST31 belonging to VNI molecular type, and ST121 belonging to VNIV molecular type were used as outgroups. Divergence time (million years) from present (number at rightside of the node) and bootstrap values (numbers in the brackets), obtained after 1000 replicates, are reported only for nodes with a bootstrap value above 70 %. The different hybrid clonal complexes (CC) are represented with circles of different colors and linked by lines to the corresponding A and D parents. The length of the branches is not proportional to the divergence time to better represent the genetic relationships between the different STs.

represents the most probable geographical origin of many AD hybrids (Cogliati et al., 2019). On the other hand, the majority of the hybrid parental haplotypes presented new allelic combinations that could suggest two possible hypotheses: a) AD hybrid parents belong to ancient genotypes that are now extinct or b) the new MLST profiles are present in environmental haploid strains but they have not yet been identified due to the limited investigations carried out so far on environmental isolates. The two hypotheses can coexist together since the results of the phylogenetic analysis obtained by this study showed that AD hybrids arise from different combinations: a) the fusion of ancient African genotypes (hybrids belonging to CC13) and originating about 1 my ago; b) the fusion of ancient and recent genotypes (hybrids belonging to CC1 and CC17); and c) the fusion of more recent genotypes (hybrids belonging to CC19, CC35, and CC37). In our population, 40 % of isolates belonged to the third combination, 39 % to the second combination and only 12 % to the first combination, suggesting that hybridization is still occurring at present and it is in continuous evolution. Our results confirm the hypothesis formulated in a previous study (Xu et al., 2002) which concluded that AD hybrids have multiple origins. However, the authors analyzed only one gene (*LAC1*) on a small hybrid population (14 strains) isolated during a brief period of time from a limited geographical area, and therefore the study was unable to reach general conclusions on phylogenetic relationships of hybrids. In contrast, the present study was carried out on a large number of AD hybrids, of different geographical origins and sources which were isolated during different time periods. They were sequenced for seven standard loci and compared with a MLST database which includes thousands of MLST profiles. Although the present phylogenetic analysis represents a first step towards the comprehension of *C. neoformans* AD hybrids evolution it remains the bias of the limited number of strains included in the analysis. A larger number of hybrids in future will surely increase the diversity here observed and will add new information to the phylogenetic analysis which could lead to new results. Also, the use of different housekeeping genes could influence the inference of the divergence time calculated in this study.

Comparison of portion A and D concatenated sequences with global STs confirmed that African STs belonging to CC633 (**a**A) and CC635 (αD) were also the most ancient ST among haploid STs. These findings agree with a previous study (Litvintseva et al., 2007) showing that the A portion of some **a**ADα hybrids originated from African **a**A parents and that they belonged to VNB molecular type. Similarly, our results showed that STs from cluster CC633 were all **a**A-VNB isolates mostly from Africa. In addition, we identified a second **a**A cluster (CC632), which originated more recently than CC633 and belonged to VNI molecular

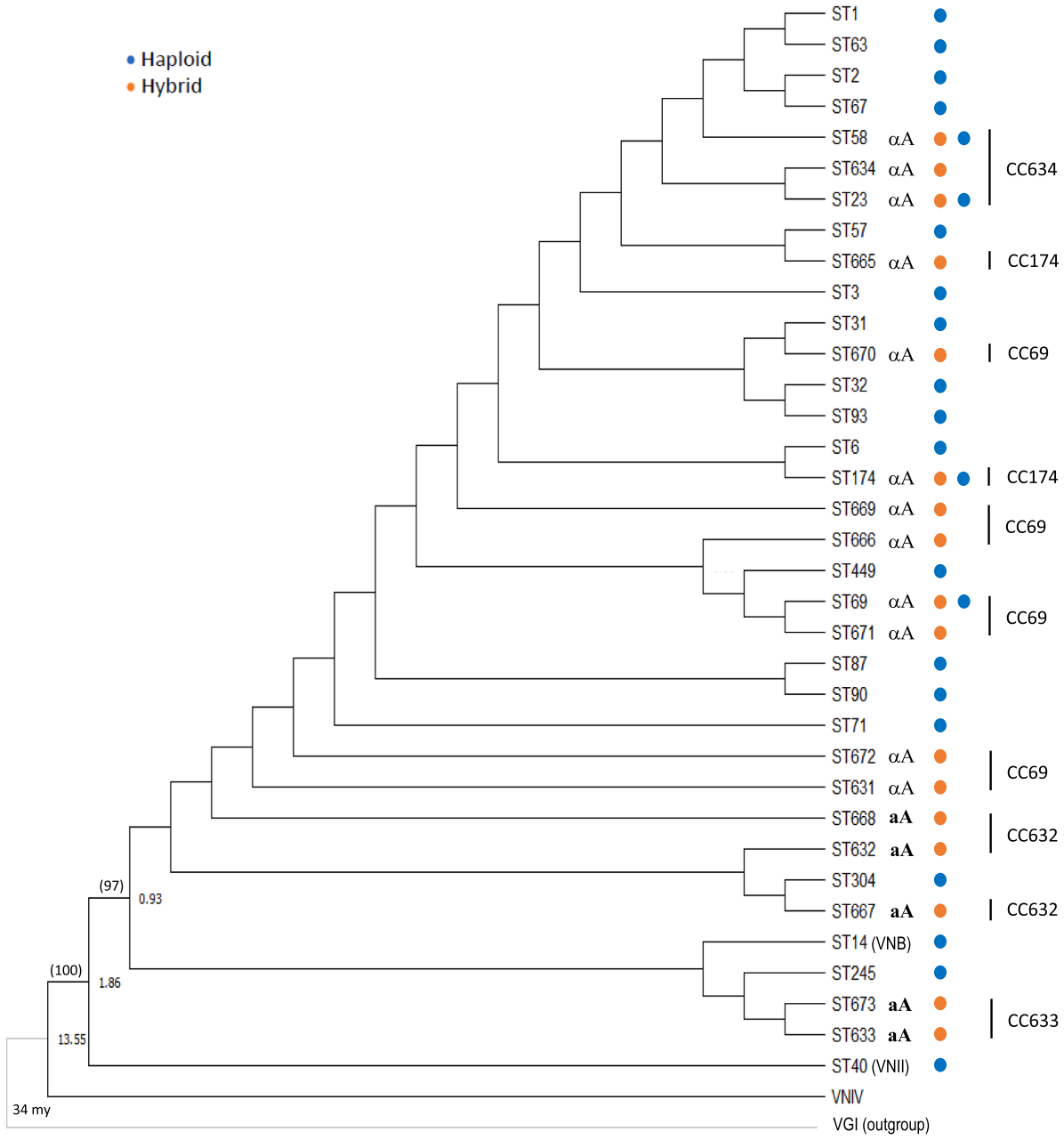

Fig. 10. Maximum likelihood phylogenetic tree inferred by aligning the seven-loci concatenated sequences of the A portion of AD hybrids with the main VNI sequence types (ST) included in the global MLST database as well as ST40 belonging to VNII molecular type and ST14 belonging to VNB molecular type. ST51 belonging to VGI molecular type and ST121 belonging to VNIV molecular type were used as outgroups. Divergence time (million years) from present (number at rightside of the node) and bootstrap values (numbers in the brackets), obtained after 1000 replicates, are reported only for nodes with a bootstrap value above 70 %. STs including only haploid isolates are marked with a blue dot whereas those including only portion A of hybrids with an orange dot. Mating types and clonal complexes are also reported. Length of the branches is not proportional to the divergence time to better represent the genetic relationships between the different STs. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

type. Therefore, a multiple origin of hybrid **a**A parents can be hypothesized again: an ancient origin from Africa, and a more recent origin from other geographical areas. However, a limitation in the inference of hybrids dispersion in our analysis is the non-homogeneous representation of hybrids from the different geographical areas, being isolates from Europe over-represented in our population.

None of the STs from the A-portion of hybrids clustered with the VNII reference ST suggesting that the genome of isolates belonging to this molecular type may not be compatible with VNIV genome and therefore they probably are not able to form hybrids. A recent study supported this hypothesis showing that VNIV vs. VNII isolates had the lowest percentage of shared mutations and the highest percentage of fixed mutations compared to the lower divergence observed for VNIV vs. VNI or VNI vs. VNII (Cogliati et al., 2019). Other alternative explanations affecting mating between VNII and VNIV isolates could be the smaller population size of VNII compared to VNI and the absence/rarity of mating type **a** isolates among VNII population.

Our results support a multiple origin also for the hybrid D parents with isolates belonging to CC635 (αD) and CC682 (**a**D) being the most ancient and the others with a more recent origin. In this latter group, αD

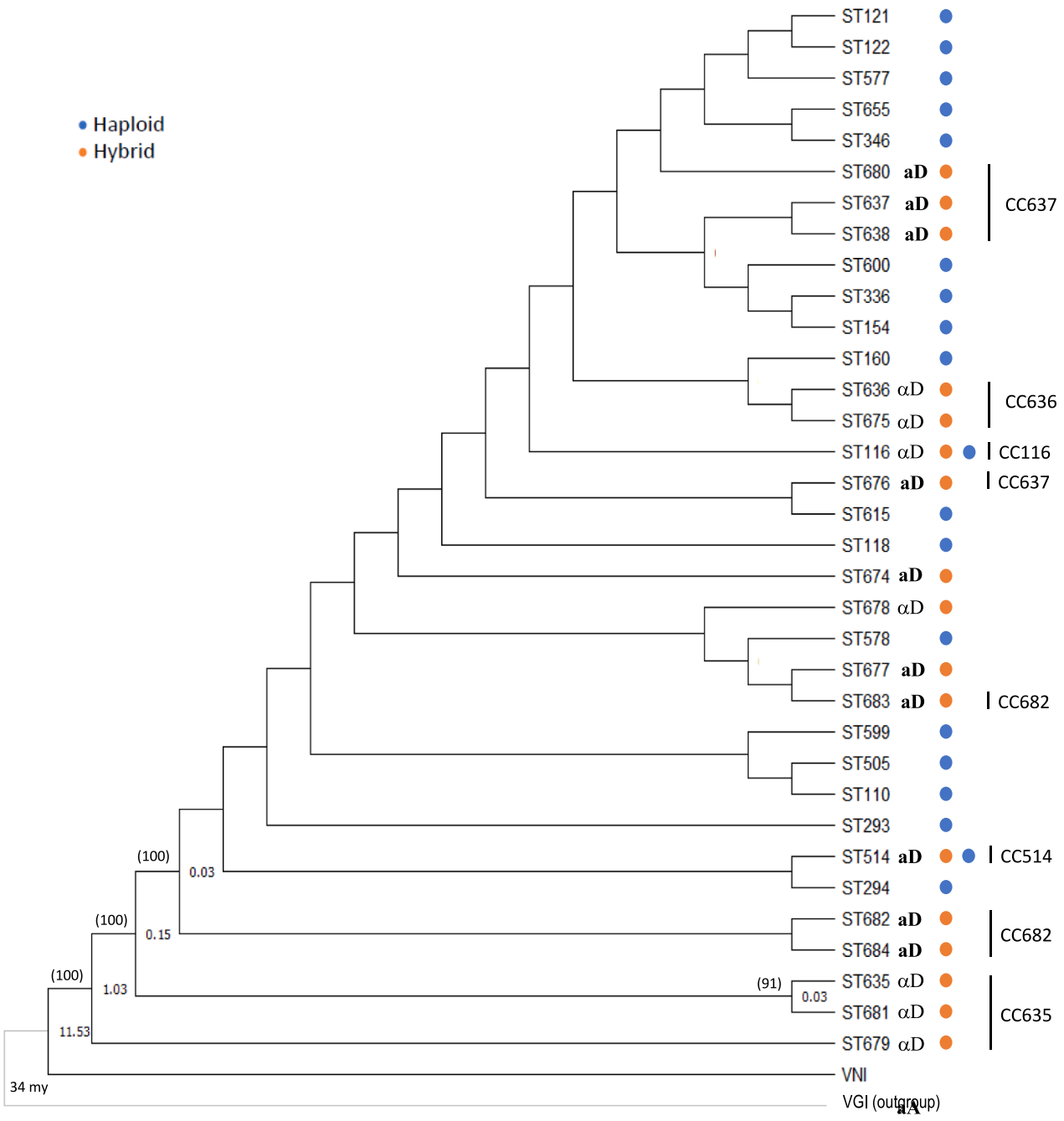

Fig. 11. Maximum likelihood phylogenetic tree inferred by aligning the seven-loci concatenated sequences of the D portion of AD hybrids with the main VNIV sequence types (ST) included in the global MLST database. ST51 belonging to VGI molecular type and ST31 belonging to VNI molecular type were used as outgroups. Divergence time (million years) from present (number at rightside of the node) and bootstrap values (numbers in the brackets), obtained after 1000 replicates, are reported only for nodes with a bootstrap value above 70 %. STs including only haploid isolates are marked with a blue dot whereas those including only portion A of hybrids with an orange dot. Mating types and clonal complexes are also reported. Length of the branches is not proportional to the divergence time to better represent the genetic relationships between the different STs. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

and **a**D STs are frequently mixed in the phylogenetic tree, confirming that sexual recombination is occurring among VNIV population (Cogliati et al., 2016; Hitchcock and Xu, 2022; Li et al., 2012).

Finally, population genetics analysis confirmed the genetic divergence between the A and D portions of hybrids and showed limited gene flow and high linkage disequilibrium within the hybrid population. These results show that A and D parents of hybrids were from specific genotypes rather than randomly in the haploid A and D populations and are in agreement with the cluster analysis discussed above. In addition, both percentage of phylogenetic compatibility and association index corroborated this conclusion as these values were significantly higher ($p < 0.001$) in A and D portion of hybrids than in the global population.

5. Conclusions

This study highlights the genetic diversity and evolutionary history of AD hybrids and provides insights into their population structure and dynamics. The results strongly support the hypothesis of a multiple

origin of AD hybrids both temporally, with ancient and recent hybridization events, and spatially with a distinct cluster from Africa and others from Europe and the other geographical areas. Hybridization did not occur randomly but between few clusters of parental strains representing therefore a limit in the genetic evolution process. Few parental haplotypes of hybrids matched with those present in global MLST database suggesting that *C. neoformans* genotypical diversity is greatly underestimated and that most of this diversity is hidden in the environment. Loss of heterozygosity was detected in most of the AD hybrids investigated confirming that it is a major evolutionary mechanism adopted by these fungi. The present study paves the way for future investigations including a larger number of AD hybrids.

6. Authors contributions

Cogliati M. contributed in conceptualizing the research, data curation and analysis, developing of methodology, coordination of the research group, and writing, reviewing and editing of the manuscript. All the other authors equally contributed in the collection and analysis of data, and writing and reviewing of the manuscript.

7. Funding

D-H Yang and KJ Kwon-Chung were supported by a research fund from the intramural program of the National Institute of Allergy and Infectious Diseases, National Institutes of Health, NIH of the US.

Min Chen was funded by the National Natural Science Foundation of China (No.82172291).

Megan Hitchcock and Jianping Xu were supported by the Natural Science and Engineering Research Council of Canada (RGPIN-2020-05732).

U. Nawrot was supported by Wroclaw Medical University statutory funding SUBZ.D230.23.030.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Appendix A. Supplementary material

Supplementary data to this article can be found online at https://doi.org/10.1016/j.fgb.2023.103861.

References

Agapow, P.M., Burt, A., 2001. Indices of multilocus linkage disequilibrium. Mol. Ecol. Notes 1 (1–2), 101–102.

Bovers, M., Hagen, F., Kuramae, E.E., et al., 2006. Unique hybrids between the fungal pathogens *Cryptococcus neoformans* and *Cryptococcus gattii*. FEMS Yeast Res. 6 (4), 599–607.

Cogliati, M., 2013. Global molecular epidemiology of *Cryptococcus neoformans* and *Cryptococcus gattii*: an atlas of the molecular types. Scientifica (cairo) 2013, 675213.

Cogliati, M., Viviani, M.A., Lin, X., 2011. Hybridization and Its Importance in the *Cryptococcus* species complex. In: Heitman, J. (Ed.), *Cryptococcus*: from Human Pathogen to Model Yeast. ASM Press, Washington, DC, pp. 359–370.

Cogliati, M., Barchiesi, F., Spreghini, E., et al., 2012. Heterozygosis and pathogenicity of *Cryptococcus neoformans* AD-hybrid isolates. Mycopathologia 173 (5–6), 347–357.

Cogliati, M., Zani, A., Rickerts, V., et al., 2016. Multilocus sequence typing analysis reveals that *Cryptococcus neoformans* var. *neoformans* is a recombinant population. Fungal Genet. Biol. 87, 22–29.

Cogliati, M., Desnos-Ollivier, M., McCormick-Smith, I., et al., 2019. Genotypes and population genetics of *Cryptococcus neoformans* and *Cryptococcus gattii* species complexes in Europe and the Mediterranean area. Fungal Genet. Biol. 129, 16–29.

Cogliati, M., Roger, F., Meyer, W., et al., 2020. New multilocus sequence typing primers to enable genotyping of AD hybrids within the *Cryptococcus neoformans* species complex. Med. Mycol. 58 (7), 1005–1009.

Dong, K., You, M., Xu, J., 2019. Genetic changes in experimental populations of a hybrid in the *Cryptococcus neoformans* species complex. Pathogens 9 (1), 3.

D'Souza, C.A., Kronstad, J.W., Taylor, G., et al., 2011. Genome variation in *Cryptococcus gattii*, an emerging pathogen of immunocompetent hosts. MBio 2 (1), e00342–e00410.

Farrer, R.A., Chang, M., Davis, M.J., et al., 2019. A new lineage of Cryptococcus gattii (VGV) discovered in the central Zambezian Miombo woodlands. MBio 10 (6), e02306–e02319.

Francisco, E.C., de Jong, A.W., Hagen, F., 2021. Cryptococcosis and *Cryptococcus*. Mycopathologia 186 (5), 729–731.

Hagen, F., Khayhan, K., Theelen, B., et al., 2015. Recognition of seven species in the *Cryptococcus gattii/Cryptococcus neoformans* species complex. Fungal Genet. Biol. 78, 16–48.

Hagen, F., Lumbsch, H.T., Arsic Arsenijevic, V., et al., 2017. Importance of resolving fungal nomenclature: the case of multiple pathogenic species in the *Cryptococcus* genus. mSphere 2 (4), e00238–e00317.

Hitchcock, M., Xu, J., 2022. Analyses of the global multilocus genotypes of the human pathogenic yeast *Cryptococcus neoformans* species complex. Genes (Basel) 13 (11), 2045.

Kwon-Chung, K.J., Bennett, J.E., Wickes, B.L., et al., 2017. The case for adopting the "species complex" nomenclature for the etiologic agents of cryptococcosis. mSphere 2 (1), e00357–e00416.

Lengeler, K.B., Cox, G.M., Heitman, J., 2001. Serotype AD strains of *Cryptococcus neoformans* are diploid or aneuploid and are heterozygous at the mating-type locus. Infect. Immun. 69, 115–122.

Li, W., Averette, A. F., Desnos-Ollivier, M., et al., 2012. Genetic diversity and genomic plasticity of *Cryptococcus neoformans* AD hybrid strains. G3 (Bethesda) 2(1), 83-97.

Litvintseva, A.P., Lin, X., Templeton, I., et al., 2007. Many globally isolated AD hybrid strains of *Cryptococcus neoformans* originated in Africa. PLoS Pathog. 3 (8), e114.

Michelotti, L.A., Sun, S., Heitman, J., et al., 2022. Clonal evolution in serially passaged *Cryptococcus neoformans* × *deneoformans* hybrids reveals a heterogenous landscape of genomic change. Genetics 220 (1), iyab142.

Monroy-Nieto, J., Bowers, J.R., Montfort, P., et al., 2022. Phylogenomic placement of American Southwest-associated clinical and veterinary isolates expands evidence for distinct *Cryptococcus gattii* VGVI. Microorganisms 10 (8), 1681.

Pappas, P.G., 2013. Cryptococcal infections in non-HIV-infected patients. Trans. Am. Clin. Climatol. Assoc. 124, 61–79.

Rajasingham, R., Govender, N.P., Jordan, A., et al., 2022. The global burden of HIV-associated cryptococcal infection in adults in 2020: a modelling analysis. Lancet Infect. Dis. 22 (12), 1748–1755.

Samarasinghe, H., Vogan, A., Pum, N., et al., 2020a. Patterns of allele distribution in a hybrid population of the *Cryptococcus neoformans* species complex. Mycoses 63 (3), 275–283.

Samarasinghe, H., Xu, J., 2018. Hybrids and hybridization in the *Cryptococcus neoformans* and *Cryptococcus gattii* species complexes. Infect. Genet. Evol. 66, 245–255.

Samarasinghe, H., You, M., Jenkinson, T.S., et al., 2020b. Hybridization facilitates adaptive evolution in two major fungal pathogens. Genes (Basel) 11 (1), 101.

Velagapudi, R., Hsueh, Y.P., Geunes-Boyer, S., et al., 2009. Spores as infectious propagules of *Cryptococcus neoformans*. Infect. Immun. 77 (10), 4345–4355.

Viviani, M.A., Wen, H., Roverselli, A., et al., 1997. Identification by polymerase chain reaction fingerprinting of *Cryptococcus neoformans* serotype AD. J. Med. Vet. Mycol. 35 (5), 355–360.

Williamson, P.R., Jarvis, J.N., Panackal, A.A., et al., 2017. Cryptococcal meningitis: epidemiology, immunology, diagnosis and therapy. Nat. Rev. Neurol. 13 (1), 13–24.

Xu, J., Luo, G., Vilgalys, R.J., et al., 2002. Multiple origins of hybrid strains of *Cryptococcus neoformans* with serotype AD. Microbiology (reading) 148 (1), 203–212.

Yang, D.H., England, M.R., Salvator, H., et al., 2021. *Cryptococcus gattii* species complex as an opportunistic pathogen: underlying medical conditions associated with the infection. MBio 12 (5), e0270821.

You, M., Lin, Y.M., Dobrin, A., et al., 2021. Genetic and phenotypic diversities in experimental populations of diploid inter-lineage hybrids in the human pathogenic *Cryptococcus*. Microorganisms 9 (8), 1579.