Research Square

Preprints are preliminary reports that have not undergone peer review.
They should not be considered conclusive, used to inform clinical practice,
or referenced by the media as validated information.

# Long-read full-length rRNA–ITS sequencing unlocks strain-level gut bacterial and fungal microdiversity in paediatric autism spectrum disorder

**Huabin Xu**
Clinical Laboratory Center, Affiliated Hospital of West Anhui Health Vocational College
**Shizhu Zheng**
Department of Pediatric Rehabilitation, Affiliated Hospital of West Anhui Health Vocational College
**Linling Chen**
Genome Decoding Institute
**Jianhe Sun**
Genome Decoding Institute
**Weiwei Zhang**
Genome Decoding Institute
**Lei Liu**
Kindstar Sequenon Biotechnology (Wuhan) Co.,Ltd
**Wen Zhang**
Department of Medical Technology, West Anhui Health Vocational College
**Guangman Zhang**
Department of Transfusion Medicine, Affiliated Hospital of West Anhui Health Vocational College
**Xiaoyan Zhang**
Department of Clinical Laboratory, Affiliated Hospital of West Anhui Health Vocational College
**Chengjun Sun**
Department of Pediatric Rehabilitation, Affiliated Hospital of West Anhui Health Vocational College
**Shaolong Jiang**
Kindstar Sequenon Biotechnology (Wuhan) Co.,Ltd
**Jumei Huang**
Genome Decoding Institute
**Fangping Dai**
daifp@genomedecoding.com

Genome Decoding Institute

**Article**

**Keywords:** long-read full-length rRNA–ITS sequencing, strain-level resolution, gut bacterial and fungal microbiota, Candida microdiversity, paediatric autism spectrum disorder

**Posted Date:** November 23rd, 2026

**DOI:** https://doi.org/10.21203/rs.3.rs-11570981/v1

**License:** This work is licensed under a Creative Commons Attribution 4.0 International License.
Read Full License

**Additional Declarations:** No competing interests reported.

# Long-read full-length rRNA–ITS sequencing unlocks strain-level gut bacterial and fungal microdiversity in paediatric autism spectrum disorder

**Huabin Xu[1,*#], Shizhu Zheng[2,*], Linling Chen[3], Jianhe Sun[3], Weiwei Zhang[3], Lei Liu[4], Wen Zhang[5], Guangman Zhang[6], Xiaoyan Zhang[1], Chengjun Sun[2], Shaolong Jiang[4], Jumei Huang[3], Fangping Dai[3#]**

1. **Clinical Laboratory Center, Affiliated Hospital of West Anhui Health Vocational College, Luan, Anhui Province, China**
2. **Department of Pediatric Rehabilitation, Affiliated Hospital of West Anhui Health Vocational College, Luan, Anhui Province, China**
3. **Genome Decoding Institute, Nantong, Jiangsu Province，China**
4. **Kindstar Sequenon Biotechnology Co.,Ltd, Wuhan, Hubei Province, China**
5. **Department of Medical Technology, West Anhui Health Vocational College, Luan, Anhui Province, China**
6. **Department of Transfusion Medicine, Affiliated Hospital of West Anhui Health Vocational College, Luan, Anhui Province, China**

***Co-first authors.**

**#Correspondent authors:**

**Prof. Dr.Huabin Xu, xuhuabin@wahvc.edu.cn**

**Prof. Dr. Fangping Dai, daifp@genomedecoding.com**

# Abstract

**Autism spectrum disorder (ASD) is a childhood neurodevelopmental disorder tightly linked to gut microbial dysbiosis, yet strain-level gut eukaryotic and bacterial signatures associated with ASD remain poorly resolved by conventional short-read sequencing. Here, we targeted the full-length bacterial 16S-rRNA–ITS–p.23S-rRNA and complete eukaryotic 18S-rRNA–ITS1–5.8S-rRNA–ITS2–p.25/28S-rRNA regions and performed high-coverage long-read sequencing on the PacBio Revio platform. Sequencing generated 10,650 bacterial operational taxonomic units (OTUs) with an average read length of 2,433 bp and 5,310 fungal OTUs with an average read length of 3,007 bp from gut samples of children with ASD and neurotypical controls. The contiguous linkage of conserved rRNA coding sequences and hypervariable ITS segments greatly improves phylogenetic resolution, enabling robust subspecies- and strain-level microbial discrimination. Comparative analysis revealed marked subspecies-level microbial differences between the two groups. Fungal subspecies profiling identified divergent alterations in eukaryotic communities: the environmental *basidiomycete Schizophyllum* was enriched in neurotypical controls, whereas multiple *Candida* subspecies—resolved from a total of 209 fungal strains, including *C. glabrata-AJB8*—were selectively overrepresented in the ASD group. Notably, the core gut commensal *C. albicans_SC5314-AJB17* was significantly enriched in children with ASD despite its ubiquitous colonization in neurotypical individuals. Our study demonstrates that full-length rRNA–ITS long-read sequencing is a powerful approach to uncover previously hidden strain-level gut bacterial and eukaryotic signatures associated with ASD, offering refined taxonomic insights for future investigations of ASD and other microbiota-associated disorders.**

## Keywords

**long-read full-length rRNA–ITS sequencing; strain-level resolution; gut bacterial and fungal microbiota; *Candida* microdiversity; paediatric autism spectrum disorder**

# Introduction

The human gut microbial dysbiosis is tightly linked to multiple human disorders, ranging from intestinal functional abnormalities, metabolic disorders to neurodevelopmental conditions such as ASD[1-4]. Benefited from the advances in high-throughput next-generation sequencing (NGS) technologies, the correlative relationships between the human gut microbiome and host disease states have been extensively explored in recent years. The large public DNA sequencing datasets of the human microbiome, which were mainly obtained from the short read NGS from the samples of patient groups and healthy controls, are now available.

Multiple microbiota profiling studies indicates that compositional dysbiosis of the gut bacteriome and mycobiome contributes to ASD pathogenesis[6, 7]. NGS revealed that both the bacteria and fungi (*candida albicans*) were altered in children with ASD[8]. The higher abundances of *Bacteroides* and *Clostridium* species were in the children with ASD compared to their unaffected siblings[9], Certain bacterial taxa were negatively associated with ASD phenotypes[10]. The commensal fungal genus *Candida* serves as a key constituent of the human gut mycobiome. Multiple *Candida* species were observed with relative higher abundance in the guts of children[11]. Commensal fungus *Candida albicans* maintains a long-term mutualistic relationship with the host to modulate gut microbiota and metabolism[12].

Although massive public short-read sequencing datasets are currently available, integrative large-scale analyses remain limited due to the lack of standardized full-length microbial reference sequences, which restricts high-resolution species- and strain-level annotation[5]. The long-read sequencing overcomes the inherent limitations of short-read technology by generating contiguous long-length sequences, which provides superior strain-level community resolution and uncovered novel taxa.[13].

In principle, the full 16S-rRNA-ITS-p.23S-rRNA; the full 18S-rRNA-ITS1-5.8S-rRNA-ITS2-p.25S/28S-rRNA regions for eukaryotes can be sequenced in long reads. We aimed to validate that as a unified barcode system for improved gut bacterial and fungal community annotation. By integrating the conserved phylogenetic signatures of rRNA genes and the hypervariable nature of ITS regions, this long-read sequencing strategy is expected to overcome the taxonomic bottlenecks, enabling

refined profiling of ASD -associated gut microbial dysbiosis at the subspecies and strain levels.

# Results

## Species annotation of the long-read sequencing data

We performed PacBio Revio long-read amplicon sequencing to characterize gut prokaryotic and eukaryotic microbial communities, including 42 children with autism spectrum disorder (ASD) and 14 age-matched healthy control (HC) children. The mean age was 4.9 ± 1.6 years (range, 1.8–8.0 years) for the ASD group and 3.6 ± 1.7 years (range, 1.8–6.0 years) for the HC group. The amplicon libraries were constructed to target the full-length bacterial 16S-rRNA–ITS–p.23S-rRNA region and the complete eukaryotic 18S-rRNA–ITS1–5.8S-rRNA–ITS2–p.28S-rRNA region. Raw long-read data were processed through systematic quality control and HiFi filtering to retain high-quality sequences for downstream analysis.

High-quality prokaryotic data were successfully obtained for all 56 samples (42 ASD and 14 HC). For eukaryotic datasets, valid high-resolution profiles were retained from 55 samples (41 ASD and 14 HC), with one ASD sample excluded due to insufficient sequencing quality. The average lengths of qualified bacterial and eukaryotic amplicons were 2,433 bp and 3,007 bp, OTU clustering and systematic taxonomic annotation identified a total of 10,650 bacterial and 5,310 fungal taxa across all samples, respectively, supporting high-confidence species- and subspecies-level taxonomic resolution. LEfSe-based differential abundance analysis further revealed 1,554 bacterial and 1,342 fungal taxa with altered abundances between the ASD and HC groups, establishing a comprehensive taxonomic dataset for subsequent cross-group microbial comparative analyses.

## The characteristics of gut bacterial communities in ASD and HC groups

After LEfSe statistical test, the 1554 of the 16S-rRNA-ITS-p.23S-rRNA bacterial OTUs were selected, which were further stratified across hierarchical taxonomic ranks, spanning phylum, class, order, family, genus and species. Of these resolved

bacterial OTUs, 868 were precisely annotated to the species/strain level. LEfSe-based linear discriminant analysis (LDA) generally revealed robust bacterial taxonomic biomarkers with intergroup differences between the ASD and HC groups at the order, family, and genus levels (Fig. 1A).

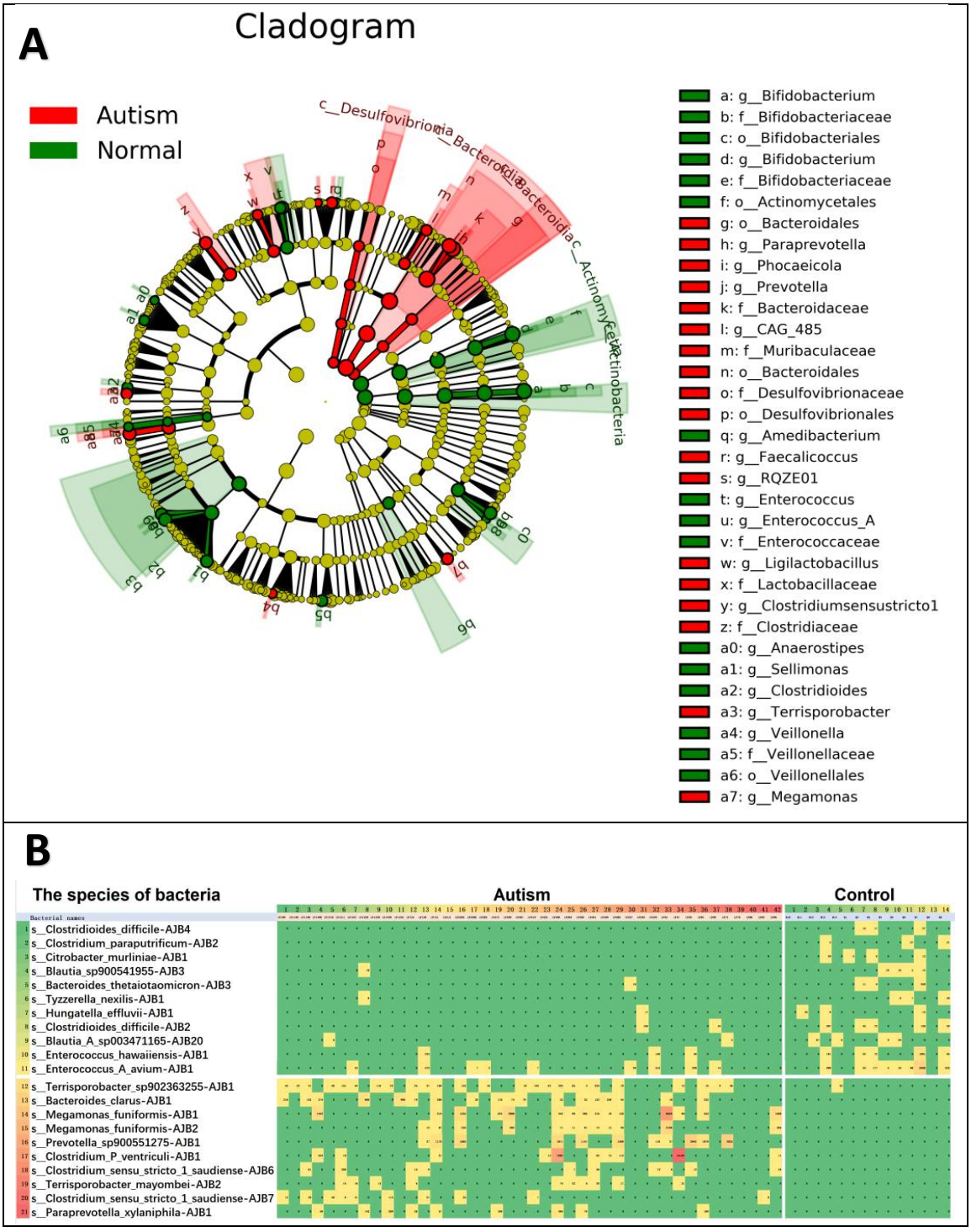
We systematically compared sequencing read abundance, sample detection prevalence, and fold changes across the finely resolved bacterial OTUs annotated at the species level between the two groups. We applied a stringent filtering framework to screen differential microbial taxa, with thresholds defined as: a minimum detection prevalence of 20% in either group,

a pairwise intergroup prevalence ratio greater than 0.2, and an absolute log2 fold change (logFC) of sequencing read abundance exceeding 2. Then, 21 bacterial species/strains were selected (Extended Data Figure. 1). Distinct clustering patterns of these differential bacterial taxa were observed between the ASD and HC groups (Fig. 1A). Taxonomic variations across screened species were visualized using volcano plots based on abundance alterations, and strain-specific enrichment profiles were further illustrated via dot-plot analysis. Several key ASD -enriched bacterial strains, including *Megamonas funiformis-AJB1*, *Bacteroides clarus-AJB1*, *Terrisporobacter sp902363255-AJB1* and *Terrisporobacter mayombei-AJB1*, , exhibited markedly elevated richness in the ASD group relative to HC controls (Extended Data Figure. 2).

We further adopted a more stringent filtering criterion, defined as a group prevalence ratio (ASD/HC) > 3, logFC > 1, logCPM > 12, and $P < 0.05$, which narrowed down the differential profile to 13 bacterial species and strains. Consistent with the above results, Consistent with the above results, *Megamonas funiformis-AJB1*, *Bacteroides clarus-AJB1*, *Terrisporobacter sp902363255-AJB1*, and *Terrisporobacter mayombei-AJB1* were significantly enriched in the ASD group. In contrast, *Blautia sp900541955-AJB3*, *Clostridioides difficile-AJB2*, *Clostridioides difficile-AJB4*, *Clostridium paraputrificum-AJB2*, and *Citrobacter murliniae-AJB1* were predominantly enriched in the HC group (Extended Data Fig. 3).

To further characterize the structural features of these differential bacterial taxa, we performed sequence clustering and comparative analysis based on full-length 16S-rRNA–ITS–p.23S-rRNA amplicons. Complete Escherichia coli 16S rRNA (J01859.1) and 23S rRNA (NR_103073.1) sequences were retrieved as genomic references for

regional variation alignment. Representative ASD -enriched strains, including *Megamonas funiformis-AJB1*, *Bacteroides clarus-AJB1*, *Terrisporobacter mayombei-AJB1*, and *Terrisporobacter sp902363255-AJB1*, were subjected to genetic variation analysis. Comparative sequence alignment demonstrated that the 16S-rRNA and partial 23S-rRNA gene regions were highly conserved across bacterial species, whereas the intervening ITS segments exhibited substantial sequence divergence. These results confirm that long-read sequencing reliably covers the full-length 16S-rRNA–ITS–p.23S-rRNA region. The remarkably higher genetic variability of bacterial ITS regions together with the relatively conserved rRNA gene sequences enables precise discrimination of closely related bacterial species and strains, substantially improving subspecies-level taxonomic resolution compared with conventional short-read strategies, which only target partial regions of the 16S rRNA gene (Extended Data Fig. 4).




**Results Fig.1**

# Figure 1 Legend

**Fig.1 | Differentially bacterial taxa between children with ASD and healthy control groups.**

**a.** Hierarchical taxonomic distribution of differentially abundant gut bacteria across successive taxonomic ranks. Concentric radial circles represent taxonomic levels ranging from phylum to genus, with node diameters scaled according to microbial relative abundance. Yellow nodes indicate non-significant taxa; red and green nodes represent ASD-enriched and HC-enriched bacterial taxa, respectively. Differential taxa were primarily resolved at the order (o), family (f), and genus (g) levels, with limited resolution at the species (s) level.

**b.** Heatmap illustrating two distinct clustering patterns of differential bacterial species in ASD and HC groups. The first cluster contained 11 species (No. 1–11) enriched in the HC group, while the second cluster included 10 species (No. 12–21) enriched in the ASD group.

## Diversity and differential gut eukaryotic species between the children with ASD and healthy control groups.

We clustered high-quality long-read amplicons spanning the 18S rRNA–ITS1–5.8S rRNA–ITS2–partial 28S rRNA region into OTUs and annotated them against curated reference databases (NCBI GenBank, GTDB and SILVA). LEfSe analysis was performed to characterize differential eukaryotic microbial profiles between the ASD and HC groups. In total, 1,342 high-confidence eukaryotic OTUs were retained, comprising 1,263 fungal, 56 plant and 23 unclassified eukaryotic taxa.

To identify gut eukaryotic taxa with group-specific differences, we first applied a prevalence threshold of >50% in at least one group and used linear discriminant analysis (LDA) to quantify the discriminatory power of individual taxa. This analysis identified 20 distinct eukaryotic signatures differentiating the two groups ($P < 0.05$). Notably, *Candida albicans*_SC5314-AJB17 (LDA = −4.10) was the most strongly enriched taxon in the ASD group relative to HCs (Extended Data Fig. 5).

Given the substantial candidate differences suggested by LEfSe, we next examined the abundance and prevalence of each species in detail. Taxa were considered differentially represented between ASD and HC groups when both abundance and prevalence differed by more than two fold. We additionally required a prevalence >20% in at least one group to exclude rare, sporadic sample detection. Increasing this prevalence cutoff reduced the number of retained taxa. At a prevalence threshold >24% in at least one group, 55 taxa remained: 8 taxa detected exclusively in HCs, 43 taxa present in both ASD and HC groups, and 4 taxa restricted to the ASD group (Fig. 2A).

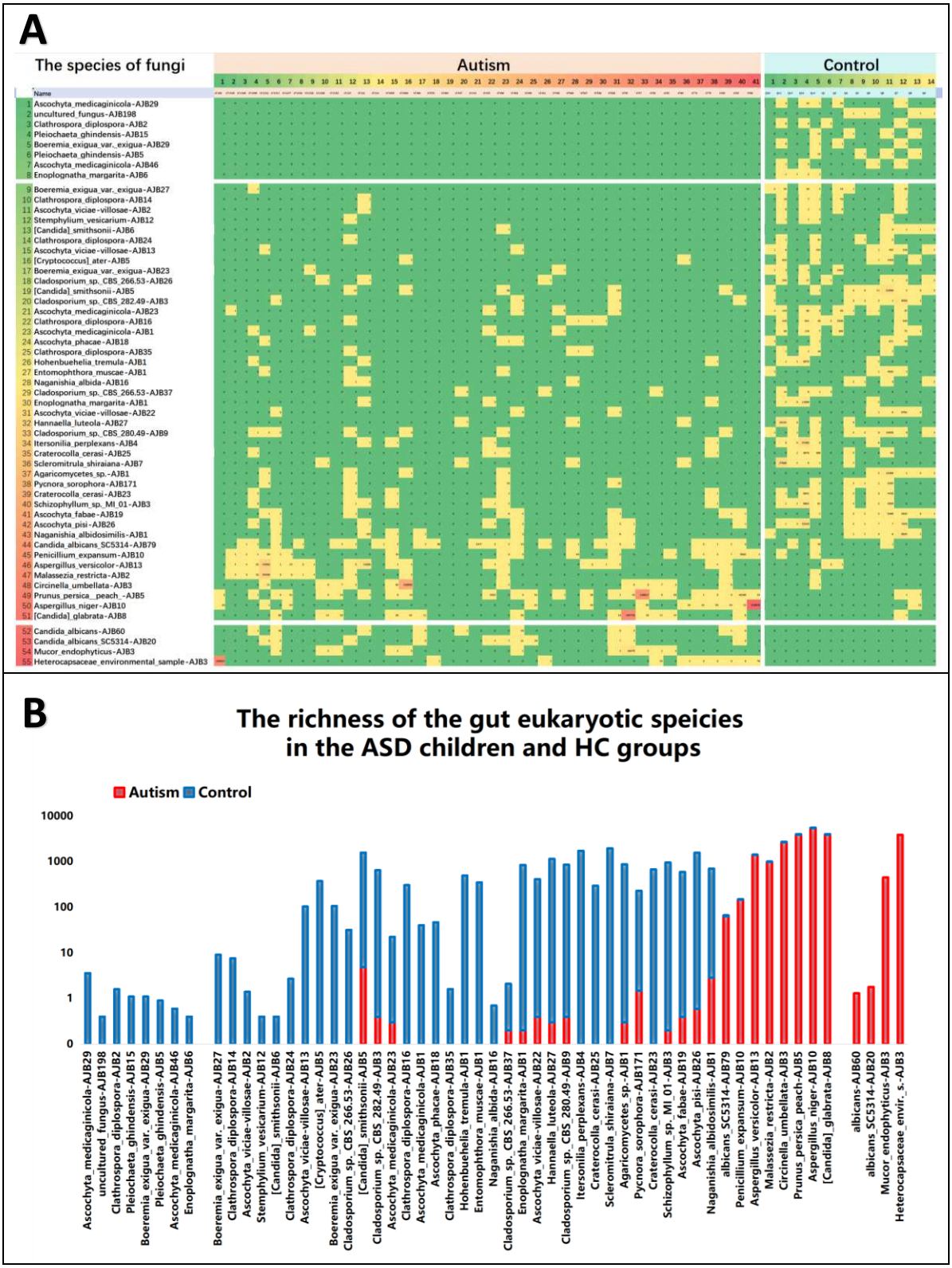
Quantitative profiling of these 55 differential taxa revealed marked group-specific enrichment patterns among gut eukaryotes. Specifically, *Candida glabrata*-AJB8 exhibited the strongest ASD-biased prevalence ratio, with a 4.44-fold higher prevalence relative to the HC group. Several fungal taxa, including *Candida albicans-AJB60* (24%), *Candida albicans_SC5314-AJB20* (27%), *Mucor endophyticus-AJB3* (27%) and *Heterocapsaceae environmental sample-AJB3* (24%), were detected only in ASD individuals and absent from all healthy controls. Conversely, *Boeremia exigua var. exigua-AJB27* represented the most prominent HC-enriched eukaryotic species with the largest prevalence ratio favoring the HC

group.

We performed rigorous comparative statistics for all 55 taxa using unified filtering thresholds (|logFC| > 2.30, logCPM > 6.97, $P < 0.05$). Of these candidates, 49 taxa also satisfied an FDR < 0.05 threshold, indicating robust enrichment or depletion associated with ASD. *Candida glabrata-AJB8*, *Heterocapsaceae environmental sample-AJB3*, *Circinella umbellata-AJB3* and *Mucor endophyticus-AJB3* showed significantly higher abundance in ASD relative to HC ($P < 0.05$, FDR < 0.05) (Fig. 2B).

Despite using identical primers, we recovered abundant long-read amplicons of the 18S rRNA–ITS1–5.8S rRNA–ITS2–partial 28S rRNA marker from diverse eukaryotic sources. These sequences originated not only from gut commensal fungi but also from eukaryotes likely derived from environmental saprotrophs and dietary plant/fruit materials. Long-read sequencing enabled accurate species-level taxonomic resolution across this broad eukaryotic spectrum.

We reconstructed phylogenetic relationships for the 55 differential eukaryotic species using the EMBL-EBI Clustal Omega web server (https://www.ebi.ac.uk/jdispatcher), which revealed distinct evolutionary clustering. As a result, one independent phylogenetic branch contained taxa originating from environmental saprotrophs, including *Schizophyllum*, *Craterocolla cerasi* and *Hohenbuehelia tremula*, while a separate clade comprised multiple differentially abundant *Candida* subspecies with pronounced intergroup differences in abundance and prevalence (Fig. 3).

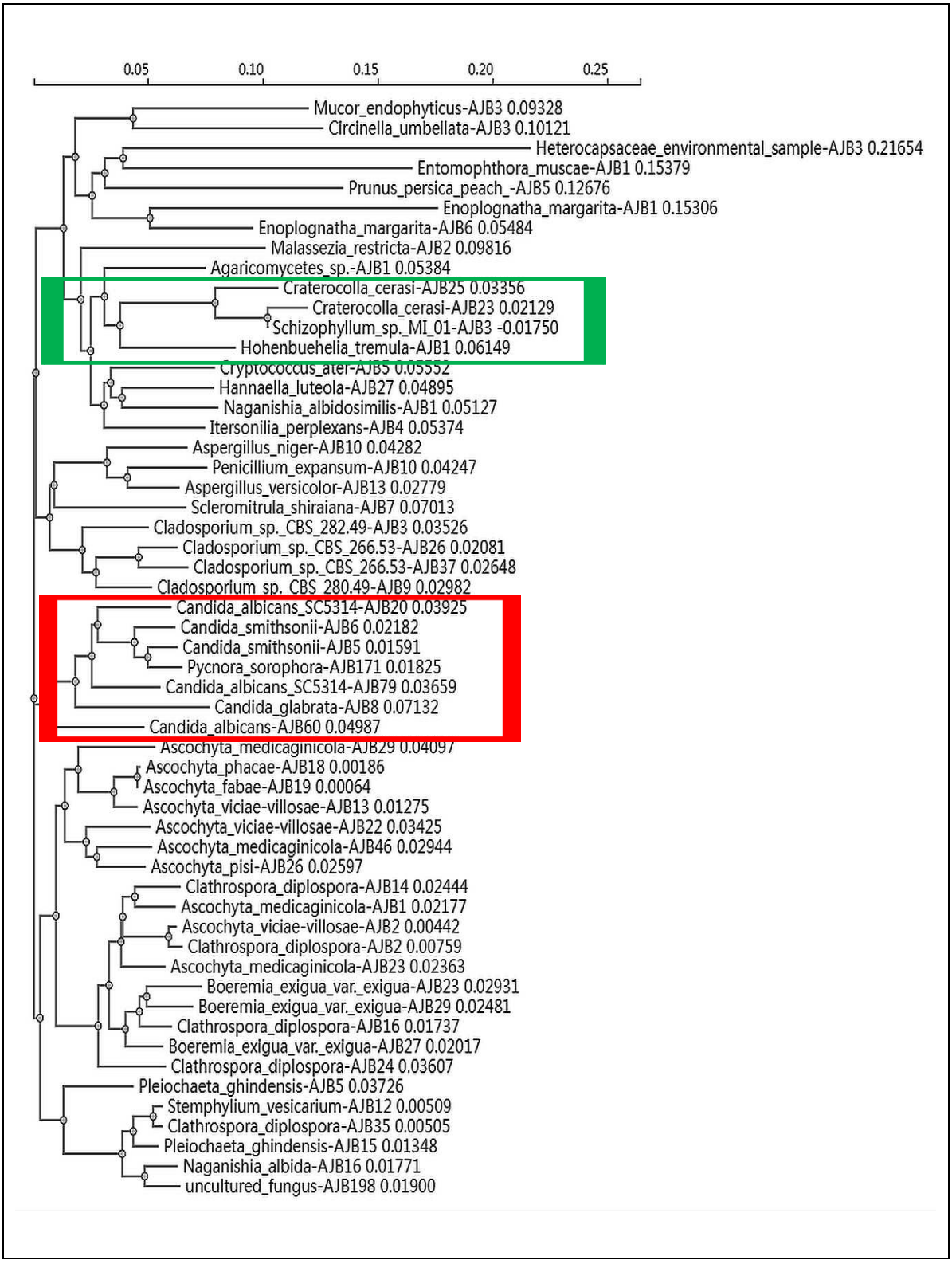

A
The species of fungi
Autism
Control
Name
1 Ascochyta_medicaginicola-AJB29
2 uncultured_fungus-AJB198
3 Clathrospora_diplospora-AJB2
4 Pleiochaeta_ghindensis-AJB15
5 Boeremia_exigua_var._exigua-AJB29
6 Pleiochaeta_ghindensis-AJB5
7 Ascochyta_medicaginicola-AJB46
8 Enoplognatha_margarita-AJB6
9 Boeremia_exigua_var._exigua-AJB27
10 Clathrospora_diplospora-AJB14
11 Ascochyta_viciae-villosae-AJB2
12 Stemphylium_vesicarium-AJB12
13 [Candida]_smithsonii-AJB6
14 Clathrospora_diplospora-AJB24
15 Ascochyta_viciae-villosae-AJB13
16 [Cryptococcus]_ater-AJB5
17 Boeremia_exigua_var._exigua-AJB23
18 Cladosporium_sp._CBS_266.53-AJB26
19 [Candida]_smithsonii-AJB5
20 Cladosporium_sp._CBS_282.49-AJB3
21 Ascochyta_medicaginicola-AJB23
22 Clathrospora_diplospora-AJB16
23 Ascochyta_medicaginicola-AJB1
24 Ascochyta_phacae-AJB18
25 Clathrospora_diplospora-AJB35
26 Hohenbuehelia_tremula-AJB1
27 Entomophthora_muscae-AJB1
28 Naganishia_albida-AJB16
29 Cladosporium_sp._CBS_266.53-AJB37
30 Enoplognatha_margarita-AJB1
31 Ascochyta_viciae-villosae-AJB22
32 Hannaella_luteola-AJB27
33 Cladosporium_sp._CBS_280.49-AJB9
34 Itersonilia_perplexans-AJB4
35 Craterocolla_cerasi-AJB25
36 Scleromitrula_shiraiana-AJB7
37 Agaricomycetes_sp.-AJB1
38 Pycnora_sorophora-AJB171
39 Craterocolla_cerasi-AJB23
40 Schizophyllum_sp._MI_01-AJB3
41 Ascochyta_fabae-AJB19
42 Ascochyta_pisi-AJB26
43 Naganishia_albidosimilis-AJB1
44 Candida_albicans_SC5314-AJB79
45 Penicillium_expansum-AJB10
46 Aspergillus_versicolor-AJB13
47 Malassezia_restricta-AJB2
48 Circinella_umbellata-AJB3
49 Prunus_persica_peach_-AJB5
50 Aspergillus_niger-AJB10
51 [Candida]_glabrata-AJB8
52 Candida_albicans-AJB60
53 Candida_albicans_SC5314-AJB20
54 Mucor_endophyticus-AJB3
55 Heterocapsaceae_environmental_sample-AJB3
B
The richness of the gut eukaryotic speicies
in the ASD children and HC groups
Autism
Control
10000
1000
100
10
1
0
Ascochyta_medicaginicola-AJB29
uncultured_fungus-AJB198
Clathrospora_diplospora-AJB2
Pleiochaeta_ghindensis-AJB15
Boeremia_exigua_var._exigua-AJB29
Pleiochaeta_ghindensis-AJB5
Ascochyta_medicaginicola-AJB46
Enoplognatha_margarita-AJB6
Boeremia_exigua_var._exigua-AJB27
Clathrospora_diplospora-AJB14
Ascochyta_viciae-villosae-AJB2
Stemphylium_vesicarium-AJB12
[Candida]_smithsonii-AJB6
Clathrospora_diplospora-AJB24
Ascochyta_viciae-villosae-AJB13
[Cryptococcus]_ater-AJB5
Boeremia_exigua_var._exigua-AJB23
Cladosporium_sp._CBS_266.53-AJB26
[Candida]_smithsonii-AJB5
Cladosporium_sp._CBS_282.49-AJB3
Ascochyta_medicaginicola-AJB23
Clathrospora_diplospora-AJB16
Ascochyta_medicaginicola-AJB1
Ascochyta_phacae-AJB18
Clathrospora_diplospora-AJB35
Hohenbuehelia_tremula-AJB1
Entomophthora_muscae-AJB1
Naganishia_albida-AJB16
Cladosporium_sp._CBS_266.53-AJB37
Enoplognatha_margarita-AJB1
Ascochyta_viciae-villosae-AJB22
Hannaella_luteola-AJB27
Cladosporium_sp._CBS_280.49-AJB9
Itersonilia_perplexans-AJB4
Craterocolla_cerasi-AJB25
Scleromitrula_shiraiana-AJB7
Agaricomycetes_sp.-AJB1
Pycnora_sorophora-AJB171
Craterocolla_cerasi-AJB23
Schizophyllum_sp._MI_01-AJB3
Ascochyta_fabae-AJB19
Ascochyta_pisi-AJB26
Naganishia_albidosimilis-AJB1
albicans_SC5314-AJB79
Penicillium_expansum-AJB10
Aspergillus_versicolor-AJB13
Malassezia_restricta-AJB2
Circinella_umbellata-AJB3
Prunus_persica_peach-AJB5
Aspergillus_niger-AJB10
[Candida]_glabrata-AJB8
albicans-AJB60
albicans_SC5314-AJB20
Mucor_endophyticus-AJB3
Heterocapsaceae_envir._s.-AJB3

# Figure 2 Legend

**Fig. 2 | Gut eukaryotic taxa with differential abundance between children with ASD and healthy control groups.**

**a.** Eukaryotic species were screened using a stringent prevalence threshold of >24% in at least one group. In total, 55 differential eukaryotic taxa were classified into three distinct blocks based on abundance distribution patterns: taxa predominantly enriched in the healthy control group (upper block), taxa shared between the two groups with divergent abundance profiles (middle block), and taxa enriched in the ASD group (lower block).

**b.** The 55 differentially abundant eukaryotic species are arranged sequentially from left to right. Sequencing read abundance is plotted on a logarithmic y-axis. All identified differential taxa satisfied rigorous statistical criteria ($|logFC| > 2$, $P < 0.05$).




# Results Fig. 3

# Figure 3 Legend

**Fig. 3 | Phylogenetic clustering and evolutionary placement of 55 differential gut eukaryotic taxa**

The phylogenetic tree illustrates the evolutionary relationships and clustering patterns of the 55 differentially abundant gut eukaryotic taxa. Multiple distinct phylogenetic branches were resolved, with two prominent clades highlighted. One branch encloses *Schizophyllum commune* (green box), while the other comprises multiple differential *Candida* subspecies and strains (red box), reflecting clear phylogenetic divergence between health-associated and ASD -enriched eukaryotic taxa.

## Multiple *Candida* subspecies/strains detected in the gut of both children with ASD and healthy control groups

To characterize the gut *Candida* population in depth, we retrieved 209 *Candida* subspecies/strains. Using the prevalence threshold >20% in one group at least, then 40 subspecies/strains were retained for subsequent comparative analysis. They were stratified into three categories according to their detection patterns: HC -enriched, ASD -enriched and ASD -exclusive taxa.

Among all qualified strains, *Candida albicans*_SC5314-AJB17 exhibited the highest overall abundance across both ASD and HC groups (Fig. 4A), with an ASD-to-HC abundance ratio of 2.52 (Extended Data Fig. 6). Beyond its dominant abundance, this strain also showed a markedly higher prevalence in the ASD group (100%) relative to healthy controls (86%), corresponding to a 1.17-fold ASD-to-HC prevalence ratio (Extended Data Fig. 7).

By comparison, *Candida glabrata-A8* exhibited extreme ASD-specific enrichment, with a relative high prevalence in the ASD group and rare detection in HCs (Extended Data Figure 6). *Candida glabrata-A8* showed the largest inter-group prevalence fold change across all identified subspecies/strains. In addition, four subspecies/strains were exclusively detected in the ASD group (Fig. 4B). Overall, *Candida albicans_SC5314-AJB17* dominated the gut *Candida* community with the highest abundance and prevalence, whereas *Candida glabrata-A8* yielded the maximal inter-group abundance variation (Fig. 4C).

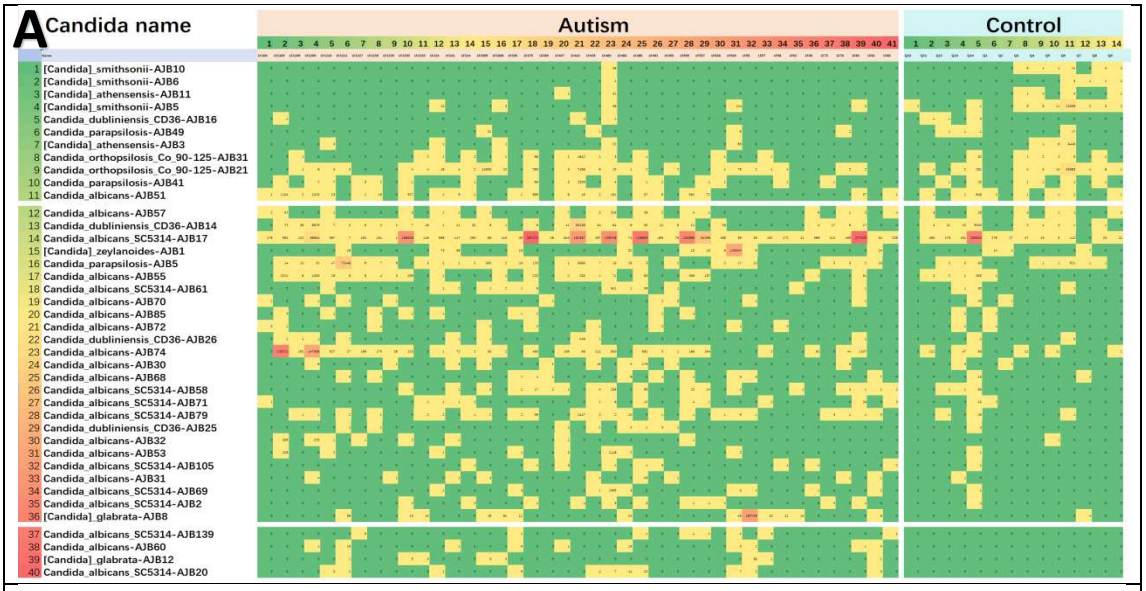
To uncover the fine-scale genetic divergence of key differential gut *Candida* strains in children with ASD, we performed pairwise genetic similarity and divergence analysis using the Hein nucleotide substitution model. Heatmap-based pairwise genetic comparison revealed clear partitioning of the examined *Candida* subspecies/strains into two genetically distinct phylogenetic clusters. Group A consisted of 16 strains, predominantly encompassing multiple *Candida albicans* variants, including *Candida albicans_SC5314-AJB17*, *Candida albicans_SC5314-AJB20*, and *Candida albicans-AJB60*. By contrast, Group B contained seven strains, with *Candida glabrata-AJB8* as the representative taxon (Fig. 4D).

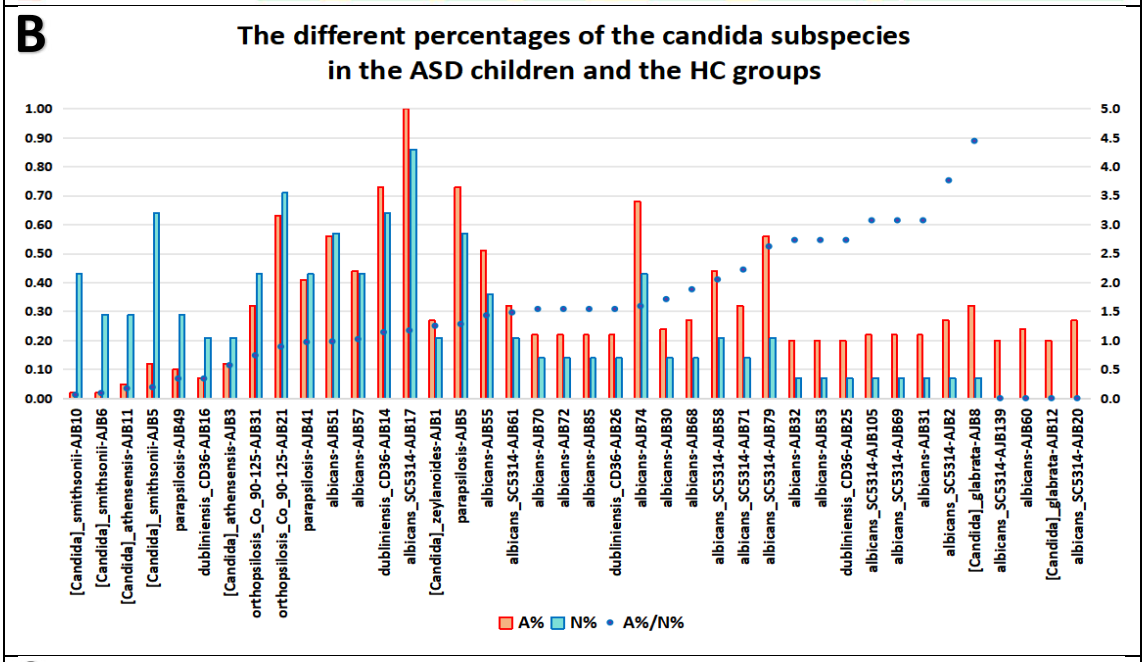
Phylogenetic reconstruction based on full-length rRNA-ITS amplicons further validated the taxonomic classification of these differentially distributed *Candida* subspecies/strains. The *Candida albicans_SC5314-AJB17* exhibited close

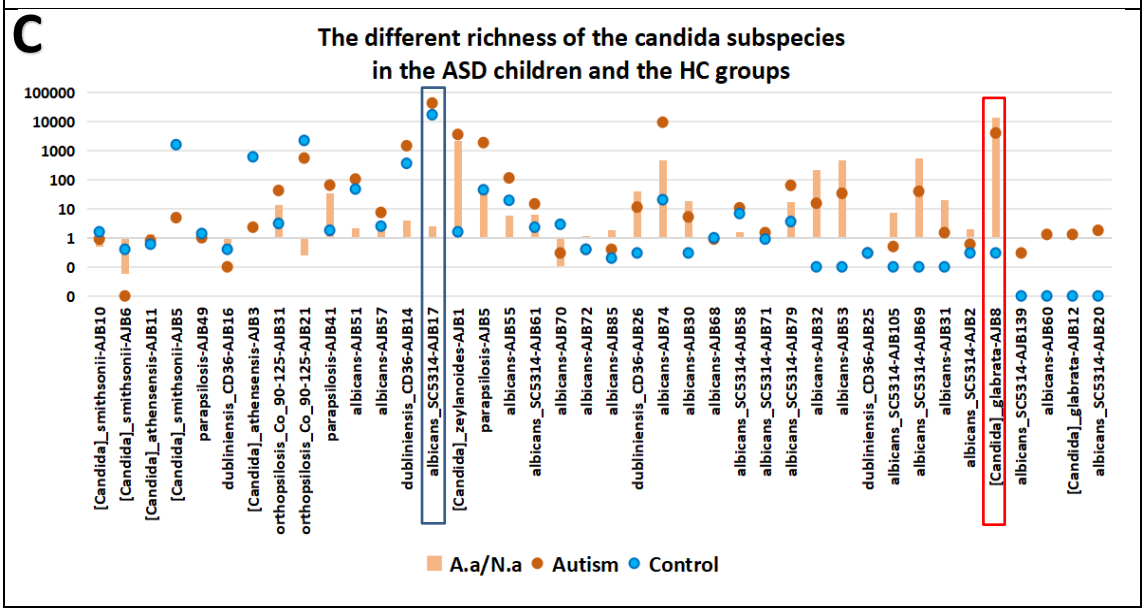
phylogenetic relatedness to the reference 18S–ITS1–5.8S–ITS2–partial 26S rRNA sequence of *Candida albicans* (MH545917.1). Meanwhile, *Candida glabrata-AJB8* clustered tightly with the reference 18S–ITS1–5.8S–ITS2–partial 26S rRNA sequence of *Nakaseomyces glabratus* (MH545922.1) and the 18S–ITS1–5.8S–ITS2–partial 25S rRNA sequence of *Candida glabrata* (AY198398.1). These phylogenetic results confirm the high accuracy of strain-level taxonomic identification achieved via long-read rRNA–ITS sequencing, and demonstrate substantial intraspecific genetic diversity and distinct phylogenetic stratification of gut *Candida* fungi in children with ASD (Fig. 5).

We further characterized the sequence conservation and genetic variation of full-length rRNA–ITS amplicons across key differential *Candida* subspecies/strains to resolve the core variable regions driving intraspecific divergence. Comparative sequence alignment demonstrated that the 18S and 5.8S rRNA regions were highly conserved among the examined *Candida* subspecies/strains, with only sporadic single-nucleotide polymorphisms and minor indels detected within the 18S rRNA gene. In contrast, the intervening ITS1 and ITS2 regions exhibited extensive sequence variability, representing the primary genetic hotspots for subspecies/strains-level differentiation.

Notably, the dominant gut strain *Candida glabrata-AJB8* possessed two characteristic insertions in the ITS1 region, which were consistent with the authoritative reference sequences of *Nakaseomyces glabratus* (MH545922.1) and *Candida glabrata strain CBS 138* (AY198398.1). Moreover, *Candida glabrata-AJB8* harbored a unique insertion in the partial 25S/28S rRNA region, which distinguished this from other gut *Candida* subspecies/strains (Fig. 6). These findings demonstrate that the hypervariable ITS regions are robust molecular markers for discriminating genetically similar *Candida* subspecies/strains, highlighting the unique advantage of long-read rRNA–ITS sequencing for uncovering fine-scale intraspecific fungal diversity.


A
Candida name
Autism
Control
1 [Candida]_smithsonii-AJB10
2 [Candida]_smithsonii-AJB6
3 [Candida]_athensensis-AJB11
4 [Candida]_smithsonii-AJB5
5 Candida_dubliniensis_CD36-AJB16
6 Candida_parapsilosis-AJB49
7 [Candida]_athensensis-AJB3
8 Candida_orthopsilosis_Co_90-125-AJB31
9 Candida_orthopsilosis_Co_90-125-AJB21
10 Candida_parapsilosis-AJB41
11 Candida_albicans-AJB51
12 Candida_albicans-AJB57
13 Candida_dubliniensis_CD36-AJB14
14 Candida_albicans_SC5314-AJB17
15 [Candida]_zeylanoides-AJB1
16 Candida_parapsilosis-AJB5
17 Candida_albicans-AJB55
18 Candida_albicans_SC5314-AJB61
19 Candida_albicans-AJB70
20 Candida_albicans-AJB85
21 Candida_albicans-AJB72
22 Candida_dubliniensis_CD36-AJB26
23 Candida_albicans-AJB74
24 Candida_albicans-AJB30
25 Candida_albicans-AJB68
26 Candida_albicans_SC5314-AJB58
27 Candida_albicans_SC5314-AJB71
28 Candida_albicans_SC5314-AJB79
29 Candida_dubliniensis_CD36-AJB25
30 Candida_albicans-AJB32
31 Candida_albicans-AJB53
32 Candida_albicans_SC5314-AJB105
33 Candida_albicans-AJB31
34 Candida_albicans_SC5314-AJB69
35 Candida_albicans_SC5314-AJB2
36 [Candida]_glabrata-AJB8
37 Candida_albicans_SC5314-AJB139
38 Candida_albicans-AJB60
39 [Candida]_glabrata-AJB12
40 Candida_albicans_SC5314-AJB20


B
The different percentages of the candida subspecies
in the ASD children and the HC groups
1.00
0.90
0.80
0.70
0.60
0.50
0.40
0.30
0.20
0.10
0.00
5.0
4.5
4.0
3.5
3.0
2.5
2.0
1.5
1.0
0.5
0.0
[Candida]_smithsonii-AJB10
[Candida]_smithsonii-AJB6
[Candida]_athensensis-AJB11
[Candida]_smithsonii-AJB5
parapsilosis-AJB49
dubliniensis_CD36-AJB16
[Candida]_athensensis-AJB3
orthopsilosis_Co_90-125-AJB31
orthopsilosis_Co_90-125-AJB21
parapsilosis-AJB41
albicans-AJB51
albicans-AJB57
dubliniensis_CD36-AJB14
albicans_SC5314-AJB17
[Candida]_zeylanoides-AJB1
parapsilosis-AJB5
albicans-AJB55
albicans_SC5314-AJB61
albicans-AJB70
albicans-AJB72
albicans-AJB85
dubliniensis_CD36-AJB26
albicans-AJB74
albicans-AJB30
albicans-AJB68
albicans_SC5314-AJB58
albicans_SC5314-AJB71
albicans_SC5314-AJB79
albicans-AJB32
albicans-AJB53
dubliniensis_CD36-AJB25
albicans_SC5314-AJB105
albicans_SC5314-AJB69
albicans-AJB31
albicans_SC5314-AJB2
[Candida]_glabrata-AJB8
albicans_SC5314-AJB139
albicans-AJB60
[Candida]_glabrata-AJB12
albicans_SC5314-AJB20
A% N% A%/N%


C
The different richness of the candida subspecies
in the ASD children and the HC groups
100000
10000
1000
100
10
1
0
0
[Candida]_smithsonii-AJB10
[Candida]_smithsonii-AJB6
[Candida]_athensensis-AJB11
[Candida]_smithsonii-AJB5
parapsilosis-AJB49
dubliniensis_CD36-AJB16
[Candida]_athensensis-AJB3
orthopsilosis_Co_90-125-AJB31
orthopsilosis_Co_90-125-AJB21
parapsilosis-AJB41
albicans-AJB51
albicans-AJB57
dubliniensis_CD36-AJB14
albicans_SC5314-AJB17
[Candida]_zeylanoides-AJB1
parapsilosis-AJB5
albicans-AJB55
albicans_SC5314-AJB61
albicans-AJB70
albicans-AJB72
albicans-AJB85
dubliniensis_CD36-AJB26
albicans-AJB74
albicans-AJB30
albicans-AJB68
albicans_SC5314-AJB58
albicans_SC5314-AJB71
albicans_SC5314-AJB79
albicans-AJB32
albicans-AJB53
dubliniensis_CD36-AJB25
albicans_SC5314-AJB105
albicans_SC5314-AJB69
albicans-AJB31
albicans_SC5314-AJB2
[Candida]_glabrata-AJB8
albicans_SC5314-AJB139
albicans-AJB60
[Candida]_glabrata-AJB12
albicans_SC5314-AJB20
A.a/N.a Autism Control

# Figure 4 Legend

**Fig. 4 | Long-read sequencing enables strain-level resolution of gut fungal variants with differential abundance and prevalence between children with ASD and healthy control groups.**

**a** In total, 40 out of 209 initial *Candida* subspecies/strains were retained following a minimum prevalence cutoff of 20% in at least one group. Qualified taxa were categorized into three parts based on cross-group detection patterns. The upper panel includes 11 strains with higher prevalence in the HC group than in the ASD group; the middle panel contains 25 ASD-enriched strains; the bottom panel shows 4 strains exclusively detected in the ASD group and absent from all HC samples. The heatmap adopts a red-to-green colour gradient corresponding to decreasing prevalence values. Notably, *Candida albicans_SC5314-AJB17* exhibited consistently strong red signals across most samples, indicating its dominant prevalence and high sequencing coverage among all detected *Candida* strains.

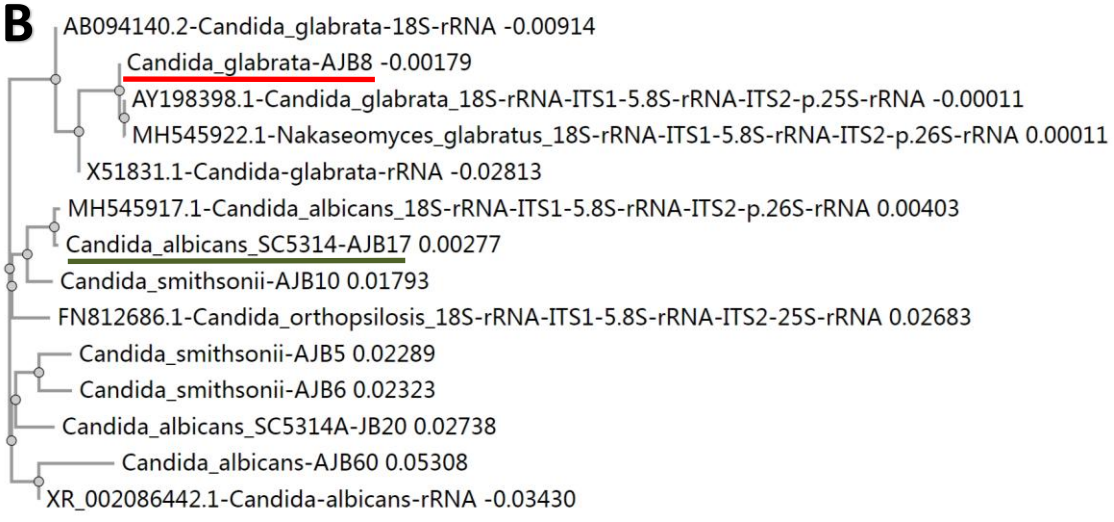
**b** The 40 qualified *Candida* strains are horizontally ranked according to their intergroup prevalence ratios. Bar plots illustrate the prevalence percentage of each strain in the HC group (light blue bars) and the ASD group (light red bars). Blue dots label strains with higher prevalence in the ASD group relative to HCs. *Candida glabrata-AJB8* showed the most prominent ASD -biased prevalence ratio among all tested strains. The 4 strains, including *Candida albicans_SC5314-A139*, *Candida albicans-A60*, *Candida glabrata-A12* and *Candida albicans_SC5314-AJB20*, were exclusively detected in the ASD group.

**c** Distinct abundance profiles of the 40 differential *Candida* subspecies/strains between the two groups. The scatter plot displays strain-level sequencing abundance on a logarithmic y-axis to resolve subtle inter-strain variations. Light blue dots represent abundance values in the HC group, while brown dots indicate abundance in the ASD group. Orange bars denote the ASD/HC abundance fold change for individual strains. *Candida albicans_SC5314-AJB17* presented the highest absolute sequencing abundance in both groups, whereas *Candida glabrata-AJB8* exhibited the maximum intergroup abundance fold change.

**A**

| | | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 | 11 | 12 | 13 | 14 | 15 | 16 | 17 | 18 | 19 | 20 | 21 | 22 | 23 |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 1 | Candida_albicans_SC5314-AJB17 | 100 | 50 | 37 | 39 | 38 | 36 | 36 | 60 | 59 | 33 | 30 | 45 | 64 | 34 | 41 | 43 | 30 | 31 | 30 | 29 | 31 | 30 | 31 |
| 2 | Candida_albicans_SC5314-AJB2 | 50 | 100 | 54 | 66 | 66 | 37 | 61 | 41 | 43 | 54 | 61 | 38 | 35 | 75 | 45 | 46 | 27 | 28 | 28 | 28 | 29 | 28 | 29 |
| 3 | Candida_albicans_SC5314-AJB20 | 37 | 54 | 100 | 51 | 51 | 34 | 51 | 40 | 38 | 53 | 51 | 39 | 42 | 50 | 32 | 40 | 27 | 27 | 28 | 29 | 28 | 27 | 29 |
| 4 | Candida_albicans_SC5314-AJB58 | 39 | 66 | 51 | 100 | 80 | 36 | 67 | 32 | 36 | 53 | 67 | 35 | 31 | 73 | 33 | 43 | 28 | 27 | 29 | 28 | 29 | 28 | 28 |
| 5 | Candida_albicans_SC5314-AJB61 | 38 | 66 | 51 | 80 | 100 | 36 | 66 | 32 | 36 | 54 | 66 | 35 | 33 | 70 | 33 | 44 | 27 | 27 | 28 | 28 | 27 | 27 | 27 |
| 6 | Candida_albicans_SC5314-AJB69 | 36 | 37 | 34 | 36 | 36 | 100 | 36 | 38 | 41 | 35 | 34 | 44 | 36 | 36 | 47 | 36 | 28 | 28 | 29 | 29 | 29 | 28 | 30 |
| 7 | Candida_albicans_SC5314-AJB79 | 36 | 61 | 51 | 67 | 66 | 36 | 100 | 32 | 37 | 53 | 66 | 35 | 31 | 67 | 33 | 43 | 28 | 28 | 28 | 29 | 28 | 28 | 28 |
| 8 | Candida_albicans-AJB31 | 60 | 41 | 40 | 32 | 32 | 38 | 32 | 100 | 58 | 35 | 33 | 45 | 56 | 32 | 41 | 42 | 28 | 30 | 29 | 29 | 28 | 28 | 31 |
| 9 | Candida_albicans-AJB53 | 59 | 43 | 38 | 36 | 36 | 41 | 37 | 58 | 100 | 40 | 36 | 49 | 55 | 36 | 40 | 46 | 29 | 29 | 29 | 29 | 29 | 29 | 31 |
| 10 | Candida_albicans-AJB55 | 33 | 54 | 53 | 53 | 54 | 35 | 53 | 35 | 40 | 100 | 54 | 36 | 34 | 52 | 33 | 40 | 28 | 27 | 28 | 28 | 27 | 26 | 28 |
| 11 | Candida_albicans-AJB60 | 30 | 61 | 51 | 67 | 66 | 34 | 66 | 33 | 36 | 54 | 100 | 34 | 31 | 66 | 33 | 43 | 31 | 29 | 28 | 29 | 27 | 27 | 28 |
| 12 | Candida_albicans-AJB70 | 45 | 38 | 39 | 35 | 35 | 44 | 35 | 45 | 49 | 36 | 34 | 100 | 58 | 36 | 38 | 55 | 28 | 29 | 29 | 29 | 30 | 29 | 30 |
| 13 | Candida_albicans-AJB85 | 64 | 35 | 42 | 31 | 33 | 36 | 31 | 56 | 55 | 34 | 31 | 58 | 100 | 31 | 38 | 50 | 28 | 29 | 29 | 29 | 30 | 30 | 31 |
| 14 | Candida_dubliniensis_CD36-AJB25 | 34 | 75 | 50 | 73 | 70 | 36 | 67 | 32 | 36 | 52 | 66 | 36 | 31 | 100 | 39 | 44 | 29 | 27 | 28 | 28 | 27 | 27 | 28 |
| 15 | Candida_dubliniensis_CD36-AJB26 | 41 | 45 | 32 | 33 | 33 | 47 | 33 | 41 | 40 | 33 | 33 | 38 | 38 | 39 | 100 | 35 | 29 | 30 | 29 | 30 | 29 | 29 | 30 |
| 16 | Candida_parapsilosis-AJB41 | 43 | 46 | 40 | 43 | 44 | 36 | 43 | 42 | 46 | 40 | 43 | 55 | 50 | 44 | 35 | 100 | 38 | 28 | 28 | 27 | 29 | 29 | 29 |
| 17 | Candida_athensensis-AJB11 | 30 | 27 | 27 | 28 | 27 | 28 | 28 | 28 | 29 | 28 | 31 | 28 | 28 | 29 | 29 | 38 | 100 | 58 | 58 | 51 | 59 | 59 | 40 |
| 18 | Candida_athensensis-AJB3 | 31 | 28 | 27 | 27 | 27 | 28 | 28 | 30 | 29 | 27 | 29 | 29 | 29 | 27 | 30 | 28 | 58 | 100 | 51 | 51 | 70 | 68 | 40 |
| 19 | Candida_glabrata-AJB12 | 30 | 28 | 28 | 29 | 28 | 29 | 28 | 29 | 29 | 28 | 28 | 29 | 29 | 28 | 29 | 28 | 58 | 51 | 100 | 67 | 52 | 52 | 39 |
| 20 | Candida_glabrata-AJB8 | 29 | 28 | 29 | 28 | 28 | 29 | 29 | 29 | 29 | 28 | 29 | 29 | 29 | 28 | 30 | 27 | 51 | 51 | 67 | 100 | 52 | 51 | 41 |
| 21 | Candida_smithsonii-AJB10 | 31 | 29 | 28 | 29 | 27 | 29 | 28 | 28 | 29 | 27 | 27 | 30 | 30 | 27 | 29 | 29 | 59 | 70 | 52 | 52 | 100 | 75 | 41 |
| 22 | Candida_smithsonii-AJB6 | 30 | 28 | 27 | 28 | 27 | 28 | 28 | 28 | 29 | 26 | 27 | 29 | 30 | 27 | 29 | 29 | 59 | 68 | 52 | 51 | 75 | 100 | 40 |
| 23 | Candida_zeylanoides-AJB1 | 31 | 29 | 29 | 28 | 27 | 30 | 28 | 31 | 31 | 28 | 28 | 30 | 31 | 28 | 30 | 29 | 40 | 40 | 39 | 41 | 41 | 40 | 100 |

**B**




# Results Fig. 5

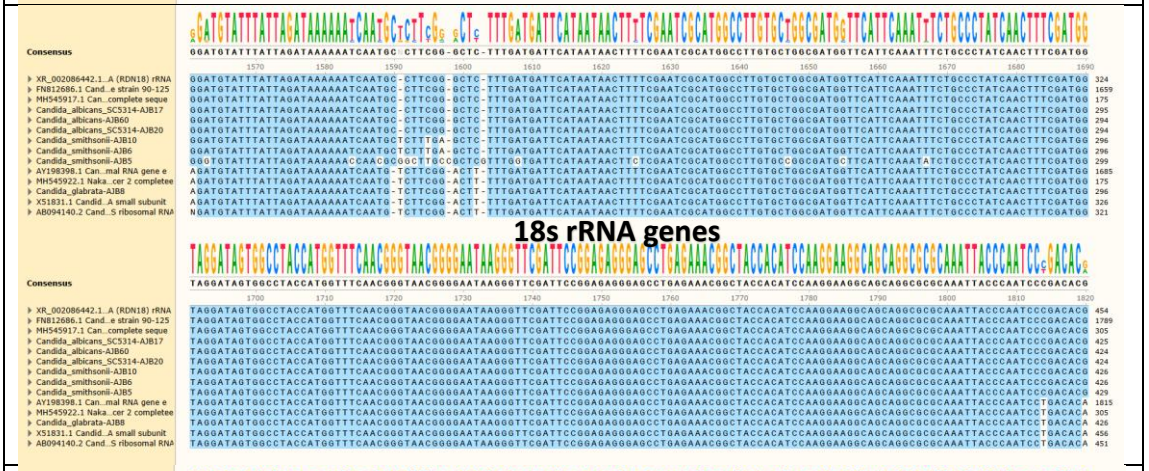
# Figure 5 Legend

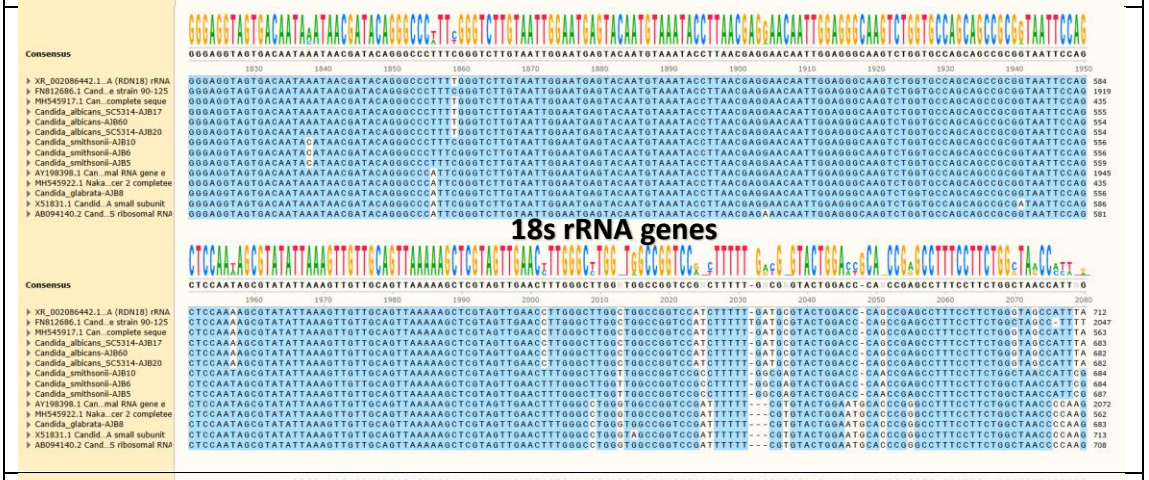
**Fig. 5 | Genetic divergence and phylogenetic validation of differential gut fungal strains resolved by long-read rRNA–ITS sequencing.**

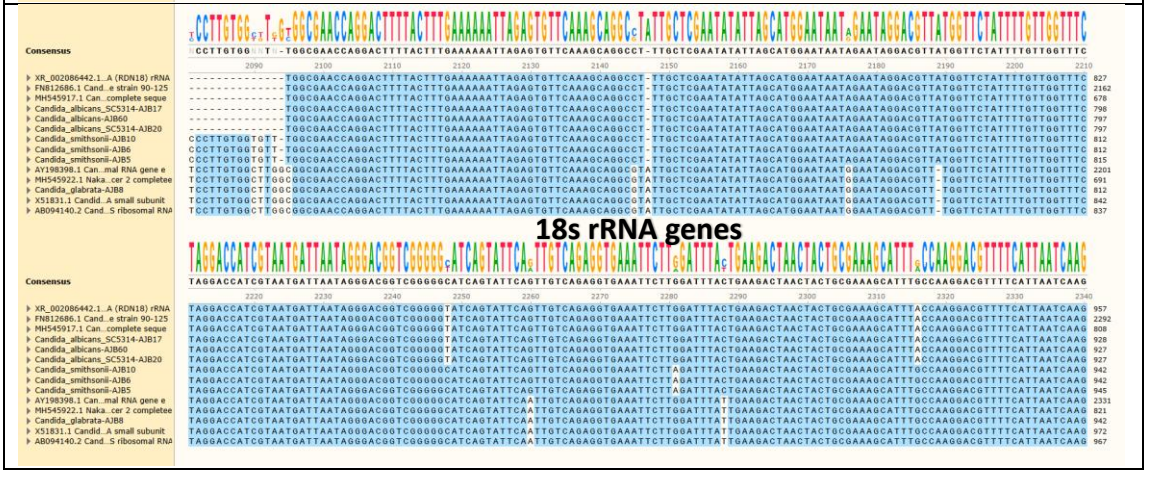
**a.** Heatmap showing pairwise sequence similarity and genetic divergence among 23 representative gut *Candida* subspecies/strains, which segregated into two well-defined phylogenetic clusters. The upper region (above the yellow line) consisted of 16 strains, including dominant *Candida albicans* variants (*Candida albicans_SC5314-AJB17*, *Candida albicans_SC5314-AJB20* and *Candida albicans-AJB60*). The lower region (below the yellow line) contains 7 strains predominantly represented by *Candida glabrata-AJB8*. Pairwise phylogenetic distances were calculated from aligned full-length 18S-rRNA–ITS1–5.8S-rRNA–ITS2–partial 26S rRNA sequences using the Hein nucleotide substitution model. The upper triangle displays pairwise sequence similarity percentages, while the lower triangle indicates corresponding genetic divergence percentages.

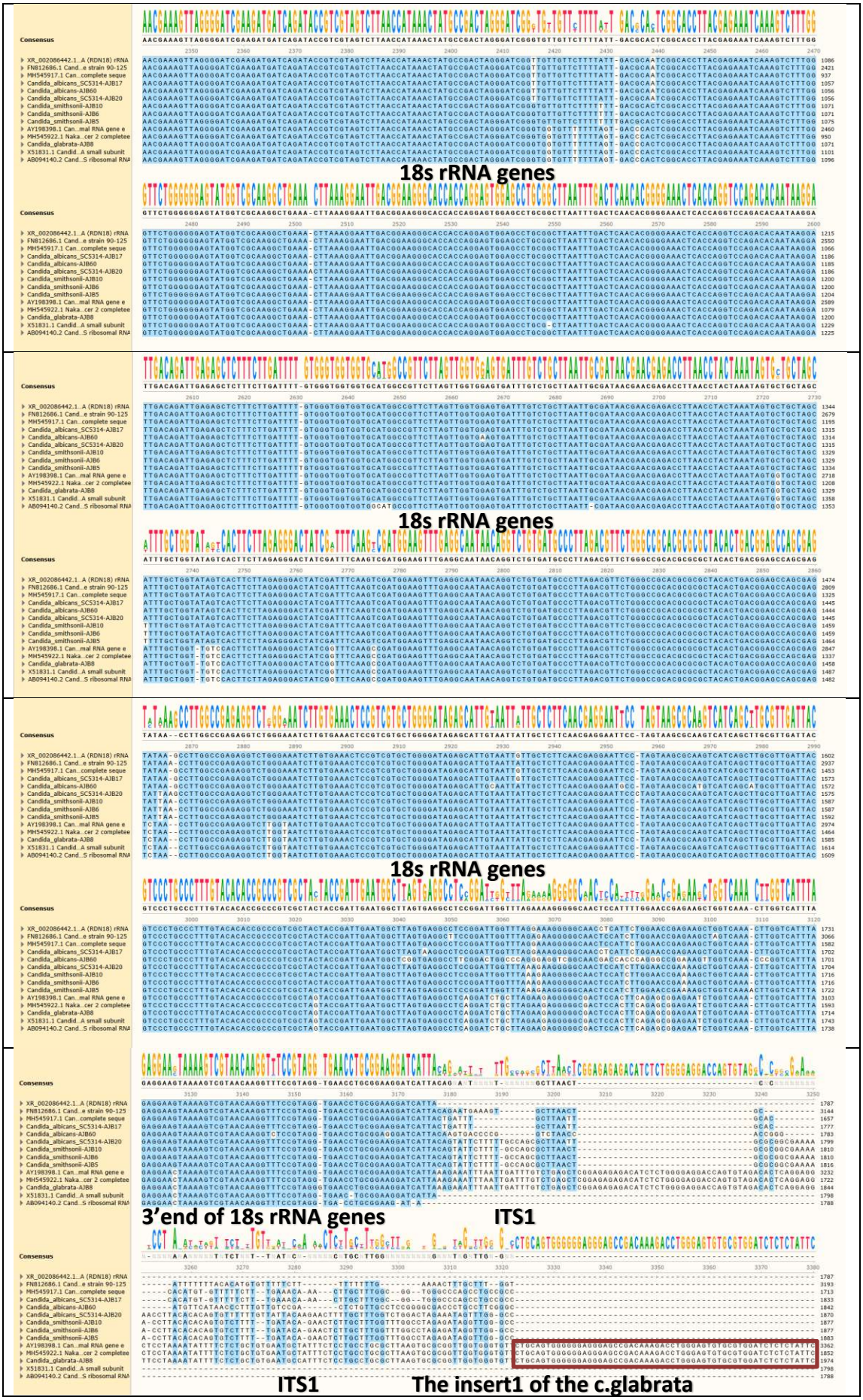
**b.** Maximum-likelihood phylogenetic tree of representative gut *Candida* subspecies/strains and publicly available reference sequences. *Candida albicans_SC5314-AJB17* clustered tightly with the curated *Candida albicans* reference sequence (MH545917.1). Similarly, *Candida glabrata-AJB8* exhibited high sequence homology and clustered with two standard reference sequences: the 18S-rRNA–ITS1–5.8S-rRNA–ITS2–partial 26S rRNA sequence of *Nakaseomyces glabratus* (MH545922.1) and the 18S-rRNA–ITS1–5.8S-rRNA–ITS2–partial 25S rRNA sequence of *Candida glabrata* (AY198398.1), thereby verifying accurate strain-level taxonomic annotation.

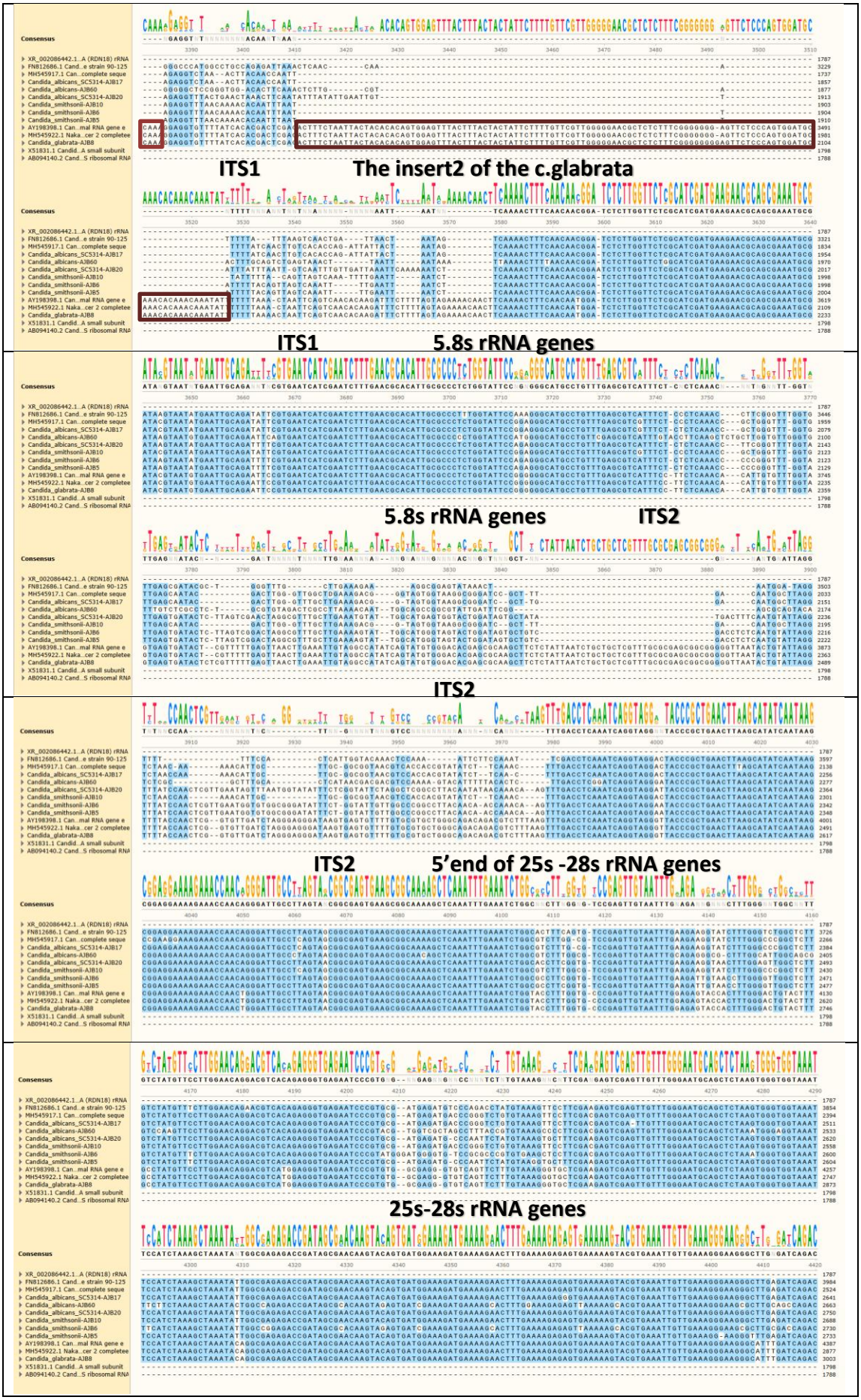

5'end of 18s rRNA genes

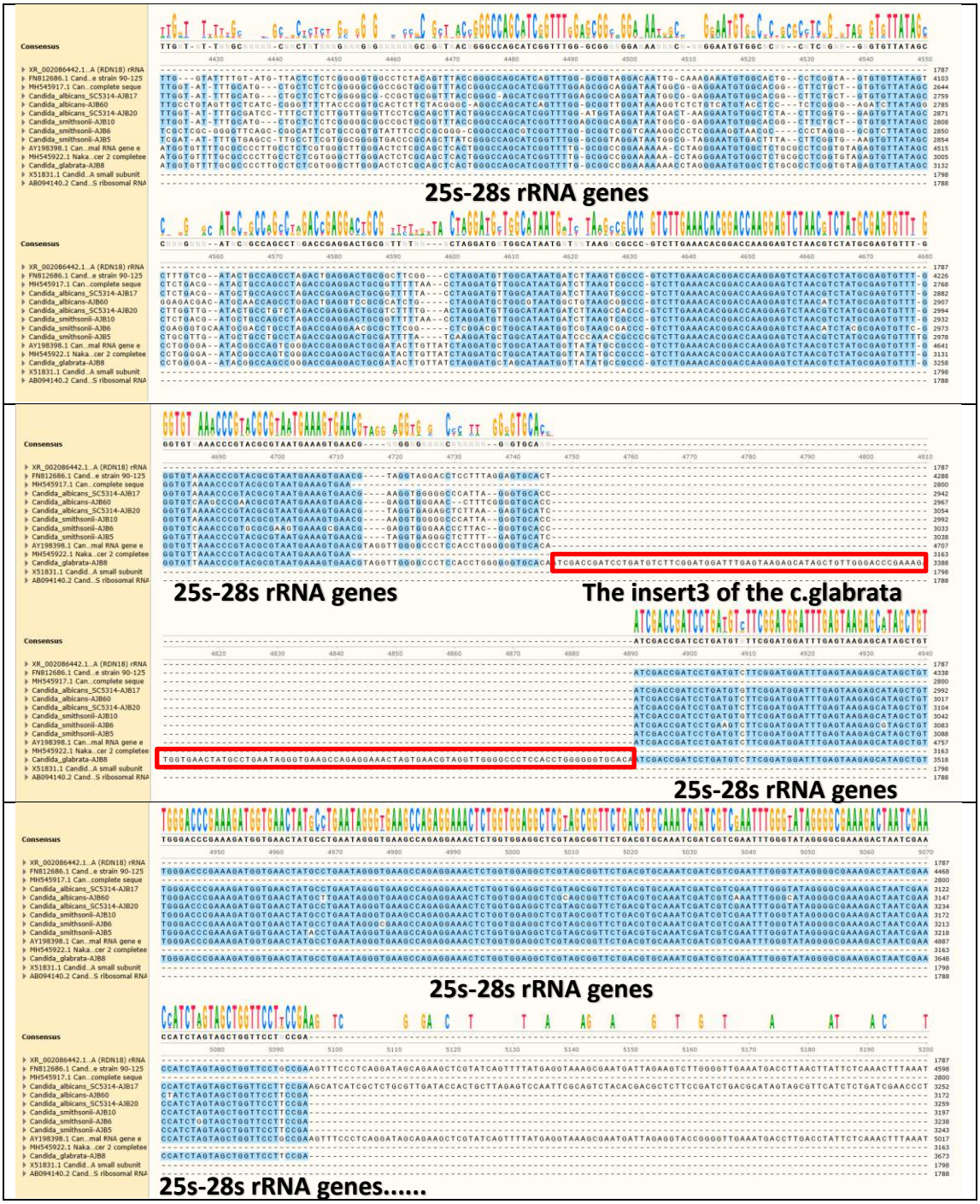

18s rRNA genes


18s rRNA genes


18s rRNA genes


18s rRNA genes
18s rRNA genes
18s rRNA genes
3'end of 18s rRNA genes
ITS1
ITS1
The insert1 of the c.glabrata


ITS1
The insert2 of the c.glabrata
ITS1
5.8s rRNA genes
5.8s rRNA genes
ITS2
ITS2
5’end of 25s -28s rRNA genes
25s-28s rRNA genes




# Results Fig. 6

# Figure 6 Legend

**Fig. 6 | Sequence conservation and variation of rRNA genes and ITS regions in key differential subspecies/strains.**

Comparative sequence analysis revealed high conservation of the 18S and 5.8S rRNA regions across the examined *Candida* subspecies, with only sporadic single-nucleotide polymorphisms and indels detected within the 18S rRNA gene. In contrast, the ITS1 and ITS2 regions exhibited substantial sequence variability, serving as the major variable regions for subspecies- and strain-level differentiation. Notably, *Candida glabrata-AJB8* harbored two characteristic insertions in the ITS1 region, which were consistent with the reference sequences of *Nakaseomyces glabratus* (MH545922.1) and *Candida glabrata strain CBS 138* (AY198398.1). In addition, *Candida glabrata-AJB8* contained a unique insertion in the partial 25S/28S rRNA region (highlighted in the red box).

All reference sequences and newly resolved fungal sequences used for comparative analysis are listed: *Candida albicans SC5314* 18S rRNA (XR_002086442.1); *Candida orthopsilosis strain 90-125* (5S rRNA, 18S rRNA, ITS1, 5.8S rRNA, ITS2, and 25S rRNA gene, FN812686.1); *Candida albicans strain CBS 5703* (partial 18S rRNA, complete ITS1–5.8S rRNA–ITS2, partial 26S rRNA, MH545917.1); *Candida albicans_SC5314-AJB17*; *Candida albicans-AJB60*; *Candida albicans_SC5314-AJB20*; *Candida smithsonii-AJB10*; *Candida smithsonii-AJB6*; *Candida smithsonii-AJB5*; *Candida glabrata strain CBS 138* (complete ITS1–5.8S rRNA–ITS2, partial 25S rRNA, AY198398.1); *Nakaseomyces glabratus strain CBS 138* (partial 18S rRNA, complete ITS1–5.8S rRNA–ITS2, partial 26S rRNA, MH545922.1); *Candida glabrata-AJB8*; *Candida glabrata* small-subunit rRNA (X51831.1); *Candida glabrata* 18S rRNA gene (AB094140.2).

# Discussion

## Advantages of Long-Read Sequencing for Gut Microorganism Identification

Historically, investigations into gut microbiota-mediated gut–brain axis perturbations in ASD have predominantly focused on bacterial communities. In this study, long-read amplicon sequencing targeting the full-length 16S-rRNA–ITS–partial 23S-rRNA fragment validated that the internal transcribed spacer (ITS) region positioned between the 16S and 23S rRNA genes exhibits markedly elevated sequence polymorphism. This high genetic variability renders the ITS segment a superior phylogenetic marker for high-resolution taxonomic identification at the species and subspecies/strain levels, resolving a key limitation of conventional short-read 16S rRNA sequencing for strain-level gut microbial profiling. Fungal contributions to ASD pathogenesis have recently emerged as a critical research frontier. Given that eukaryotic 18S rRNA and 5.8S rRNA genes are highly conserved across eukaryotic species, whereas the ITS1 and ITS2 spacers harbour extensive sequence divergence, we performed full-length sequencing of the 18S-rRNA–ITS1–5.8S-rRNA–ITS2–partial 28S-rRNA cassette to profile gut eukaryotic communities. Capitalizing on these divergent genetic features, our full-length rRNA–ITS long-read sequencing strategy enabled precise strain-level characterization of both gut bacteriome and mycobiome, thereby uncovering previously unrecognized bacterial and fungal dysbiosis associated with ASD.

## Dominant Bacterial Taxa in the Healthy Control Group

*Blautia*is a common genus of anaerobic bacteria widely distributed in the feces and intestinal tracts of mammals. We observed a marked enrichment of *Blautia* in HC groups. *Blautia* abundance was 3.6-fold higher in HC subjects, while the dominant strain *Blautia* sp900541955-AJB3 exhibited a 12-fold HC-specific enrichment. *Blautia* contributes to host metabolism for acetate, which serves as an important energy source for intestinal epithelial cells and brain tissues, exerting inhibitory effects against pathogens and displaying anti-inflammatory properties[14]. Reduced levels of *Blautia* may associate with the gut microbial dysbiosis in ASD. Further, our long-read sequencing data clearly revealed four strains of *Clostridioides*. The broad

taxonomic labels "*Clostridioides*" or "*Clostridium*" encompassed strain level genetic and functional diversity. The strain-level resolution is essential for understanding potential roles of the bacteria in gut-brain axis communication.

## Dominant Bacterial Taxa in the ASD Group

Conversely, the ASD gut was enriched with multiple opportunistic and metabolism-disrupting bacterial taxa. At the strain level, the several taxa were significant enriched in the ASD group than HC group. The *Clostridium sensu stricto 1 saudiense-AJB5* was markedly enriched in the ASD group (50%) compared to the HC group (7%). The abundance of *Megamonas* from a human fecal microbiota[15] is elevated compared to normal-weight individuals[16], pediatric diarrhea[17] and epilepsy[18]. A higher abundance of *Megamonas* has been observed in ASD children compared to healthy controls[19]. Our long-read sequencing data revealed a striking enrichment of *Megamonas funiformis* in the ASD group. It accounted for 36% of prevalence, whereas not detected in the HC group. Bacteroides uniformis was identified from human fecal samples[20]. A significant reduction in *Bacteroides clarus* following the low-fat diet[21]. We observed that *Bacteroides_clarus-AJB1* was present at the abundance of 40% in the ASD group but was undetectable in healthy controls. *Terrisporobacter mayombei* was characterized with their metabolic capabilities, vitamins, cofactors, and acetogenesis[22]. Significantly enriched two strains, *Terrisporobacter mayombei-AJB1* and *Terrisporobactersp 902363255-AJB1*, were observed in the ASD group compared to the HC group ($p < 0.05$). *Prevotella* represent a key constituent of the human gut microbiota[23] involved in intestinal health and disease[24]. In the present study, *Prevotella* was detected in the ASD group, rather than in the healthy control group.

## Dominant Eukaryotic *Candida* Subspecies/Strains in the ASD Group.

Immune dysregulation and gastrointestinal symptoms are highly prevalent among children with ASD, who frequently exhibit marked alterations in gut microbiota composition and ecological balance. While most research has focused on bacterial constituents, studies examining the fungal microbiome remain limited. In our analysis, several *Candida* taxa showed greater abundance in the ASD group based on long-read sequencing data, including *Candida glabrata-AJB8* (logFC = 15.21, P =

0.002), *Candida albicans-AJB60* (logFC = 4.67, P = 0.029), and *Candida albicans SC5314-AJB20* (logFC = 4.49, P = 0.023). These findings suggest that specific *Candida* species or strains may be associated with ASD. Notably, the strain *Candida albicans SC5314-AJB17* exhibited the highest abundance among all 209 *Candida* variants and was present in 100% of ASD samples (41/41) and 86% of HC samples (12/14).

These findings align with previous reports linking *Candida albicans* to ASD. The higher positive *Candida albicans* have detected in ASD compared to HC with ELISA-based studies[25,26]. Mechanistically, intestinal overgrowth of *Candida albicans* may impair carbohydrate and mineral absorption while increasing toxin production[27]. It can synthesize cyclic dipeptides that may affect neurodevelopment via apoptotic pathways[28], and its gut metabolism linked to GABAergic excess and proposed as a factor in autism etiology[29]. Probiotic supplementation may help alleviate *Candida*-associated intestinal symptoms[30]. Importantly, the high-abundance *candida* subspecies/strains stably colonized in the gut of children with ASD and healthy controls, indicating that certain *Candida* strains represent core commensal eukaryotes in the children gut, rather than purely pathogenic microbes. Accumulating evidence supports the dual role of intestinal *Candida* species, acting as both harmless commensals and opportunistic pathogens depending on the gut microenvironment [31–34]. The strain-level resolution achieved by long-read sequencing allowed us to detect subspecies or strain level enrichment patterns that may distinguish stable commensal colonization from potentially pathogenic overgrowth, a distinction that would be obscured by genus-level profiling.

## Differential Abundance and prevalent of Other Eukaryotes

Interestingly, several species of fungi (*schizophyllum, Craterocolla cerasi-AJB23, Craterocolla cerasi-AJB25* and *Hohenbuehelia_tremula*) were identified during the analyzing of the 18s-rRNA-ITS1-5.8s-rRNA-ITS2-p.25s/28s-rRNA sequences. All these fungal taxa exhibited consistent enrichment in the HC group with robust statistical significance (all logFC < −10, P < 0.001) (Extended Data Figure. 8). *Hohenbuehelia, Schizophyllum* and *Craterocolla* are environmental saprotrophic fungi, and their detection in human stool likely originates from dietary intake or environmental spore exposure. The medicinal fungus *Schizophyllum commune* produces bioactive metabolites with antioxidant, anti-inflammatory, and metabolic

regulatory properties. Depletion of these protective fungal taxa in ASD may impair intestinal barrier integrity and exacerbate gastrointestinal dysfunction. Consistently, mushroom-containing diets have been reported to induce region-specific remodeling of gut microbiota in animal model[35].

# Conclusions

Using full-length rRNA–ITS long-read sequencing, we systematically profiled gut bacterial and fungal communities at subspecies and strain resolution in children with ASD and healthy controls (Extended Data Fig. 9). Our study demonstrates that full-length rRNA–ITS long-read profiling unlocks previously hidden gut microbial microdiversity, refining the high-resolution landscape of ASD-associated gut dysbiosis and providing insights for future microbiota-targeted mechanistic investigations.

# Limitation

Although we performed full-length rRNA–ITS long-read sequencing to study gut bacterial and fungal communities, we acknowledge that the sample size should be expanded and that samples should not be restricted to a single local area; future studies should include participants from multiple geographic regions.

# Methods

## Sample collection and DNA extraction

Faecal samples were collected from paediatric patients with a clinical diagnosis of autism spectrum disorder (ASD) and age-matched healthy controls. Participants were recruited from Luan Hospital and local communities in Luan City, Anhui Province, China. All sample collection procedures were conducted under ethical approval from the Medical Ethics Committee, Affiliated Hospital of West Anhui Health Vocational College (approval ID: KYXM-2025-003). The study adhered to the principles of the Declaration of Helsinki. Written informed consent was obtained voluntarily from the parents or legal guardians of all minor participants prior to faecal sampling. Guardians were fully informed of the study objectives, sample usage, potential risks and benefits, and all procedures complied with relevant ethical and legal requirements. Fresh faecal specimens were immediately stored at −80 °C upon collection until further processing. Microbial genomic DNA was extracted from faecal samples using the QIAamp Rapid DNA Stool Kit (Qiagen, Hilden, Germany), following the manufacturer's standard protocol. DNA integrity and purity were evaluated by agarose gel electrophoresis, while DNA concentration was quantified using a Qubit 2.0 Fluorometer with the Qubit dsDNA HS Assay Kit (Life Technologies, Carlsbad, CA, USA). All qualified DNA samples were stored at −20 °C for subsequent Library construction and sequencing.

## Library construction and third-generation long-read sequencing

Purified genomic DNA was fragmented using the Megaruptor 3 system and concentrated with AMPure PB magnetic beads for targeted amplification. Conserved rRNA gene regions were used to design specific PCR primers for capturing full-length prokaryotic and eukaryotic ribosomal sequences spanning hypervariable internal transcribed spacer (ITS) regions. For bacterial community profiling, primer pairs targeting the 16S-rRNA–ITS–partial 23S rRNA region were applied: forward 16S-For (5′-AGRGTTYGATYMTGGCTCAG-3′) and reverse 16S-Rev (5′-CCDTTCCYTCRCRGTACT-3′). The resulting amplicons covered the entire 16S rRNA gene, the full-length ITS region, and partial sequences of the 23S rRNA gene.

For eukaryotic microbial profiling, primer pairs targeting the 18S-rRNA–ITS1–5.8S-rRNA–ITS2–partial 25S/28S rRNA region were used: forward 18S-For (5′-CTTGTCTCAAAGATTAAGCCATGCAT-3′) and reverse 18S-Rev (5′-TCGGAAGGAACCAGCTACTA-3′), generating amplicons encompassing the full 18S rRNA gene, ITS1, 5.8S rRNA, ITS2, and partial 25S/28S rRNA sequences. All target fragments were amplified from faecal genomic DNA via standard PCR protocols. For long-read library preparation, PCR products were processed using the SMRT Bell Preparation Kit 3.0 (Pacific Biosciences) following the official PacBio Revio workflow. Briefly, primer annealing was conducted, and SMRTcell templates were bound to polymerases using the Revio Polymerase Kit (Pacific Biosciences). Final sequencing libraries were constructed and sequenced on the PacBiio Revio platform at Kindstar Sequenon Biotechnology (Wuhan, China).

## Sequencing data analysis

Long-read sequencing data derived from prokaryotic 16S-rRNA–ITS–partial 23S-rRNA and eukaryotic 18S-rRNA–ITS1–5.8S-rRNA–ITS2–partial 25S/28S rRNA amplicons were processed through a standardized analytical pipeline, including quality control, operational taxonomic unit (OTU) clustering, taxonomic annotation, and differential abundance comparison between the ASD and healthy control (HC) groups. Raw sequencing reads were first subjected to rigorous quality control to remove low-quality sequences, adapter and primer contaminants, redundant reads, and chimeric sequences, ensuring high accuracy for downstream analyses. High-quality clean reads were then clustered into OTUs based on sequence similarity for quantifying the relative abundance of distinct microbial taxa across all samples. All OTUs were further annotated against three authoritative reference databases, including GTDB, SILVA and GenBank, to enable high-resolution taxonomic classification at species and subspecies levels. Subsequently, the Linear Discriminant Analysis Effect Size (LEfSe) algorithm was applied to identify microbial taxa with significantly differential abundance and distinct distribution patterns between the two groups.

## Species- and subspecies-level statistical analysis

Fine-scale quantitative statistical analyses were performed at the species and

subspecies levels based on annotated OTU profiles. Briefly, the mean abundance and standard deviation of each OTU were calculated for the ASD and HC groups, and intergroup abundance fold changes were determined with a threshold set at >2.0. The detection prevalence of each OTU in the two groups was further statistically evaluated. OTUs with a detection prevalence exceeding 20% in at least one group were retained to eliminate low-frequency microbial features with negligible biological relevance. Taxa with an intergroup fold change >2.0 and $P \leq 0.05$ were defined as significantly differential taxa. The screened differential OTUs were used to systematically characterize fine-scale species- and subspecies-level microbial variations between the ASD and HC gut microbiota.

## Structural analysis of prokaryotic and eukaryotic rRNA genes

Qualified long-read sequences were clustered and subjected to structural analysis. Multiple sequence alignment was performed using MAFFT software, and structural visualization of aligned sequences was implemented via Mag5 software. Standard full-length reference sequences of 16S-rRNA, 23S-rRNA, 18S-rRNA and 25S-rRNA were retrieved from the GenBank database and used for sequence alignment, structural comparison and regional characterization of prokaryotic and eukaryotic rRNA amplicons.

# Data availability

All raw long-read rRNA–ITS sequencing data generated in this study have been deposited in the Genome Sequence Archive (GSA). The associated BioProject (PRJCA059830/ subPRO087183) and GSA accession (CRA039994/ subCRA065381) have passed official data validation. Raw sequencing reads are accessible at https://ngdc.cncb.ac.cn/gsa/s/mwtxuUGg. No restricted-access datasets were utilized in this work.

# References


1. Bezawada N, Phang TH, Hold GL, Hansen R.Autism Spectrum Disorder and the Gut Microbiota in Children: A Systematic Review. *Ann Nutr Metab*. **76** (1):16-29. (2020) https://doi:10.1159/000505363
2. Sharon, G., Cruz, N.J., Kang, D.W., Gandal, M.J., Wang, B., Kim, Y.M., Zink, E.M., Casey, C.P., Taylor, B.C., Lane, C.J., Bramer, L.M., Isern, N.G., Hoyt, D.W., Noecker, C., Sweredoski, M.J., Moradian, A., Borenstein, E., Jansson, J.K., Knight, R., Metz, T.O., Lois, C., Geschwind, D.H., Krajmalnik-Brown, R. & Mazmanian, S.K. Human gut microbiota from autism spectrum disorder promote behavioral symptoms in mice. *Cell* **177**, 1600–1618.e17 (2019). https://doi.org/10.1016/j.cell.2019.05.004
3. Petropoulos, A., Stavropoulou, E., Tsigalou, C. & Bezirtzoglou, E. Microbiota gut–brain axis and autism spectrum disorder: mechanisms and therapeutic perspectives. *Nutrients* **17**, 2984 (2025). https://doi.org/10.3390/nu17182984
4. Ahmed, H., Leyrolle, Q., Koistinen, V., Kärkkäinen, O., Layé, S., Delzenne, N. & Hanhineva, K. Microbiota-derived metabolites as drivers of gut–brain communication. *Gut Microbes* **14**, 2102878 (2022). https://doi.org/10.1080/19490976.2022.2102878
5. Manghi P, Antonello G, Schiffer L, Golzato D, Wokaty A, Beghini F et al. Meta-analysis of 22,710 human microbiome metagenomes defines an oral-to-gut microbial enrichment score and associations with host health and disease. *Nat Commun* **17**(1), 196. (2025). https://doi.org/10.1038/s41467-025-66888-1
6. Młynarska E, Barszcz E, Budny E, Gajewska A, Kopeć K, Wasiak J, Rysz J, Franczyk B. The Gut-Brain-Microbiota Connection and Its Role in Autism Spectrum Disorders. *Nutrients.* **17** (7):1135. (2025) https://doi: 10.3390/nu17071135.
7. Zou R, Wang Y, Duan M, Guo M, Zhang Q, Zheng H. Dysbiosis of Gut Fungal Microbiota in Children with Autism Spectrum Disorders. *.J Autism Dev Disord.* **51**(1):267-275. (2021). https://doi: 10.1007/s10803-020-04543-y.
8. Su Q, Wong OWH, Lu W, Wan Y, Zhang L, Xu W et al. Multikingdom and functional gut microbiota markers for autism spectrum disorder. *Nat Microbiol* **9** (9), 2344–2355. (2024). https://doi.org/10.1038/s41564-024-01739-1
9. Chen Q, Xu J, Yang J, Qin X, Fan J, Ke H et al. Gut microbiota analysis in children with autism spectrum disorder and their family members. *Sci Rep* **15** (1), 44282. (2025). https://doi.org/10.1038/s41598-025-27822-z
10. Lu W, Wong OWH, Zhu J, Chen S, Tun HM, Wan Y et al. Gut microbiome

composition and strain-sharing in multiplex autism spectrum disorder families. *Nat Commun* **17** (1), 3255, (2026). https://doi.org/10.1038/s41467-026-70142-7

11. Auchtung TA, Stewart CJ, Smith DP, Triplett EW, Agardh D, Hagopian WA et al. Temporal changes in gastrointestinal fungi and the risk of autoimmunity during early childhood: the TEDDY study. *Nat Commun* **13** (1), 3151. (2022). https://doi.org/10.1038/s41467-022-30686-w
12. Peroumal D, Sahu S R, Kumari P, Utkalaja B G, Acharya N. Commensal fungus Candida albicans maintains a long-term mutualistic relationship with the host to modulate gut microbiota and metabolism. *Microbiol Spectr* **10**(6), e02462-22 (2022). https://doi.org/10.1128/spectrum.02462-22
13. Gehrig JL, Portik DM, Driscoll MD, Jackson E, Chakraborty S, Gratalo D et al. Finding the right fit: evaluation of short-read and long-read sequencing approaches to maximize the utility of clinical microbiome data. *Microb Genom* **8**(3), 000794 (2022). https://doi.org/10.1099/mgen.0.000794
14. De Sales-Millán A, Reyes-Ferreira P, Aguirre-Garrido JF, Corral-Guillé I, Barrientos-Ríos R, Velázquez-Aragón JA. Comprehensive analysis of gut microbiota composition and functional metabolism in children with autism spectrum disorder and neurotypical children: implications for sex-based differences and metabolic dysregulation. *Int J Mol Sci* **25**(12), 6701 (2024). https://doi.org/10.3390/ijms25126701
15. Maaloum M, Diop A, Ndongo S, Nguyen TT, Cadoret F, Raoult D, Fournier PE. Draft genome sequence of Megamonas funiformis strain Marseille-P3344, isolated from a human fecal microbiota. *Genome Announc* **6**(2), e01459-17 (2018). https://doi.org/10.1128/genomeA.01459-17
16. Wu C, Yang F, Zhong H, Hong J, Lin H, Zong M et al. Obesity-enriched gut microbe degrades myo-inositol and promotes lipid absorption. *Cell Host Microbe* **32**(8), 1301–1314.e9 (2024). https://doi.org/10.1016/j.chom.2024.06.012
17. Gallardo P, Izquierdo M, Viver T, Bustos-Caparros E, Piras D, Vidal RM et al. A metagenomic approach to unveil the association between fecal gut microbiota and short-chain fatty acids in diarrhea caused by diarrheagenic Escherichia coli in children. *Microb Cell* **11**, 116–127 (2024). https://doi.org/10.15698/mic2024.04.820
18. Türay S, Cangür Ş, Kahraman G, Kayabaşı E, Çetiner ÖF, Aydın B et al. Can the gut microbiota serve as a guide to the diagnosis and treatment of childhood epilepsy? *Pediatr Neurol* **145**, 11–21 (2023). https://doi.org/10.1016/j.pediatrneurol.2023.04.006
19. Zou R, Xu F, Wang Y, Duan M, Guo M, Zhang Q et al. Changes in the gut microbiota of children with autism spectrum disorder. *Autism Res* **13**(9), 1614–1625 (2020). https://doi.org/10.1002/aur.2358

20. Watanabe Y, Nagai F, Morotomi M, Sakon H, Tanaka R. Bacteroides clarus sp. nov., Bacteroides fluxus sp. nov. and Bacteroides oleiciplenus sp. nov., isolated from human faeces. *Int J Syst Evol Microbiol* **60**(8), 1864–1869 (2010). https://doi.org/10.1099/ijs.0.015107-0

21. Cuevas-Sierra A, Romo-Hualde A, Aranaz P, Goni L, Cuervo M, Martínez JA et al. Diet- and sex-related changes of gut microbiota composition and functional profiles after 4 months of weight loss intervention. *Eur J Nutr* **60**(6), 3279–3301 (2021). https://doi.org/10.1007/s00394-021-02508-0

22. Böer T, Bengelsdorf FR, Bömeke M, Daniel R, Poehlein A. Genome-based metabolic and phylogenomic analysis of three Terrisporobacter species. *PLoS One* **18**(10), e0290128 (2023). https://doi.org/10.1371/journal.pone.0290128

23. Blanco-Míguez A, Gálvez EJC, Pasolli E, De Filippis F, Amend L, Huang KD et al. Extension of the Segatella copri complex to 13 species with distinct large extrachromosomal elements and associations with host conditions. *Cell Host Microbe* **31**(11), 1804–1819.e9 (2023). https://doi.org/10.1016/j.chom.2023.09.013

24. Fernandez-Cantos MV, Babu AF, Hanhineva K, Kuipers OP. Identification of metabolites produced by six gut commensal Bacteroidales strains using non-targeted LC-MS/MS metabolite profiling. *Microbiol Res* **283**, 127700 (2024). https://doi.org/10.1016/j.micres.2024.127700

25. Hughes HK, Ashwood P. Anti-Candida albicans IgG antibodies in children with autism spectrum disorders. *Front Psychiatry* **9**, 627 (2018). https://doi.org/10.3389/fpsyt.2018.00627

26. Ozgur EG, Balci S. Might the fungus candida albicans a risk factor for autism? A meta-analysis study. *Afr Health Sci* **23**(4), 230–235 (2023). https://doi.org/10.4314/ahs.v23i4.25

27. Kantarcioglu AS, Kiraz N, Aydin A. Microbiota-gut-brain axis: yeast species isolated from stool samples of children with suspected or diagnosed autism spectrum disorders and in vitro susceptibility against nystatin and fluconazole. *Mycopathologia* **181**(1–2), 1–7 (2016). https://doi.org/10.1007/s11046-015-9949-3

28. Semon BA. Dietary cyclic dipeptides, apoptosis and psychiatric disorders: a hypothesis. *Med Hypotheses* **82**(6), 740–743 (2014). https://doi.org/10.1016/j.mehy.2014.03.016

29. Burrus CJ. A biochemical rationale for the interaction between gastrointestinal yeast and autism. *Med Hypotheses* **79**(6), 784–785 (2012). https://doi.org/10.1016/j.mehy.2012.08.029

30. Severance EG, Gressitt KL, Stallings CR, Katsafanas E, Schweinfurth LA,

Savage CLG et al. Probiotic normalization of Candida albicans in schizophrenia: a randomized, placebo-controlled, longitudinal pilot study. *Brain Behav Immun* **62**, 41–45 (2017). https://doi.org/10.1016/j.bbi.2016.11.019

31. Alonso-Monge R, Gresnigt MS, Román E, Hube B, Pla J. Candida albicans colonization of the gastrointestinal tract: A double-edged sword. *PLoS Pathog*. **17**(7):e1009710 (2021). https://doi: 10.1371/journal.ppat.1009710
32. Bays DJ, Savage HP. Candida albicans gastrointestinal colonization resistance: a host-microbiome balancing act. *Infect Immun* **93**(8), e00610-24 (2025). https://doi.org/10.1128/iai.00610-24
33. Kumamoto CA, Gresnigt MS, Hube B. The gut, the bad and the harmless: Candida albicans as a commensal and opportunistic pathogen in the intestine. *Curr Opin Microbiol*. **56**:7–15 (2020). https://doi: 10.1016/j.mib.2020.05.006
34. Shao TY, Haslam DB, Bennett RJ, Way SS. Friendly fungi: symbiosis with commensal Candida albicans. *Trends Immunol.* **43**(9):706-717 (2022). https://doi: 10.1016/j.it.2022.07.003.
35. Saromines CJ, Reinoso S, Ruiz A, Torrecillas S, Tello Martín ML, Pérez Clavijo M et al. Mushroom stem-based diets elicit region-specific shifts in rainbow trout gut microbiota. *Sci Rep* **15**(1), 42896 (2025). https://doi.org/10.1038/s41598-025-27707-1

# Acknowledgement

We express many thanks to all clinical colleagues who participated in the clinical diagnosis and care of paediatric ASD patients and assisted with faecal sample collection for this study. We are grateful to clinical staff and research team members for their valuable support regarding sample collection, clinical data collation and experimental processing. We also acknowledge the technical support provided for PacBio high-throughput long-read sequencing. This work was financially supported by the Anhui Health and Medicine Research Grants (LARX-20240304).

# Author contributions

Prof. Dr. Huabin Xu conceived the project and wrote the research grant proposal. Prof. Dr. Huabin Xu and Dr. Shizhu Zhen coordinated clinical activities and drafted the manuscript. Prof. Dr. Fangping Dai oversaw all experimental workflows and bioinformatic analyses, and co-drafted the manuscript. Dr. Guangman Zhang, Dr. Xiaoyan Zhang and Dr. Chengjun Sun performed clinical diagnosis and management of paediatric participants with ASD. Associate researcher Weiwei Zhang designed PCR primers for targeted capture of prokaryotic and eukaryotic rRNA–ITS regions. Dr. Lei Liu and Shaolong Jiang carried out library preparation and long-read sequencing on the PacBio Revio platform. Bioinformatic analysts Jianhe Sun and Linling Chen performed all bioinformatic analyses and contributed to manuscript writing and revision. Dr. Jumei Huang provided input on experimental design and revised the manuscript. All authors have read and approved the final version of the manuscript.

# Competing interests

The authors declare no competing financial or non-financial interests. The funders had no role in study design, data collection and analysis, manuscript preparation, or publication.

# Ethics statement

The study was performed in accordance with the principles of the Declaration of Helsinki. All procedures involving human participants were reviewed and approved by the Medical Ethics Committee, Affiliated Hospital of West Anhui Health Vocational College (Approval ID: KYXM-2025-003), for the project “Multi-omics Integrative Mechanism Study on the Involvement of Gut Microbiota in the Progression of Childhood Autism”. All participants were minor children. Written informed consent was obtained voluntarily from their parent(s)/legal guardian(s) prior to faecal sample collection and data acquisition. Guardians were fully informed of the study objectives, sample usage, potential risks and benefits. All consent and sample handling procedures complied with relevant ethical standards and national regulations.

# Additional information

Extended Data supporting the findings of this study will be available either as downloadable PDF files or via the official journal website upon final publication.

## Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- ExtendedDataLegendsforlongread20261031.docx