



OPEN

Phylogenomic and biochemical analysis reassesses temperate marine yeast *Yarrowia lipolytica* NCIM 3590 to be *Yarrowia bubula*

Prashant Gaikwad^{1,2}, Swanand Joshi^{1,2}, Akshay Mandlecha¹ & Ameeta RaviKumar¹

Yarrowia clade contains yeast species morphologically, ecologically, physiologically and genetically diverse in nature. *Yarrowia lipolytica* NCIM 3590 (NCIM 3590), a biotechnologically important strain, isolated from Scottish sea waters was reinvestigated for its phenotypic, biochemical, molecular and genomic properties as it exhibited characteristics unlike *Y. lipolytica*, namely, absence of extracellular lipolytic activity, growth at lower temperatures (less than 20 °C) and in high salt concentrations (10% NaCl). Molecular identification using ITS and D1/D2 sequences suggested NCIM 3590 to be 100% identical with reference strain *Yarrowia bubula* CBS 12934 rather than *Y. lipolytica* CBS 6124 (87% identity) while phylogenetic analysis revealed that it clustered with *Y. bubula* under a separate clade. Further, whole genome sequencing of NCIM 3590 was performed using Illumina NextSeq technology and the draft reported here. The overall genome relatedness values obtained by dDDH (94.1%), ANiB/ANIm (99.41/99.42%) and OrthoANI (99.47%) indicated proximity between NCIM 3590 and CBS 12934 as compared to the reference strain *Y. lipolytica*. No extracellular lipase activity could be detected in NCIM 3590 while *LIP2* gene TBLASTN analysis suggests a low 42% identity with e value 2×10^{-77} and 62% coverage. Hence molecular, phylogenetic, genomics, biochemical and microbial analyses suggests it belongs to *Yarrowia bubula*.

Yeasts from the *Yarrowia* clade differ distinctly from conventional yeasts and are widespread geographically in different ecological niches. Species within this clade are diverse and exist in both yeast and mycelial forms in order to adapt and colonize various environments. The most extensively studied and biotechnologically relevant species in the *Yarrowia* genus is *Yarrowia lipolytica* but other species have also been reported from different ecological niches such as animals, fresh/sea water and dairy. Species isolated from animal/insect related niche such as bacon, chicken and meat are *Yarrowia divulgata*¹ from Denmark, USA and Hungary; *Yarrowia galli*² (USA), *Yarrowia bubula* and *Yarrowia porcina*³ (Hungary), *Yarrowia (Candida) alimentaria*⁴ (Norway) while *Yarrowia yakushimensis*⁵ (Japan) and *Yarrowia parophonii*⁶ (Bulgaria) from insect guts. Species isolated from aquatic ecosystem, are *Yarrowia (Candida) phangngensis*⁷ (Thailand) and *Yarrowia keelungensis*⁸ (Taiwan). *Yarrowia (Candida) oslonensis*⁴ and *Yarrowia brassicae*⁹ were isolated from yogurt (Norway) and traditional Chinese sauerkraut, respectively. All the above mentioned species are non-fermentative and assimilate a wide variety of carbon sources^{1–8} suggesting phenotypic diversity and variability amongst them¹⁰. All the species in the *Yarrowia* clade are oleaginous and exhibit specific growth patterns while differing in their ability to accumulate the amount and type of lipid¹¹. A number of lipases have been explored in the *Yarrowia* clade, the most prominent being extracellular lipase, Lip2p, encoded by the *LIP2* gene¹². This gene is present in all the species of the *Yarrowia* clade but may not be expressed in active form in all the species¹³. Tolerance to NaCl (10%, w/v) differs among species with *Y. hollandica*, *Y. keelungensis*, *Y. divulgata*, *Y. porcina* and *Y. bubula* able to grow in high saline conditions^{1,3,4,8}. Similarly, diversity in optimal growth temperature is observed, ranging from 21 °C for *Y. alimentaria* to 37 °C for *Y. phangngensis* with other *Yarrowia* species growing between 25 and 30 °C¹¹. *Y. lipolytica* has often found to inhabit marine ecosystems including high saline waters^{14,15} as well as in contaminated milk, dairy products¹⁶, poultry¹⁷ and meat products¹⁸ in USA and Europe. This non-conventional strictly aerobic ascomycetous yeast generally regarded as safe (GRAS) by the American Food and Drug Administration FDA, USA has been exploited in several biotechnological, environmental and industrial applications. Some examples

¹Institute of Bioinformatics and Biotechnology, Savitribai Phule Pune University, Ganeshkhind, Pune, Maharashtra 411 007, India. ²These authors contributed equally: Prashant Gaikwad and Swanand Joshi. ✉email: ameeta@unipune.ac.in

include heterologous host for production of pharmaceutically and industrially relevant proteins, enzymes, organic acids, biofuels, bioremediation of industrial and environmental wastes^{19–22} and is also considered as a very good model for dimorphism studies^{23,24}. The strain NCIM 3590 obtained from National Collection of Industrial Microorganisms (NCIM), Pune, India is a psychotropic marine isolate from Scottish seawater²⁵ and deposited in National Collection of Yeast Cultures as NCYC 789 and in Microbial Type Culture Collection, Chandigarh as MTCC 35. This strain has been studied for its potential in biodiesel production²⁶, bioabsorption and bioremediation of heavy metals^{27,28}, in nanoparticle synthesis and melanin production²⁹. A survey of literature suggests that scarce information regarding its basic microbiology, growth and physiology exist and is the only *Y. lipolytica* strain reported to grow at lower temperatures (20 °C and less) and did not exhibit any extracellular lipase activity, a characteristic of *Y. lipolytica* strains. The classical identification of yeast species has traditionally been based on colony/cellular morphology and physiological characteristics with respect to their sugar assimilation, fermentation and oxidation capabilities³⁰. These morphological and physiological characteristics are not sufficient to identify *Yarrowia* species as very often similar profiles between species are obtained. In this context, correct identification and classification of yeast is essential for basic biological and application based research and as the genomes of many type strains of species belonging to the *Yarrowia* clade have been sequenced, this information would be useful for identification. Hence, we re-examine this biotechnologically relevant strain using microbiological, biochemical, molecular biological and genome based studies.

Results and discussion

ITS and D1/D2 rDNA identification and phylogenetic analysis. The sequenced lengths of ITS and D1/D2 regions of NCIM 3590 showed that it consisted of 311 and 446 nucleotides, respectively which have been submitted to GenBank (NCBI, USA) with accession numbers MK411246 and MK411222, respectively. BLAST analysis of ITS regions of NCIM 3590 showed 100% identity with the reference strain *Y. bubula* CBS 12934 (KY105958.1) with query coverage of 95%. The ITS sequence of NCIM 3590 when compared with *Y. bubula* CBS 12934 (consensus length 290 nt) showed 100% similarity index with gap number and gap length as 0 and divergence as 0%. In contrast, comparison with reference *Y. lipolytica* CBS 6124 (consensus length 311 nt), 83% similarity with gap number (5), gap length (11), divergence (14.2%) and sequence identity (87%) was seen. For D1/D2 sequences, NCIM 3590 showed 100% identity, query coverage, and sequence similarity with 0% divergence, gap length and gap number (0) with the reference *Y. bubula* CBS 12934 (NG_059943.1). Sequence comparison with *Y. lipolytica* CBS 6124, showed a similarity index of 89.5% with gap number and gap length (1), divergence (13%) and identity (87.5%). Generally, ITS region is selected as the standard fungal barcode for identification³¹ but, for yeasts it is recommended to use D1/D2 together with ITS region for identification and establishment of evolutionary relationships³². It has also been reported that strains showing more than 1% difference or with changes in more than 3 nucleotides in ITS and D1/D2 domains are likely to represent different species⁴. Thus, the molecular identification using ITS and D1/D2 sequences suggests that NCIM 3590 could belong to *Y. bubula*. Further the phylogenetic analysis of ITS and D1/D2 sequences from GenBank, NCBI showed that *Yarrowia* sp. could be separated into 13 clades (Fig. 1a,b and Table S1). All the seven different *Y. lipolytica* strains used in the study were grouped together in a single clade with 100% bootstrap support while the ITS and D1/D2 sequences of NCIM 3590 clustered with CBS 12934 suggesting its close relatedness. Since a close similarity between NCIM 3590 with *Y. bubula* and its distinct difference from *Y. lipolytica* strains was seen, the sequence comparisons of the strains from these clades/groups were analyzed and the results for ITS and D1/D2 given in supplementary information (Table S2). The ITS and D1/D2 genetic distances of NCIM 3590 with other *Y. lipolytica* species ranged from 0.392 to 0.407 and ~0.21 base substitutions per site, respectively (Table 1). The percent divergence over sequence pairs between NCIM 3590 with different *Yarrowia* groups was calculated using *p*-distance method and the results given in supplementary information (Table S4). NCIM 3590 was considered as a different group for comparison and establishing the evolutionary divergence. The base differences per site in ITS and D1/D2 sequences between each *Yarrowia* species was ascertained and averaged forming a dataset of 14 groups and divergence given (Table S4). The divergence for ITS and D1/D2 rDNA sequences with *Y. lipolytica* was 13.35 and 9.19%, respectively and no divergence seen with *Y. bubula* (Table S4). The variations found in their D1/D2 and ITS sequences are adequate to separate the *Yarrowia* species from one another^{4,11}. Thus, as per delineation of species since less than 1% and no nucleotide difference was observed between NCIM 3590 and *Y. bubula* CBS 12934, the phylogenetic analysis suggests that they belong to the same species and is distinctly separated from the *Y. lipolytica* clade.

NGS and de novo genome assembly. In order to validate the above results, whole genome sequencing of NCIM 3590 was undertaken. The initial genome de novo assembly was carried out using SPAdes assembler (optimized with 10 million reads, *k*-mers: 21, 33, 55 and 77) enabling assembly of 2485 contigs representing with N50 of 39,245 bp (Table 2). Contigs (521) were found to be more than 10 kb length with an average contigs length of 8592 bp. The SPAdes³³ assembler relies on “paired *de-Bruijn* graphs” (PDBG) approach which utilises a *k*-bimer based adjustment strategy for creating *de-Bruijn* graph using the paired-end reads. The genome is assembled using multiple *k*-mer sizes and eventually combining the reads into contigs. The assembler was initially designed for assembling prokaryotic genomes but later developed to accommodate large eukaryotic genomes. Assembled contigs were further scaffolded de novo using SSPACE that resulted in 2074 scaffolds with the N50 of 69,096 bp. SSPACE (SSAKE-based Scaffolding of Pre-Assembled Contigs after Extension) program is used to scaffold pre-assemblies produced by SPAdes. SSPACE³⁴ requires paired-end data from next-generation sequencing technology, read orientation information, mean values and standard deviations of the insert sizes used in library preparation. Using paired-read sequencing data the assembler assess the order, distance and orientation of contigs and combines them into scaffolds. Based on the alignments the contigs are linked into scaffolds

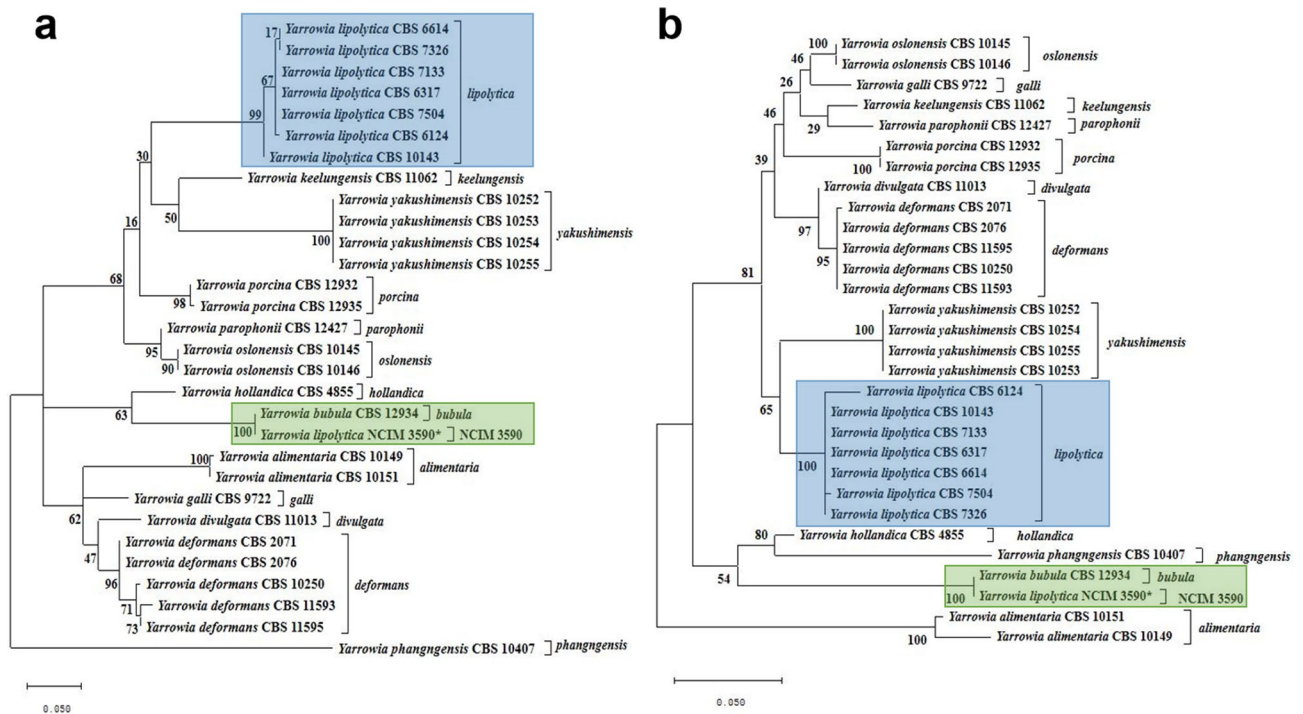


Figure 1. Phylogeny of NCIM3590. The phylogenetic tree analysis by maximum likelihood (ML) method showing the placement of NCIM 3590 and related species based on the analysis of (a) internal transcribed spacer (ITS) regions and (b) D1/D2 domains of the large subunit rRNA gene. The ML tree was obtained using the GTR + G + I model. The tree with the highest log likelihood is shown. The tree is drawn to scale with branch lengths measured in the number of base substitutions per site. All positions containing gaps and missing data were eliminated. Blue box represents *Y. lipolytica* as single clade/group and green as *Y. bubula* as single clade/group.

Organisms	NCIM 3590 (base substitutions per site)	
	D1/D2	ITS
<i>Yarrowia lipolytica</i> CBS 6124	0.209	0.407
<i>Yarrowia lipolytica</i> CBS 6317	0.197	0.403
<i>Yarrowia lipolytica</i> CBS 6614	0.192	0.407
<i>Yarrowia lipolytica</i> CBS 7133	0.192	0.403
<i>Yarrowia lipolytica</i> CBS 7326	0.192	0.407
<i>Yarrowia lipolytica</i> CBS 7504	0.194	0.403
<i>Yarrowia lipolytica</i> CBS 10143	0.192	0.392
<i>Yarrowia bubula</i> CBS 12934	0.000	0.000
<i>Yarrowia lipolytica</i> CBS 6124	0.209	0.407
<i>Yarrowia lipolytica</i> CBS 6317	0.197	0.403

Table 1. Evolutionary distances between NCIM 3590 and various *Y. lipolytica* and *Y. bubula* strains using GTR + G + I model^a. ^aEvolutionary distances between pairs of sequences are shown. Analyses were conducted using the General Time Reversible model. This analysis involved 30 nucleotide sequences. There were a total of 403 positions (sites) for D1D2 and 314 (sites) for ITS in the final dataset. Evolutionary analyses were conducted in MEGA X.

and N-characters (gaps) are placed between the connected scaffolds. As per assembly statistics, scaffolds (730) greater than 1 kb were considered for super-scaffolding/gap closing along with paired end data using SOAPdenovo2 with asm flags (3, 4) parameter (<https://github.com/aquaskyline/SOAPdenovo2>). SOAPdenovo2³⁵ utilizes six modules namely, read error correction, *de Bruijn* graph (DBG) construction, contig assembly, paired-end (PE) reads mapping, scaffold construction, and gap closure. As a result of super-scaffolding method applied on greater than 1 kb scaffolds, the total number of scaffolds and non-ATGC count was decreased. The minimum scaffold length was increased from 1000 to 1032 bases (Table 2). The de novo genome assembly strategy was selected as an unbiased approach since it does not consider prior knowledge of the source DNA sequence length,

	Contig	Scaffold	Draft genome
Contigs generated	2485	2074	485
Maximum contig length	277,850	406,748	406,748
Minimum contig length	500	500	1032
Average contig length	8592	10,333	41,258
Total contigs length	21,352,539	21,431,617	20,010,574
Total number of non-ATGC characters	346	35,429	33,863
Percentage of non-ATGC characters	0.002	0.165	0.169
Contigs \geq 1 Kbp	1069	730	485
Contigs \geq 10 Kbp	521	380	376
Contigs \geq 1 Mbp	0	0	0
N50 value	39,245	69,096	73,209

Table 2. Draft de novo genome assembly statistics of NCIM 3590.

Sr. no	Analysis	Formula/parameters	Query Genome	NCIM 3590	NCIM 3590	NCIM 3590
			Reference Genome	CBS12934	CLIB 89	CLIB 122
1	dDDH	Formula 1 (HSP length/total length)	DDH (%)	98.8	17	16.9
			Distance	0.209	0.8045	0.8051
			Prob DDH \geq 70%	99.36	0	0
		Formula 2 (identities/HSP lengths)	DDH (%)	94.1	25.68	25.6
			Distance	0.0077	0.1698	0.1696
			Prob DDH \geq 70%	96.7	0.01	0.01
		Formula 3 (identities/total lengths)	DDH (%)	99.1	17.1	17
			Distance	0.0284	0.8377	0.8382
			Prob DDH \geq 70%	99.97	0	0
				Diff in GC%	–	0.32
2	ANI	ANiB (%)	–	99.41	80.02	80.04
		ANIm (%)	–	99.42	86.14	86.16
		OrthoANI (%)	–	99.47	80.25	80.18
	Interpretation		Same species	Distinct species	Distinct species	

Table 3. Genomic relatedness using dDDH and ANI of NCIM 3590 with CBS 12934 and with two reference strains CLIB 122 and CLIB 89.

layout or composition. Super-scaffolding and gap closure in the assembled scaffolds resulted in 485 super-scaffolds with N50/N90 values of 73,209/20,565 bp, respectively. The draft NCIM 3590 genome contains approximately 20,010,574 bp and has a GC content of 47%.

The availability of completely annotated genome assembly is a significant advantage for the study of any organism. CLIB 122³⁶ was the first genome reported to be fully sequenced and annotated. Additionally WSH-Z06, H222 (CLIB 80), W29 (CLIB 89), IBT 446 and PO1f. were fully sequenced and annotated at the chromosome level while few strains were assembled upto contigs and scaffolds level (<https://www.ncbi.nlm.nih.gov/genome/browse#!/eukaryotes/yarrowia>). The reference genomes, namely, *Y. bubula* CBS 12934³⁷ and two *Y. lipolytica* strains CLIB 122 and CLIB 89 (W29)³⁸, were selected for comparison of whole genome sequencing data as these strains not only had a genome size similar to NCIM 3590 for assembly purpose and also the availability of complete genomic and annotation information.. In general, *Y. lipolytica* has a GC content reported to lie between 49 and 50%. The two reference strains CLIB 122 and CLIB 89 exhibited a GC content of 49.1 and 49.0% while for NCIM 3590 and *Y. bubula* CBS 12934 was 47 and 46.2%, respectively which was 2% lower than the *Y. lipolytica* strains (Table S5).

Overall genome relatedness index (OGRI) and genome comparison. The digital DDH (dDDH) tools calculate inter-genomic distances using three formulas and convert these distances into percent-wise dDDH similarities. For distance calculation, Formula 1 utilizes HSP (High-scoring Segment Pairs) length and total lengths of the genome; Formula 2 uses identities and HSP length while Formula 3 uses identities and total lengths^{39–41}. The calculated dDDH values for CBS 12934 with NCIM 3590 were found to be 98.8, 94.1 and 99.1% with Formulas 1, 2 and 3, respectively (Table 3) indicating that all three formulas suggest a high degree of similarity or relatedness. In contrast, the dDDH similarity values for NCIM 3590 were found to be ~ 17, 26 and 17% with Formulas 1, 2 and 3, respectively suggesting low level of relatedness. A dDDH similarity score of greater than or equal to 70% is a criterion for assigning two strains to the same species^{40,42} and hence the values obtained

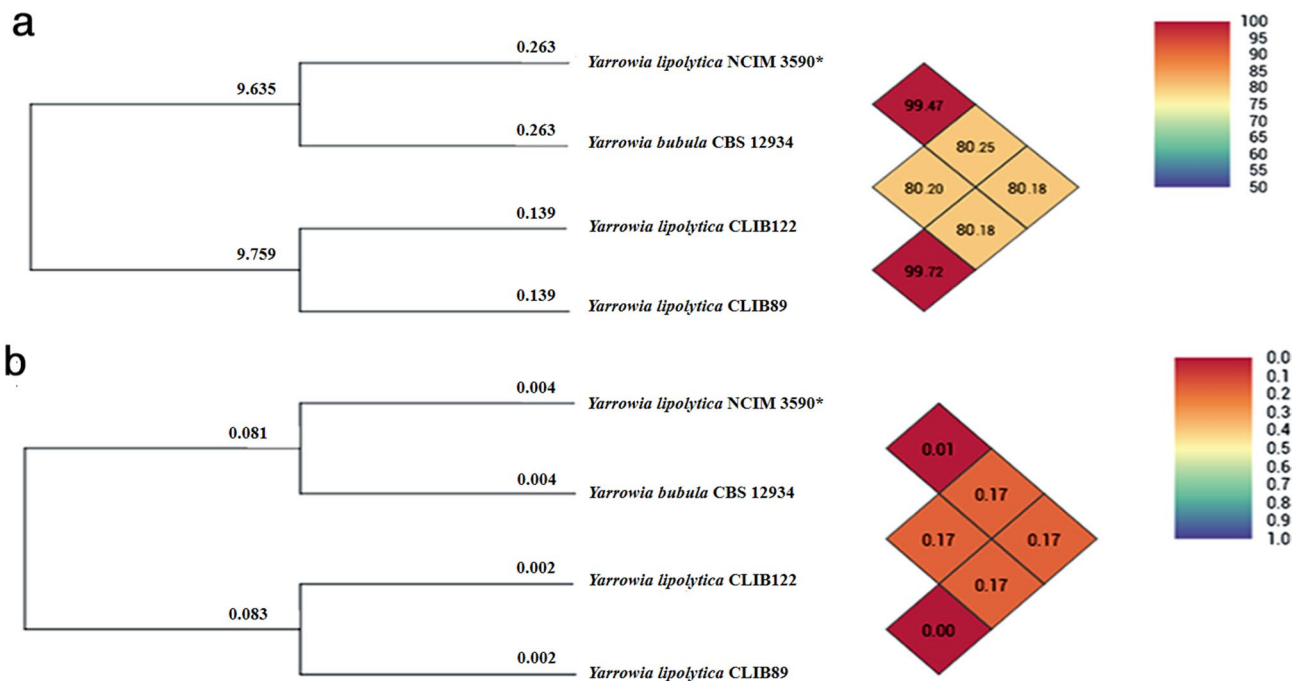


Figure 2. Dendrogram of the inter-species genomic relatedness for four *Yarrowia* strains. Heat map generated using (a) Overall orthologous average nucleotide identity values (OrthoANI) and (b) GGDC distance among *Yarrowia* genomes was calculated using Orthologous Average Nucleotide Identity Tool (OAT) software version 0.90. *Y. lipolytica* CLIB 122 (GenBank accession GCA_000002525.1), *Y. lipolytica* CLIB 89 (W29) (GenBank accession GCA_001761485.1), NCIM 3590 (GenBank accession GCA_003571375.1) and *Y. bubula* CBS 12934 (GenBank accession GCA_900519075.1). Values in color matrix boxes indicate the similarity percentage among the genomes.

suggests that NCIM 3590 and *Y. bubula* CBS 12934 are related and belong to the same species. The distance between NCIM 3590 and *Y. bubula* CBS 12934 was 0.0209, 0.0077 and 0.0284 while between *Y. lipolytica* strains and NCIM 3590 was ~0.80, 0.169 and 0.83 with Formulas 1, 2 and 3, respectively (Table 3). Two genomes are considered to belong to the same species when they have a genome distance value (GD) of less than 0.0284^{39,41} and thus, based on the distance values also the two genomes of NCIM 3590 and *Y. bubula* CBS 12934 seems to be closely related. Another parameter for relatedness using dDDH, calculates the difference in G + C content (%) which was found to be 0.32% between NCIM 3590 and reference *Y. bubula* CBS 12934 while it was 1.94% and 1.87% for CLIB 122 and CLIB 89, respectively (Table 3). Meier-Kolthoff et al.⁴³ suggested that a value greater than 1.0% indicates distinct species⁴³. Since the difference in G + C content is greater than 1% between NCIM 3590 and CLIB 122 and CLIB 89, it suggests that NCIM 3590 does not belong to the same species as CLIB 122 and 89.

Though the GGDC tool of dDDH has been effectively used for species delineation in prokaryotes, very few reports exist on its application in fungal and yeast systems. Mefteh et al.⁴⁴ used this tool to show that the two strains of *Penicillium citrinum* genomes belong to the same species (DDH 97.3% and distance 0.004), the two strains of *Geotrichum candidum* are genetically distant (DDH 18.3% and distance 0.236) using Formula 2⁴⁴. Similarly, relatedness between *Saccharomyces cerevisiae* S288c and four *Candida* species namely *C. auris* 6684, *C. albicans* (SC-5314 and WO-1), *C. lusitaniae* ATCC 42720 and *C. glabrata* CBS-138 has been studied⁴⁵. Genomic relatedness was also determined using Average Nucleotide Identity (ANI) and OrthoANI (OANI) which are the mean of nucleotide identity values between the two organisms and have been widely used indices under OGRI⁴⁶. The ANIb/ANIIm/OANI values between CLIB 122 and CLIB 89 is 99.71%/99.70%/99.72% indicating that they belong to the same species (Fig. 2a). The ANIb/ANIIm/OANI values between NCIM 3590 and CBS 12934 is ~99.41% while between CLIB 122 and CLIB 89 were ~80.00%, respectively (Table 3). The inter-genomic distance between NCIM 3590 and *Y. bubula* CBS 12934 was calculated to be 0.01 and between two *Y. lipolytica* species was 0, while between NCIM 3590 and *Y. lipolytica* species was 0.17 (Fig. 2b). Values greater than 96% and genome distance closer 0 indicate that strains belong to the same species⁴⁶. The data obtained for NCIM 3590 and CBS 12934 using ANI, OrthoANI and dDDH suggests that both strains are genomically related to each other and corroborates with the established boundaries for genomic species delineation (95–96% for ANI and OANI, 70% for digital DDH)^{46–48}. Our results suggest that the NCIM 3590 and *Y. bubula* CBS 12934 are closely related and belong to the same species.

Based on ITS and D1/D2 rDNA sequencing, whole genome comparison and OGRI, NCIM 3590 is likely to be *Y. bubula*. The strain has tremendous potential for biotechnological applications and scarce information with respect to its basic microbiological aspects is available and hence, it was deemed necessary to investigate these aspects.

Sr. no	Organism	Colony	Cellular morphology/ Budding	Temperature (°C)	NaCl	Sugar assimilation	GC %	References
	Type strain							
01	<i>Yarrowia lipolytica</i>	Butyrous, hyphal, tannish white	Spherical, ellipsoidal/multilateral	25	–	NAG +, Man +, Galv, VFM –	48.98	11
	CBS 6124							
02	<i>Yarrowia divulgata</i>	Butyrous, smooth, cream	Spherical, ellipsoidal/polar	25.30	+	NAG +, Man +, Gal –, VFM –	50.20	1
	CBS 11013							
03	<i>Yarrowia (Candida) galli</i>	Butyrous, cream, hirsute	Spherical, ellipsoidal/multilateral	30	+	NAG –, Man +, Gal +, VFM +	49.10	2
	CBS 9722							
04	<i>Yarrowia keelungensis</i>	Smooth, entire, tannish white	Ellipsoid to elongate/multilateral	25, 30, 35	+	NAG +, Man +, Gal +, VFM –	48.10	8
	CBS 11062							
05	<i>Yarrowia (Candida) phangngensis</i>	Butyrous, cream, hirsute	Spherical, ellipsoidal/multilateral	25, 30, 35, 37	+	NAG –, Man +, Gal –, VFM –	43.30	7
	CBS 10407							
06	<i>Yarrowia (Candida) oslonensis</i>	Not reported	Ovoid to globose/multilateral	25, 27, 30	–	NAG +, Man –, Gal +, VFM –	50.60	4
	CBS 10146							
07	<i>Yarrowia (Candida) hollandica</i>	Not reported	Ovoid to globose/multilateral	25, 27, 30	+	NAG +, Man +, Gal +	47.80	4
	CBS 4855							
08	<i>Yarrowia (Candida) alimentaria</i>	Not reported	Ovoid to globose/multilateral	21, 25, 27	–	NAG +, Man –, Gal +, VFM –	49.20	4
	CBS 10151							
09	<i>Yarrowia bubula</i>	Butyrous, smooth, cream	Ellipsoid/multilateral	25	+	NAG +, Man +, Gal +, VFM –	46.20	3
	CBS 12934							
10	<i>Yarrowia porcina</i>	Butyrous, smooth, cream	Ellipsoid/multilateral	25, 30	+	NAG +, Man +, Gal +, VFM –	43.70	3
	CBS 12935							
11	<i>Yarrowia yakushimensis</i>	Butyrous cream	Ovoid to elongate/multilateral	25	ND	NAG –, Man +, Gal –, VFMnd	48.30	5
	CBS 10254	Undulate margin						
12	<i>Yarrowia deformans</i>	Mucoid, cream, wrinkled	Ellipsoid, elongate/multilateral	25	ND	NAG +, Man +, Gal –, VFM –	49.50	5
	CBS 2071							
13	<i>Yarrowia parophonni</i>	Butyrous, smooth, cream	Ovoid to globose/multilateral	25	ND	NAG +, Man +, Gal +, VFM –	ND	6
	CBS 12427							
14	<i>Yarrowia lipolytica</i>	Umbonate, Opaque, wrinkled	Spherical, ellipsoidal/bilateral	10, 15, 20, 25, 28	+	NAG +, Man +, Gal +, VFMnd	47.00	Present study
	NCIM 3590							

Table 4. Comparison of NCIM 3590 with different *Yarrowia* sp. (+) Positive, (–) Negative, *v* Variable, ND not determined, NAG *N*-Acetyl-D-glucosamine, *Man* D-Mannitol, *Gal* D-galactose, *VFM* vitamin free medium; GC%—<https://www.ncbi.nlm.nih.gov/genome/browse#!/eukaryotes/yarrowia>. CBS Central Bureau voor Schimmelcultures, Utrecht, The Netherlands, NCIM National Collection of Industrial Microorganisms, Pune, India.

Colony morphology. As no information regarding phenotypic characterization, namely, size, form, elevation, margin edge, surface, opacity and chromogenesis of NCIM 3590 is available in literature, the yeast growth on different culture media was undertaken⁴⁹. The marine yeast NCIM 3590 was able to grow on all the eight different growth media studied (Fig. S1; Table S6), and compared with the reference strains *Y. lipolytica* CBS 6124¹⁴ and *Y. bubula* CBS 12934³ (Table 4). Of these, five media (MEA, MGY, YPG, PDA and TA) exhibited similar colony characteristics and were opaque, white in colour, circular, 2–5 mm in diameter, entire margins and edge, umbonate, dome-shaped elevation with shiny, wrinkled surface. In contrast, the reference strain when grown on MEA (5%, w/v) for 72 h showed tannish-white butyrous colonies¹⁴ whereas *Y. bubula* showed cream coloured butyrous colonies³.

The yeast when grown on YLDM, YES and YPO showed varied colony morphologies (Table S6). YLDM is a selective medium used to differentiate *Y. lipolytica* from other yeasts as its colonies produce a unique deep brown colour after 24 h due to the presence of pigments^{49,50}. On YLDM, NCIM 3590 showed circular, white coloured, hat-shaped colonies (Fig. S1f) in contrast to the reference *Y. lipolytica* strain. On YES, media, no growth was seen when grown on minimal medium YNB-sucrose (1%, w/v) suggesting its inability to utilize the sugar while, it was able to assimilate sucrose when grown on complete medium, with white wrinkled surface, erose margin and fuzzy growth (Fig. S1g). YES agar, is routinely used for sporulation tests in yeasts³. No sporulation was seen in NCIM 3590 upto 15 days. Most *Yarrowia* strains are haploid (the only known exception is (CBS 6124) and therefore cannot sporulate unless being mated. On YPO, the colony morphology is given in Table S6 and no zone of clearance was observed in 72 h suggesting the absence of extracellular lipolytic activity (Fig. S1h). This is unlike *Y. lipolytica* which is a known producer of extracellular lipase and shows a clear zone of clearance when grown on YPO. Hence, growth patterns for NCIM 3590 were found to be different from the reference *Y. lipolytica* strain.

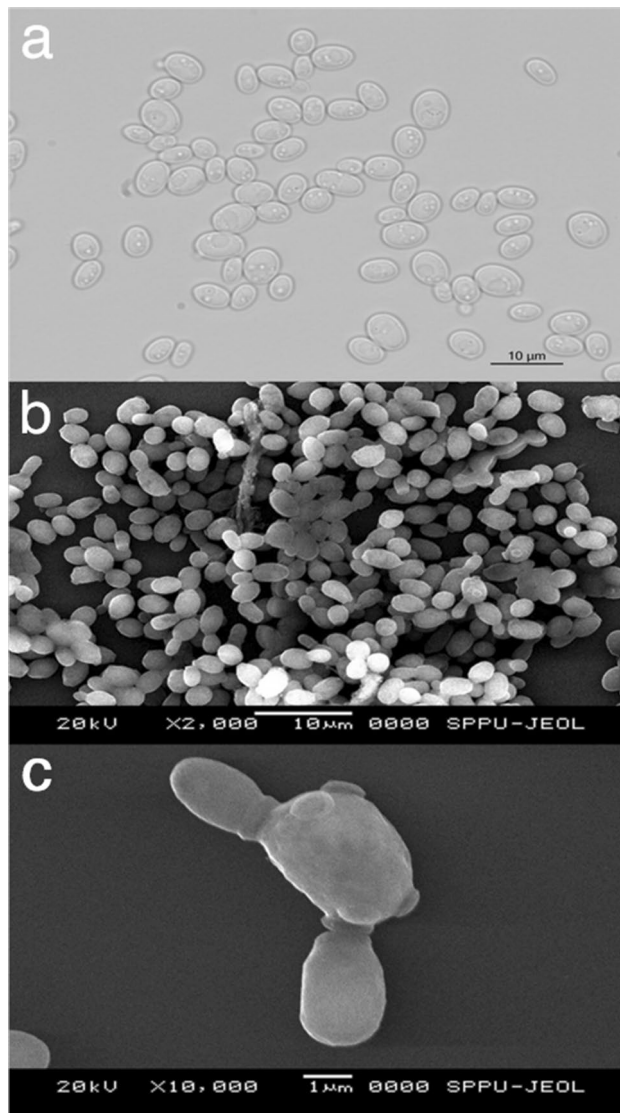


Figure 3. Cell morphology of NCIM 3590. (a) Light microscopy and (b) SEM and micrographs of NCIM 3590 strain grown on YNBG broth for 72 h. Micrographs showing Bar 10 μm . (c) shows budding and the arrows indicates bud scar. Bar 1 μm .

Cellular morphology. The light microscopic image (Fig. 3a) and scanning electron micrographs (Fig. 3b,c) of cells grown for 72 h in YNBG liquid media were spherical to ellipsoidal, 3–6 μm in size and displayed unipolar or bipolar budding. Elongated yeast forms were observed after 5 days of incubation and no filaments were observed even when grown for more than 7 days. *Yarrowia* sp. are dimorphic exhibiting yeast cells, pseudo-hyphae and hyphae depending on growth conditions.

A comparison of the cellular morphology and budding pattern between different *Yarrowia* sp. is given in Table 4. It is to be noted that while most of *Yarrowia* sp. reported so far show multilateral budding, NCIM 3590 shows bilateral budding pattern. In YNB medium, *Y. lipolytica* grows as the yeast form with a polar budding while hyphal growth can be induced either by *N*-acetyl-D-glucosamine (NAG) or by adding serum to the culture medium⁵¹. No growth of NCIM 3590 upto 5 days could be seen in vitamin free media and serum (1%, v/v). In NAG (1%, w/v) and serum (10%, v/v) the cells grew in yeast form and no transition to the filamentous form was observed. This is in agreement with the earlier report by Bankar et al.²⁵ wherein, only yeast forms were observed for NCIM 3590²⁴.

Effect of temperature and salt on growth of NCIM 3590. The effect of different temperatures and salt concentrations on growth of NCIM 3590 strain was evaluated. The yeast grew at temperatures between 10 and 28 $^{\circ}\text{C}$ with optimal growth at 20 $^{\circ}\text{C}$ and no growth on YNBG was observed at 30 $^{\circ}\text{C}$ and beyond as shown in Fig. 4a. Being a marine isolate, the effect of salt was studied on the yeast NCIM 3590 upto 72 h (Fig. 4b). The

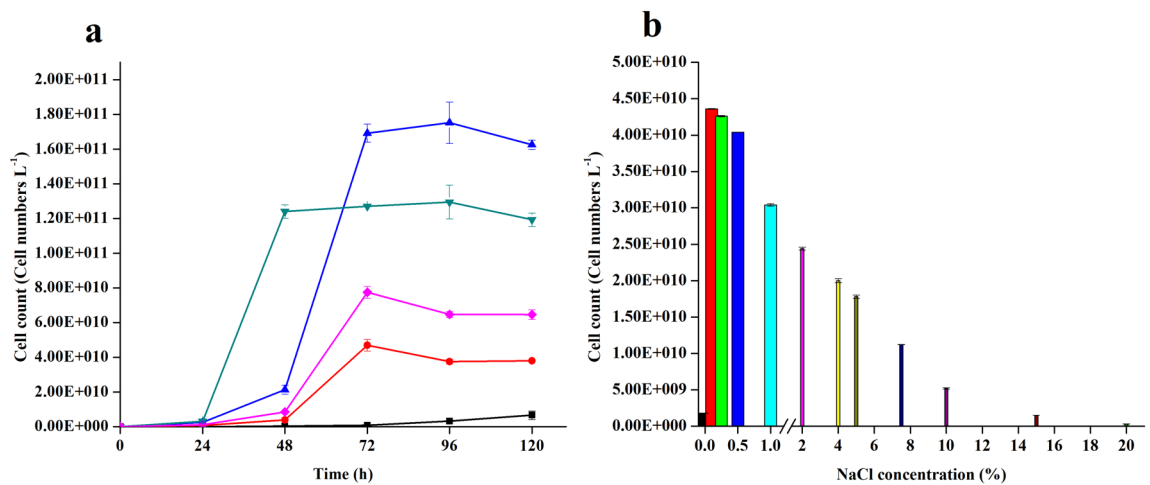


Figure 4. Effect of temperature and salt on growth of NCIM 3590. **(a)** Effect of temperature on the growth of NCIM 3590. The yeast was grown on YNBG at different temperatures for 120 h. Temperature symbols—10 °C (black filled square), 15 °C (red filled circle), 20 °C (blue triangle), 25 °C (dark cyan inverted triangle), 28 °C (magenta filled square). Data points are mean values from triplicate experiments with their standard deviation. **(b)** Effect of sodium chloride concentration on growth of NCIM 3590. Salt legends – 0.0% (black filled square), 0.1 (red filled square), 0.25% (green filled square), 0.5% (blue filled square), 1% (cyan filled square), 2% (magenta filled square), 4% (yellow filled square), 5% (dark yellow filled square), 7.5% (navy filled square), 10.0% (purple filled square), 15.0% (wine filled square), 20.0% (olive filled square). All experiments were carried out in duplicates and the mean values were recorded and the standard deviations (n=2) are indicated as error bars. Different coloured bar indicates different concentration of NaCl used.

yeast was able to grow well upto 10% NaCl/5% glucose. Lesser growth was seen at higher salt concentration (15 and 20% NaCl/5% glucose) after 120 h.

On comparison with other *Yarrowia* sp. (Table 4), diversity in growth temperature and salt concentration is seen. In general, most of the *Yarrowia* sp. reported to date grow optimally between 25 and 30 °C. For example, *Y. lipolytica* CBS 6124¹⁴, *Y. divulgata* CBS 11013¹, *Y. galli* CBS 9722², *Y. oslonensis* CBS 10146⁴ and *Y. hollandica* CBS 4855⁴ grew upto 30 °C. Growth of *Y. porcina* CBS 12935³ and *Y. keelungensis* CBS 11062⁸ is reported upto 35 °C while *Y. phangngensis* CBS 10407⁷ could grow upto 37 °C. As of date, *Y. alimentaria* CBS 10151⁴ is the only reported strain able to grow between 15 and 25 °C with an optimal growth at 21 °C¹¹ while NCIM 3590 was the only *Yarrowia* strain with an optimal growth temperature of 20 °C. Amongst the *Yarrowia* clade, strains reported to grow in 10% NaCl were *Y. divulgata* CBS 11013¹, *Y. galli* CBS 9722², *Y. bubula* CBS 12934³, *Y. porcina* CBS 12935³, *Y. hollandica* CBS 4855⁴, *Y. phangngensis* CBS 10407⁷ and *Y. keelungensis* CBS 11062⁸ while *Y. lipolytica* CBS 6124¹⁴, *Y. keelungensis* CBS 11062⁸ and *Y. alimentaria* CBS 10151⁴ were unable to grow at this concentrations. Thus, a variation amongst *Yarrowia* species was observed as they could have adapted to the diverse ecological niches they were isolated from. Hence, based on growth at low temperature and high salt concentrations, it suggests that NCIM 3590 clearly differs phenotypically from *Y. lipolytica* CBS 6124.

Biochemical characterization. *Sugar assimilation studies.* The assimilation and fermentation studies of NCIM 3590 were carried out on differing sugars and sugar alcohols (Table 4). The test for glucose fermentation was found to be negative as no gas production could be detected which corroborates the results obtained with other *Yarrowia* sp.^{1–8}. Assimilation test for NCIM 3590 at 72 h were analysed using the metabolic fingerprinting database in Biolog. Carbon compounds strongly assimilated by NCIM 3590 were D-glucose, i-erythritol, D-gluconic acid, 2-keto D-gluconic acid and N-acetyl-D-glucosamine (NAG). Weak assimilation was seen for D-galactose, salicin, L-sorbose, D-xylose, L-arabinose, D-arabinose, D-ribose, glycerol, D-mannitol, succinate while no assimilation could be seen for inulin, sucrose, D-raffinose, D-melibiose, D-trehalose, maltose, α-Methyl L-D-glucoside, D-cellobiose, L-rhamnose and D-glucosamine (Table S7). The results from Biolog suggests NCIM 3590 as *Y. lipolytica* with probability score 1.0, similarity value (SIM) of 0.896 and distance of 1.543. A species to be considered acceptable for identification must have a distance value greater than 5.0 and a SIM of greater than 0.75. Based on the probability score and SIM value NCIM 3590 can be identified as *Y. lipolytica*. However, the distance value is much lower than the acceptable value of greater than 5. Also to be noted that the Biolog database has only *Y. lipolytica* and no other *Yarrowia* species listed in it and hence the best hit amongst the available database with *Y. lipolytica* is seen.

Sugar assimilation and utilization patterns also illustrate the diversity and adaptation amongst strains thereby, offering a convenient key for yeast identification. Upon comparison (Table 4), all *Yarrowia* species assimilate glucose, glycerol and i-erythritol while strains with differential assimilation of sugars such as NAG, D-mannitol and D-galactose have been reported. NAG could not be assimilated by *Y. galli* CBS 9722², *Y. yakushimensis* CBS 10254⁵ and *Y. phangngensis* CBS 10407⁷. D-mannitol assimilation was not observed in *Y. alimentaria* CBS 10151⁴ and *Y. oslonensis* CBS 10146⁴ while D-galactose could not be utilized by *Y. divulgata* CBS 11013¹, *Y. yakushimensis*

CBS 10254⁵, *Y. deformans* CBS 2071⁵ and *Y. phangngensis* CBS 10407⁷. However, all the three sugars could be assimilated by NCIM 3590, suggesting that it behaved differently in its sugar assimilation profile. The three strains *Y. porcina* CBS 12935³, *Y. hollandica* CBS 4855⁴ and *Y. keelungensis* CBS 11062⁸ are reported to grow at 30 °C and above while NCIM 3590 is known to grow at lower temperature. As of date *Y. alimentaria* CBS 10151⁴ is the only reported strain able to grow at lower temperature¹¹ but its inability to assimilate mannitol differentiates it from NCIM 3590.

Extracellular lipase activity. *Yarrowia lipolytica* is a known producer of high amounts of extracellular lipase and hence NCIM 3590 was screened for lipolytic activity by culturing on TO and YPO at 20 °C for 48 h with no clearance zone observed after 96 h incubation. The extracellular lipolytic activity in NCIM 3590 was also investigated on different broth media namely, YNBG, YPG, YPGO, YPGTr and YPGTw. Negligible amount of extracellular lipase activity was detected in supernatant of YNBG, YPG, YPGO, YPGTr and YPGTw under the given assay conditions while good cell growth was seen after 72 h incubation at 20 °C (Table S8). *Y. lipolytica* PO1d, a known producer of extracellular lipase produced 20 and 50 U/mL of lipase activity in YNBO (Yeast nitrogen base containing 1% (w/v) olive oil) and YPGO, respectively while *Y. lipolytica* strains, ATCC20460 and IMUFRJ 50682, produced up to 30 U/L on unsupplemented olive mill wastewater (OMW)⁵². *Y. lipolytica* CECT 1240 (ATCC 18942)⁵³ and *Y. lipolytica* W29 (ATCC 20255)⁵⁴ showed higher lipolytic activity of 700 and 770 U/L with YNBO and YPDO.

Thus, though variability in the levels of extracellular levels of lipase are seen in *Y. lipolytica*, all strains produced it unlike as that seen in case of NCIM 3590 which did not show any significant levels of activity.

LIP2 gene analysis. As no significant extracellular lipase activity was detected in the crude supernatant, bioinformatic analysis was carried out to determine the presence of *LIP2* gene in NCIM 3590. Extracellular lipase, Lip2, encoded by *LIP2* gene (Gene Id YALI0A20350g) is a 334-amino acid precursor protein containing the putative 13-amino acid signal sequence¹³. The TBLASTN of *LIP2* gene showed 42% identity and 60% coverage with e value 2×10^{-77} with scaffold 57 of NCIM 3590 (Table S9). Further, the matched coordinates 100,268 to 101,158 from scaffold 57 (NKYT01000426.1) were used to identify the ORFs using the NCBI ORF finder. The generated 8 ORFs were used as query to carry out BLASTP (Reverse BLAST) against the *LIP2* gene. Out of 8 ORFs generated only one ORF showed 44% identity (113/254 aa) with e value 9×10^{-79} and 62% coverage (158/254 aa). To validate the results, a similar TBLASTN for *LIP2* gene was carried out with reference CLIB 122 and CLIB 89 which showed 100% identity and coverage while in reverse blast, one ORF out of 9, resulted in 100% identity and coverage with *LIP2*. Thus it seems that the ORF obtained from NCIM 3590 shows low homology to *LIP2*. According to Meunchan et al.¹⁴, while *LIP2* gene is likely to be present in all members of the clade it has been suggested that the gene has undergone a number of evolutionary events with a high number of duplications. Differing degrees of homologies amongst them exist and 11 lipases homologous to YLLip2 seen of which many were found to be transcriptionally inactive while others were actively expressed as in *Y. lipolytica*, *Y. galli* and *Y. phangngensis*¹³. Hence, low e value, percent identity and coverage in case of NCIM 3590 suggests that the lipase from NCIM 3590 may be transcriptionally inactive or may not belong to the *LIP2* family.

In conclusion, this study reassesses the strain NCIM 3590 based on molecular, phylogenetic, genomic, biochemical and microbiological data. Based on this, we suggest that NCIM 3590 and *Y. bubula* CBS 12934 belong to same species. The availability of NCIM 3590 genome will help in providing a platform for elucidating its potential applications and contribute to the understanding of this unusual *Yarrowia* strain with optimum growth temperature at 20 °C.

Materials and methods

Media and culture conditions. NCIM 3590 obtained from National Collection of Industrial Microorganisms (NCIM), NCL, Pune, India was maintained on Yeast extract Peptone Dextrose agar (YPD; Yeast extract, 0.3%; Peptone, 0.5%; Dextrose, 1%) at 20 °C for 48 h.

ITS and D1/D2 sequencing and phylogenetic analysis. Colonies of NCIM 3590 were suspended in saline and genomic DNA was isolated using geneO-spin Microbial DNA isolation kit (geneOmbio Technologies, Pune, India) in accordance with the manufacturers instructions. The Internal transcribed spacer, ITS1 (5'TCC GTAGGTGAACCTGCGG-3') and ITS4 (5'-TCCTCCGCTTATTGATATGC-3')⁵⁵ and for the D1/D2 domain, primers NL-1 (5'-GCATATCAATAAGCGGAGGAAAAG-3') and NL-4 (5'-GGTCCGTGTTTCAAGAGG-3')⁵⁶ were used and amplified using standard PCR reaction with an annealing temperature of 54 and 55 °C, respectively. The products were purified by using a ExoSAP-IT PCR product Purification kit (Invitrogen Bioservices, India) and sequenced using an ABI PRISM Big Dye Terminator V3.1 kit (Applied Biosystems, USA) on 3130 Genetic analyser Automated DNA sequencing machine. The sequences were analyzed using Sequencing Analysis 5.2, ChromasPro v3.1 and BLAST analysis performed using BLASTN at NCBI server (<http://www.ncbi.nlm.nih.gov/BLAST>).

Phylogenetic analysis was performed with ITS and D1/D2 sequences of different *Yarrowia* species taken from NCBI database. All sequences were aligned separately using Clustal W⁵⁷ with default parameters in MEGA X⁵⁸ (Molecular Evolutionary Genetics Analysis) software. For both ITS and D1/D2 rDNA set the best fit nucleotide substitution model was determined using Maximum likelihood (ML) criterion. ML tree was constructed using general time-reversible model with gamma-distributed rates of variation among sites and a proportion of invariable sites (GTR + G + I model). The reliability of the trees was tested by bootstrapping with 1000 replicates. The tree is drawn to scale with branch lengths in the same units as those of evolutionary distances (number of base substitutions per site) which is used to infer the phylogenetic tree. Percent divergence at nucleotide level was calculated using *p*-distance method in MEGA X (www.megasoftware.net).

NGS and de novo genome assembly. The genomic DNA from NCIM 3590 was extracted by CTAB method followed by NEXTflex DNA library preparation and Illumina NextSeq 500 paired-end sequencing according to the manufacturer's instructions. In total, ~41 Million paired-end reads (150 bp) were generated (estimated coverage ~600 ×). The generated paired-end reads of 150 bps were processed further for de novo genome assembly. Quality control of reads was checked with FastQC v2.2 (<http://www.bioinformatics.babraham.ac.uk/projects/fastqc/>) and trimming carried out using an in-house perl script. Reads were assembled into contigs using SPAdes v3.1³³ and were further assembled into scaffolds with SSPACE³⁴. Scaffolds were checked for the length and the scaffolds which are above 1000 bases length were used for super-scaffolds and gap closure using SOAPdenovo2 tool³⁵. The reads obtained were submitted to Sequence Read Archive (SRA) and scaffolds to Genbank in NCBI repository.

Overall genome relatedness index (OGRI) and genome comparison. The OGRI between genome of NCIM 3590 and the reference genomes of CLIB 122, W29 (CLIB 89) and CBS 12934 was determined using digital DNA-DNA hybridization (dDDH) and Average Nucleotide Identity (ANI). The dDDH was calculated using the web-based DSMZ service (<http://ggdc.dsmz.de>) available with Genome-to-Genome Distance Calculator (GGDC 2.0) with BLAST method. The ANI was determined using JspeciesWS⁴⁸ while orthoANI was calculated using Orthologous Average Nucleotide Identity Tool⁴⁶ (OAT).

Colony morphology of NCIM 3590. The colony morphology was studied by growing the yeast on different media as described by Kurtzman et al.⁴⁹. The strain was grown for 96 h at 20 °C on different agar media (w/v) namely; (a) Malt extract agar (ME), (b) Malt extract Glucose Yeast extract Peptone agar (MGYP), (c) Potato Dextrose agar (PD), (d) Yeast extract Peptone Glucose agar (YPG), (e) Tributyrin agar (T), (f) *Yarrowia lipolytica* Differential Medium (YLDM), (g) Yeast extract Sucrose agar (YES) and (h) Yeast extract peptone olive oil agar (YPO). Colony morphology of the strain was ascertained microscopically by using trinocular microscope (Model no. NSZ-606) Lawrence and Mayo, London. Media composition details given in supplementary information.

Microscopy studies. Bright field and Scanning Electron Microscopy (SEM) were carried out to determine the morphology. Wet mounts of the washed *Yarrowia* cells grown for 72 h under shaking conditions (120 rpm) at 20 °C in Yeast Nitrogen base containing Glucose (1%, w/v) (YNBG) were prepared. Cell morphology was ascertained on a slide by use of Axioskope 40 light microscope attached with a photographic equipment and images acquired at 100 × magnification (ProgRes Capture Pro 2.7 and the software AxioVison Rel. 4.8)⁵⁹.

For SEM, the separated cells were fixed in a 3% glutaraldehyde solution in 0.1 M Phosphate Buffered Saline (pH 7.2) for 30 min and washed three times in the same buffer. After fixation the cells were dehydrated in a series of alcohols of increasing concentrations (35, 50, 75, 95% and absolute ethanol) for 20 min at each stage and then with Hexamethyldisilazane for 10 min. The specimens were dried overnight on cleaned grease free silicon wafer and sputtered with a platinum-carbon mixture in JFC-1600 (JEOL, Japan) vacuum evaporator and observed under the SEM (JEOL JSM-6360A)⁶⁰.

Effect of temperature and salt on growth of NCIM 3590. For pre-inoculum, the yeast cells were grown in liquid YNBG and incubated on a rotary shaker (120 rpm) at 20 °C for 48 h, cells harvested and washed twice with autoclaved distilled water and centrifuged at 10,000g for 10 min. The cell pellet was re-suspended in water and 1 OD cells ($1 \text{ OD} \sim 2 \times 10^7 \text{ CFU/mL}$) inoculated per 100 mL media as mentioned below. To study the effect of temperature on growth of NCIM 3590, batch experiments were carried out. Cells (0.5 OD per 50 mL media) were inoculated into YNBG and incubated at different temperatures (5, 10, 15, 20, 25, 28, 30 and 35 °C) on a rotary shaker (120 rpm) for varying time intervals (24–120 h) and growth assessed every 24 h. The cells were spun at 10,000g for 10 min at 4 °C and the pellet was washed twice with sterile distilled water, vortex mixed to separate the cells, OD₆₀₀ taken and cell dry weight determined by freeze-drying the cells. Similarly, to study the effect of salt, different concentration of sodium chloride (0, 0.1, 0.25, 0.5, 1, 2, 4, 5, 7.5, 10, 15 and 20%, w/v NaCl) in 5% glucose on NCIM 3590 growth was determined as mentioned above.

Sugar assimilation. The sugar assimilation profile of NCIM 3590 was determined using Biolog system (Biolog MicroStation with Microlog System, Release 4.20, Biolog, Hayward, CA, USA, (<http://www.biolog.com/microID.html>)). Culture was grown on YPD agar for 72 h at 20 °C and the yeast suspension was prepared in 15 mL sterile water at inoculum density transmittance level $47 \pm 2\%$ (corresponds to 0.33 Optical density (OD₆₀₀ nm) or 6×10^6 colony forming units per mL (CFU/mL)). The Biolog YT MicroPlate was inoculated with 100 µL of cell suspension and incubated for 24, 48, 72 and 96 h at 20 °C. Colorimetric change in each well was referenced against control wells and scored as mentioned in the protocol.

The fermentation test was carried out manually by Kurtzman et al.⁴⁹ wherein 0.1 OD cells were inoculated in YNB containing 1% (w/v) glucose in medium size tubes containing inverted durhams tube⁴⁹. Bromothymol blue was used as an indicator which on acid production changes the medium color from blue to green or yellow. Gas production is evidenced visually by presence of visible air bubbles trapped inside the Durham tube.

Lipase activity. Pre-inoculum was carried out as above and washed cell pellet (0.5 OD per 50 mL) was inoculated into different media and incubated at 20 °C for 72 h. The media used were as follows: (a) Yeast Nitrogen Base containing Glucose (YNBG), (b) Yeast extract Peptone Glucose; (YPG), (c) Yeast extract Peptone Glucose Olive oil (YPGO), (d) Yeast extract Peptone Glucose Tributyrin (YPGTr) and (e) Yeast extract Peptone

Glucose Tween 80 (YPGTw). Stock solutions of fatty acid (10% olive oil and 10% tributyrin) were subjected to sonication three times for 1 min each on ice, autoclaved separately and added into liquid media buffered with 50 mM phosphate buffer, pH 6.8. The samples were removed after 72 h to determine lipase activity, soluble protein content and cell wet weight. The cells were harvested by centrifugation at 5000g for 10 min at 4 °C, the supernatant obtained was used as extracellular lipase source. Spectrophotometric method using *p*-nitrophenyl palmitate (*p*-NPP) as a substrate was used with slight modification for measurement of lipase activity^{12,61}. Protein concentration was estimated by the method of BCA with bovine serum albumin as a standard⁶². The media composition, lipase enzyme assay have been provided in supplementary material.

Lipase gene (*LIP2*) analysis. *Yarrowia lipolytica* (YALI) Lip2 protein sequence (YALI0A20350g) was used as query for TBLASTN on the 14 different *Yarrowia* species namely CBS 10151 (GCA_900518985.1), CBS 12934 (GCA_900519075.1), CBS 2071 (GCA_900519085.1), CBS 11013 (GCA_900519045.1), CBS 9722 (GCA_900519055.1), CBS 4855 (GCA_900519065.1), CBS 11062 (GCA_900519035.1), CLIB 89 (GCA_001761485.1), CLIB 122 (GCA_000002525.1), NCIM 3590 (GCA_003571375.1), CBS 10146 (GCA_900519015.1), CBS 10407 (GCA_900519005.1), CBS 10407 (GCA_900519025.1) and CBS 10253 (GCA_900518995.1). The best hit of NCIM 3590 from TBLASTN was used for orf generation and further used for reciprocal blast against YALI Lip2.

Data availability

The whole genome sequencing data can be accessed through BioProject Accession Number PRJNA328405. The respective BioSample Accession Numbers is SAMN05170375. The SRA reference numbers of the whole genome sequencing are SRX1850030 (Illumina NextSeq500 short paired-end reads). This Whole Genome sequencing data has been deposited at DDBJ/EMBL/GenBank under the Accession NKYT00000000. The version described in this paper is version NKYT01000000. The datasets generated and analysed during the current study for ITS and D1/D2 rDNA sequences are available at NCBI with Genbank Accession Nos. MK411246 and MK411222 respectively.

Received: 9 July 2020; Accepted: 3 February 2021

Published online: 09 March 2021

References

- Nagy, E. *et al.* *Yarrowia divulgata* f.a., sp. nov., a yeast species from animal-related and marine sources. *Int. J. Syst. Evol. Microbiol.* **63**, 4818–4823. <https://doi.org/10.1099/ijms.0.057208-0> (2013).
- Peter, G., Dlauchy, D., Vasdineyi, R., Lehoczki, J. T. & Deak, T. *Candida galli* sp. nov., a new yeast from poultry. *Antonie Van Leeuwenhoek* **86**, 105–110. <https://doi.org/10.1023/B:ANTO.0000036117.03847.bd> (2004).
- Nagy, E., Dlauchy, D., Medeiros, A., Peter, G. & Rosa, C. A. *Yarrowia porcina* sp. nov. and *Yarrowia bubula* f.a. sp. nov., two yeast species from meat and river sediment. *Antonie Van Leeuwenhoek* **105**, 697–707. <https://doi.org/10.1007/s10482-014-0125-4> (2014).
- Knutsen, A. K. *et al.* Polyphasic re-examination of *Yarrowia lipolytica* strains and the description of three novel *Candida* species: *Candida oslonensis* sp. nov., *Candida alimentaria* sp. nov. and *Candida hollandica* sp. nov.. *Int. J. Syst. Evol. Microbiol.* **57**, 2426–2435. <https://doi.org/10.1099/ijms.0.65200-0> (2007).
- Groenewald, M. & Smith, M. T. The teleomorph state of *Candida deformans* Langeron & Guerra and description of *Yarrowia yakushimensis* comb. nov. *Antonie Van Leeuwenhoek* **103**, 1023–1028. <https://doi.org/10.1007/s10482-013-9882-8> (2013).
- Gouliamova, D. E., Dimitrov, R. A., Gueorguiev, B. V., Smith, M. T. & Groenewald, M. *Yarrowia parophonii*. *Fungal Planet Descr. Sheets* **39**, 270–467. <https://doi.org/10.3767/persoonia.2017.39.11> (2017).
- Limtong, S., Youngmanitchai, W., Kawasaki, H. & Seki, T. *Candida phangngensis* sp. nov., an anamorphic yeast species in the *Yarrowia* clade, isolated from water in mangrove forests in Phang-Nga Province, Thailand. *Int. J. Syst. Evol. Microbiol.* **58**, 515–519. <https://doi.org/10.1099/ijms.0.65506-0> (2008).
- Chang, C. F., Chen, C. C., Lee, C. F. & Liu, S. M. Identifying and characterizing *Yarrowia keelungensis* sp. nov., an oil-degrading yeast isolated from the sea surface microlayer. *Antonie Van Leeuwenhoek* **104**, 1117–1123. <https://doi.org/10.1007/s10482-013-0033-z> (2013).
- Liu, K. F., Li, X. H. & Hui, F. L. *Yarrowia brassicae* f.a., sp. nov., a new yeast species from traditional Chinese sauerkraut. *Int. J. Syst. Evol. Microbiol.* **68**, 2024–2027. <https://doi.org/10.1099/ijsem.0.002783> (2018).
- Groenewald, M. *et al.* *Yarrowia lipolytica*: safety assessment of an oleaginous yeast with a great industrial potential. *Crit. Rev. Microbiol.* **40**, 187–206. <https://doi.org/10.3109/1040841X.2013.770386> (2014).
- Michely, S., Gaillardin, C., Nicaud, J. M. & Neugeglise, C. Comparative physiology of oleaginous species from the *Yarrowia* clade. *PLoS ONE* **8**, e63356. <https://doi.org/10.1371/journal.pone.0063356> (2013).
- Pignede, G. *et al.* Characterization of an extracellular lipase encoded by LIP2 in *Yarrowia lipolytica*. *J. Bacteriol.* **182**, 2802–2810. <https://doi.org/10.1128/jb.182.10.2802-2810.2000> (2000).
- Meunchan, M. *et al.* Comprehensive analysis of a yeast lipase family in the *Yarrowia* clade. *PLoS ONE* **10**, e0143096. <https://doi.org/10.1371/journal.pone.0143096> (2015).
- Kurtzman, C. P. *Yarrowia* van der Walt & von Arx., 1980. The Yeasts, a Taxonomic Study Chapter 82 (Vol. 2) 2011 Elsevier B.V.
- Butinar, L., Santos, S., Spencer-Martins, I., Oren, A. & Gunde-Cimerman, N. Yeast diversity in hypersaline habitats. *FEMS Microbiol. Lett.* **244**, 229–234. <https://doi.org/10.1016/j.femsle.2005.01.043> (2005).
- Garnier, L. *et al.* Diversity of spoilage fungi associated with French dairy products. *Int. J. Food Microbiol.* **241**, 191–197. <https://doi.org/10.1016/j.ijfoodmicro.2016.10.026> (2017).
- Deak, T., Chen, J. & Beuchat, L. R. Molecular characterization of *Yarrowia lipolytica* and *Candida zeylanoides* isolated from poultry. *Appl. Environ. Microbiol.* **66**, 4340–4344. <https://doi.org/10.1128/aem.66.10.4340-4344.2000> (2000).
- Gardini, F. *et al.* A survey of yeasts in traditional sausages of southern Italy. *FEMS Yeast Res.* **1**, 161–167. <https://doi.org/10.1111/j.1567-1364.2001.tb00027.x> (2001).
- Bankar, A. V., Kumar, A. R. & Zinjarde, S. S. Environmental and industrial applications of *Yarrowia lipolytica*. *Appl. Microbiol. Biotechnol.* **84**, 847–865. <https://doi.org/10.1007/s00253-009-2156-8> (2009).
- Coelho, M. A. Z., Amaral, P. F. F. & Belo, I. *Yarrowia lipolytica* an industrial workhorse. *Curr. Res. Technol. Educ. Top. Appl. Microbiol. Microb. Biotechnol.* 930–944 (2010).
- Harzevili, F. D. *Biotechnological Applications of the Yeast Yarrowia lipolytica*. *Springer Briefs in Microbiology*, Heidelberg. <https://doi.org/10.1007/978-3-319-06437-6> (2010).

22. Coelho, M. A. Z., Amaral, P. F. F. & Belo, I. *Yarrowia lipolytica* an industrial workhorse. *Curr. Res. Technol. Educ. Top. Appl. Microbiol. Microb. Biotechnol.* **2**, 930–944 (2010).
23. Morales-Vargas, A. T., Domínguez, A. & Ruiz-Herrera, J. Identification of dimorphism-involved genes of *Yarrowia lipolytica* by means of microarray analysis. *Res. Microbiol.* **163**, 378–387. <https://doi.org/10.1016/j.resmic.2012.03.002> (2012).
24. Bankar, A., Zinjarde, S., Shinde, M., Gopalghare, G. & Ravikumar, A. Heavy metal tolerance in marine strains of *Yarrowia lipolytica*. *Extremophiles* **22**, 617–628. <https://doi.org/10.1007/s00792-018-1022-y> (2018).
25. Ross, S. S. & Morris, E. O. An investigation of the yeast flora of marine fish from Scottish coastal waters and a fishing ground off Iceland. *J. Appl. Bact.* **28**, 224–234. <https://doi.org/10.1111/j.1365-2672.1965.tb02147.x> (1965).
26. Katre, G., Joshi, C., Khot, M., Zinjarde, S. & Ravikumar, A. Evaluation of single cell oil (SCO) from a tropical marine yeast *Yarrowia lipolytica* NCIM 3589 as a potential feedstock for biodiesel. *AMB Express* **2**, 36. <https://doi.org/10.1186/2191-0855-2-36> (2012).
27. Vatsal, A., Zinjarde, S. S. & Ravikumar, A. Phenol is the initial product formed during growth and degradation of bromobenzene by tropical marine yeast, *Yarrowia lipolytica* NCIM 3589 via an early dehalogenation step. *Front. Microbiol.* **8**, 1165. <https://doi.org/10.3389/fmicb.2017.01165> (2017).
28. Bankar, A. V., Kumar, A. R. & Zinjarde, S. S. Removal of chromium(VI) ions from aqueous solution by adsorption onto two marine isolates of *Yarrowia lipolytica*. *J. Hazard. Mater.* **170**, 487–494. <https://doi.org/10.1016/j.jhazmat.2009.04.070> (2009).
29. Apte, M. *et al.* Psychrotrophic yeast *Yarrowia lipolytica* NCYC 789 mediates the synthesis of antimicrobial silver nanoparticles via cell-associated melanin. *AMB Express* **3**, 32. <https://doi.org/10.1186/2191-0855-3-32> (2013).
30. Pitt, J. I. & Hocking, A. D. *Fungi and Food Spoilage* (Springer, New York, 2009). <https://doi.org/10.1007/978-0-387-92207-2>.
31. Schoch, C. L. *et al.* Fungal Barcoding Consortium. Nuclear ribosomal internal transcribed spacer (ITS) region as a universal DNA barcode marker for Fungi. *PNAS* **109**, 6241–6246. <https://doi.org/10.1073/pnas.111701810> (2012).
32. Kurtzman, C. P. Use of gene sequence analyses and genome comparisons for yeast systematics. *Int. J. Syst. Evol. Microbiol.* **64**, 325–332. <https://doi.org/10.1099/ijs.0.054197-0> (2014).
33. Bankevich, A. *et al.* SPAdes: a new genome assembly algorithm and its applications to single-cell sequencing. *J. Comput. Biol.* **19**, 455–477. <https://doi.org/10.1089/cmb.2012.0021> (2012).
34. Boetzer, M., Henkel, C. V., Jansen, H. J., Butler, D. & Pirovano, W. Scaffolding pre-assembled contigs using SSPACE. *Bioinformatics* **27**, 578–579. <https://doi.org/10.1093/bioinformatics/btq683> (2011).
35. Luo, R. *et al.* SOAPdenovo2: an empirically improved memory-efficient short-read de novo assembler. *Gigascience*. **27**(1), 18. <https://doi.org/10.1186/2047-217X-1-18> (2012).
36. Dujon, B. *et al.* Genome evolution in yeasts. *Nature* **430**, 35–44. <https://doi.org/10.1038/nature02579> (2004).
37. Cervenak, F. *et al.* Identification of telomerase RNAs in species of the *Yarrowia* clade provides insights into the co-evolution of telomerase, telomeric repeats and telomere-binding proteins. *Sci. Rep.* **9**, 13365. <https://doi.org/10.1038/s41598-019-49628-6> (2019).
38. Magnan, C. *et al.* Sequence assembly of *Yarrowia lipolytica* strain W29/CLIB89 shows transposable element diversity. *PLoS ONE* **11**(9), e0162363. <https://doi.org/10.1371/journal.pone.0162363> (2016).
39. Meier-Kolthoff, J. P., Auch, A. F., Klenk, H. P. & Goker, M. Genome sequence-based species delimitation with confidence intervals and improved distance functions. *BMC Bioinf.* **14**, 60. <https://doi.org/10.1186/1471-2105-14-60> (2013).
40. Auch, A. F., Klenk, H. P. & Goker, M. Standard operating procedure for calculating genome-to-genome distances based on high-scoring segment pairs. *Stand. Genom. Sci.* **2**, 142–148. <https://doi.org/10.4056/sigs.541628> (2010).
41. Auch, A. F., vonJan, M., Klenk, H. P. & Goker, M. Digital DNA–DNA hybridization for microbial species delineation by means of genome-to-genome sequence comparison. *Stand. Genom. Sci.* **2**, 117–134. <https://doi.org/10.4056/sigs.531120> (2010).
42. Tindall, B. J., Rossello-Mora, R., Busse, H. J., Ludwig, W. & Kämpfer, P. Notes on the characterization of prokaryote strains for taxonomic purposes. *Int. J. Syst. Evol. Microbiol.* **60**, 249–266. <https://doi.org/10.1099/ijs.0.016949-0> (2010).
43. Meier-Kolthoff, J. P., Klenk, H. P. & Goker, M. Taxonomic use of DNA G + C content and DNA–DNA hybridization in the genomic age. *Int. J. Syst. Evol. Microbiol.* **64**, 352–356. <https://doi.org/10.1099/ijs.0.056994-0> (2014).
44. Mefteh, F. B. *et al.* Fungal root microbiome from healthy and brittle leaf diseased date palm trees (*Phoenix dactylifera* L.) reveals a hidden untapped arsenal of antibacterial and broad spectrum antifungal secondary metabolites. *Front. Microbiol.* **8**, 307. <https://doi.org/10.3389/fmicb.2017.00307> (2017).
45. Chatterjee, S. *et al.* Draft genome of a commonly misdiagnosed multidrug resistant pathogen *Candida auris*. *BMC Genom.* **16**, 686. <https://doi.org/10.1186/s12864-015-1863-z> (2015).
46. Lee, L., Kim, Y. O., Park, S. C. & Chun, J. OrthoANI: an improved algorithm and software for calculating average nucleotide identity. *Int. J. Syst. Evol. Microbiol.* **66**, 1100–1103. <https://doi.org/10.1099/ijssem.0.000760> (2016).
47. Goris, J. *et al.* DNA–DNA hybridization values and their relationship to whole-genome sequence similarities. *Int. J. Syst. Evol. Microbiol.* **57**, 81–91. <https://doi.org/10.1099/ijs.0.64483-0> (2007).
48. Richter, M., Mora, R. R., Glockner, F. O. & Peplies, J. J. SpeciesWS: a web server for prokaryotic species circumscription based on pairwise genome comparison. *Bioinformatics* **32**, 929–931. <https://doi.org/10.1093/bioinformatics/btv681> (2015).
49. Kurtzman, C. P., Fell, J. W., Boekhout, T. & Robert, V. Methods for isolation, phenotypic characterization and maintenance of yeasts. In *The Yeasts A Taxonomy Study* (eds Kurtzman, C. P. *et al.*) 87–110 (Elsevier, Amsterdam, 2011). <https://doi.org/10.1016/B978-0-444-52149-1.00007-0>.
50. Carreira, A., Ferreira, L. M. & Loureiro, V. Production of brown tyrosine pigments by the yeast *Yarrowia lipolytica*. *J. Appl. Microbiol.* **90**, 372–379. <https://doi.org/10.1046/j.1365-2672.2001.01256.x> (2001).
51. Nicaud, J. M. *Yarrowia lipolytica*. *Yeast* **29**, 409–418. <https://doi.org/10.1002/yea.2921> (2012).
52. Lopes, M. *et al.* The use of olive mill wastewater by wild type *Yarrowia lipolytica* strains: medium supplementation and surfactant presence effect. *J. Chem. Technol. Biotechnol.* **84**, 533–537. <https://doi.org/10.1002/jctb.2075> (2008).
53. Deive, F. J., Sanroman, M. A. & Longo, M. A. A comprehensive study of lipase production by *Yarrowia lipolytica* CECT 1240 (ATCC 18942): from shake flask to continuous bioreactor. *J. Chem. Technol. Biotechnol.* **85**, 258–266. <https://doi.org/10.1002/jctb.2301.2301> (2009).
54. De Felice, B., Pontecorvo, G. & Carfagna, M. Degradation of waste waters from olive oil mills by *Yarrowia lipolytica* ATCC 20255 and *Pseudomonas putida*. *Acta Biotechnol.* **17**, 231–239. <https://doi.org/10.1002/abio.370170306> (1997).
55. White, T. J., Bruns, T. & Taylor, J. Amplification and direct sequencing of fungal ribosomal RNA Genes for phylogenetics. in *PCR Protocols* 315–322 (Academic Press, New York, 1990).
56. Kurtzman, C. P. & Robnett, C. J. Identification and phylogeny of ascomycetous yeasts from analysis of nuclear large subunit (26S) ribosomal DNA partial sequences. *Antonie Van Leeuwenhoek* **73**, 331–371. <https://doi.org/10.1023/a:1001761008817> (1998).
57. Thompson, J. D., Higgins, D. G. & Gibson, T. J. CLUSTAL W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice. *Nucleic Acids Res.* **22**, 4673–4680. <https://doi.org/10.1093/nar/22.22.4673> (1994).
58. Kumar, S., Stecher, G., Li, M., Knyaz, C. & Tamura, K. MEGA X: molecular evolutionary genetics analysis across computing platforms. *Mol. Biol. Evol.* **35**, 1547–1549. <https://doi.org/10.1093/molbev/msy096> (2018).
59. Vatsal, A., Zinjarde, S. S. & Ravikumar, A. Growth of a tropical marine yeast *Yarrowia lipolytica* NCIM 3589 on bromoalkanes: relevance of cell size and cell surface properties. *Yeast* **28**, 721–732. <https://doi.org/10.1002/yea.1901> (2011).
60. Murtey, M. D. & Ramasamy, P. Sample preparation for scanning electron microscopy—lifesciences. *Intech Open* **2**, 161–185. <https://doi.org/10.5772/61720> (2016).

61. Yadav, K. N. S. *et al.* Differential induction, purification and characterization of cold active lipase from *Yarrowia lipolytica* NCIM 3639. *Bioresour. Technol.* **102**, 10663–106670. <https://doi.org/10.1016/j.biortech.2011.09.013> (2011).
62. Stoscheck, C. M. Quantitation of protein. *Methods Enzymol.* **182**, 50–68. [https://doi.org/10.1016/0076-6879\(90\)82008-p](https://doi.org/10.1016/0076-6879(90)82008-p) (1990).

Acknowledgements

The authors gratefully acknowledge Dr. Sarah Cherian and Mr. Atul M. Walimbe, Bioinformatics group ICMR—National Institute of Virology, Pune, for guidance and important suggestions regarding phylogenetic analysis and Genotypic technologies, Bangalore for the technical support. Figures 1 and 2 were generated using MEGA X (www.megasoftware.net) and the tool OAT <https://www.ezbiocloud.net/tools/orthoani> by Prashant Gaikwad and Akshay Mandlecha. Other data bases used in the study. For GGDC 2.0—<https://ggdc.dsmz.de/ggdc>. JspeciesWS—<http://jspecies.ribohost.com/jspeciesws/>. Soapdenovo 2.04—<https://github.com/aquaskyline/SOAPdenovo2>. Spades v3.1—<https://cab.spbu.ru/files/release3.1.0/manual.html#sec2.1>. SSPACE—https://github.com/nsoranzo/sspace_basic. Genbank—<http://www.ncbi.nlm.nih.gov/>.

Author contributions

P.G. and S.J. contributed equally by carrying out the experiments, analysis of data and drafting the manuscript and A.M. contributed in carrying out SEM, temperature and salt tolerance experiments. A.R.K. conceived the idea, formulated the experiments, analysed the data and drafted the manuscript. All authors were involved in revising the manuscript.

Funding

This work was financially supported by University Grant's Commission (UGC) under the University of Potential for Excellence (UPE II) program (UGC 262(A)(1)) and Department of Science and Technology; under DST-PURSE programme (GOI-A-570). SJ thanks DST-PURSE for financial support.

Competing interests

The authors declare no competing interests.

Additional information

Supplementary Information The online version contains supplementary material available at <https://doi.org/10.1038/s41598-021-83914-6>.

Correspondence and requests for materials should be addressed to A.R.

Reprints and permissions information is available at www.nature.com/reprints.

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.



Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

© The Author(s) 2021