Adaptive evolution of engineered yeast for squalene production improvement and its genome-wide analysis

Paramasivan, Kalaivani; Aneesha, A.; Gupta, Nabarupa; Mutturi, Sarma

Published in:
Yeast

Link to article, DOI:
10.1002/yea.3559

Publication date:
2021

Document Version
Peer reviewed version

Citation (APA):
Adaptive evolution of engineered yeast for squalene production improvement and its genome-wide analysis

Kalaivani Paramasivan1,2,3 | Aneesha A1,2 | Nabarupa Gupta1 | Sarma Mutturi1,2

1Microbiology and Fermentation Technology Department, CSIR-Central Food Technological Research Institute, Mysuru, India
2AcSIR-Academy of Scientific & Innovative Research, Ghaziabad, Uttar Pradesh, 201002, India
3The Novo Nordisk Foundation Center for Biosustainability, Technical University of Denmark, Lyngby, Denmark

Correspondence
Sarma Mutturi, Microbiology and Fermentation Technology Department, CSIR-Central Food Technological Research Institute, Mysuru, India.
Email: sarma.mutturi@gmail.com; smutturi@cftri.res.in

Funding information
University of Graz, Grant/Award Number: KLN1; Department of Science and Technology, Government of India; Department of Biotechnology; (SERB); Science and Engineering Research Board, Grant/Award Number: CRG/2019/003323

Abstract
In the present study, the adaptive evolution of a metabolically engineered Saccharomyces cerevisiae strain in the presence of an enzyme inhibitor terbinafine for enhanced squalene accumulation via serial transfer leads to the development of robust strains. After adaptation for nearly 1500 h, a strain with higher squalene production efficiency was identified at a specific growth rate of 0.28 h⁻¹ with a final squalene titer of 193 mg/L, which is 16.5-fold higher than the BY4741 and 3-fold higher over the metabolically engineered SK22 strain. Whole-genome sequencing comparison between the reference strain and the evolved variant SK23 has led to the identification of 462 single-nucleotide variants (SNVs) between both strains, with 102 SNVs affecting metabolism-related genes. It was also established that F420I mutation of ERG1 in S. cerevisiae improves squalene synthesis. Further, the effect of increased squalene on lipid droplet and neutral lipid pattern in the evolved mutant strains was investigated by fluorescent techniques proving that the neutral lipid content and clustering of lipid droplets increase with an increase in squalene.

Take Away
- Terbinafine-based adaptive evolution improves squalene in Saccharomyces cerevisiae
- Whole-genome sequence of adapted strain revealed several mutations
- ERG1F420I has been established as causal mutation for squalene improvement
- Increase in squalene accumulation is associated with lipid droplet increase

KEYWORDS
adaptive evolution, lipid droplet, Nile red, Saccharomyces cerevisiae, squalene, whole-genome sequencing

1 | INTRODUCTION

Saccharomyces cerevisiae is one of the significant industrial platform strains for biochemical production and is one of the most sought-after organisms for terpene production (Paramasivan & Mutturi, 2017a). It is emerging as an alternative platform for product synthesis to be used in food, feed, pharmaceutical, and cosmetic industries. Terpenes constitute a diversified group of compounds and are usually harnessed from natural sources. Squalene belongs to the class of naturally occurring C30 terpenoid compounds with potential applications as an anti-cancer agent, vaccine adjuvant, dietary supplement due to its cardiovascular protective property, and moisturizer owing to its antioxidant activity (Kohno et al., 1995; Rao et al., 1998; Spanova & Daum, 2011). Squalene has also exhibited antiviral property against the hepatitis C virus (Saito et al., 2015). Shark liver oil is the most abundant source of squalene, but due to the endangerment of the shark species, biotechnological production in microbial hosts seems to be a promising alternative source for sustainable squalene production.
Adaptive evolution involves exploiting the organism’s inherent ability to undergo mutations towards the desired phenotype using a selective pressure (Conrad et al., 2014). Evolutionary engineering involves the process of prolonged growth with a selective advantage towards the strain that expresses the desirable trait (Çakar et al., 2005; Sonderegger & Sauer, 2003). With the significant reduction in costs and ease in genome sequencing technologies, whole-genome sequencing is performed on evolved strains to characterize the genetic basis underlying the phenotypic change. The divergence of the evolved strain with an altered genotype from the ancestor strain provides an insight into the functionality of the genes and evolutionary mechanisms. Adaptive laboratory evolution is a critical strategy in improving squalene content and can be achieved by evolutionary engineering for bio-sustainable molecule synthesis and rational strain engineering (Conrad et al., 2014). Reducing the downstream enzyme activity of squalene, squalene epoxide, is a crucial strategy in improving squalene content and can be achieved by a specific non-competitive enzyme inhibitor, terbinafine (Garaiová et al., 2014; Ta et al., 2012). Terbinafine inhibits squalene epoxide activity coded by ERG1, which converts squalene to squalene epoxide in the presence of oxygen molecules (Ryder, 1992). In the previous study, a squalene-producing strain SK22 has been generated by over-expressing tHMG1 and POS5 via integration while deleting ADK1 involved in nucleotide metabolism and LYS1 involved in lysine metabolism. tHMG1 coding for HMG-CoA reductase involved in sterol biosynthesis and POS5 coding for NADH kinase involved in NADPH regeneration are intuitive targets, whereas ADK1 and LYS1 are identified by genome-scale modeling (Paramasivan et al., 2019). Jhonne (2011) also observed in silico-based targets such as LYS1, LYS9, LYS12, ADK1, and ZWF1 for improving dihydrolaetosaminic acid (diterpene) in S. cerevisiae (Jhonne, 2011). In the present study, terbinafine is used as a selective pressure to gradually evolve an engineered S. cerevisiae strain for further enhancement of squalene flux. The study aims to evolve the strain for squalene increase adaptively and identify the genetic changes in the evolved high-squalene-accumulating strain followed by qualitative and quantitative analysis of lipid storage in yeast.

Increased squalene gets accumulated in yeast cells as lipid droplets along with triacylglycerols (TAGs) and sterol esters (SEs) (Spanova et al., 2010). Lipid droplet is a lipid storage organelle composed of triacylglycerols, sterol esters, and squalene present in all eukaryotic cells. Neutral lipids in yeast are comprised of triacylglycerols, sterol esters, free sterols, and free fatty acids such as palmitic acid, oleic acid, linoleic acid. Whether an increase in squalene improves lipid droplet or not is a matter of debate in the past decades. Squalene improvement is found to be associated with aggregation of lipid droplets (Garaiová et al., 2014; Spanova et al., 2012) whereas a few other studies state that lipid droplets remain unaffected in yeast by an increase in squalene (Spanova et al., 2010; Wei et al., 2018). Hence, the lipid droplets were examined for squalene accumulating evolved S. cerevisiae. The results suggest that the formation of lipid droplets and the neutral lipid content increases with increased squalene accumulation.

2 MATERIALS AND METHODS

2.1 Chemicals

Terbinafine hydrochloride, standards (ergosterol and squalene), and Nile red were obtained from Sigma-Aldrich (Bangalore, India). Growth media components (yeast extract, peptone, agar) were from HiMedia (India). Organic solvents (HPLC grade) were procured from Merck (Germany). Other standard chemicals of the highest purity available were acquired from various suppliers.

2.2 Adaptive evolution of engineered strain

Adaptive evolution was a long-term serial transfer using terbinafine as a stress inducer. SK22 (BY4741; MATa; ura3Δ0; leu2Δ0; his3Δ1; met15Δ0; YIR034c::kanMX4;YDR226w::KILEU2;X2::loxp-KIURA3-loxpxPETF1-tHMG1-PPGK1-POS5-[pCFB-TPHP], derivative of SK21) was used as a background strain for the construction of evolved SKE strains designated as E1–E60. The cell populations were evolved in 20 ml YPD liquid medium (1% yeast extract, 2% peptone, and 2% glucose) in 100 ml Erlenmeyer flasks at 30°C with gradually increasing terbinafine concentration from 0.01–300 μg/L during 60 subcultures. Strains were serially transferred every 12 h to a fresh medium for 30 days (~500 generations), at which a stable growth rate of 0.21–0.25 was achieved. The cultures from the evolved population were harvested in the stationary phase of growth and stored at −80°C after the addition of glycerol (50% v/v) for subsequent analysis. From different stages of adaptive evolution, colonies were isolated on solid YPD medium and labeled as E1–E60. The strains E1–E60 grown in the absence of terbinafine were further screened for squalene production and neutral lipid analysis. The evolved population was characterized by measuring biomass concentration (A500) and calculating the specific growth rate (μ). Absorbance to cellular dry weight correlation was determined by spinning down known A500 cells followed by drying at 65°C to a constant weight and fitting a linear curve. The doubling time was calculated from the growth rate: td = ln (2)/μ. Serial transfers were done during the mid-log phase of growth at (A500 = 0.5) using an adjusted inoculum volume. Time of inoculation and A500 of the culture were recorded, and colony morphology or signs of foreign contamination were observed during the experiments.

2.3 Cultivation for squalene production

The standard SD (synthetically defined) medium composition was followed as described in Amberg et al. (2005). The SD medium
contained 7.5 g of (NH₄)₂SO₄, 3.5 g of KH₂PO₄, 0.74 g of MgSO₄7H₂O, 1.8 ml of trace elements, and 1 ml of vitamins per liter. The trace metal solution consisted of the following (per liter): EDTA (sodium salt), 6.0 g; CaCl₂ 7H₂O, 1.8 g; ZnSO₄ 7H₂O, 1.8 g; CoCl₂ 2H₂O, 0.12 g; CuSO₄ 5H₂O, 0.12 g; Na₂MoO₄ 7H₂O, 0.16 g; CaCl₂ 7H₂O, 1.8 g; FeSO₄ 7H₂O, 1.2 g; H₂BO₃, 0.4 g and KI, 0.04 g. The vitamin solution contained (per liter): biotin, 0.05 g; p-aminobenzoic acid, 0.2 g; nicotinic acid, 1 g; Ca-pantothenate, 1 g; pyridoxine–HCl, 1 g; thiamine–HCl, 1 g and Myo-inositol, 25 g. thawed aliquots of the frozen stock cultures were plated in the YPD medium and incubated overnight, and it served as inoculum for fresh shake-flask cultivation on the SD medium, after another 12 h, was used to inoculate production culture flasks. Cell growth of the culture was monitored as A₆₀₀−C under vacuum for 4 h. The lyophilized cells were dispersed in chloroform/methanol (2:1, v/v) mixture followed by sonication. Subsequently, the extract was filtered and subjected to evaporation at 50°C. The samples dispersed in chloroform were subjected to centrifugation for removal of cell debris prior to HPLC analysis. HPLC was performed using a semi-preparative reversed-phase C-18 column (Phenomenex Kinetex, Hyderabad, India) (particle size 5 μm, 200 x 4.6 mm i.d.) maintained at 35°C. Detection and quantification of squalene were carried out at 195 nm using an UV-vis diode array multiple wavelength detector. The flow rate was 1.5 ml/min with 100% acetonitrile as the mobile phase.

### 2.4 Squalene extraction and analysis

The extraction and analysis of squalene is according to Paramasivan et al. (2018). In brief, cells were harvested by centrifugation to obtain cell pellets, and the cell pellets were then subjected to lyophilization at −50°C under vacuum for 4 h. The lyophilized cells were dispersed in chloroform/methanol (2:1, v/v) mixture followed by sonication. Subsequently, the extract was filtered and subjected to evaporation at 50°C. The samples dispersed in chloroform were subjected to centrifugation for removal of cell debris prior to HPLC analysis. HPLC was performed using a semi-preparative reversed-phase C-18 column (Phenomenex Kinetex, Hyderabad, India) (particle size 5 μm, 200 x 4.6 mm i.d.) maintained at 35°C. Detection and quantification of squalene were carried out at 195 nm using an UV–VIS spectrophotometer (UV-1601, Shimadzu, Japan). Specific growth rates were calculated during the exponential growth phase of each culture. Samples were collected, centrifuged, and the obtained cell pellets were stored at −20°C for squalene analysis.

### 2.5 Genomic DNA extraction and sequencing

Genomic sequencing of the evolved strain E47 (later designated as SK23) was carried to predict the genetic changes in the evolved high-squalene-producing strain. SK23 was cultured in 10 ml YPD medium at 30°C, and the cells were harvested during the mid-exponential growth phase for DNA extraction. Genomic DNA was extracted using the standard glass bead method (Amberg et al., 2005). The cells were vortexed with glass beads followed by brief centrifugation to remove glass beads and cell debris. The DNA from the aqueous phase was extracted after the addition of an organic solvent. Pure DNA was isolated by ethanol precipitation, followed by library preparation. The paired-end sequencing libraries were prepared using the ligation Sequencing Kit 1D (SQK-LSK108) and Library Loading Bead Kit (EXP-LLB001) following the manufacturer’s instructions with an insert size of 1000–20,000 bp (Oxford nanopore technologies, UK). Fragmentation of DNA by using an adaptive focused acoustic technology (AFA; Covaris), leads to the generation of double-stranded DNA (dsDNA) fragments with 3’ or 5’ overhang. The fragments were then subjected to end repair followed by adapter ligation to generate 5’-phosphorylated, blunt-ended dsDNA fragments and DNA size selection using a bead-based method. According to the protocol guide, the purified libraries were quantified using Qubit Fluorometric Quantitation and Qubit™ 4 NGS Starter Kit (Thermo Fischer Scientific, India). The libraries were qualified using the high-sensitivity DNA chip (Agilent Technologies, Waldbronn, Germany). Indexed libraries were then sequenced using the MiniION Nanopore sequencing platform (Oxford Nanopore Technologies, UK).

### 2.6 Genomic sequence analysis

The bioinformatic analysis tool, Nano plot, was used for the sequence quality check. The raw sequence data were filtered and trimmed to remove adaptor sequences, ambiguous reads, and low-quality sequences using Porkchop Tool. These reads were trimmed to obtain high-quality clean reads. The filtered reads were mapped to the reference genome using MarginAlign with Minimap2 to detect single-nucleotide variants, chromosomal and mitochondrial regional deletion. Single-nucleotide variations (SNV) in coding sequences representing non-synonymous mutations have been detected. The number of reads ranged from 55,000 to 60,000 resulting in a coverage of approximately 20x to 30x. Variant calling (single-nucleotide variants), INDEL detection and annotation of the detected variants were performed using Geneious pro. The genes harboring mutations were subjected to gene set enrichment analysis using YeastEnrichr tool (Chen et al., 2013; Kuleshov et al., 2016). Grantham score calculations for protein variants were carried out using ENVISION tool (Gray et al., 2018). Sequence data of strain SK23 was submitted at the NCBI Sequence Read Archive (https://www.ncbi.nlm.nih.gov/sra) under BioProject ID PRJNA590882.

### 2.7 Homology model of Erg1p and ligand interaction

Amino acid sequence of Erg1p (NP_011691.1) was used to predict the homology model for squalene monoxygenase. I-TASSER suite (https://zhanglab.ccmb.med.umich.edu/I-TASSER/) was used for determining the structure of this protein. Based on TM-align program of I-TASSER, the enzyme was observed to be closer to 6c6n (Human squalene epoxidase) in PDB database. A model with highest confidence score (TM-score of 0.73 ± 0.11 and RMSD of 7.0 ± 4.1 Å) was selected for structural analysis and docking studies. Initially the protein model was docked with its cofactor FAD, later the Erg1p-FAD complex was for docking terbinafine. Docking was carried using AutoDockTools (v.1.5.6) (http://www.scripps.edu/mb/olson/doc/autodock) and the complex was processed using Discovery Studio 2017.
2.8 Establishment of causal mutation ERG1\textsuperscript{F420I}

S. cerevisiae KLN1 strain (MAT\textalpha, ERG1::URA3 leu2 ura3 trp1) was kindly gifted by Prof. Friederike Turnowsky (University of Graz, Austria). This strain was used for expressing the ERG1 gene variant from SK23 which harbors F420I mutation. KLN1 was grown inYPD medium containing ergosterol and Tween 80 under oxygen-limiting conditions. Ergosterol and Tween 80 were dissolved in pure ethanol and steamed at 100°C for 10 min before being added to the medium to final concentrations of 10 and 420 mg 1\(^{-1}\), respectively. Genomic DNA was isolated from SK23 and BY4741 (control) using standard procedure. Using the isolated genomic DNA as template, ERG1 was amplified by PCR with Phusion DNA Polymerase (Thermo fisher) having (5’-GTCCAGTATTGAACAAATAC AGGTT-3’) as forward primer and (5’-TTGACGGTTCTATCCTCTCT-3’) as reverse primer. Later the amplicons were verified for F420I mutation using sequence analysis. The obtained amplicons were transformed into S. cerevisiae KLN1 using Gietz and Schiestl (2007) method. Successful integration of fragments by homologous recombination at the ERG1 locus was selected based on growth of the strain in aerobic conditions. Later the transformants were cultivated for squalene synthesis.

2.9 Nile red staining and fluorescence microscopy

Lipid droplets in evolved yeast strains were visualized by fluorescence microscopy after Nile red staining. The cells were grown in the absence of selection pressure, terbinafine, and harvested by centrifugation at 5000 rpm for 1 min. To remove the background fluorescence of the media components and extracellular metabolites, the cell pellets were washed with PBS and are resuspended to a density of 2 \times 10^7 cells/ml (\textit{A}_{600}-1.0). Two hundred microliters of the cells were resuspended in a 1.5 ml Eppendorf tube followed by 20 μl DMSO. Twenty microliters of 100 μg/ml Nile red in acetone was added to a final concentration of 10 μg/ml. The fluorescence signal was captured by channel (450–490 nm band-pass excitation filter, a 510 nm dichromatic mirror, and a 515 nm long-pass emission filter) to view lipid droplets. Ten microliters of yeast cell suspension pre-stained with Nile red was mounted on a microscopic slide and observed in the microscope supplemented with immersion objective. Micrographs were recorded using an IMT-2 inverted microscope fluorescence microscope (Olympus BX 51, Japan) under blue light excitation. Real-time representative fluorescent images were acquired with a digital Olympus camera and processed. The content and morphology of lipid droplets in control and squalene accumulating cells were evaluated by visual inspection of five independent images for each experimental condition.

2.10 Flow cytometry and spectrofluorometry

Cell suspension pre-stained with Nile red was gently mixed and incubated in the dark for 20 min at room temperature and measured immediately by flow cytometry. Flow cytometer was used to monitor the increase in neutral lipid content during the evolution process. Samples were acquired using Guava EasyCyte 6-2 L Benchtop Flow Cytometer (EMD Millipore Corporation, Billerica, MA, USA) and analyzed with a standard procedure using InCyte software. For each sample, a minimum of 5000 cells was acquired per sample on the Guava instrument. The Guava EasyCyte 6-2 L is equipped with two lasers whose excitation wavelengths are 488 and 640 nm. Nile red exhibits yellow-gold fluorescence when dissolved in neutral lipids, which are detected by the FL2 channel. Neutral lipid content was measured at an excitation wavelength of 480 nm and an emission wavelength of 535 nm, using a Shimadzu spectrofluorophotometer (model RF-5301PC, Shimadzu, Japan). The relative fluorescent intensity of Nile red in the samples was measured after reducing autofluorescence and the fluorescent intensity of Nile red in the buffer.

3 RESULTS AND DISCUSSION

3.1 Adaptive evolution for developing high-squalene-accumulating strain

An evolutionary engineering approach based on long-term adaptation in terbinafine to select strains for enhanced squalene titer was envisaged in the present study. Squalene epoxidation is the first and foremost downstream step of squalene to sterol conversion, and it has been targeted in the present study. Terbinafine inhibits the activity of squalene monooxygenase coded by the ERG1 gene. ERG1 gene has been targeted for mutational studies by several researchers towards squalene improvement (Garaiová et al., 2014; Hull et al., 2014). The strain SK22 derived from our previous study was evolved in the presence of a specific enzyme inhibitor terbinafine at a concentration ranging from low to high (0.01–300 μg/L) (Paramasivan et al., 2019). Laboratory evolution experiments with SK22 have yielded several evolved strain populations with the specific growth rate similar to control in the presence of terbinafine. Because the population is non-homogenous, cultures were streaked onto plates to isolate pure colonies from each evolution experiment and were subjected to squalene and lipid droplet analysis. The selected colony from each subculture was considered as a phenotypic and genotypic representative within the population. Squalene was analyzed at the end of every 10 generations up to 600 generations in the absence of terbinafine to study the squalene accumulation during the evolution experiment.

Evolutionary engineering was initiated with 20 g/L glucose as the carbon source in shake-flask cultures in the presence of terbinafine. The cultures were maintained in prolonged exponential growth by sub-culturing in fresh medium before reaching the stationary phase. The growth was monitored throughout the evolution by collecting culture aliquots periodically. The growth of treated cells remained unaffected, whereas the evolved strain has shown significantly higher squalene levels. There were no adverse effects found in the growth as the cells were acclimatized to evolution with a gradual increase in terbinafine concentration. SK22 variants with enhanced ability to
produce squalene have been identified. During initial transfers, the strain has shown a squalene titer of 20–30 mg/L (Figure 1). Terbinafine concentration during the evolution of SK22 has been increased to 1 μg/L and 300 μg/L at the end of 30th and 60th subcultures, respectively (Table S1). After ten transfers (0.4 μg/L terbinafine) (Table S1), 30 transfers (1 μg/L terbinafine) (Table S1) and 45 transfers (150 μg/L terbinafine) (Table S1), the strain showed a squalene titer of 20, 67, and 188 mg/L, respectively (Figure 1). High-level production of 193 mg/L has been achieved in E47 strain under shake-flask conditions, which is 3.5-fold higher over the reference strain (SK22) which was reported to synthesize a maximum titer of 55 mg/L under shake-flask conditions (Figure 1). The evolution of SK22 resulted in a high-squalene-accumulating strain, named SK23, with increased squalene yield. Starting with an initial squalene titer of 25 mg/L, a continuously increasing trend of squalene to 193 mg/L has been observed (Figure 1; Figure 2). SK23 strain significantly improved in biomass production, which could be one of the main reasons for the high squalene titer of SK23. SK23 in the presence of terbinafine at a concentration of 100 μg/L has given a maximum titer of 420 mg/L in shake-flask conditions (Figure 2).

WT and SK22 are the original starting strains which represent wild-type and engineered S. cerevisiae strains, respectively. WTE and SK23, respectively, WTE is the strain isolated by plating the cells after 600 generations at the end of 60th subculture. The squalene levels in both WT and SK22 were doubled in the presence of terbinafine. A higher fold increase of squalene has been observed in the evolved strain, whereas the ergosterol levels are relatively constant in the range of 7.5–8 mg/g dry cell weight (DCW) which implies that the cells have undergone a selection process. The higher fold increase in squalene is consistent with the results obtained by Garaiová et al. (2014), where there was a 30-fold increase in squalene in the presence of terbinafine. A 35-fold increase in squalene by an engineered strain over the control strain was achieved in a study by Rasool et al. (2016) in terbinafine. Although anaerobic conditions are more suitable for squalene synthesis, the strains were evolved in aerobic conditions as anaerobiosis can lead to impaired sterol synthesis in the absence of external supplementation of sterols/tween-80 (Verduyn et al., 1992). To further understand the physiological changes due to adaptive evolution and analyze the mutations incorporated, the strain SK23 was subjected to whole-genome sequencing.

### 3.2 | Whole-genome sequence analysis

All the strains obtained after 300 generations produced high squalene in comparison to reference strain SK22. However, one of the most promising hyper producer SK23 has been subjected to whole-genome sequencing. In this study, the strain sequence of S288c, parent strain of BY4741 was compared with that of SK23. Summary statistics of whole-genome sequencing data and the methods used in the analysis are described in Table 1. The mutations were found to be in functional genes, regulatory genes as well as in intergenic regions. A total of 103 INDELs (insertions and deletions) and 462-point mutations among which 102 metabolism specific SNVs were identified in the evolved strain relative to the reference strain. The 102 metabolism specific SNVs were tabulated (Table S2). Among the mutations, the genes found in in silico studies (unpublished results) and based on
published reports were analyzed further (Table 2) (Choi et al., 2010). The genes were mostly localized in mitochondria followed by lipid droplets and microbodies when subjected to bioinformatic functional analysis of pathways (Figure S1a). The mutation harboring genes were mainly involved in molecular functions such as prenyl transferase activity followed by carboxylic acid transmembrane transporter activity and hydroxymethyl formyl and related transferase activity (Figure S1b). The genes were primarily involved in biological processes involved in fermentative metabolism and stress resistance upon pheromone analysis (Figure S1c). The genes harboring mutations were found to be largely involved in biological processes such as squalene enhancement and is found to harbor a missense mutation at position 420, where the phenylalanine (F) is converted to the squalene epoxidase and is found to harbor a missense mutation. L251F, F402L, F420L, P430S, or F433S substitutions were observed in Erg1p by Leber et al. (2003) during generation of terbinafine-resistant S. cerevisiae mutants. These mutations were also validated by replacement of wild-type ERG1 gene with mutated alleles to prove that single point mutations of Erg1p using L251F, F402L, F420L, P430S, or F433S substitutions are sufficient for S. cerevisiae to confer high terbinafine resistance. In another study by Klobučníková et al. (2003), P430A substitution was identified as the partial cause of terbinafine resistance in S. cerevisiae. They have also demonstrated that squalene accumulation was higher along with reduced ergosterol synthesis in the mutant strain during early exponential phase when it is cultivated in the presence of 50 μg/ml of terbinafine. Rocha et al. (2006) have identified F389L substitution in ErgA (squalene epoxide gene) as the reason for terbinafine resistance in Aspergillus nidulans. Later this mutation was validated in Aspergillus fumigatus. F389L mutation of ErgA corresponds to F402L mutation of ERG1 in S. cerevisiae. Ruckenstuhl et al. (2007) have carried homology model of Erg1p based on crystal structure of PHBH from P. fluorescens. In their study, FAD interactions with Erg1p residues were observed to coincide with experimental results where terbinafine sensitive mutants were generated. Garaiová et al. (2014) used ERG1 as the target gene for site-directed mutagenesis to improve squalene. They found L37P and Q443UAG mutations to improve squalene accumulation and the cells were hypersensitive to terbinafine with reduced squalene epoxide activity.

In the present study, homology model of Erg1p was generated based on crystal structure of human squalene epoxidase as reported by Padyana et al. (2019). Docking was carried with both FAD and terbinafine. Initially, Erg1p-FAD complex was created as shown in Figure 3a. It was interesting to observe that the interaction between Erg1p-FAD complex (Figure 3b) were similar to those observed by Ruckenstuhl et al. (2007). Also, this complex was used as a template to dock terbinafine. The results from Figure 3c indicates that terbinafine binds to a domain away from FAD to Erg1p. Also, from Figure 3d it can be observed that F420 along with Y90, V249, L259, L343, T344, L394, L398, F402, P430, F433, and L434 residues interact with terbinafine. Based on the above studies, it can be concluded that the missense mutation, F420I could be seemingly a causative mutation for both terbinafine resistance and squalene accumulation. In order to confirm the same, studies were conducted as described in Section 2.8 and the results are provided in Table S3. It can be seen that KLN1 strain (erg1 deletion strain) harboring SK23’s ERG1 gene with F420I mutation has resulted in approximately 6.8-fold increase in squalene synthesis. This indicates F420I is one of the seemingly causative mutation in the evolved strain SK23 for squalene improvement.

<table>
<thead>
<tr>
<th>TABLE 1</th>
<th>Summary statistics of whole-genome sequencing data and the methods used in the analysis</th>
</tr>
</thead>
<tbody>
<tr>
<td>Parameter</td>
<td>Value/Method</td>
</tr>
<tr>
<td>Average read length</td>
<td>3880</td>
</tr>
<tr>
<td>Total Reads</td>
<td>59,993</td>
</tr>
<tr>
<td>Mapped read %</td>
<td>99.19%</td>
</tr>
<tr>
<td>Sequenced bases</td>
<td>232,811,987</td>
</tr>
<tr>
<td>Coverage</td>
<td>30x</td>
</tr>
<tr>
<td>GC %</td>
<td>38</td>
</tr>
<tr>
<td>Assembly</td>
<td>Reference Genome Assembly</td>
</tr>
<tr>
<td>Assembler tool</td>
<td>MarginAlign with Minimap2</td>
</tr>
<tr>
<td>Reference genome</td>
<td>Saccharomyces cerevisiae S288C</td>
</tr>
<tr>
<td>Annotation tool</td>
<td>Geneious pro</td>
</tr>
<tr>
<td>Sequencing method</td>
<td>Nanopore sequencing platform</td>
</tr>
</tbody>
</table>

- **ERG1** gene has been inhibited by the addition of terbinafine leading to the squalene enhancement and is found to harbor a missense mutation at position 420, where the phenylalanine (F) is converted to isoleucine (I) (Table 2). The bioinformatics functional analysis has been carried out for the altered protein sequence. F420I has a low grantham score of 21 which indicates that the fundamental property of the protein is not changed upon substitution as grantham score represents the chemical dissimilarity of the protein. F420I (21), F420L (22), and F420T (22) are the three amino acid substitutions with the lowest grantham score. F420I has also been predicted to retain the hydrophobicity of the residue. As the genome-sequencing results revealed the presence of missense mutation (F420I) in Erg1p, we were interested to understand the role of this residue for imparting terbinafine resistance in S. cerevisiae and squalene accumulation. L251F, F402L, F420L, P430S, and F433S substitutions were observed in Erg1p by Leber et al. (2003) during generation of terbinafine-resistant S. cerevisiae mutants. These mutations were also validated by replacement of wild-type ERG1 gene with mutated alleles to prove that single point mutations of Erg1p using L251F, F402L, F420L, P430S, or F433S substitutions are sufficient for S. cerevisiae to confer high terbinafine resistance. In another study by Klobučníková et al. (2003), P430A substitution was identified as the partial cause of terbinafine resistance in S. cerevisiae. They have also demonstrated that squalene accumulation was higher along with reduced ergosterol synthesis in the mutant strain during early exponential phase when it is cultivated in the presence of 50 μg/ml of terbinafine. Rocha et al. (2006) have identified F389L substitution in ErgA (squalene epoxide gene) as the reason for terbinafine resistance in Aspergillus nidulans. Later this mutation was validated in Aspergillus fumigatus. F389L mutation of ErgA corresponds to F402L mutation of ERG1 in S. cerevisiae. Ruckenstuhl et al. (2007) have carried homology model of Erg1p based on crystal structure of PHBH from P. fluorescens. In their study, FAD interactions with Erg1p residues were observed to coincide with experimental results where terbinafine sensitive mutants were generated. Garaiová et al. (2014) used ERG1 as the target gene for site-directed mutagenesis to improve squalene. They found L37P and Q443UAG mutations to improve squalene accumulation and the cells were hypersensitive to terbinafine with reduced squalene epoxide activity.

In the present study, homology model of Erg1p was generated based on crystal structure of human squalene epoxidase as reported by Padyana et al. (2019). Docking was carried with both FAD and terbinafine. Initially, Erg1p-FAD complex was created as shown in Figure 3a. It was interesting to observe that the interaction between Erg1p-FAD complex (Figure 3b) were similar to those observed by Ruckenstuhl et al. (2007). Also, this complex was used as a template to dock terbinafine. The results from Figure 3c indicates that terbinafine binds to a domain away from FAD to Erg1p. Also, from Figure 3d it can be observed that F420 along with Y90, V249, L259, L343, T344, L394, L398, F402, P430, F433, and L434 residues interact with terbinafine. Based on the above studies, it can be concluded that the missense mutation, F420I could be seemingly a causative mutation for both terbinafine resistance and squalene accumulation. In order to confirm the same, studies were conducted as described in Section 2.8 and the results are provided in Table S3. It can be seen that KLN1 strain (erg1 deletion strain) harboring SK23’s ERG1 gene with F420I mutation has resulted in approximately 6.8-fold increase in squalene synthesis. This indicates F420I is one of the seemingly causative mutation in the evolved strain SK23 for squalene improvement.

- **ERG3** coding for C-5 sterol desaturase, **ERG8** coding for phosphomevalonate kinase, and **ERG27** coding for 3-keto sterol reductase involved in sterol biosynthesis harbors synonymous mutations, whereas **ERG4**, **ERG7**, **ERG10**, **ERG24**, **ERG26**, and **ERG29** genes have missense mutations. **ERG5** and **ERG25** have nonsense mutations at positions 515 and 155, respectively (Table 2). It is interesting to observe that
<table>
<thead>
<tr>
<th>S. No</th>
<th>Gene (Essentiality\textsuperscript{a})(Uniprot ID\textsuperscript{b})</th>
<th>Description</th>
<th>Nucleotide change</th>
<th>Resultant change (Grantham score\textsuperscript{c})</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>ERG1 (E) (P32476)</td>
<td>Squalene epoxidase</td>
<td>T1258A</td>
<td>Missense mutation (Phe to Ile) (F420I) (21)</td>
</tr>
<tr>
<td>2</td>
<td>POS5 (NE) (Q6892)</td>
<td>NADH kinase</td>
<td>C539T</td>
<td>Missense mutation (Ser to Leu) (S178L) (NA)</td>
</tr>
<tr>
<td>3</td>
<td>ADH5 (NE) (P38113)</td>
<td>Alcohol dehydrogenase</td>
<td>T784C</td>
<td>Missense mutation (Tyr to His) (Y262H) (83)</td>
</tr>
<tr>
<td>4</td>
<td>ICL2 (NE) (Q12031)</td>
<td>Isocitrate Lyase</td>
<td>C126T</td>
<td>Synonymous mutation (Phe to Phe) (F420F) (0)</td>
</tr>
<tr>
<td>5</td>
<td>ASN1 (NE) (P49089)</td>
<td>Asparagine synthetase</td>
<td>G1714A</td>
<td>Missense mutation (Glu to Lys) (E572K) (NA)</td>
</tr>
<tr>
<td>6</td>
<td>ERG3 (NE) (P32353)</td>
<td>C-5 sterol desaturase</td>
<td>T54C; T1002C</td>
<td>Synonymous mutation (Ala to Ala) (A18A) (0); Synonymous mutation (Ala to Ala) (A334A) (0)</td>
</tr>
<tr>
<td>7</td>
<td>ERG4 (NE) (P25340)</td>
<td>C-24 sterol desaturase</td>
<td>G644A</td>
<td>Missense mutation (Gly to Leu) (G215L) (138)</td>
</tr>
<tr>
<td>8</td>
<td>ERG5 (NE) (P54781)</td>
<td>C-22 sterol desaturase</td>
<td>G1374A; A1543T</td>
<td>Missense mutation (Lys to Phe) (K458F) (102); Nonsense mutation (Thr to Stop codon) (T515*) (NA)</td>
</tr>
<tr>
<td>9</td>
<td>ERG7 (E) (P38604)</td>
<td>2,3-oxidosqualene-lanosterol cyclase</td>
<td>G28A</td>
<td>Missense mutation (Gly to Ser) (G10S) (56)</td>
</tr>
<tr>
<td>10</td>
<td>ERG8 (E) (P24521)</td>
<td>Phosphomevalonate kinase</td>
<td>A1329G; 1330-31 GA to AG</td>
<td>Missense mutation (Lys to Lys) (K443K) (0); Missense mutation (Asp to Ser) (D444S) (65)</td>
</tr>
<tr>
<td>11</td>
<td>ERG10 (E) (P41338)</td>
<td>Acetoacetyl CoA thiolase</td>
<td>T473C; G745A</td>
<td>Missense mutation (Leu to Pro) (L158P) (98); Missense mutation (Ala to Thr) (A249T) (58)</td>
</tr>
<tr>
<td>12</td>
<td>ERG24 (E) (P32462)</td>
<td>C-14 sterol reductase</td>
<td>T322C</td>
<td>Missense mutation (Tyr to His) (Y108H) (83)</td>
</tr>
<tr>
<td>13</td>
<td>ERG25 (E) (P53045)</td>
<td>C-4 sterol methyl oxidase</td>
<td>G448A; G464A</td>
<td>Missense mutation (Val to Ile) (V150I) (29); Nonsense mutation (Trp to Stop codon) (W155*) (NA)</td>
</tr>
<tr>
<td>14</td>
<td>ERG26 (E) (P53199)</td>
<td>C-3 sterol dehydrogenase</td>
<td>T638C</td>
<td>Missense mutation (Leu to Pro) (L213P) (NA)</td>
</tr>
<tr>
<td>15</td>
<td>ERG27 (E) (Q12452)</td>
<td>3-keto sterol reductase</td>
<td>A708G</td>
<td>Synonymous mutation (Thr to Thr) (T236T) (0)</td>
</tr>
<tr>
<td>16</td>
<td>ERG29 (E) (P40207)</td>
<td>Unknown protein</td>
<td>362–63 AA to GG; 373–74 GA to AG</td>
<td>Missense mutation (Glu to Gly) (E121G) (NA); Missense mutation (Glu to Arg) (E125R) (NA)</td>
</tr>
<tr>
<td>17</td>
<td>YEH2 (NE) (Q07950)</td>
<td>Yeast steryl ester hydrolase</td>
<td>C422T</td>
<td>Missense mutation (Ala to Val) (A141V) (NA)</td>
</tr>
<tr>
<td>18</td>
<td>AAT1 (NE) (Q18082)</td>
<td>Aspartate AminoTransferase</td>
<td>G1149A</td>
<td>Synonymous mutation (Gly to Gly) (G383G) (0)</td>
</tr>
<tr>
<td>19</td>
<td>PDA1 (NE) (P16387)</td>
<td>Pyruvate dehydrogenase</td>
<td>G92A; A506G</td>
<td>Missense mutation (Arg to His) (R31H) (29); Missense mutation (Lys to Arg) (K169R) (26)</td>
</tr>
<tr>
<td>20</td>
<td>ARO9 (NE) (P38840)</td>
<td>Aromatic aminotransferase II</td>
<td>G424A</td>
<td>Missense mutation (Gly to Arg) (G142R) (NA)</td>
</tr>
<tr>
<td>21</td>
<td>PYC1 (NE) (P11154)</td>
<td>Pyruvate carboxylase</td>
<td>A79G; 1763–64 CG to GC; C2339T; G2827A</td>
<td>Missense mutation (Arg to Gly) (R27G) (NA); Missense mutation (Arg to Ala) (R588A) (NA); Missense mutation (Ala to Val) (A780V) (NA); Missense mutation (Gly to Ser) (G943S) (NA)</td>
</tr>
</tbody>
</table>

(Continues)
<table>
<thead>
<tr>
<th>S. No</th>
<th>Gene (Essentiality) (Uniprot ID)</th>
<th>Description</th>
<th>Nucleotide change</th>
<th>Resultant change (Grantham score)</th>
</tr>
</thead>
<tbody>
<tr>
<td>22</td>
<td>ADE16 (NE) (P54113)</td>
<td>Inosine monophosphate cyclohydrolase</td>
<td>G106A; G783T; G798A; A1309G; G1332A</td>
<td>Missense mutation (Gly to ser) (G365) (56); Synonymous mutation (Ala to Ala) (A261A) (0); Synonymous mutation (Lys to Lys) (K226K) (0); Missense mutation (Asn to Asp) (N437D) (23); Synonymous mutation (Gly to Gly) (G444G) (0)</td>
</tr>
<tr>
<td>23</td>
<td>ASP1 (NE) (P38986)</td>
<td>L-asparaginase</td>
<td>C144T</td>
<td>Synonymous mutation (Cys to Cys) (C48C) (0)</td>
</tr>
<tr>
<td>24</td>
<td>HAP1 (NE) (P0CE41)</td>
<td>Heme activator protein</td>
<td>A1618G; 1762–63 GA to AG</td>
<td>Missense mutation (Thr to Ala) (T540A) (58); Missense mutation (Glu to Arg) (E588R) (54)</td>
</tr>
<tr>
<td>25</td>
<td>FAT1 (NE) (P38225)</td>
<td>Fatty acyl-CoA synthetase</td>
<td>G622C</td>
<td>Missense mutation (Ala to Pro) (A208P) (NA)</td>
</tr>
<tr>
<td>26</td>
<td>AAT2 (NE) (P23542)</td>
<td>Aspartate amino transferase</td>
<td>G396C; T397T</td>
<td>Missense mutation (Lys to Asn) (K132N) (94); Missense mutation (Pro to Ser) (P133S) (74)</td>
</tr>
<tr>
<td>27</td>
<td>ACO1 (NE) (P19414)</td>
<td>Aconitase</td>
<td>C201T; C511T</td>
<td>Missense mutation (Tyr to Thr) (Y67Y) (0); Nonsense mutation (Gln to Stop codon) (Q171) (NA)</td>
</tr>
<tr>
<td>28</td>
<td>HMG2 (NE) (P12684)</td>
<td>HMG-CoA reductase</td>
<td>557–558 TC to CT; T2790C</td>
<td>Missense mutation (Glu to Ala) (Q186A) (NA); Synonymous mutation (Glu to Asp) (D930D) (0)</td>
</tr>
<tr>
<td>29</td>
<td>NDI1 (NE) (P32340)</td>
<td>NADH dehydrogenase</td>
<td>C716T</td>
<td>Missense mutation (Thr to Ile) (T2391) (89)</td>
</tr>
<tr>
<td>30</td>
<td>PDR5 (NE) (P33302)</td>
<td>Pleiotropic Drug Resistance</td>
<td>C242T; C2493T</td>
<td>Missense mutation (Pro to Leu) (P811L) (98); Missense mutation (Pro to Pro) (P831P) (0)</td>
</tr>
<tr>
<td>31</td>
<td>PDR6 (NE) (P32767)</td>
<td></td>
<td>C528T; C2604T</td>
<td>Synonymous mutation (Tyr to Tyr) (Y176Y) (0); Synonymous mutation (Cys to Cys) (C668C) (0)</td>
</tr>
<tr>
<td>32</td>
<td>PDR11 (NE) (P40550)</td>
<td></td>
<td>G1185A</td>
<td>Synonymous mutation (Gln to Gln) (Q395Q) (0)</td>
</tr>
<tr>
<td>33</td>
<td>PDR12 (NE) (Q02785)</td>
<td></td>
<td>197–98 CT to TC; C2059T; T3235C; C4221T</td>
<td>Missense mutation (Ser to Phe) (S66F) (155); Missense mutation (Leu to Phe) (L687F) (22); Missense mutation (Tyr to His) (Y1079H) (83); Synonymous mutation (Thr to Thr) (T1407T) (0)</td>
</tr>
<tr>
<td>34</td>
<td>PDR15 (NE) (Q04182)</td>
<td></td>
<td>T2138C</td>
<td>Missense mutation (Leu to Pro) (L713P) (98)</td>
</tr>
<tr>
<td>35</td>
<td>PDR18 (NE) (P53756)</td>
<td></td>
<td>C2687T</td>
<td>Missense mutation (Pro to Leu) (P896L) (98)</td>
</tr>
<tr>
<td>36</td>
<td>GND1 (NE) (P38720)</td>
<td>6-phospho gluconate dehydrogenase</td>
<td>C897A</td>
<td>Missense mutation (Ser to Ser) (S299S) (0)</td>
</tr>
<tr>
<td>37</td>
<td>HAP3 (NE) (P13434)</td>
<td>Heme Activator Protein</td>
<td>T401C</td>
<td>Missense mutation (Leu to Pro) (L134P) (98)</td>
</tr>
<tr>
<td>38</td>
<td>HEM12 (NE) (P32347)</td>
<td>Uroporphyrinogen decarboxylase</td>
<td>A875G</td>
<td>Missense mutation (Asn to Ser) (N292S) (NA)</td>
</tr>
<tr>
<td>39</td>
<td>HEM14 (NE) (P40012)</td>
<td>Protoporphyrinogen oxidase</td>
<td>A1584G; G1585A</td>
<td>Synonymous mutation (Val to Val) (V528V) (0); Missense mutation (Asp to Asn) (D529N) (NA)</td>
</tr>
<tr>
<td>40</td>
<td>HXT3 (NE) (P32466)</td>
<td>Hexose Transporter</td>
<td>CT323TC; G691T; A1064T; C1067T</td>
<td>Missense mutation (Ser to Phe) (S108F) (NA); Missense mutation (Trp to Leu) (W232L) (NA); Missense mutation (Asp to Val) (D355V) (NA); Missense mutation (Ser to Phe) (S356F) (NA)</td>
</tr>
<tr>
<td>41</td>
<td>LDH1 (NE) (P38139)</td>
<td>Serine hydrolase</td>
<td>C123G</td>
<td>Synonymous mutation (Arg to Arg) (R41R) (0)</td>
</tr>
</tbody>
</table>
most of these mutations are seen in the genes, ERG1, ERG3, ERG4, ERG5, ERG7, ERG10, ERG24, ERG25, ERG26, ERG27, and ERG29 which are downstream of squalene synthesis in the ergosterol pathway of *S. cerevisiae*. The synergistic effect of these mutations could be one of the key reasons for SK23 strain to accumulate higher squalene.

As the strain SK23 could withstand high terbinafine concentrations (>150 μg/L), PDR (pleiotropic drug resistance) genes were analyzed for mutations. PDR genes code for ATP-binding cassette (ABC) transporters, which are multidrug transporters that facilitate drug removal from the cells. PDR5, PDR12, PDR15, and PDR18 have missense mutations, whereas PDR6 and PDR11 have synonymous mutations (Table 2). Besides, THR1 coding for homoserine kinase involved in threonine biosynthesis, ASN1 coding for Asparagine synthetase involved in asparagine biosynthesis and ARO9 coding for aromatic aminotransferase II involved in aromatic amino acid catabolism also have missense mutations. LDH1 coding for Serine hydrolase, GND1 coding for 6-phospho gluconate dehydrogenase, ICL2 coding for isocitrate lyase and ASP1 coding for L-asparaginase has synonymous mutations in its proteins sequence (Table 2).

**TABLE 2**

<table>
<thead>
<tr>
<th>S. No</th>
<th>Gene (Essentiality[^a]) (Uniprot ID[^b])</th>
<th>Description</th>
<th>Nucleotide change</th>
<th>Resultant change (Grantham score[^c])</th>
</tr>
</thead>
<tbody>
<tr>
<td>42</td>
<td>MDH3 (NE) (P32419) Malate dehydrogenase</td>
<td>C627T; C690T</td>
<td>Synonymous mutation (Tyr to Tyr) (Y209Y) (0); Synonymous mutation (Ala to Ala) (A230A) (0)</td>
<td></td>
</tr>
<tr>
<td>43</td>
<td>PFK2 (NE) (P16862) Phospho fructokinase</td>
<td>A2638G</td>
<td>Missense mutation (Lys to Glu) (K880E) (NA)</td>
<td></td>
</tr>
<tr>
<td>44</td>
<td>THR1 (NE) (P17423) Homoserine kinase</td>
<td>G338A</td>
<td>Missense mutation (Gly to Asp) (G113D) (94)</td>
<td></td>
</tr>
</tbody>
</table>

[^a]: Essentiality is defined as NE—non-essential gene and E—essential gene.
[^b]: The values in parentheses indicate Uniprot ID.
[^c]: Grantham score: conservative (0–50), moderately conservative (51–100), moderately radical (101–150), or radical (≥151).

**FIGURE 3** Homology modeling and docking studies for Erg1p from *S. cerevisiae*. (a) Homology structure of Erg1p (gray) docked to FAD (yellow). (b) Interaction of FAD with Erg1p including sites such as D335, G27, G210, and D209. (c) Homology structure of Erg1p-FAD complex docked to terbinafine (red). (d) Interaction of terbinafine to Erg1p-FAD complex including site F420 among others.
mutation in AAT1 has not altered the amino acid (Table 2). HMG2 coding for HMG-CoA reductase involved in the HMG-CoA conversion has been identified to harbor mutations (Table 2). HMG-CoA reductase is involved in the ergosterol biosynthesis pathway and has been targeted for overexpression in the HMG-CoA pathway studies as HMG-CoA is a critical precursor for the terpene and sterol biosynthesis pathway (Mantzouridou & Tsimidou, 2010). FAT1 gene coding for fatty acyl-CoA synthetase in fatty acid biosynthesis and also involved in fatty acid transport has a missense mutation. HXT3 coding for hexose transporter involved in glucose transport has several missense mutations (Table 2).

HAP1 gene codes for a transcription factor and is involved in regulating ERG9 and heme-mediated regulation of HMG1 (Garaiová et al., 2014; Kennedy et al., 1999). HAP1 gene has been found to harbor missense mutations in its protein sequence (Table 2), leading to altered gene expression. HAP1 could be considered a potential gene target as it is involved in the regulation of squalene synthase coded by ERG9, which is a critical enzyme in squalene synthesis. NDI1 gene coding for NADH dehydrogenase involved in central carbon metabolism has been identified to harbor mutation that leads to altered protein expression (Table 2). PDA1 gene codes for pyruvate dehydrogenase enzyme, which converts pyruvate to acetyl-CoA. PDA1 harbors missense mutations, which in turn may lead to altered protein formation. HEM12 codes for uroporphyrinogen decarboxylase enzyme involved in heme biosynthesis process. Heme mutant enhance ERG9 gene expression, and it also mediates the regulation of the HMG1 gene (Kennedy et al., 1999). HEM12 gene harbors a mutation leading to altered protein expression. HEM14 involved in heme biosynthesis has also been found to harbor a synonymous mutation (Table 2). YEH2 gene coding for steryl ester hydrolase involved in sterol homeostasis has been found to harbor a missense mutation. ADH5 is another critical gene involved in ethanol metabolism that harbors a missense mutation in its protein sequence. Overproduction of squalene has proven to reduce the ethanol production where the gene expression levels of the ADH5 gene was low compared to the wild-type strain (Rasool et al., 2016). Further, the downregulation of ADH genes possibly reduces the conversion of acetyl-CoA to ethanol which could be redirected to terpene biosynthesis. Further, few other mutations were also observed, as listed in Table 2. Though whole-genome sequencing has led to identifying mutations, independent in vitro evolution experiments are required to enable the random and causal mutation separations (Guadalupe-Medina et al., 2014; Oud et al., 2012). Analyzing individual mutation with enzyme-substrate docking studies might lead to further understanding the rationale behind improved squalene synthesis and tolerance to terbinafine. Such kind of study could also differentiate between causal and random mutations. However, the findings from the current study could potentially facilitate such reverse metabolic engineering for squalene improvement in S. cerevisiae.

3.3 Quantitative and qualitative analysis of neutral lipid droplets using Nile red-stained yeast cells

Based on Nile red staining, it was observed that terbinafine adapted cells of S. cerevisiae accumulating higher squalene had enhanced neutral lipid and lipid droplet accumulation (Figures 4–6). ERG1 gene is inhibited by evolving the engineered strains in the presence of terbinafine at various concentrations, which leads to the development of several squalene hyper producing strains (Figure 1). To assess the effect of increased squalene on lipid droplets, the resulting strains were qualitatively screened for lipid droplet clustering using fluorescence microscopy based on Nile red assay (Figure 4). Strains, E15, E30, E45, and E58, were subjected to fluorescent microscopy for lipid droplet clustering analysis. Cells were stained with Nile red, and their images were taken by fluorescence microscopy and evaluated as described in methods. Figure 4 shows the fluorescence of Nile red-stained S. cerevisiae cells when viewed under red fluorescence (e.g., 545/Em.572 nm).

Lipid droplets detected in the wild-type culture were found to be significantly lower in comparison to the evolved strain. The aggregation of droplets was higher in all the evolved strains (Figure 4b–f). Moreover, the intensity of lipid droplets is directly correlated to squalene. For instance, E15 (Figure 4b) synthesizes a moderate level of
FIGURE 5  Neutral lipid fluorescence intensity of Nile red-stained *S. cerevisiae* wild-type and evolved mutants grown in the absence of terbinafine using excitation of 488 nm and emission wavelengths of 535 nm. Data represent the mean of duplicates ± SE.

FIGURE 6  Flow cytometric histograms of Nile-red-stained *S. cerevisiae* cells which represent the effect of terbinafine treatment on lipid droplet content in the evolved strains. (a) Control. (b) E15. (c) E45. (d) E50. (e) E55. (f) E58. (g) Overlay. Cell fluorescence was measured at the laser excitation wavelength of 488 nm and an emission wavelength of 515–560 nm.
squalene, whereas E45, E50, E55, and E58 (Figure 4c–f) synthesize a high level of squalene. Lipid droplet clustering was quantified as a percentage of cells based on the number of cells with lipid droplet clusters to the cells without clusters. The clusters for lipid droplets were identified using fluorescence microscopy as shown in Figure 4. Here the cells with more fluorescence were considered to have higher number of lipid droplets and vice versa. The control strain shows approximately 18% of clustering, whereas the E15, E45, E50, E55, and E58 (Figure 4b–f) show approximately 48%, 73%, 86%, 93%, and 94%, respectively. Results were calculated as the average of two independent experiments.

Further, the neutral lipid content and pattern were monitored during the evolution experiment by sampling at various time intervals and analyzed using a spectrophotometer and a flow cytometer (Figures 5 and 6). To signify the association of lipid droplet increase with squalene enhancement, Nile red-stained cells were subjected to spectrofluorometric analysis. The relative fluorescence of neutral lipids using the Nile red assay was measured with a spectrophotometer. The neutral lipid content has increased to 18.7-fold in SK23 strain when compared to the control strain. The hyper-accumulator strains E46, E52, E53, E54, and E59, have shown a 16.9, 18.6, 15.5, 18.4, and 16.6-fold increase in neutral lipid content (Figure 5).

All the evolved strains after 300 generations have shown a higher level of neutral lipid content (Figure 5). Further qualitative characterization of neutral lipid patterns was carried out using a flow cytometer, and the results were found to agree with the fluorescent microscopic analysis. The lipid droplets of evolved strains after 15, 30, 45, and 58 subcultures have been qualitatively analyzed with flow cytometry. The graphs represent the different lipid-rich cells by their increased fluorescence intensity relative to control cells analyzed for yellow–gold fluorescence (Figure 6). The plot 01 of Figure S2 represents the size and granularity of the cells, and plot 02 of Figure S2 represents the fluorescent intensity of the gated cells. Flow cytometric analysis of unstained cells with autofluorescence (yellow peaks) was compared with the fluorescence (colored peaks) of evolved cells (Figure 6). An increase in fluorescence indicates an increase in neutral lipid content. In summary, the above data represents that the increase in squalene leads to an increase in neutral lipid content, leading to lipid droplet formation.

**ERG1** gene codes for squalene epoxidase (squalene mono-oxygenase) convert squalene to squalene epoxide in the presence of oxygen molecules during sterol synthesis (Jahnke & Klein, 1983). Erg1p is localized in lipid particles and microsomal fraction, whereas ERG9 coding for squalene synthase is localized in the endoplasmic reticulum (Leber et al., 1998; Popják & Agnew, 1979). Erg9p is involved in the formation of squalene from farnesyl pyrophosphate. Lipid droplet clustering, which refers to the aggregation of six or more droplets along with squalene accumulation, has been observed in ERG1 mutant cells defective in sterol synthesis (Ta et al., 2012). A similar phenomenon was also observed in the cells in which squalene mono-oxygenase is inhibited by terbinafine. Ta et al. (2012) concluded that squalene enhancement by ERG1 mutation or inhibition is associated with lipid droplet clustering in yeast and mammalian cells. The study also states that squalene is responsible for the clustering of lipid droplets, not the other intermediates of the sterol pathway proving the role of squalene in lipid droplet dynamics.

According to Spanova et al. (2010), squalene synthesized in hem1 mutant was localized in lipid droplets. Another study by Spanova et al. (2012) also states that the increased squalene need not necessarily get accumulated in the lipid droplets, but instead, it is accommodated in the cell membrane when lipid droplets are absent. Interestingly, a recent study states that squalene synthesis could also be compartmentalized into peroxisomes (Liu et al., 2019). Nile red staining is an efficient method to determine the lipid phenotype of the yeast cells (Greenspan et al., 1985). It has been widely used to analyze lipid droplets and estimate relative amounts of neutral lipids in yeast (Ta et al., 2012). Lipid droplet is a storage depot for neutral lipids in yeast. Neutral lipids can be rapidly measured by spectrofluorimetric and flow cytometric methods. Granularity can also be used as a parameter to measure intracellular neutral lipids.

In the study by Ta et al. (2012), it has been proven that squalene accumulation is associated with lipid droplet clustering using different yeast strains based on fluorescent microscopy. This hypothesis is proven in the present study using different methods such as fluorescent microscopy, spectrofluorimetry, and flow cytometry analysis. Though squalene is involved in lipid droplet accumulation and clustering, it cannot initiate the formation of lipid droplets solely by itself in the absence of storage molecules TAGs and SEs. In the absence of TAGs and SEs, squalene is sequestered in microsomes and is not lipotoxic to the cells even at higher concentrations (Spanova et al., 2010). Moreover, the mechanism by which squalene increases the clustering of lipid droplets is presently unclear (Spanova et al., 2012). Yeast cells were subjected to Nile-red staining in the mid-exponential phase as the cells in the stationary phase undergoes stress such as carbon source exhaustion, high acidic environment, and higher ethanolic content which might cause leakage of lipid droplets due to compromised membrane integrity (Lopes da Silva et al., 2017).

**4 | CONCLUSION**

The previous metabolic engineering strategies have enabled us to achieve a production level of 65 mg/L of squalene under shake-flask conditions. Medium composition, extraction, kinetic analysis, and cultivation conditions have also been optimized in the previous studies. Apart from rational strain engineering, evolution has proven to be a powerful tool in strain engineering. Adaptive laboratory evolution has enhanced the squalene titer to 193 and 420 mg/L in the absence and presence of terbinafine respectively under shake-flask conditions, which are 4-fold and 8-fold higher than the engineered parent strain respectively. Genome-wide modifications in the evolved strain (SK23) were analyzed using whole-genome sequencing. The squalene yield increased from 0.4% of the theoretical yield in the SK22 strain (0.00128 g/g glucose) to 2.2% of the theoretical yield (0.007 g/g glucose) in the evolved strain (SK23). Among the
different identified mutations through the genome sequencing analysis, ERG1 was established as causal mutation, and also HMG2, ACO1, PG1, and HAP1 could be some of the potential leads for inverse metabolic engineering. Also, transcriptomic and metabolomic analysis of the strains can be carried out to study the flux flow towards squalene and understand the basis of complex phenotypes. The neutral lipid content and the lipid droplet accumulation were increasing with the increase in squalene content. Hence, the present study confirms that there is a strong interrelation between squalene improvement and lipid droplets accumulation. However, further investigation on the clustering of lipid droplets by squalene increase needs to be carried out.

ACKNOWLEDGEMENTS
SM acknowledges the Science and Engineering Research Board (SERB) for funding the project [CRG/2019/003323]. KP acknowledges the Department of Biotechnology, India and the Department of Science and Technology (DST), India, for awarding the fellowship. Special thanks to Prof. Friederike Turnowsky and Ms. Gertrude Zisser (University of Graz, Austria) for providing the KLN1 strain. The Director, CSIR – Central Food Technological Research Institute (CFTRI), Mysuru, India, is acknowledged for supporting this work. We thank Ms. Debjani and Mr. Anbalagan for experimental support during flow cytometric and fluorescence microscopic analysis, respectively. We also thank Yaazh Xenomics for the support in genomic sequence retrieval and analysis.

CONFLICT OF INTEREST
The authors report no conflict of interest.

AUTHOR CONTRIBUTIONS
SM and KP designed the study. SM supervised the study. KP, AA, and NG performed the experiments. SM, KP, and AA analyzed the data and wrote the manuscript. The datasets generated during the current study will be made available upon request.

ORCID
Kalaivani Paramasivan https://orcid.org/0000-0001-6743-0003
Sarma Murturi https://orcid.org/0000-0002-3251-8643

REFERENCES
https://aem.asm.org/content/76/10/3097, https://doi.org/10.1128/AEM.00115-10


**SUPPORTING INFORMATION**

Additional supporting information may be found online in the Supporting Information section at the end of this article.