

SCIENTIFIC REPORTS



OPEN

Complete genome sequence and comparative genomics of the probiotic yeast *Saccharomyces boulardii*

Indu Khatri, Rajul Tomar , K. Ganesan , G. S. Prasad & Srikrishna Subramanian

The probiotic yeast, *Saccharomyces boulardii* (*Sb*) is known to be effective against many gastrointestinal disorders and antibiotic-associated diarrhea. To understand molecular basis of probiotic-properties ascribed to *Sb* we determined the complete genomes of two strains of *Sb* i.e. Biocodex and unique28 and the draft genomes for three other *Sb* strains that are marketed as probiotics in India. We compared these genomes with 145 strains of *S. cerevisiae* (*Sc*) to understand genome-level similarities and differences between these yeasts. A distinctive feature of *Sb* from other *Sc* is absence of Ty elements Ty1, Ty3, Ty4 and associated LTR. However, we could identify complete Ty2 and Ty5 elements in *Sb*. The genes for hexose transporters *HXT11* and *HXT9*, and asparagine-utilization are absent in all *Sb* strains. We find differences in repeat periods and copy numbers of repeats in flocculin genes that are likely related to the differential adhesion of *Sb* as compared to *Sc*. Core-proteome based taxonomy places *Sb* strains along with wine strains of *Sc*. We find the introgression of five genes from *Z. bailii* into the chromosome IV of *Sb* and wine strains of *Sc*. Intriguingly, genes involved in conferring known probiotic properties to *Sb* are conserved in most *Sc* strains.

The probiotic yeast *Saccharomyces boulardii* (*Sb*) has unique physiological properties such as tolerance to variations in pH, temperature and local stresses like the presence of GI enzymes, bile salts, and organic acids¹. Several clinical studies have been performed to present *Sb* as a unique organism that inhibits pathogens^{2–5} and restores gut flora and improves digestion^{6,7}. The mechanism of elimination of the pathogenic bacteria is mainly attributed to the adhesion proteins of *Sb* that binds to bacteria and inhibits their adhesion to the mucous-intestinal membrane^{8,9}. Certain proteins in *Sb* have been reported previously for their key roles in providing protection against *Escherichia coli*, *Clostridium difficile*, *Vibrio cholera*, and *Helicobacter pylori* infections¹⁰. A 63 kDa protein phosphatase of *Sb* dephosphorylates the *E. coli* endotoxin¹¹, a 54 kDa serine protease provides protection against *Clostridium difficile* infections by cleaving toxins A and B^{12,13} and a 120 kDa heat and trypsin-labile non-proteolytic protein of *Sb* neutralizes the secretions induced by cholera toxin by possibly reducing cyclic Adenosine Monophosphate (cAMP) levels¹⁴.

A French Scientist Henri Boulard initially isolated *Sb* from the fruits lychee and mangosteen in 1923 and the organism was characterized as '*Saccharomyces boulardii*' a novel species of genus *Saccharomyces* possibly to differentiate its probiotic effects and application from other yeast species¹⁵. Characterization of *Sb* as a separate species was further supported by the lack of galactose utilization and sporulation as compared to *S. cerevisiae* (*Sc*)¹⁶. Molecular phylogenetic and typing techniques suggested that *Sb* forms a separate cluster but belong to species *Sc*¹⁷. Comparative genomic hybridization experiments also established that *Sc* and *Sb* are different strains of the same species but the loss of all intact Ty1/2 elements was reported only in *Sb*. The loss of Ty elements was hypothesized to be related to *Sb*'s non-sporulation and diploidy¹⁸ as the transcription of these mobile elements is under diploid control^{19,20}. The numbers of Ty elements are maintained via transposition during sporulation and haploid mitotic growth, but the absence of these stages can lead to the loss of Ty elements^{21,22}.

The unique formulations of *Sb*, isolated from different sources like lychee, mangosteen, pineapple, etc., explored for treating different disorders have been patented by different companies and laboratories^{23–26}. The mode of action of the probiotic yeast is not completely known¹⁰, however the beneficial effects of the yeast have

CSIR-Institute of Microbial Technology, Chandigarh, India. Correspondence and requests for materials should be addressed to S.S. (email: krishna@imtech.res.in)

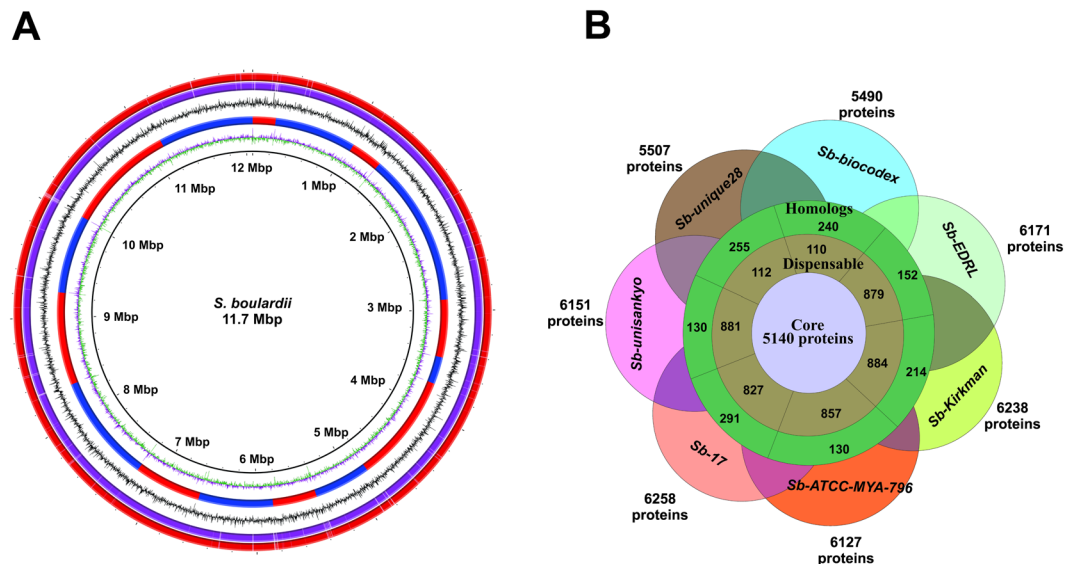


Figure 1. (A) Circular representation of the *Sb-biocodex* and *Sb-unique28* complete genome. Circles (from inside to outside): circle 1 (GC content), circle 2 (*Sc* S288C complete chromosomes with alternative blue and red color); circle 3 (GC skew); circle 4 (*Sb-biocodex* complete chromosomes); circle 5 (*Sb-unique28* complete chromosomes). BRIG 0.95 was used to build the circular representation. Mapping studies were done using BLASTn with an E-value cut-off $1e-5$. (B) Core orthologous proteins among all *Sb* proteome represented by orthology diagram depicting dispensable proteome and homologs among these strains.

| <i>Sb</i> strains | Number of Contigs | N50 (bp) | Genome Size (Mbp) |
|---------------------|-------------------|----------|-------------------|
| <i>Sb-biocodex</i> | 31 | 792,172 | 12.0 |
| <i>Sb-unique28</i> | 30 | 909,172 | 12.1 |
| <i>Sb-EDRL</i> | 77 | 306,308 | 11.5 |
| <i>Sb-kirkman</i> | 115 | 416,209 | 11.7 |
| <i>Sb-unisankyo</i> | 164 | 262,146 | 11.6 |

Table 1. Genome Assembly statistics of all *Sb* strains.

been established through various clinical studies^{27–33}. However, the probiotic yeast has also been found to be associated with fungemia in immune-compromised patients^{34–37}. Our group reported the first genome of *Sb-EDRL* strain (Econorm from Dr. Reddy's Laboratory) and tried to trace the genomic reasons for the probiotic behavior of this yeast³⁸. There we explored that the proteins appeared to be specifically present in *Sb* were also present in the *Sc* strains³⁸. However, a single draft genome may not be sufficient to study the probiotic properties associated with the organism. Whole genome sequence of more *Sb* strains including complete genomes will be required to understand the evolution and quantitative variations among *Sb* strains.

We have sequenced whole genomes from five *Sb* strains marketed by Laboratories Biocodex, Kirkman Labs, Unisankyo Ltd. (Now Sanzyme Ltd.) and Unique Biotech to find the reasons for probiotic properties of *Sb*. Two strains isolated from *Sb* sachets marketed by Laboratories Biocodex and Unique Biotech was assembled to completion for comparison and to address biologically relevant differences. We report the complete genomes of *Sb-biocodex* and *Sb-unique28* in this paper along with draft genomes of *Sb-kirkman* and *Sb-unisankyo* and updated version of the *Sb-EDRL* genome. We have analyzed genomes of seven different strains of *Sb* to find if they are different and can account for different probiotic properties or are species specific. We have compared the *Sb-biocodex* with *Sb-unique28* genome to find differences among these probiotics at the strain level. To address the differences between probiotic yeasts and other strains of *Sc* we compared all five *Sb* genomes with all available genomes of *Sb* and *Sc* strains. This genomic study for probiotic yeast relates to the previous clinical and molecular studies and reports variations at genome level among strains of *Sb* and *Sc*.

Results

Complete genomes of *Sb* strains and their characteristics. The complete genomes were obtained for *Sb-biocodex* and *Sb-unique28* (Fig. 1A) sequenced using PacBio P6C4 chemistry at ~200x coverage. The final assembly of *Sb-biocodex* (12 Mbp genome and N50 792,172 bp) comprises of 16 complete chromosomes and 14 unplaced contigs. Similarly, *Sb-unique28* was finalized with 14 complete chromosomes, two chromosomes (Chromosome 5 and 9) comprising of two contigs each and nine unplaced contigs (12.1 Mbp genome and N50 929,172 bp) (Table 1; Supplementary File I). Detailed genomic features of *Sb-biocodex* and *Sb-unique28* are discussed in Supplementary methods (Supplementary Figure 1). The complete chromosomes of *Sb-biocodex*

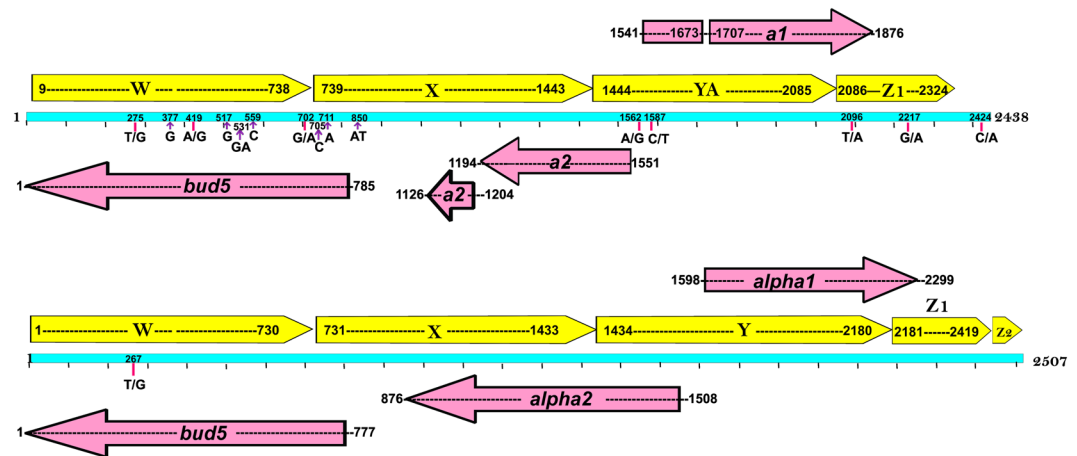


Figure 2. *MAT* Locus of *Sb biocodex* depicting the annotated regions. (A) *MATa* locus with yellow arrows depicting the *MAT* region divided in W, X, Y_A and Z₁ regions with coded *MATa1* and *MATa2* regulatory proteins. (B) *MATα* locus with yellow arrows depicting the *MAT* region divided in W, X, Y, Z₁ and Z₂ regions with coded *MATα1* and *MATα2* regulatory proteins.

and *Sb-unique28* were identified with either the telomeres or the telomeric and sub-telomeric regions. The Chromosome I of *Sb-biocodex* had a shorter length as compared to *Sb-unique28* and *Sc S288C* where we found that the sub-telomeric region of approximately 0.02 Mbp on the right arm with genes annotated as dubious or uncharacterized ORFs (YAR053W to YAR071W) in *Sc S288C* genome (<http://www.yeastgenome.org>) is absent.

Sb-EDRL, sequenced using 454 sequencing data, was assembled in 107 contigs (Genome Size: 11.5 Mb and N50: 271,789 bp) and was further scaffolded into 77 gapless contigs (Genome Size: 11.5 Mb and N50: 819,652 bp) using Illumina HiSeq PE and MP shotgun data (Table 1). The shotgun reads of *Sb-kirkman* and *Sb-unisankyo* were assembled in 115 contigs (Genome Size: 11.7 Mb; N50: 621,720 bp) and 164 contigs (Genome Size: 11.6 Mb; N50: 262,146 bp), respectively (Table 1).

Complete 2-micron plasmid was retrieved from *Sb-biocodex* and *Sb-unique28* matching to *Sc YJM993* plasmid (Length: 6318 bp; Genbank identifier: CP004528.1). Similarly, the complete circular plasmid was also retrieved from *Sb-EDRL*, *Sb-kirkman*, and *Sb-unisankyo* by mapping their reads to the *Sb-biocodex* plasmid sequence. The complete circular plasmid obtained from all the *Sb* strains was found to be 100% identical. G₅₂₉₇ → A and A₅₅₈₂ → G polymorphisms in the *rep2* gene were observed in *Sb* plasmids as compared to *Sc YJM993* plasmid. The mutation A₅₅₈₂ → G was non-synonymous but G₅₂₉₇ → A corresponds to A₂₉₆ → V amino acid change in the Rep2 protein (Supplementary Figure 2).

Approximately 5500 CDS and 300 tRNAs were predicted in all the *Sb* strains (Supplementary File I). The *Sc S288C* genome was also re-annotated using a similar method. All these ORFs were characterized functionally based on the gene names and description provided in Saccharomyces Genome Database (SGD)³⁹. The core proteome comprised of 5140 proteins for all *Sb* strains (Fig. 1B) and approximately 200 proteins are found to be unique in each *Sb* strain (Supplementary File II). The unique proteins in all the strains of *Sb* were extracted and subjected to BLASTp against the proteome of all other *Sb* to find if any homologs to those proteins are present in other strains of *Sb*. All these unique proteins have homologs in other strains of *Sb*. All the *Sb* genomes assembled and sequenced in our study were compared with two already available *Sb* genomes (*Sb-17* and *Sb-ATCC-MYA-796*) and 145 *Sc* genomes (Supplementary File III).

Mating Locus and Sporulation in *Sb*. Mating type of yeast was determined by the two different alleles of Mating-type (*MAT*) Locus *MATa* and *MATα*⁴⁰. *Sb* has been suggested to be diploid in previous studies^{1, 41, 42} and should comprise both *MATa* and *MATα* sequences on the Chromosome III at a heterozygous locus. The characterized locus in *Sc* was used as a query to search the *Sb-biocodex* genome and both *MATa*, and *MATα* sequences were retrieved from *Sb-biocodex* at 99% identity in chromosome III and unplaced scaffold, respectively. *MATa* sequence in *Sb-biocodex* is 2438 bp in length which is 99.7% identical to *Sc MATa* locus (GI: V01313) with eight substitutions and seven insertions (Fig. 2). The region was divided into W, X, Y_A and Z₁ regions based on the alignment to *MATa* locus of *Sc* (GI: V01313). Similarly, the *MATα* gene was retrieved by subjecting the *MATα* region of *Sc S288C* from SGD to BLASTn against *Sb* genomes. The *MATα* gene was 2507 bp with only one substitution T₂₆₇ → G and could be divided into W, X, Y, Z₁ and Z₂ regions based on the alignment to *Sc S288C MATα* gene. The ORFs coded by *MATa* and *MATα* were identified through Augustus, and very short ORFs were predicted using DNA to protein translation tool (<http://insilico.ehu.es/translate/>). In yeast the *MATα* locus codes for *bud5*, *MATα1*, and *MATα2* genes and *MATa* locus codes for *bud5*, *MATa1*, and *MATa2* genes. Homothallic switching endonuclease (*HO*) gene is required for gene conversion at *MAT* Locus in haploid cells⁴³. In heterothallic strains T189A, G223S, L405S, H475L substitutions and deletion of 36 amino acids (524–559) in *HO* protein results in loss of endonuclease activity^{44, 45} whereas homothallic cells express *HO*⁴⁵. The *HO* gene in all strains of *Sb* did not have any of the above-mentioned substitutions or deletions, and hence, the probiotic yeast *Sb* is homothallic diploid wherein both the *MAT* loci are present in the genome.

As it is known that both the heterothallic and homothallic diploid strains sporulate under conditions of nutrient deficiency⁴⁶, we performed sporulation phenotype assay (detailed in Supplementary Methods), to investigate the sporulation in *Sb*. We found that the *Sb* did not sporulate on sporulation media even after one week of incubation concordant with previous studies stating non-sporulation behavior of the probiotic yeast⁴¹. To investigate the role of sporulation proteins in the non-sporulating behavior of *Sb* the presence and the absence of all the sporulation proteins mentioned in SGD were identified using BLASTp in *Sb* proteome. Further, to rule out the possibility of the lack of any meiotic and mitotic genes hindering the sporulation pathway, we searched for 110 meiotic genes and 56 mitotic genes in *Sb* genome and found all were present (Supplementary File IV). An earlier report suggesting the divergence in CDC16, DMC1, and MND2 sequences as a possible reason for the defective sporulation was also investigated⁴¹; but all these proteins of *Sb* were >99% identical to the respective proteins in *Sc*. Also, *Sb* was grown on non-fermentable carbon source (Glycerol) to find if the non-sporulation of *Sb* is governed by respiration-sensing pathway⁴⁷ (detailed in Supplementary Methods). The growth of *Sb* was found on the non-fermentable carbon source (Supplementary Figure 3). Thus, the non-sporulation phenotype of *Sb* is unlikely to be due to a mutation in sporulation and respiration pathway genes. Since functional *MAT* loci are also critical for sporulation⁴⁸, we speculate that mutations at these loci, as seen in *MATa* sequence in *Sb-biocodex* with eight substitutions and seven insertions compared to *Sc MATa* (Fig. 2), might be responsible for the non-sporulation phenotype of *Sb*.

Gene copy number variations in *Sb* strains as compared to *Sc*. *Genes absent.* The mapping studies to *Sc* genes of yeastmine database⁴⁹ revealed that 50 genes had no reads mapped onto them. Also, we found that 94 genes apart from these 50 genes had read coverage <20 reads. The absence of these genes was confirmed by subjecting these genes as a query to BLASTn against the *Sb-biocodex* and *Sb-unique28* PacBio assembly with complete chromosomes. Out of these 144 genes, 85 were dubious ORFs, 32 were uncharacterized genes, and 27 were functionally verified genes. The verified genes include two maltase utilization genes (*MAL11* and *MAL13*), two hexose transporters (*HXT9* and *HXT11*), four asparagine catabolism gene (*ASP3-1*, *ASP3-2*, *ASP3-3* and *ASP4-4*), three palatinose utilization genes (*IMA2*, *IMA3*, *IMA4*), two putative membrane glycoprotein (*VTH1* and *VTH2*) and *ARN2*, *REE1*, *AYT1*, *AIF1*, *COS10*, *ENB1* and *BDS1* (Table 2). Except *ASP3* locus, all these absent genes belong to telomeric or subtelomeric regions of chromosome.

Also, the yeastmine database genes were subjected to BLASTn against all strains of *Sc* included in this study to find genes unique to *Sb* (Supplementary File V). *MAL11*, *MAL13*, and *ARN2* were present in more than 70% of the strains of different subgroups of *Sc* strains but were absent in all the probiotic strains. We found that the *BDS1* gene was present in tree isolates, laboratory strains, and environmental samples but was absent in wine strains, beer strains, clinical, fruit derived, bakery and bioethanol producing strains. Similarly, the *REE1* gene was present in tree isolates, laboratory strains and environmental samples and a few strains belonging to subgroup wine, beer and clinical but was absent in probiotic *Sb* strains.

The *ASP3* locus was present in more than 80% of the laboratory or industrial strains and bioethanol producing strains and *Sc* strain YJM1383, a fruit derived strain; *Sc* strain CLIB324, a bakery strain, and a few clinical strains (YJM248, YJM339, YJM451, YJM693, YJM1078, and YJM1311). It was absent in the wine, distillery, and probiotic strains. The ORFs coding for this locus on chromosome XII is adjacent to the ribosomal DNA locus. The hexose transporter family is large and comprises of *HXT1-17* genes⁵⁰ of which *HXT11* and *HXT9* were absent from all strains of *Sb*. *HXT11* and/or *HXT9* null mutants of *Sc* are resistant to cycloheximide, sulfomethuron methyl, and 4-NQO (4-nitroquinoline-N-oxide)⁵¹ indicating *Sb* strains also could be resistant to these chemicals.

Multi-copy genes. The variation in the copy number of genes in a genome can have phenotypic and physiological differences⁵². Genes for PAU proteins, a member of the seripauperin multigene family, were found to be present in 18–20 copies in the genome, and *gag-pol fusion* genes were present in 15 copies in the whole genomes of *Sb-biocodex* and *Sb-unique28* (Table 3). *THI13* is present in five copies at the sub-telomeric regions and *IMD3* and *COS3* are present in four copies at the telomeric regions of *Sb* chromosomes. *Imd3* catalyzes the rate-limiting step in the *de novo* synthesis of GTP⁵³ and *Cos3* is involved in salt resistance⁵⁴ in *Sc*. The clusters of duplicated and triplicated genes mostly encode stress-related proteins, elongation factors, ribosomal proteins, kinases and transporters, fluoride export and altering replication stress tolerance. These duplicated genes could be helping in better adaptation of *Sb* to the harsh conditions of the mammalian host.

Ty elements in *Sb*. Ty1, Ty3, and Ty4 elements were absent in all the *Sb* genomes whereas Ty2 elements were present in *Sb-biocodex*, *Sb-kirkman* and *Sb-unique28* and one Ty5 element was present in all *Sb* strains except *Sb-17* and *Sb-MYA-796* (Supplementary Figure 4). Also, the presence of genes encoding *gag-pol* and *gag-co-pol* fusion proteins was confirmed by read mapping and coverage analysis. These were either contained in the Ty2 or Ty5 elements. The presence of Ty elements was compared in between complete genomes of *Sb viz. Sb-biocodex* and *Sb-unique28*; where the elements are present in the same chromosomes with some deviations in positions (Table 4; Supplementary File VI). Ty1–4 elements integrate near tRNA or RNA polymerase III genes⁵⁵, but analyzing the neighboring genes of these Ty elements, we found only one Ty element in each of *Sb-unique28* and *Sb-biocodex* has Ribosomal 40S subunit protein upstream of it.

Flocculation and adhesion. *FLO1*, *FLO5*, *FLO8*, *FLO9*, *FLO10*, *FLO11*, *FIG2*, and *AGA1* encode flocculation proteins that belong to yeast adhesin families, and their sufficient expression leads to flocs, flor, biofilms or filaments formation by either binding to other yeast cell receptors or foreign surfaces⁵⁶. The protein sequences of these flocculins obtained from SGD were subjected to BLASTp against *Sb* and *Sc* proteomes (Supplementary File VII). Flocculation genes are characterized by the presence of a large number of repetitive sequences with linear correlation to their size⁵⁷. *FLO1* and *FLO5* are paralogs that arose from segmental duplication⁵⁸ of which we

| Systematic | Genes | Gene Function |
|------------|---------------|--|
| YOL165C | <i>AAD15</i> | Aryl-Alcohol Dehydrogenase |
| YNR074C | <i>AIF1</i> | Mitochondrial cell death effector |
| YHL047C | <i>ARN2</i> | Transporter |
| YLR155C | <i>ASP3-1</i> | Cell-wall L-asparaginase II involved in asparagine catabolism |
| YLR157C | <i>ASP3-2</i> | Cell-wall L-asparaginase II involved in asparagine catabolism |
| YLR158C | <i>ASP3-3</i> | Cell-wall L-asparaginase II involved in asparagine catabolism |
| YLR160C | <i>ASP3-4</i> | Cell-wall L-asparaginase II involved in asparagine catabolism |
| YLL063C | <i>AYT1</i> | Acetyltransferase |
| YOL164W | <i>BDS1</i> | Bacterially-derived sulfatase |
| YLR465C | <i>BSC3</i> | Bypass of Stop Codon |
| YNR075W | <i>COS10</i> | Protein of unknown function |
| YGR295C | <i>COS6</i> | Protein of unknown function |
| YOL158C | <i>ENB1</i> | Endosomal ferric enterobactin transporter |
| YOL156W | <i>HXT11</i> | Putative hexose transporter that is nearly identical to Hxt9p |
| YJL219W | <i>HXT9</i> | Putative hexose transporter that is nearly identical to Hxt11p |
| YOL157C | <i>IMA2</i> | Isomaltase |
| YIL172C | <i>IMA3</i> | Isomaltase |
| YJL221C | <i>IMA4</i> | Isomaltase |
| YGR289C | <i>MAL11</i> | High-affinity maltose transporter (alpha-glucoside transporter) |
| YGR288W | <i>MAL13</i> | MAL-activator protein |
| YIR041W | <i>PAU15</i> | seripauperin |
| YKL224C | <i>PAU16</i> | seripauperin |
| YJL217W | <i>REE1</i> | Cytoplasmic protein involved in the regulation of enolase (ENO1) |
| YAL064C-A | <i>TDA8</i> | Topoisomerase I Damage Affected |
| YOR068C | <i>VAM10</i> | Vacuolar Morphogenesis |
| YIL173W | <i>VTH1</i> | Putative membrane glycoprotein |
| YJL222W | <i>VTH2</i> | Putative membrane glycoprotein |

Table 2. Absent genes in *Sb-biocodex* and *Sb-unique28*.

| | <i>Sb biocodex</i> | <i>Sb unique28</i> | <i>Sc S288C</i> |
|--|--------------------|--------------------|-----------------|
| Total clusters | 163 | 150 | 146 |
| Clusters with two proteins | 148 | 137 | 127 |
| Clusters with more than two proteins | | | |
| Genes in multiple copies | <i>Sb biocodex</i> | <i>Sb unique28</i> | <i>Sc S288C</i> |
| Seripauperin PAU | 18 | 20 | 7 |
| gag-pol fusion proteins | 16 | 16 | 49 |
| Thi13 | 5 | 6 | 1 |
| IMP dehydrogenase IMD3 | 4 | 2 | 1 |
| Cos3p | 4 | 5 | 1 |
| YIL169C-like protein | 3 | 4 | 2 |
| Aad4p | 3 | 2 | 1 |
| Fex1p | 3 | 3 | 1 |
| ribosomal 60S subunit protein L2B | 3 | 4 | 1 |
| Hsp32p | 3 | 3 | 1 |
| Y' element ATP-dependent helicase protein 1 copy 1 | 2 | 5 | 5 |

Table 3. Multicopy genes in *Sb-biocodex*, *Sb-unique28* and *Sc S288C*.

| | <i>Sb-17</i> | <i>Sb-biocodex</i> | <i>Sb-EDRL</i> | <i>Sb-kirkman</i> | <i>Sb-MYA-796</i> | <i>Sb-unique28</i> | <i>Sb-unisankyo</i> | Sc S288C |
|-----------------|--------------|--------------------|----------------|-------------------|-------------------|--------------------|---------------------|----------|
| gag | 1 | | | 1 | 1 | 1 | 1 | 1 |
| gag co pol | | 3 | 5 | 3 | | 3 | 1 | 45 |
| gag pol | | 2 | | | | 3 | | 45 |
| pseudo | 2 | | 2 | 2 | 2 | 2 | 2 | 2 |
| Suppressor SPT7 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 |
| Ty1 element | | | | | | | | 31 |
| Ty1 LTR | 147 | 191 | 173 | 190 | 151 | 200 | 169 | 483 |
| Ty2 element | | 10 | | 1 | | 7 | | 13 |
| Ty2 LTR | 5 | 4 | 12 | 9 | 8 | 14 | 11 | 32 |
| Ty3 element | | | | | | | | 2 |
| Ty3 LTR | 26 | 46 | 40 | 45 | 29 | 45 | 33 | 42 |
| Ty4 element | | | | | | | | 3 |
| Ty4 LTR | 15 | 16 | 17 | 17 | 15 | 17 | 15 | 35 |
| Ty5 element | | 1 | 1 | 1 | | 1 | 1 | 1 |
| Ty5 LTR | 5 | 3 | 3 | 4 | 4 | 3 | 3 | 8 |
| Ty A | | | | | | | | 1 |
| Ty B | | | | | | | | 1 |

Table 4. Distribution of Ty elements in *Sb* strains.

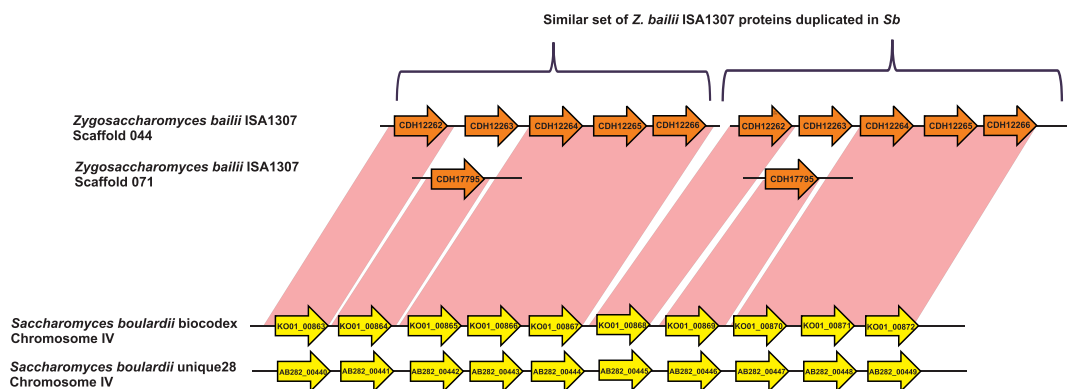


Figure 3. Introgression of five genes in two copies from *Zygosaccharomyces bailii* to Chromosome IV of *Sb* genomes.

could trace the presence of *FLO1* gene in our *Sb* genomes. *FLO8*, *FLO10*, *FLO11*, *FIG2*, and *AGA1* were also present in all strains of *Sb*. *FLO1*, *FLO5*, *FLO8*, *FLO9*, *FLO10*, and *FLO11* are telomeric genes and have repeats^{59, 60}. These were searched in other *Sc* genomes, but we found that most of them encode truncated proteins possibly owing to limitations of sequencing technology. In *Sb* genomes, we found that these genes encode either truncated proteins or could not be traced in *Sb-EDRL*, *Sb-kirkman*, *Sb-unisankyo*, *Sb-17*, and *Sb-MYA-796*. The complete sequences of flocculins from *Sc* and *Sb* strains were retrieved and were aligned using MUSCLE⁶¹ and were edited manually to trace variation in the number and period of the repeats in sequences (Supplementary Figure 5). Seven copies of 45 residues were identified in *FLO1* protein of *Sb* and *Sc* YJM1385 (fruit borne) and *Sc* YJM1129 (brewery strain) whereas only one copy of this period was present in other *Sc* strains. The *FLO8* protein of *Sb* was identical to that in most of the *Sc* strains except *Sc* strains S288C, BY4741, BY4742, FY1679, JK9-3d, SEY6210, W303, X2180-1A and YPH499 where a point mutation has resulted in the truncation of the protein⁶². *FLO10* and *FLO11* domains are present in all strains of *Sb* with repeats where the copy number and a period length of the repeats were similar to all the *Sc* strains.

Introgression of *Zygosaccharomyces bailii* proteins into *Sb* and *Sc* wine strains. Ten genes on chromosome IV of *Sb-biocodex* and *Sb-unique28* were found to encode proteins >90% identical to *Zygosaccharomyces bailii* ISA1307 proteins (Fig. 3). On investigating this region, we found that five genes of *Z. bailii* had introgressed and further undergone duplication. Three of these genes were annotated as encoding uncharacterized proteins; one encodes a probable 5-oxoprolinase, and one is an allantoin transporter. These genes were searched in the sequenced *Sc* strains too, where a single copy of these genes was present in similar fashion in *Sc* UFMG A-905 strain with probiotic properties and wine strains YJM339, RM11-1a, L1528, and YS9. Three copies of these genes were also present in *Sc* wine strain BC187. Few of these introgressed genes were also present in bioethanol producing strain *Sc* JAY291; wine strains *Sc* EC1118, Vin13, VL3, AWRI796 in single copy and *Sc* wine strain LalvinQA23 has two copies of these *Z. bailii* genes.

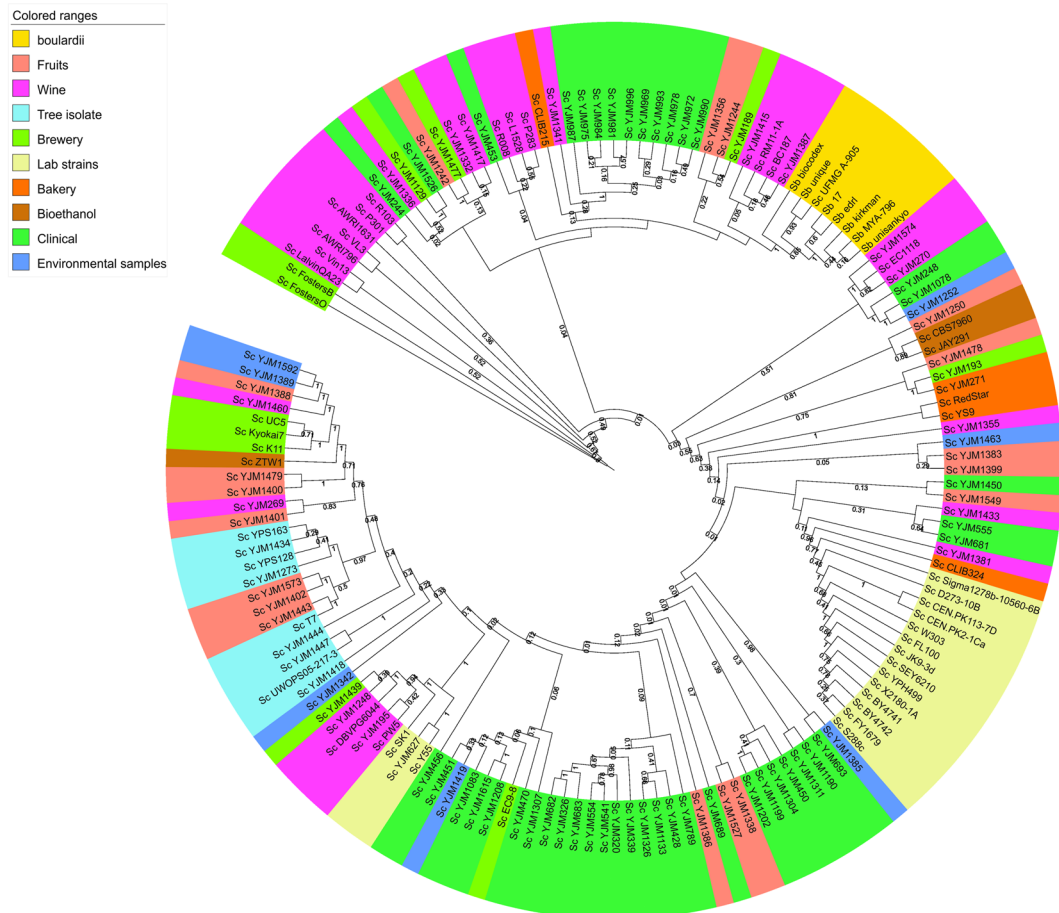


Figure 4. Maximum Evolution tree for 182 orthologous proteins of all strains of *Sb*, *Sc* and *S. kudriavzevii* as outgroup species. The taxa are shaded based on the isolation source of strains as boulardii: gold; Fruits: lightsalmon; Wine: hotpink; Tree isolate: skyblue; Brewery: lawngreen; Lab strains: palegoldenrod; Bakery: darkorange; Bioethanol: peru; Clinical: mediumseagreen; Environmental samples: cornflowerblue.

Taxonomic position of *Sb*. *Sb* and *Sc* shares more than 99% genomic relatedness as determined by Average Nucleotide identity (ANI) (Supplementary File VIII). Thus we retrieved the core proteome from all 145 *Sc* strains and 7 *Sb* strains and a outgroup species *S. kudriavzevii*. 182 proteins for which orthologs could be obtained in all 153 organisms were retrieved and concatenated to find the taxonomic position of *Sb* with comparison to *Sc*. The ML based tree rooted the tree at *S. kudriavzevii*, the outgroup species, which further clustered all the *Sc* strains (Fig. 4). The *Sc* strains in the phylogenetic tree were grouped as per their isolation source. All the *Sb* strains were grouped in a clade, where *Sc* UFMG A-905 strain is closer to *Sb-unique28*. *Sb-biocodex* was present at the root of the *Sb* clade that groups *Sb-17*, *Sb-EDRL*, *Sb-unisankyo*, *Sb-kirkman*, and *Sb-MYA-796*. The *Sb* strains share the clade with wine strains *Sc* strains BC187, YJM1387, YJM1417, YJM1332 and R008, brewery strains *Sc* YJM1477 and *Sc* strain YJM1242 isolated from fruits. Separate clusters were observed in case of laboratory strains whereas clinical isolates were grouped into three distinct clusters in the tree. *Sc* strains isolated from fruits were scattered across tree but were closer to *Sc* wine strains or tree isolates. In terms of taxonomy, it is clear that the *Sb* strains are closely related to the *Sc* wine strains, and it would be of interest to explore the probiotic potential of these wine strains.

Discussion

The complete genomes of *Sb* determined in this study are the best assemblies of the yeast known, as the long PacBio reads used in the study assisted in the identification of complete chromosomes, telomeres and complete structures of Ty elements which could not be identified using Illumina HiSeq short read data. The comparative genomic hybridization analysis for *Sb* revealed at the first place that the Ty1-2 elements were absent from *Sb*^{18,41} which could not be identified in *Sb-EDRL* and *Sb-unisankyo* draft assemblies but in complete genomes of *Sb-biocodex* and *Sb-unique-28*, we were able to identify 15 complete Ty elements. Chromosome IX has been mentioned to exhibit trisomy¹, but such event could not be traced instead we found that the chromosome XII had double the read coverage (400x) as compared to the other chromosomes (200–270x) which could be owing to the aneuploidy of the chromosome XII. The *rep2* gene of 2-micron circle plasmid had G₅₂₉₇->A synonymous mutation resulting in A₂₉₆->V amino acid change as reported earlier⁶³. *Sb* strains share ~3100 orthologous proteins with one or more *Sc* strains suggesting large conserved protein repertoire between *Sc* and *Sb*. The *Sb* proteome

had high conservation level among the strains of *Sb* owing to the presence of the protein homologs in one or the other strains of *Sb*. *Sb* strains are likely to be homothallic and diploid due to the presence of both *MAT* alleles and *HO* gene. Being diploid the yeast should sporulate, but the sporulation was absent in *Sb* as revealed by sporulation assay which has already been established in a previous study¹. The respiration efficiency of *Sb* strains was determined by its growth on non-fermentable carbon sources⁴⁷ and hence the non-sporulation behavior of the organism is not owing to the respiration deficiency. Even the absence or divergence of any sporulation gene was not observed in *Sb* proteome in comparison to that of *Sc*. Possibly the deficiency of the mating type genes⁴⁸ in *Sb* could be leading to the non-sporulation behavior.

Sb can prevent antibiotic-associated diarrhea, recurrent *Clostridium difficile*-associated diarrhea and colitis, Traveller's diarrhea, acute bacterial and viral diarrhea, anti-inflammatory bowel diseases by various mechanisms¹⁰. *Sb*, with antimicrobial properties, secretes 54 kDa^{12,13}, 63 kDa¹¹, and 120 kDa⁶⁴ proteins that exhibit the protection of gut microflora against pathogenic bacteria either through cleaving the toxin or by reducing the cAMP level. These proteins were searched in *Sb* and *Sc*, and it was established that these proteins are not unique to *Sb* but were present in *Sc* too³⁸ (Supplementary Methods). *Sb* has been shown to hinder the biofilm formation by pathogenic strains because of steric hindrance caused by its larger size as compared to bacteria⁶⁵. Also, *Sb* exerts its antimicrobial effect by adhering to intestinal mucus membrane and eliminating pathogens by flow preventing their adhesion to the intestine⁸. Adhesion to other foreign surfaces has been reported to be a critical step for pathogenic as well as a probiotic organism⁵⁷. Yeast also chooses its lifestyle according to its environment and can form different colonies such as non-adhesive colonies, self-adhesive non-dissolvable colonies, biofilms during foreign adhesion or flocs or flor⁵⁶. All flocculin genes harbor a large number of repeats that tend to increase or decrease in copy numbers affecting the degree of flocculation and sensitivity to stress conditions⁵⁶. We could identify complete flocculin genes in the whole genomes of *Sb* where the repeats and their copies were varying even within *Sb* genomes but were consistent in between *Sb-biocodex* and *Sb-unique28*. *Sb* harbors all the flocculin genes required for protection against environmental stress as ethanol and fungicides (*FLO1*)⁶⁶, floc formation (*FLO10* and *FLO11*)⁶⁷, and biofilm formation (*FLO10*, *FLO11*, and *FIG2*)^{67,68}. *FLO8* gene required for expression of *FLO1* and *FLO11* was present in *Sb*, but was truncated in laboratory strains, impairing their flocculation and adhesion to the foreign surfaces⁶². These genes are located at telomeres and are highly repetitive, and the maximum number of repeats identified in the strains of *Sb* could be conferring higher adhesive properties to the organism.

During evolution, the *Saccharomyces* yeast can undergo a process of gene duplication, polyploidy, chromosomal rearrangements, interspecific hybridization, and introgression⁶⁹. The process of eukaryote-to-eukaryote gene transfer events and introgression in *Sc* strains have been validated through genetic experiments and certain regions among *Sc* strains have been reported to have similarity with *Z. bailii*, *S. arboricola*, *S. bayanus* and other yeasts^{70,71}. A particular region in wine yeast *Sc* EC1118 has been proposed earlier to be transferred from *Z. bailii* type strain CBS680⁷⁰. Similarly, a specific region of 10 proteins in chromosome IV of *Sb* was syntenic to the five proteins in *Z. bailii* ISA1307 which has got duplicated in *Sb* and few wine strains of *Sc*.

The taxonomic position of *Sb* as a separate species has been controversial^{10,41,72-75}. It was initially considered as a separate species of genus *Saccharomyces*, but CGH analysis characterized it as a strain of *Sc*. The core proteome based phylogeny, obtained from the *Sb*, *Sc*, and *S. kudriavzevii* could resolve the clades. All *Sb* strains got clustered together in a clade along with *Sc* UFMG A-905 strain, a *Sc* strain with probiotic properties where *Sb-unique28* shared the sister clade with *Sc* UFMG A-905. All the phylogenies drawn revealed concordantly that the wine strains of *Sc* are closer to the *Sb*.

There is no doubt that the *Sb* belongs to *Sc* species and is a strain of *Sc*, but the probiotic features of *Sb* make it a yeast with beneficial effects in gastrointestinal disorders^{3,5,7,12,14,24,26,41,76-79}. The genomic perspective of the organism with relevance to its probiotic features was examined in the study where we couldn't identify any specific and unique proteins in *Sb*, since the *Sb* proteins are homologous with one or the other proteins of *Sc* strains. The genomic perspective in this study revealed that the *Sb* probiotic strains are closer to wine strains of *Sc* than industrial or baking strains; as revealed by the absence of *ASP3* locus, the introgression of *Z. bailii* proteins and the core proteome based taxonomic placement of probiotic *Sb* and wine *Sc* strains. *Sc* BC187 a wine strain shows maximum similarity with the *Sb* strains and might be explored for its probiotic properties similar to *Sc* strain UFMG A-905^{80,81}.

Methods

Isolation and purification of *Sb* genomic DNA. The lyophilized yeast *Sb* available in the market as probiotics in sachets were used for isolation of source DNA for *Sb-biocodex* (Florastor), *Sb-EDRL* (Dr. Reddy's Laboratories) and *Sb-unique28* (Unique Biotech). Two cultures of *Sb* from Unisankyo Ltd. (Now Sanzyme Ltd.) and Kirkman Labs that were maintained at MTCC (IMTECH, India) since 2003 were obtained for DNA isolation and sequencing. DNA isolation was performed using the ZR Fungal/Bacterial DNA miniprep kit (Zymogen), and purity index was checked as the ratio of OD at 260/280 nm was >1.8 as observed by NanoDropND-1000 spectrophotometer. Detailed isolation and purification methods have been provided in supplementary material.

Genome Sequencing. *Sb-unique28* and *Sb-biocodex* were sequenced using PacBio P6C4 chemistry using eight and nine SMRT cells, respectively. 101-bp paired-end (PE) shotgun data from Illumina HiSeq-1000 high-throughput sequencing technology was also obtained for *Sb-biocodex*. The sequencing depth for PacBio sequencing was approximately 200x for both *Sb-biocodex* and *Sb-unique28*. *Sb-EDRL* was sequenced earlier using Roche 454 sequencing technology which now has also been sequenced on Illumina HiSeq 1000 platform to obtain 101-bp PE shotgun data twice and one 2 K and one 8 K mate-pair (MP) library shotgun data. *Sb-kirkman* was sequenced using Illumina HiSeq1000 to get 101-bp PE shotgun data along with 2 K and 8 K MP reads. For *Sb-unisankyo*, only 101 bp shotgun data was obtained from Illumina HiSeq 1000 sequencing platforms. The Illumina HiSeq-1000 sequencing was performed at C-CAMP, Bangalore, India and PacBio P6C4 chemistry

SMRT cells sequencing data was obtained from Genome Quebec Centre, McGill University, Canada. Detailed Sequencing protocols have been provided in supplementary material.

Genome Assembly and Annotation. *Sb-biocodex* and *Sb-unique28* SMRT cells were assembled using Hierarchical Genome Assembly Process (HGAP) v2.0⁸² pipeline of the SMRT Portal. The *Sb-biocodex* and *Sb-unique28* contigs were aligned to *Sc* reference strain S288C using Mauve aligner⁸³ and the contigs completely mapping to the chromosomes were submitted to NCBI as complete chromosomes. Some unplaced contigs were obtained which were subjected to BLASTn⁸⁴ against the finalized chromosomes of respective strains and the NT database to find if any contig belongs to a plasmid or mitochondrial genomes.

Sb-EDRL, assembled earlier with Roche 454 shotgun data and submitted to NCBI (ATCS01000000), was further scaffolded and gapfilled using SSPACE v3.1⁸⁵ and GapFiller v1.10⁸⁶ with Illumina PE and MP data. *Sb-kirkman* and *Sb-unisankyo* were assembled *de novo* using SPAdes v3.1⁸⁷. Further, these assemblies were scaffolded using SSPACE v3.0 and gapfilled using GapFiller v1.10. Also, the Illumina Next-Generation Sequencing (NGS) data for four *Sb*: *Sb-biocodex*, *Sb-EDRL*, *Sb-kirkman* and *Sb-unisankyo* were mapped on to the 16 chromosomes of *Sb-biocodex* to find the gaps that were not covered by Illumina reads. Final assemblies were submitted to GenBank under accession numbers *Sb-biocodex*: LIIL01, *Sb-unique28*: LIOO01, *Sb-kirkman*: LOMX01, *Sb-unisankyo*: LNQF01 and *Sb-EDRL*: ATCS02.

All the *Sb* strains were annotated using Augustus⁸⁸ as gene predictor with species model 'Sc S288C' and tRNA was predicted by tRNAscan-SE 1.23⁸⁹. Features thus annotated were subjected to BLASTp against Saccharomyces Genome Database (SGD)³⁹ protein dataset and non-redundant (NR) protein sequence database for functional characterization of the proteins with an E-value cutoff of $1e^{-5}$.

Gene copy number variations. The complete set of *Sc* genes present in yeastmine database⁴⁹ were mapped by the Illumina shotgun reads of *Sb* strains *Sb-biocodex*, *Sb-kirkman*, *Sb-unisankyo*, and *Sb-EDRL*. Genes with no read coverage were checked in the complete PacBio genome assembly of *Sb-biocodex* for their absence. The genes duplicated within *Sb-biocodex* and *Sb-unique28* genomes with >90% identity, and >90% query coverage were identified.

Presence-Absence Variations (PAVs). The proteins or genes involved in adhesion, flocculins, sporulation, meiotic, mitotic, galactose utilization and palatinose utilization were downloaded from SGD and were subjected to BLASTp or BLASTx against the proteome of all *Sb* and *Sc* and the hits thus obtained were filtered at >50% query coverage and >30% identity. The duplicate hits were removed, and the PAVs were plotted as matrix across all 7 *Sb* and 145 *Sc* genomes.

Genome datasets used for comparison. The *Sb* genomes were compared to 145 strains of *Sc* reported at SGD and NCBI. The annotations for all strains of *Sc* were obtained from SGD and NCBI. NCBI reports 168 genomes as on 01-May-2015 of which 50 were reported in SGD. Five strains (*Sc* FL100, *Sc* RM11-1a, *Sc* Sigma1278b, *Sc* W303 and *Sc* YPS163) were excluded as their updated versions were considered for the analysis. Two strains (*Sc* CLIB382 and *Sc* M22) were excluded from annotations as the number of scaffolds was more than 6000. *Sc* strain T73 and Y10 were also eliminated as the number of annotated features was less than 3000. Fourteen strains with no annotations available at NCBI and SGD were not used for the comparative analysis. Unannotated *Sc* strain UFMG A-905 available from NCBI was utilized in the comparative study as the strain has been mentioned to exhibit probiotic effects. Two *Sb-17* and *Sb-ATCC-MYA-796* were also used for the comparison. The features prediction was made using Augustus⁸⁸ for these strains. All these strains were divided into nine subgroups based on the origin of the strains as Fruit-derived, Wine, Tree-isolates, Beer, Laboratory, Bioethanol, Bakery, Clinical and Environmental.

Identification of Ty elements. Retrotransposons sequences downloaded from SGD database were subjected to BLASTn against the *Sb* genomes. The BLAST results obtained were further filtered with query coverage of 90% and best hits were retrieved. Further, the matched regions were screened manually.

Core-proteome Analysis. The orthologous pairs of proteins across all *Sb* and *Sc* proteomes were extracted using Proteinortho v2.3⁹⁰ PERL script and was considered as the core proteome of all *Sb* strains. The homologs of proteins of one *Sb* strain were identified by subjecting the proteins to BLASTp against proteome of other *Sb* strains.

Taxonomy. The core proteins retrieved from proteome of *Sb* and *Sc* and outgroup species *S. kudriavzevii* IFO 1802 were concatenated and aligned using MAFFT⁹¹. The alignment was fed to MEGA v6.0⁹² for generation of Minimum Evolution (ME) tree at 100 bootstrap values. The evolutionary distances were calculated using Dayhoff matrix based method.

Data access. This Whole Genome Shotgun project has been deposited at DDBJ/EMBL/GenBank along with the plasmid and ITS sequences retrieved from all *Sb*. The version described in this paper is the first version for *Sb-biocodex*- LIIL01000000, *Sb-unique28*- LIOO01000000, *Sb-kirkman*- LOMX01000000, *Sb-unisankyo*- LNQF01000000 and second version of *Sb-EDRL*- ATCS02000000.

Conclusions

Two complete genomes and three draft genomes of *Sb* were sequenced and assembled. The complete genomes revealed the presence of Ty2 elements and gag-co-pol genes in *Sb* unlike the complete absence of Ty1/2 elements in *Sb* suggested previously. Homozygous diploid probiotic yeast *Sb* had non-sporulation phenotype for which the

absence/divergence of sporulation genes or respiration efficiency is not responsible, but the deficiency in mating genes may be playing a role.

The physiological and molecular differences making *Sb* different from *Sc* were explored through genome analysis. We found that the *HXT11* and *HXT9* hexose transporter genes were absent in *Sb* only but were present in all *Sc* strains. The asparagine utilization (*ASP3*) locus was absent in *Sb* and *Sc* wine and distillery strains and were only present in *Sc* laboratory and bioethanol and some clinical strains. All flocculins except FLO5 protein and adhesins were present across all *Sb*, and we found that these have a larger number of repeats comparable to most of the *Sc* strains probably enabling adhesive properties to the organism.

The introgression of five genes, related to transporters, from *Z. bailii* ISA1307 to *Sb* was found to be present in similar fashion in *Sc* wine strains too. The taxonomic position of *Sb* was derived using 182 core proteins as the high genomic relatedness between *Sb* and *Sc* did not allow a few molecular markers to resolve the phylogeny. Taxonomically the probiotic yeast shares the clade with *Sc* UFMG A-905 and wine strains. In light of the genomic and taxonomic evidence we found that the probiotic yeast is closer to the wine strains as compared to industrial strains.

References

- Fietto, J. L. *et al.* Molecular and physiological comparisons between *Saccharomyces cerevisiae* and *Saccharomyces boulardii*. *Canadian journal of microbiology* **50**, 615–621, doi:10.1139/w04-050 (2004).
- Berg, R., Bernasconi, P., Fowler, D. & Gautreaux, M. Inhibition of *Candida albicans* translocation from the gastrointestinal tract of mice by oral administration of *Saccharomyces boulardii*. *J Infect Dis* **168**, 1314–1318 (1993).
- Dahan, S. *et al.* *Saccharomyces boulardii* interferes with Enterohaemorrhagic *Escherichia coli*-induced signalling pathways in T84 cells. *Infection and Immunity* **71**, 766–773 (2003).
- Tasteyre, A., Barc, M., Karjalainen, T., Bourlioux, P. & Collignon, A. Inhibition of *in vitro* cell adherence of *Clostridium difficile* by *Saccharomyces boulardii*. *Microb Pathog* **32**, 219–225 (2002).
- Zbar, N. S., Nashi, L. F. & Saleh, S. M. *Saccharomyces boulardii* as effective probiotic against *Shigella flexneri* in mice. *International Journal of Materials, Methods and Technologies* **1**, 17–21 (2013).
- Kelesidis, T. & Pothoulakis, C. Efficacy and safety of the probiotic *Saccharomyces boulardii* for the prevention and therapy of gastrointestinal disorders. *Therapeutic advances in gastroenterology* **5**, 111–125, doi:10.1177/1756283X11428502 (2012).
- Zaouche, A. *et al.* Effects of oral *Saccharomyces boulardii* on bacterial overgrowth, translocation, and intestinal adaptation after small-bowel resection in rats. *Scandinavian journal of gastroenterology* **35**, 160–165 (2000).
- Tiago, F. C. *et al.* Adhesion to the yeast cell surface as a mechanism for trapping pathogenic bacteria by *Saccharomyces* probiotics. *Journal of medical microbiology* **61**, 1194–1207, doi:10.1099/jmm.0.042283-0 (2012).
- Gedek, B. R. Adherence of *Escherichia coli* serogroup O 157 and the *Salmonella typhimurium* mutant DT 104 to the surface of *Saccharomyces boulardii*. *Mycoses* **42**, 261–264 (1999).
- Lukaszewicz, M. In *Probiotics* (ed Everlon, Rigobelo) Ch. 16, (InTech, 2012).
- Buts, J.-P. *et al.* *Saccharomyces boulardii* produces in rat small intestine a novel protein phosphatase that inhibits *Escherichia coli* endotoxin by dephosphorylation. *Pediatric Research* **60**, 24–29 (2006).
- Castagliuolo, I., LaMont, J. T., Nikulasson, S. T. & Pothoulakis, C. *Saccharomyces boulardii* protease inhibits *Clostridium difficile* toxin A effects in the rat ileum. *Infect Immun* **64**, 5225–5232 (1996).
- Castagliuolo, I., Riegler, M. F., Valenick, L., Lamont, J. T. & Pothoulakis, C. *Saccharomyces boulardii* protease inhibits the effects of *Clostridium difficile* Toxins A and B in Human Colonic Mucosa. *Infection and Immunity* **67**, 302–307 (1998).
- Czerucka, D., Roux, I. & Rampal, P. *Saccharomyces boulardii* inhibits secretagogue-mediated adenosine 3',5'-cyclic monophosphate induction in intestinal cells. *Gastroenterology* **106**, 65–72 (1994).
- van der Aa Kuhle, A. & Jespersen, L. The taxonomic position of *Saccharomyces boulardii* as evaluated by sequence analysis of the D1/D2 domain of 26S rDNA, the ITS1–5.8S rDNA–ITS2 region and the mitochondrial cytochrome-c oxidase II gene. *Systematic and applied microbiology* **26**, 564–571, doi:10.1078/072320203770865873 (2003).
- McFarland, L. V. *Saccharomyces boulardii* is not *Saccharomyces cerevisiae*. *Clinical infectious diseases: an official publication of the Infectious Diseases Society of America* **22**, 200–201 (1996).
- Mitterdorfer, G., Mayer, H. K., Kneife, W. & Viernstein, H. Clustering of *Saccharomyces boulardii* strains within the species *S. cerevisiae* using molecular typing techniques. *Journal of Applied Microbiology* **93**, 521–530 (2002).
- Edwards-Ingram, L. C. *et al.* Comparative genomic hybridization provides new insights into the molecular taxonomy of the *Saccharomyces sensu stricto* complex. *Genome research* **14**, 1043–1051, doi:10.1101/gr.2114704 (2004).
- Dubois, E., Jacobs, E. & Jauniaux, J. C. Expression of the ROAM mutations in *Saccharomyces cerevisiae*: involvement of trans-acting regulatory elements and relation with the Ty1 transcription. *The EMBO journal* **1**, 1133–1139 (1982).
- Warmington, J. R. & Oliver, S. G. A downstream activator sequence regulates the expression of the yeast transposon. *Current genetics* **13**, 357–362 (1988).
- Warmington, J. R. *et al.* A 'hot-spot' for Ty transposition on the left arm of yeast chromosome III. *Nucleic acids research* **14**, 3475–3485 (1986).
- Warmington, J. R., Green, R. P., Newlon, C. S. & Oliver, S. G. Polymorphisms on the right arm of yeast chromosome III associated with Ty transposition and recombination events. *Nucleic acids research* **15**, 8963–8982 (1987).
- Hublot, B. & Levy, R. H. Method for preventing or treating pseudo-membranous colitis (1986).
- Line, J. E., Stern, N. J., Bailey, J. S. & Cox, N. A. *Saccharomyces boulardii* treatment to diminish *Campylobacter* and *Salmonella* in poultry (2000).
- Girard, P., Guern, M.-E. L., Verleye, M. & Hublot, B. Medicinal Product for Treating Gastrointestinal Ulcers (2011).
- Hublot, B., Groux, H., Bernasconi, P., Levy, R. & Le, G. M.-E. Treating weight loss in patients suffering from inflammatory bowel diseases with *S. boulardii* (2014).
- Plein, K. & Hotz, J. Therapeutic effects of *Saccharomyces boulardii* on mild residual symptoms in a stable phase of Crohn's disease with special respect to chronic diarrhoea—a pilot study. *Zeitschrift für Gastroenterologie* **31**, 129–134 (1993).
- Riaz, M., Alam, S., Malik, A. & Ali, S. M. Efficacy and safety of *Saccharomyces boulardii* in acute childhood diarrhoea: a double blind randomised controlled trial. *Indian journal of pediatrics* **79**, 478–482, doi:10.1007/s12098-011-0573-z (2012).
- Sudha, M. R. Safety assessment studies of probiotic *Saccharomyces boulardii* strain Unique 28 in Sprague-Dawley rats. *Beneficial microbes* **2**, 221–227, doi:10.3920/BM2011.0013 (2011).
- Sudha, M. R., Bhonagiri, S. & Kumar, M. A. Oral consumption of potential probiotic *Saccharomyces boulardii* strain Unique 28 in patients with acute diarrhoea: a clinical report. *Beneficial microbes* **3**, 145–150, doi:10.3920/BM2011.0055 (2012).
- Surawicz, C. M., McFarland, L. V., Elmer, G. & Chinn, J. Treatment of recurrent *Clostridium difficile* colitis with vancomycin and *Saccharomyces boulardii*. *The American journal of gastroenterology* **84**, 1285–1287 (1989).
- Villar-Garcia, J. *et al.* Effect of probiotics (*Saccharomyces boulardii*) on microbial translocation and inflammation in HIV-treated patients: a double-blind, randomized, placebo-controlled trial. *Journal of acquired immune deficiency syndromes* **68**, 256–263, doi:10.1097/QAI.000000000000468 (2015).

33. Wu, D., Teng, D., Wang, X., Dai, C. & Wang, J. *Saccharomyces boulardii* prevention of the hepatic injury induced by *Salmonella* Enteritidis infection. *Canadian journal of microbiology* **60**, 681–686, doi:10.1139/cjm-2014-0259 (2014).
34. Bassetti, S., Frei, R. & Zimmerli, W. Fungemia with *Saccharomyces cerevisiae* after treatment with *Saccharomyces boulardii*. *The American journal of medicine* **105**, 71–72 (1998).
35. Cassone, M. *et al.* Outbreak of *Saccharomyces cerevisiae* subtype *boulardii* fungemia in patients neighboring those treated with a probiotic preparation of the organism. *Journal of clinical microbiology* **41**, 5340–5343 (2003).
36. Graf, C. & Gavazzi, G. *Saccharomyces cerevisiae* fungemia in an immunocompromised patient not treated with *Saccharomyces boulardii* preparation. *The Journal of infection* **54**, 310–311, doi:10.1016/j.jinf.2006.01.019 (2007).
37. Lherm, T. *et al.* Seven cases of fungemia with *Saccharomyces boulardii* in critically ill patients. *Intensive care medicine* **28**, 797–801, doi:10.1007/s00134-002-1267-9 (2002).
38. Khatri, I. *et al.* Gleaning evolutionary insights from the genome sequence of a probiotic yeast *Saccharomyces boulardii*. *Gut pathogens* **5**, 30, doi:10.1186/1757-4749-5930 (2013).
39. Cherry, J. M. *et al.* SGD: *Saccharomyces* Genome Database. *Nucleic acids research* **26**, 73–79 (1998).
40. Haber, J. E. Mating-type genes and MAT switching in *Saccharomyces cerevisiae*. *Genetics* **191**, 33–64, doi:10.1534/genetics.111.134577 (2012).
41. Edwards-Ingram, L. *et al.* Genotypic and physiological characterization of *Saccharomyces boulardii*, the probiotic strain of *Saccharomyces cerevisiae*. *Applied and environmental microbiology* **73**, 2458–2467, doi:10.1128/AEM.02201-06 (2007).
42. Wang, T. *et al.* The establishment of *Saccharomyces boulardii* surface display system using a single expression vector. *Fungal genetics and biology: FG & B* **64**, 1–10, doi:10.1016/j.fgb.2013.11.006 (2014).
43. Nasmyth, K. A. Molecular genetics of yeast mating type. *Annual review of genetics* **16**, 439–500, doi:10.1146/annurev.ge.16.120182.002255 (1982).
44. Ekino, K., Kwon, I., Goto, M., Yoshino, S. & Furukawa, K. Functional analysis of HO gene in delayed homothallism in *Saccharomyces cerevisiae* wy2. *Yeast* **15**, 451–458, doi:10.1002/(SICI)1097-0061(199904)15:6<451::AID-YEA383>3.0.CO;2-K (1999).
45. Meiron, H., Nahon, E. & Raveh, D. Identification of the heterothallic mutation in HO-endonuclease of *S. cerevisiae* using HO/hochimeric genes. *Current genetics* **28**, 367–373 (1995).
46. Kurtzman, C., Fell, J. W., Boekhout, T. & Robert, V. *The Yeasts: A Taxonomic Study*. 5 edn, Vol. 2 (Elsevier, 2011).
47. Jambhekar, A. & Amon, A. Control of meiosis by respiration. *Current biology: CB* **18**, 969–975, doi:10.1016/j.cub.2008.05.047 (2008).
48. Tsuboi, M. & Takahashi, T. Genetic analysis of the non-sporulating phenotype of Brewer's yeast. *Journal of Fermentation Technology* **66**, 605–613 (1988).
49. Balakrishnan, R. *et al.* YeastMine—an integrated data warehouse for *Saccharomyces cerevisiae* data as a multipurpose tool-kit. *Database: the journal of biological databases and curation* **2012**, bar062, doi:10.1093/database/bar062 (2012).
50. Boles, E. & Hollenberg, C. P. The molecular genetics of hexose transport in yeasts. *FEMS microbiology reviews* **21**, 85–111 (1997).
51. Nourani, A., Wesolowski-Louvel, M., Delaveau, T., Jacq, C. & Delahodde, A. Multiple-drug-resistance phenomenon in the yeast *Saccharomyces cerevisiae*: involvement of two hexose transporters. *Molecular and cellular biology* **17**, 5453–5460 (1997).
52. Landry, C. R., Oh, J., Hartl, D. L. & Cavalieri, D. Genome-wide scan reveals that genetic variation for transcriptional plasticity in yeast is biased towards multi-copy and dispensable genes. *Gene* **366**, 343–351, doi:10.1016/j.gene.2005.10.042 (2006).
53. Hyle, J. W., Shaw, R. J. & Reines, D. Functional distinctions between IMP dehydrogenase genes in providing mycophenolate resistance and guanine prototrophy to yeast. *The Journal of biological chemistry* **278**, 28470–28478, doi:10.1074/jbc.M303736200 (2003).
54. Mitsui, K. *et al.* A novel membrane protein capable of binding the Na⁺/H⁺ antiporter (Nha1p) enhances the salinity-resistant cell growth of *Saccharomyces cerevisiae*. *The Journal of biological chemistry* **279**, 12438–12447, doi:10.1074/jbc.M310806200 (2004).
55. Bolton, E. C. & Boeke, J. D. Transcriptional interactions between yeast tRNA genes, flanking genes and Ty elements: a genomic point of view. *Genome research* **13**, 254–263, doi:10.1101/gr.612203 (2003).
56. Bruckner, S. & Mosch, H. U. Choosing the right lifestyle: adhesion and development in *Saccharomyces cerevisiae*. *FEMS microbiology reviews* **36**, 25–58, doi:10.1111/j.1574-6976.2011.00275.x (2012).
57. Verstrepen, K. J. & Klis, F. M. Flocculation, adhesion and biofilm formation in yeasts. *Molecular microbiology* **60**, 5–15, doi:10.1111/j.1365-2958.2006.05072.x (2006).
58. Hodgson, J. A., Berry, D. R. & Johnston, J. R. Discrimination by heat and proteinase treatments between flocculent phenotypes conferred on *Saccharomyces cerevisiae* by the genes FLO1 and FLO5. *Journal of general microbiology* **131**, 3219–3227, doi:10.1099/00221287-131-12-3219 (1985).
59. Goossens, K. & Willaert, R. Flocculation protein structure and cell-cell adhesion mechanism in *Saccharomyces cerevisiae*. *Biotechnology letters* **32**, 1571–1585, doi:10.1007/s10529-010-0352-3 (2010).
60. Soares, E. V. Flocculation in *Saccharomyces cerevisiae*: a review. *J Appl Microbiol* **110**, 1–18, doi:10.1111/j.1365-2672.2010.04897.x (2011).
61. Edgar, R. C. MUSCLE: multiple sequence alignment with high accuracy and high throughput. *Nucleic acids research* **32**, 1792–1797, doi:10.1093/nar/gkh340 (2004).
62. Liu, H., Styles, C. A. & Fink, G. R. *Saccharomyces cerevisiae* S288C has a mutation in FLO8, a gene required for filamentous growth. *Genetics* **144**, 967–978 (1996).
63. Douradinha, B. *et al.* Novel insights in genetic transformation of the probiotic yeast *Saccharomyces boulardii*. *Bioengineered* **5**, 21–29, doi:10.4161/bioe.26271 (2014).
64. Brandao, R. L. *et al.* Intracellular signal triggered by cholera toxin in *Saccharomyces boulardii* and *Saccharomyces cerevisiae*. *Applied and environmental microbiology* **64**, 564–568 (1998).
65. Pontier-Bres, R. *et al.* Modification of *Salmonella Typhimurium* motility by the probiotic yeast strain *Saccharomyces boulardii*. *PLoS one* **7**, e33796, doi:10.1371/journal.pone.0033796 (2012).
66. Smukalla, S. *et al.* FLO1 is a variable green beard gene that drives biofilm-like cooperation in budding yeast. *Cell* **135**, 726–737, doi:10.1016/j.cell.2008.09.037 (2008).
67. Guo, B., Styles, C. A., Feng, Q. & Fink, G. R. A *Saccharomyces* gene family involved in invasive growth, cell-cell adhesion, and mating. *Proceedings of the National Academy of Sciences of the United States of America* **97**, 12158–12163, doi:10.1073/pnas.220420397 (2000).
68. Lo, W. S. & Dranginis, A. M. The cell surface flocculin Flo11 is required for pseudohyphae formation and invasion by *Saccharomyces cerevisiae*. *Molecular biology of the cell* **9**, 161–171 (1998).
69. Barrio, E., Gonza, I. S., Arias, A., Belloch, C. & Querol, A. In *The Yeast Handbook* Vol. 2 (eds A., Querol & G. H., Fleet) 153–174 (Springer, Berlin, 2006).
70. Novo, M. *et al.* Eukaryote-to-eukaryote gene transfer events revealed by the genome sequence of the wine yeast *Saccharomyces cerevisiae* EC1118. *Proceedings of the National Academy of Sciences of the United States of America* **106**, 16333–16338, doi:10.1073/pnas.0904673106 (2009).
71. Strobe, P. K. *et al.* The 100-genomes strains, an *S. cerevisiae* resource that illuminates its natural phenotypic and genotypic variation and emergence as an opportunistic pathogen. *Genome research* **25**, 762–774, doi:10.1101/gr.185538.114 (2015).
72. Cardinali, G. & Martini, A. Electrophoretic karyotypes of authentic strains of the sensu stricto group of genus *Saccharomyces*. *Int J Syst Bacteriol* **44**, 791–797 (1994).

73. Czerucka, D., Piche, T. & Rampal, P. Review article: yeast as probiotics – *Saccharomyces boulardii*. *Alimentary pharmacology & therapeutics* **26**, 767–778, doi:10.1111/j.1365-2036.2007.03442.x (2007).
74. Posteraro, B. *et al.* Molecular tools for differentiating probiotic and clinical strains of *Saccharomyces cerevisiae*. *International journal of food microbiology* **103**, 295–304, doi:10.1016/j.ijfoodmicro.2004.12.031 (2005).
75. van den Brink, J. *et al.* Energetic limits to metabolic flexibility: responses of *Saccharomyces cerevisiae* to glucose-galactose transitions. *Microbiology* **155**, 1340–1350, doi:10.1099/mic.0.025775-0 (2009).
76. Buts, J. P., Bernasconi, P., Vaerman, J. P. & Dive, C. Stimulation of secretory IgA and secretory component of immunoglobulins in small intestine of rats treated with *Saccharomyces boulardii*. *Digestive diseases and sciences* **35**, 251–256 (1990).
77. Buts, J. P., De Keyser, N. & De Raedemaeker, L. *Saccharomyces boulardii* enhances rat intestinal enzyme expression by endoluminal release of polyamines. *Pediatr Res* **36**, 522–527, doi:10.1203/00006450-199410000-00019 (1994).
78. Diezmann, S. & Dietrich, F. S. Oxidative stress survival in a clinical *Saccharomyces cerevisiae* isolate is influenced by a major quantitative trait nucleotide. *Genetics* **188**, 709–722, doi:10.1534/genetics.111.128256 (2011).
79. Fuller, R. Probiotics in man and animals. *Journal of Applied Bacteriology* **66**, 365–378 (1989).
80. Martins, F. S. *et al.* Oral treatment with *Saccharomyces cerevisiae* strain UFMG 905 modulates immune responses and interferes with signal pathways involved in the activation of inflammation in a murine model of typhoid fever. *International journal of medical microbiology: IJMM* **301**, 359–364, doi:10.1016/j.ijmm.2010.11.002 (2011).
81. Tiago, F. C. *et al.* Effect of *Saccharomyces cerevisiae* strain UFMG A-905 in experimental model of inflammatory bowel disease. *Beneficial microbes* **6**, 807–815, doi:10.3920/BM2015.0018 (2015).
82. Chin, C. S. *et al.* Nonhybrid, finished microbial genome assemblies from long-read SMRT sequencing data. *Nature methods* **10**, 563–569, doi:10.1038/nmeth.2474 (2013).
83. Rissman, A. I. *et al.* Reordering contigs of draft genomes using the Mauve aligner. *Bioinformatics* **25**, 2071–2073, doi:10.1093/bioinformatics/btp356 (2009).
84. Altschul, S. F., Gish, W., Miller, W., Myers, E. W. & Lipman, D. J. Basic local alignment search tool. *Journal of molecular biology* **215**, 403–410, doi:10.1016/S0022-2836(05)80360-2 (1990).
85. Boetzer, M., Henkel, C. V., Jansen, H. J., Butler, D. & Pirovano, W. Scaffolding pre-assembled contigs using SSPACE. *Bioinformatics* **27**, 578–579, doi:10.1093/bioinformatics/btq683 (2011).
86. Nadalin, F., Vezzi, F. & Policriti, A. GapFiller: a de novo assembly approach to fill the gap within paired reads. *BMC bioinformatics* **13** Suppl 14, S8, doi:10.1186/1471-2105-13-S14-S8 (2012).
87. Bankevich, A. *et al.* SPAdes: a new genome assembly algorithm and its applications to single-cell sequencing. *Journal of computational biology: a journal of computational molecular cell biology* **19**, 455–477, doi:10.1089/cmb.2012.0021 (2012).
88. Stanke, M. & Morgenstern, B. AUGUSTUS: a web server for gene prediction in eukaryotes that allows user-defined constraints. *Nucleic acids research* **33**, W465–W467, doi:10.1093/nar/gki458 (2005).
89. Schattner, P., Brooks, A. N. & Lowe, T. M. The tRNAscan-SE, snoscan and snoGPS web servers for the detection of tRNAs and snoRNAs. *Nucleic acids research* **33**, W686–W689, doi:10.1093/nar/gki366 (2005).
90. Lechner, M. *et al.* Proteinortho: detection of (co-)orthologs in large-scale analysis. *BMC bioinformatics* **12**, 124, doi:10.1186/1471-2105-12-124 (2011).
91. Katoh, K., Misawa, K., Kuma, K. & Miyata, T. MAFFT: a novel method for rapid multiple sequence alignment based on fast Fourier transform. *Nucleic acids research* **30**, 3059–3066 (2002).
92. Tamura, K., Stecher, G., Peterson, D., Filipinski, A. & Kumar, S. MEGA6: Molecular Evolutionary Genetics Analysis version 6.0. *Molecular biology and evolution* **30**, 2725–2729, doi:10.1093/molbev/mst197 (2013).

Acknowledgements

This work was supported by the Council of Scientific Research (CSIR) Network project on Man as a super-organism: Understanding the Human Microbiome (HUM-CSIR-BSC-0119) and project “Expansion and modernization of Microbial Type Culture Collection and Gene Bank (MTCC) jointly supported by Council of Scientific and Industrial Research (CSIR) Grant No. BSC0402 and Department of Biotechnology (DBT) Govt. of India Grant No. BT/PR7368/INF/22/177/2012”. IK is supported by a research fellowship from the University Grants Commission, and RT is supported by a research fellowship from the Council for Scientific and Industrial Research. We thank the C-CAMP (<http://www.ccamp.res.in/>) and Genome Quebec Centre (<http://www.genomequebec.com/>) next-generation genomics facility for help in obtaining Illumina and PacBio data, respectively. We would like to acknowledge Unique Biotech Ltd. for providing us the *Sb-unique28* sample.

Author Contributions

S.S. conceived the idea; R.T. and G.S.P. isolated genomic DNA and carried out strain identification, I.T.S. sequencing and sporulation phenotype assay. I.K. carried out the quality control, assembly, and annotation of the genomic data and performed the comparative analysis. I.K., S.S. and K.G. performed interpretation of the data. K.G. guided us throughout the experiments and provided insightful comments. I.K., S.S., and R.T. wrote the manuscript. All authors have read and approved the manuscript.

Additional Information

Supplementary information accompanies this paper at doi:10.1038/s41598-017-00414-2

Competing Interests: The authors declare that they have no competing interests.

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



This work is licensed under a Creative Commons Attribution 4.0 International License. The images or other third party material in this article are included in the article's Creative Commons license, unless indicated otherwise in the credit line; if the material is not included under the Creative Commons license, users will need to obtain permission from the license holder to reproduce the material. To view a copy of this license, visit <http://creativecommons.org/licenses/by/4.0/>

© The Author(s) 2017