Assessing genetic diversity in Brettanomyces yeasts using DNA fingerprinting and whole genome sequencing

Running title: Genetic diversity in Brettanomyces

Sam Crauwels¹, Bo Zhu², Jan Steensels³, Pieter Busschaert¹, Gorik De Samblanx¹,³, Kathleen Marchal⁴,⁵, Kris A. Willems¹, Kevin J. Verstrepen²,*, and Bart Lievens¹,*

¹ Laboratory for Process Microbial Ecology and Bioinspirational Management, Cluster for Bioengineering Technology (CBeT), Department of Microbial and Molecular Systems (M2S), Leuven Food Science and Nutrition Research Centre (LFoRCE), Leuven Institute for Beer Research (LIBR), KU Leuven, Campus De Nayer, B-2860 Sint-Katelijne-Waver, Belgium

² VIB Lab for Systems Biology & Centre of Microbial and Plant Genetics (CMPG) Lab for Genetics and Genomics, M2S, LFoRCE, LIBR, KU Leuven, B-3001 Leuven, Belgium

³ Electrical Engineering (ESAT) Technology Cluster, KU Leuven, Campus De Nayer, B-2860 Sint-Katelijne-Waver, Belgium

⁴ Department of Plant Biotechnology and Bioinformatics & Department of Information Technology, IMinds, Ghent University, B-9052 Gent

⁵ CMPG, M2S, KU Leuven, B-3001 Leuven, Belgium

* Both authors equally contributed as corresponding author.

Corresponding authors:
BL: Laboratory for Process Microbial Ecology and Bioinspirational Management, KU Leuven, Campus De Nayer, Fortseesteenweg 30A, B-2860 Sint-Katelijne-Waver, Belgium. Phone: +32 15 305590. Fax: +32 15 305099. E-mail: bart.lievens@kuleuven.be

KV: VIB Lab for Systems Biology & CMPG Lab for Genetics and Genomics, KU Leuven, Bio-Incubator, Gaston Geenslaan 1, B-3001 Leuven, Belgium. Phone: +32 16 751390. Fax: +32 16 751391. E-mail: kevin.verstrepen@biw.vib-kuleuven.be
ABSTRACT

Brettanomyces yeasts, with the species Brettanomyces (Dekkera) bruxellensis as the most important one, are generally reported as spoilage yeasts in the beer and wine industry due to production of phenolic off-flavors. However, B. bruxellensis is also known as a beneficial contributor in certain fermentation processes, such as the production of certain specialty beers. Nevertheless, despite its economic importance, Brettanomyces yeasts have remained poorly understood at the genetic and genomic level so far. In this study, the genetic relationship between more than 50 Brettanomyces strains from all presently known species and isolated from several sources was studied using a combination of DNA fingerprinting techniques. This revealed an intriguing correlation between the B. bruxellensis fingerprints and their respective isolation source. To further explore this relationship, we sequenced a (beneficial) beer isolate of B. bruxellensis (VIB X9085; ST05.12/22) and compared its genome sequence with that of two wine spoilage strains (AWRI 1499 and CBS 2499). ST05.12/22 was found to be substantially different from both wine strains, especially at the level of single nucleotide polymorphisms (SNPs). In addition, there were major differences in the genome structure between the strains investigated, including the presence of large duplications and deletions. Gene content analysis revealed the presence of 20 genes which were present in both wine strains but absent in the beer strain, including many genes involved in carbon and nitrogen metabolism. Vice versa, no genes were found in ST05.12/22 that were missing in both AWRI 1499 and CBS 2499. Together, this study provides tools to discriminate Brettanomyces strains and provides a first glimpse at the genetic diversity and genome plasticity of B. bruxellensis.
Key words: *Brettanomyces bruxellensis*, comparative genomics, DNA fingerprinting, (off-) flavor, intraspecific heterogeneity.
INTRODUCTION

Brettanomyces species, with *B. bruxellensis* as the most important one, are generally reported as spoilage yeasts that produce off-flavors in beer and wine. The aroma characteristics of their spoilage-causing metabolites are typically described as burnt plastic, barnyard, horse sweat and leather, amongst some other unpleasant odors (1-4), resulting in wines and beers that are less-preferred by consumers. Spoilage of wine by *B. bruxellensis* is, in fact, considered the most important microbiological issue in the wine industry (5). However, the same species is a beneficial and even crucial contributor to the production of certain specialty beers such as lambic and gueuze which are typified by flavors generated during secondary fermentation by this yeast (6, 7). Additionally, the species is of increasing relevance for the biofuel industry (8). Apart from isolations from beer and wine, *Brettanomyces* species have been detected and isolated in other foods such as cider, soft drinks, dairy products and olives (9-20).

Despite its economic importance as either a spoilage contaminant in wine and non-alcoholic beverages, or as a vital component of the fermentation biota in the production of certain beers, the physiology and ecology of *Brettanomyces* yeasts have only recently been the subject of intensive research (1, 21-25). However, little is still known about the level of genomic inter-strain variation within *B. bruxellensis*, or within the genus *Brettanomyces*.

*Brettanomyces* currently encompasses five species, including the anamorphs *B. anomalus*, *B. bruxellensis*, *B. custersianus*, *B. naardenensis* and *B. nanus*, with teleomorphs existing for the first two species, *Dekkera anomalana* and *D. bruxellensis* (26). So far, most genetic studies on *Brettanomyces* have focused on rapid fingerprinting using ribosomal RNA (rRNA) sequencing (26, 27), PCR-restriction fragment length polymorphism (RFLP) (28), random amplified polymorphic DNA
(RAPD) (29), amplified fragment length polymorphism (AFLP), arbitrarily primed PCR (AP-PCR) and microsatellite fingerprinting (25, 30). However, so far these studies have been mostly performed for strains belonging to the same species, most often *B. bruxellensis* (17, 20, 29). In most cases, only a limited set of isolates were investigated, or studies were performed on isolates from only a single origin, e.g. from wineries (17, 31, 32). Consequently, this may limit our view on the genetic diversity within this genus. Nevertheless, most of these studies suggest great inter-strain variability within *Brettanomyces*, especially for *B. bruxellensis*.

Recent advances in whole genome sequencing technology have led to an increasing number of completely sequenced microbial genomes providing the opportunity to compare different species or strains of the same species on a genomic scale (33). Woolfit *et al.* reported a partial genome sequence of a *B. bruxellensis* wine contaminant (CBS 2499) and identified approximately 3,000 genes (22). Recently, the full genome sequence of this strain was determined and used to deduce the genetic background of some food-relevant properties and the evolutionary history of this yeast (23). The authors found that this yeast is phylogenetically distant to food-related yeasts like *Saccharomyces* and is most related to *Pichia (Komagataella) pastoris*, which is, a poor ethanol producer, unlike *B. bruxellensis* (23). Additionally, the full genome sequence of another *B. bruxellensis* wine spoilage strain (AWRI 1499) has revealed a triploid genome enriched in genes that may aid survival in the challenging environment of wine (24).

Comparative genomics of four wine isolates, including CBS 2499, AWRI 1499 and two newly sequenced *B. bruxellensis* isolates, revealed differences in nutrient utilization and ploidy level within *B. bruxellensis*, with some strains being diploid and others being triploid (34). Triploid isolates were found to possess a core diploid
genome and a distantly related third genomic complement (34). Further, the authors presented evidence suggesting that this form of triploidy has arisen more than once in the evolutionary history of *B. bruxellensis*, and that it confers a selective advantage for strains from wineries (34). Whereas the sequencing of these wine spoilage *B. bruxellensis* strains has increased our understanding of this species, our knowledge on how these strains behave in comparison with strains from another niche remains fairly limited. For example, no genome sequence is available for a *Brettanomyces* strain from an industry where its presence is desirable, such as the fermentation of Belgian gueuze and lambic beers.

Here, the genetic relationship between 50 *Brettanomyces* strains belonging to all species presently classified within the genus and isolated from several food-related sources was studied using a combination of established fingerprinting techniques. This revealed an intriguing correlation between the *B. bruxellensis* fingerprints and the niches where the respective strains were isolated. Additionally, we sequenced a (beneficial) beer isolate of *B. bruxellensis* (VIB X9085; ST05.12/22) and compared its genome sequence with that of two wine spoilage strains. Emphasis was put on describing single nucleotide polymorphisms (SNPs), small insertions and deletions (InDels), copy number variations (CNVs), and the presence of unique genes.

**MATERIALS AND METHODS**

**Yeast collection, DNA extraction and nitrate assimilation tests.** A total of 50 strains isolated from different food products and beverages, representing the different *Brettanomyces* species, i.e. *B. anomalus* (*D. anomala*), *B. bruxellensis* (*D. bruxellensis*), *B. custersianus*, *B. naardenensis* and *B. nanus*, was used in this study. A subset of these strains was isolated from three lambic casks from the Cantillon...
brewery (Anderlecht, Belgium; February 2012) as described previously (3, 19, 20). Additional strains were obtained from several culture collections or kindly provided by colleagues (Table 1). Obtained beer samples from the Cantillon brewery were diluted ten times in Wort Extract Medium (WEM) (10 ml) and incubated at 21°C for 7 days with vigorous shaking. WEM was prepared by stirring 400 g freshly grounded malt in 500 ml distilled water for 60 min at 65°C. Following filtration (50 µm) and autoclave sterilization the wort was supplemented with 150 mg/l (NH₄)₂SO₄ and 10 g/l glucose. The medium was then depleted of simple fermentable sugars such as glucose, fructose, sucrose, maltose and maltotriose by inoculating it with *Saccharomyces cerevisiae* (to mimic the main lambic wort fermentation period (19)), followed by a 96 h incubation at 25°C. Subsequently, *S. cerevisiae* cells were removed by centrifugation (15 min, 4,000 g, 4°C), and 20 g/l glucose, 10 mg/l cycloheximide, 40 mg/l oxytetracycline and 60 mg/l chloroamphycocil were added to the supernatant, which resulted in WEM suitable for *Brettanomyces* isolation. After growing the *Brettanomyces* cultures in WEM, a tenfold dilution series of each sample was plated (100 µl) in duplo on either Wallerstein Laboratory Nutrient (WLN) agar (20), *Dekkera/Brettanomyces* Differential Medium (DBDM) according to Rodrigues *et al.* (3) but without adding ethanol, and Universal Beer Agar (UBA) supplemented with the antibiotics cycloheximide (10 mg/l), oxytetracycline (10 mg/l) and chloramphenicol (50 mg/l) (19). Plates were incubated for 5-14 days at 25°C. From each countable plate (containing less than 300 colonies) five colonies were randomly selected and further subcultivated to obtain pure cultures. Identification based on partially sequencing the nuclear large subunit (LSU) rRNA gene resulted in five different *Brettanomyces* isolates, all of them belonging to the *B. bruxellensis* species (Table 1). Following incubation of 5 days at 25°C on Yeast Peptone Glucose (YPG)
agar genomic DNA was isolated using the phenol/chloroform extraction method described by Lievens et al. (35). DNA yields were determined spectrophotometrically at 260 nm and diluted to 10 ng/µl. For nitrate assimilation tests, strains were grown on YPG agar for 5 days at 25°C, and then inoculated into 5 ml medium as described by Conterno et al. (1), and supplemented with either 0.1% (w/v) nitrate or 0.1% (w/v) ammonium sulfate as a positive control, or no nitrogen source as a negative control. After 7 days of aerobic incubation under agitation at 25°C, growth of the different strains was evaluated by means of visual inspection. Isolates were stored at -80°C in Yeast extract Peptone Dextrose (YPD) broth containing 26.1% glycerol. DNA extracts were stored at -20°C.

DNA fingerprinting and phylogenetic analysis. DNA extracted from all isolates listed in Table 1 was amplified using the primer pair NL1 and NL4, amplifying the divergent D1/D2 domains of the LSU rRNA gene (8). Amplification was performed in a reaction volume of 20 µl, containing 312.5 μM of each dNTP, 1.0 μM of each primer, 1.25 units TaKaRa Ex Taq polymerase, 1× Ex Taq Buffer (Clontech Laboratories, Palo Alto, CA, USA), and 1 µl genomic DNA. Amplification was performed using a Bio-Rad T100 thermal cycler according to the following thermal profile: initial denaturation at 95°C for 2 min, followed by 30 cycles of 95°C for 1 min, 59°C for 45 s and 72°C for 1 min. A final 10-min extension step at 72°C concluded the protocol. Sequencing was performed using the same primers as for the amplification. Subsequently, SeqTrace software (36) was used to identify, align and compute consensus sequences with the same start and end motifs (457-471 bp) from matching forward and reverse sequences. Remaining ambiguous bases were manually edited according to the paired electropherograms. Blast analysis (37) of the obtained
sequences against GenBank (38) confirmed the identity of the isolates as they were purchased or received. Following alignment of the different consensus sequences a maximum likelihood tree was constructed with Mega 5.2 (39) to assess the phylogenetic relationships of the different isolates. In addition, all DNA samples were subjected to three fingerprinting techniques previously used successfully to type \textit{Brettanomyces} strains, including RAPD-PCR, AP-PCR and rep-PCR. With regard to the RAPD analysis, first 10 decamer oligonucleotides, randomly chosen from the Operon primer kits (Operon Technologies Inc, Alameda, CA, USA), were screened on a subset of ten yeast isolates to select the most discriminative and reliable RAPD primers. Three primers resulting in distinct, reproducible polymorphic bands were selected for further analysis, including OPC20 (5’-ACTTCGCCAC-3’), OPD19 (5’-CTGGGGACTT-3’) and OPK03 (5’-CCAGCTTAGG-3’). Likewise, for the rep-PCR analysis, two primers and one primer set were first tested on a few isolates, including the BOXA1R primer (5’-CTACGGCAAGGCGACGCTGACG-3’), the (GTG)\textsubscript{5} primer (5’-GTGGTGGTGGTGGTGGTGGT-3’) and the primer pair REP1R-1 (5’-IIIICGICGICATCIGGC-3’) and REP2-I (5’-ICGICTTATICGCGCTAC-3’) (40). As the BOXA1R primer yielded only one to three bands and primer (CTG)\textsubscript{5} resulted in some migration problems during gel electrophoresis, only REP1R-I and REP2-I, yielding 2-13 clear bands, was maintained for analysis of the whole collection. For the AP-PCR, the M13 universal primer (5’-TTATGAAACGACGGCCAGT-3’) (41) as well as two other 15-mer primers derived from the microsatellite core-sequence of the wild-type phage M13, including 5’-GAGGGTGCGGTGTCT-3’ and 5’-GAGGGTTGGGCGCTT-3’ (41) were used. All amplifications were performed using a Bio-Rad T100 thermal cycler in a total volume of 20 µl containing 0.5 µM of each primer, 0.15 mM of each dNTP, 1.0 unit Titanium Taq DNA polymerase, 1× Titanium
Taq PCR buffer (Clontech Laboratories, Palo Alto, CA, USA), and 1 µl genomic DNA. Before amplification, DNA samples were denatured at 94°C for 2 min. Subsequently, 35 cycles were run of 1 min at 94°C, 1 min at 35°C (RAPD), 40°C (rep-PCR) or 49°C (AP-PCR), and 2 min at 72°C, with a final extension step for 10 min at 72°C. Obtained PCR products were separated by loading 7.5 µl of the reaction volume on 1.5% agarose gels followed by gel electrophoresis in 1× Tris/acetate EDTA (TAE) buffer at 120 V for 110 min. Gels were stained with ethidium bromide and visualized with UV light. A 1-kb DNA ladder (Smartladder, Eurogentec, Seraing, Belgium) was used as size marker for comparison. The BioChemi System (UVP, Upland, CA, USA) was used to acquire image data. All reactions were performed three times to check reproducibility, and yielded identical results, demonstrating the robustness of our methods. In all analyses, a *S. cerevisiae* isolate (69240, Novagen, USA) was used as a reference. Sterile distilled water was used as a negative control. Obtained images were processed using GelCompar software version 6.6.4 (Applied Maths, Sint-Martens-Latem, Belgium) and analyses were performed on the combined data sets of all fingerprint results obtained in this study. Following normalization and background subtraction, fingerprint similarities were calculated using the Pearson correlation coefficient (42). Cluster analysis was performed by the unweighted pair group method with arithmetic averages (UPGMA) (42). In addition, non-metric multidimensional scaling (nMDS) plots were constructed to create a two-dimensional representation of the relationships among the different isolates (43, 44). Since nMDS ordination is an iterative algorithm that involves a “goodness of fit” estimate, an important component of an nMDS plot is a measure of the goodness of fit of the final plot, also called the “stress” of the plot. A stress value greater than 0.2 indicates that the plot is close to random. Stress less than 0.2 indicates a useful two-dimensional representation and less
than 0.1 corresponds to an ideal ordination with no real prospect of misinterpretation (44). For our analysis, stress was calculated using the R package vegan (45). All ordinations in our analysis were computed following 10,000 random starts. In addition to the graphical representation, it was determined whether significant differences between groups of objects could be observed using analysis of similarity (ANOSIM (46)), based on the earlier obtained distance matrix (47). This non-parametric method compares the average rank similarity between objects within a group with the rank similarity between objects of different groups and produces a test statistic $R$ which can range from 0 to 1. An $R$ value of 1 states complete separation of the groups, while an $R$ value equal to 0 indicates that no separation occurs (46, 47). Multiresponse permutation procedure (MRPP) was used to confirm the results obtained with ANOSIM. MRPP calculates the chance-corrected within group agreement $A$, varying from 0 to 1. $A = 0$ when within-group heterogeneity equals expectation by chance. When $A = 1$, all items are identical within each group. In ecology, values for $A$ are commonly below 0.1, even when there are apparent differences in groups. An $A > 0.3$ is fairly high (48, 49). ANOSIM, MRPP, and nMDS procedures were performed using the Vegan package in R v12.2.1 (45, 50).

High-coverage genome sequencing, de novo assembly, scaffolding and annotation.
Following DNA purification, one paired-end library (2x100 bp, 500 bp inserts) and two mate-pair libraries (2x100 bp, 2 kb inserts and 2x100 bp, 5 kb inserts) for Illumina (Illumina, San Diego, CA, USA) were prepared for *B. bruxellensis* strain ST05.12/22 (VIB X9085), originally isolated from lambic beer, according to the manufacturer’s instructions. Libraries were sequenced on the Illumina HiSeq 2000 platform at the Beijing Genomics Institute (BGI, Shenzhen, China). Reads were subjected to quality

After removing the adaptors and low quality reads (below a Phred quality score of 30), trimmed reads were de novo assembled using SOAPdenovo v1.05 (51), providing a pseudo-haploid assembly. Assembled contigs were subsequently scaffolded by SSPACE (52) using the Illumina mate-pair information. Gaps inside the scaffolds were closed by GapCloser (53) based on the paired-end read data. As suggested by Curtin et al. (24), coding sequences (CDSs) were predicted by AUGUSTUS v2.5.5 (54) with the S. cerevisiae gene models as reference. All predicted genes were annotated by using the NCBI KOG database (55) and BLAST2GO (56) with AWRI 1499 (= ST05.12/62 in our study) as a reference (24).

Genome comparison with AWRI 1499 and CBS 2499. Short read sequences obtained for ST05.12/22 were mapped to the B. bruxellensis AWRI 1499 (24) and CBS 2499 (23) genome assemblies as a reference using NovoAlign version 3.00.5 a(www.novocraft.com) with default settings. MarkDuplicates command in Picard (http://picard.sourceforge.net/) was used to remove the reads that mapped to the same positions in the reference genomes (PCR duplications). Single-nucleotide variations (SNPs, single nucleotide polymorphisms) and small insertions and deletions (InDels) were called for each contig using SAMtools and GATK (57-58). Default settings were used except the maximum read depth in SAMtools was set to 200X (-D 200). The generated SNPs and InDels were then filtered using custom Perl scripts to minimize false positive mutation calls. First, mutations with a total read depth below 20X were discarded. Second, SNPs and InDels with a Phred quality score below 30 were removed. Third, mutation calls were only kept when at least 80% of the reads was positive for the homogeneous sites and at least 30% of the reads was positive for the
heterogeneous sites. The lists of SNPs/InDels were then annotated by in-house Perl scripts with the *B. bruxellensis* AWRI 1499 and CBS 2499 assemblies as reference. SNP and InDel density was calculated using in-house R (50) and Python (59) scripts. Homologous genes were determined by comparing the whole genome assemblies using Blastn (37). Only genes that showed similarity lower than 1e-10 (E-value) were considered as homologues. Structural genome variation in ST05.12/22, including large duplications and deletions (>1 kb), was assessed using CNVnator (60) to identify copy-number variations through read-depth analysis. Default parameters were used except bin size was set to 100. As suggested by the developer q0<0.5 was used as cutoff. The analysis was not performed for the wine strains as we did not possess the raw, unassembled sequence data for AWRI 1499.

### RESULTS

**Phylogenetic relationships within the genus *Brettanomyces***. In order to examine the genetic relationships between the investigated *Brettanomyces* strains, a phylogram was constructed based on partial LSU rRNA gene sequences. This perfectly divided the different species into distinct clades (Fig. 1). In general, low sequence divergence was observed between strains belonging to the same species (0 to 1.4 %). Most divergence was observed for *B. bruxellensis* and *B. custersianus*, each displaying a subcluster of a few isolates within the species clades. Remarkably, within the *B. bruxellensis* clade, all soft drink isolates grouped separately, having 1 to 3 SNPs in comparison with the other *B. bruxellensis* isolates (Fig. 1). Obtained LSU rRNA gene sequences were deposited in GenBank under the accession numbers KF790763 - KF790811. Next, all isolates were subjected to a number of DNA fingerprinting assays, including three RAPD-PCR analyses, three AP-PCR analyses and one rep-PCR analysis. The
UPGMA dendrogram derived from Pearson correlation based on the combined data sets showed high congruence with the LSU rDNA-based phylogenetic tree (Fig. 2). However, the discriminative power displayed was considerably higher with these fingerprinting methods. Based on a DNA fingerprint similarity level of 66%, UPGMA clustering perfectly matched species delineation (Fig. 2), corroborating the results from the nMDS ordination (Fig. 3A; stress = 0.15). Additionally, an ANOSIM R test statistic equal to 0.9998 was found (p<0.00001; combined data sets), indicating that the tested groups were significantly different. When increasing the cut-off fingerprint similarity level to 80%, the five species clusters could be further divided into 13 subclusters (Table 1; Fig. 2), among which the *B. bruxellensis* subclusters generally represented different groups of strains isolated from a similar environment. More specifically, subcluster II-A contained only isolates from soft drinks (and one bantu beer strain (ST05.12/18)), while subcluster II-B harbored beer strains. Separate subclusters were formed for the wine strains. Interestingly, all isolates obtained from the Cantillon brewery, except ST05.12/48 (II-B), fell in a separate subcluster (II-D) (Table 1). nMDS ordination also grouped the *B. bruxellensis* isolates together according to the niche they were isolated from: Cantillon brewery, beer, wine and soft drinks (Fig. 3B, stress=0.08). In this case, the ANOSIM R test statistic was 0.684 (p<0.00001; combined data sets), supporting a large (but not complete) and statistically significant separation of the different groups, which was also supported by the MRPP results (A=0.071, p<0.000001; combined data sets). Altogether, these results suggest that *B. bruxellensis* strains isolated from similar niches are genetically more related than strains from different niches. This relation between genotype and niche is supported by evidence that strains isolated from similar niches in different locations clustered together. On the contrary, isolates obtained from, for example, the
same geographic region but different niches did not cluster together. For example, *B.
bruxellensis* strains ST05.12/21, ST05.12/30 and ST05.12/33 were all isolated in the
Netherlands, the first two from soft drinks, the latter from a Dutch stout beer. Based on
both LSU rRNA gene sequencing and DNA fingerprinting, the soft drink isolates
clustered together with other isolates from soft drinks, while the beer isolate was more
related to the rest of the beer isolates, irrespective of the year of isolation. Indeed,
whereas, for example, strains ST05.12/33 and ST05.12/34 were both isolated in 1939,
they grouped closely together with beer strains that were isolated several years later
(Table 1; Fig. 2 and 3B). To further support this correlation between genotype and
niche, we expanded our collection with seven additional *B. bruxellensis* strains from
wine (CBS 1940, CBS 1941, CBS 1942, CBS 1943, CBS 2336, MUCL 54012 and
MUCL 54015) and subjected them to M13 fingerprinting (using primer 5’-
TTATGAAACGACGGCCAGT-3’), together with the other *B. bruxellensis* strains.
Again, a correlation could be observed between source of isolation and genetic pattern
(Fig. S1), supporting our findings.

**Genome sequencing of a *B. bruxellensis* strain isolated from lambic fermentation.**

As detailed above, the few recent studies focusing on genome sequencing of
*Brettanomyces* yeasts have focused on wine spoilage isolates of *B. bruxellensis*. To
obtain a more complete view on the *Brettanomyces* genomes, and further investigate
the association between strains and isolation source, we sequenced the genome of a *B.
bruxellensis* strain originating from a spontaneous Belgian lambic fermentation
(ST05.12/22; genotype cluster II-B). Comparisons were made with the genome
sequences available for the wine *B. bruxellensis* strains AWRI 1499 (ST05.12/62;
genotype cluster II-E) and CBS 2499 (ST05.12/56; genotype cluster II-B), from
Australia and France, respectively. Comparison with these two strains is especially interesting as they not only originate from a different niche (wine), but also belong to different genotype clusters (II-B and II-E, representing the same and another genotype as our beer isolate). Therefore, genomic comparison of our beer strain with these two wine strains should provide us more insight into the genomic landscape of *B. bruxellensis*. De novo assembly of the ST05.12/22 sequence reads yielded 85 scaffolds with N50 of 257.6 kb at 100 to 110-fold coverage, and an assembly length of 13.0 Mb (Table 2), which is comparable with the assembly length obtained for AWRI 1499 (12.7 Mb (24)) and CBS 2499 (13.4 Mb (23)). In total, 5,255 gene models were predicted by AUGUSTUS for ST05.12/22 with the *S. cerevisiae* based model as reference, 36 and 17 of which had no homologues in AWRI 1499 and CBS 2499, respectively, and 30 and 16 *vice versa* (but see further). Sequence data for ST05.12/22 have been deposited in the NCBI short-read archive under the Bioproject accession SRP041023. The assembly and listing of annotations can be found via http://dx.doi.org/10.6084/m9.figshare.1007637.

**Variant analysis.** Single nucleotide variation and InDel analysis was performed by mapping the ST05.12/22 reads to both the AWRI 1499 and CBS 2499 genome assemblies. Compared to AWRI 1499 a total of 65,535 SNPs were found, or 5.04 SNPs per kb, of which 50.4 % were homozygous and 49.6 % were heterozygous in ST05.12/22 (Data set S1). Further, a total of 14,092 InDels (< 8 bp) were called, with an average density of 1.08 InDels per kb (Data set S2). The majority of these InDels represented deletions (92.7 %). Compared to the CBS 2499 genome assembly, 82,676 nucleotide variations were found, among which 79,421 SNPs (6.11 per kb; 87.7 % heterozygous, 22.3 % homozygous) (Data set S3) and 3,255 InDels (0.25 per kB; 68.2
% deletions) (Data set S4). SNPs and InDels were not uniformly distributed across the ST05.12/22 genome assembly, with some regions showing much higher SNP or InDel densities than others (Fig. S2).

**Ploidy level and allelic relationships.** Given the triploid nature of AWRI 1499, comprising a moderately heterozygous diploid and a third divergent haploid (24; 34), and the fact that CBS 2499 was recently confirmed to be a diploid (34), it was of interest to investigate the genomic organization of our strain in relation to both reference strains. First, the ploidy level of ST05.12/22 was estimated by taking advantage of allele proportions. In a diploid genome, it is expected that the average frequency of a particular allele at a heterozygous site will be around 0.5, while this would be closer to 0.66 for heterozygous sites in a triploid. As can be seen in Fig. 4, ST05.12/22 showed a maximum average allele frequency consistent with a diploid state, suggesting ST05.12/22 being diploid. In order to determine whether the diploid strains contained the divergent haplotype of AWRI 1499, five loci that displayed three clearly defined haplotypes in AWRI 1499 (24; 34) were investigated. Maximum-likelihood phylogenies were constructed based on the corresponding individual haplotype sequences for AWRI 1499 (3 sequences for each locus), CBS 2499 (2 sequences for each locus) and ST05.12/22 (2 sequences for each locus). Two of the three alleles of AWRI 1499 and both alleles of ST05.12/22 and CBS 2499 formed a highly related clade (Fig. 5). The third allele from AWRI 1499, on the other hand, was always divergent from the conserved clade. This thus confirms that *B. bruxellensis* has a core diploid genome, with some strains having a divergent third haploid complement of chromosomes (34). Moreover, ST05.12/22 and CBS 2499 had identical sequences for locus g1822.t1 (Fig. 5B) and exhibited only a few differences
for g2560.t1 (1 identical allele, 1 having 99.8 % identity; Fig. 5D) and g3222.t1 (1 identical allele, 1 having 99.9 % identity; Fig. 5E). For locus g1851.t1 (Fig. 5C), the three yeasts had one identical allele, and one showing differences across the three isolates (between 99.5 and 99.8 % identity).

**Structural genome variation.** Structural genome variation between ST05.12/22 and both reference strains was further investigated by CNV determination. Compared to AWRI 1499, CNVnator analysis enabled the identification of 61 CNVs (>1 kb), including 44 genomic duplications and 17 large deletions. Regarding the duplications, at least four genomic regions, encoding for a total of 69 genes (AHIQ01000029 (29 kb), AHIQ01000031 (20 kb), AHIQ01000102 (43 kb) and AHIQ01000195 (87 kb)) displayed a doubled copy number (4n or greater; normalized read depth ≥2) in ST05.12/22. Normalized read depths of 0 were obtained for two deletions (AHIQ01000315 and AHIQ01000316), suggesting no copy was retained in ST05.12/22 (Data set S5). Blast analysis of the AWRI 1499 CDS sequences corresponding to the 17 predicted deletions against the ST05.12/22 assembly yielded a number of genes belonging to five deletions that could not be clearly aligned to the ST05.12/22 genome (“no hit” or worse match; AHIQ01000211, AHIQ01000280, AHIQ01000303, AHIQ01000315 and AHIQ01000316; 26 genes in total; Data set S5), suggesting the presence of five completely deleted regions. Compared to CBS 2499, 40 regions with deletions and 40 duplications were found, seven having a normalized read depth score ≥2 (in total harboring 11 genes; scaffold 1 (4 kb), scaffold 2 (4.7 kb), scaffold 3 (2.6 kb), scaffold 6 (1.4 kb and 1.2 kb), scaffold 9 (3.6 kb) and scaffold 10 (2.5 kb)). Regarding the deletions, several regions were found with normalized read depths close to 0 (Data set S6). However, Blast analysis of the CBS 2499 translated...
sequences corresponding to the deleted regions was carried out and reduced the number of deletions to four regions (scaffolds 17, 18, 20 and 24; 42 genes) (Data set S6). PCR analysis followed by amplicon sequencing of at least one gene corresponding to each of the deletions predicted by CNVnator confirmed the loss of five (26 genes) and four regions (42 genes) compared to AWRI 1499 and CBS 2499, respectively (Data set S5; Data set S6; for primers see Table S1).

Blast analysis of the AWRI 1499 genome sequence using the ST05.12/22 assembly as a reference revealed a total of 30 genes that were uniquely found in AWRI 1499 (E-value <e-10). Ten out of these 30 genes were also found by the CNV analysis and were confirmed by PCR to be present in AWRI 1499 and missing in ST05.12/22. The other 20 genes represented ORFs encoding putative proteins. However, their absence in ST05.12/22 could not be confirmed by additional Blast analysis of these genes against the ST05.12/22 assembly nor by PCR analysis, except for two genes (EIF47553 and EIF48003) (Data set S7; for primers see Table S1). Likewise, Blast analysis of the CBS 2499 assembly revealed 16 genes that were present in CBS 2499 but missing in ST05.12/22, among which 10 were found by the CNV analysis and/or PCR confirmed (Data set S8). Vice versa, 36 and 17 genes were missing in AWRI 1499 and CBS 2499, respectively, and present in ST05.12/22. Two genes, , for which a function has not been determined yet, were confirmed using PCR to only be present in the beer strain when compared to AWRI 1499 (Data set S9; for primers see Table S1). No genes were confirmed to be present in ST05.12/22 and absent in CBS 2499 (Data set S10; for primers see Table S1).

Altogether, these analyses resulted in a total of 20 genes, clustered in four islands, that are present in both wine strains (AWRI 1499 and CBS 2499) but missing in the beer strain (ST05.12/22) (Table 3; also illustrated in Fig. S3 (61)). The presence
and absence of these 20 genes in AWRI 1499/CBS 2499 and ST05.12/22, respectively, was confirmed by subjecting the different strains to a PCR screen (reciprocally tested) (Data set S5 and S6; for primers see Table S1). No genes were found that occurred in ST05.12/22, but did not exist in both AWRI 1499 and CBS 2499. The gene clusters present in both wine strains but missing in ST05.12/22 represented two genes on AWRI 1499 reference contig AHIQ01000211 (CBS 2499 scaffold 18), twelve on AHIQ01000280 (CBS 2499 scaffold 17), four on AHIQ01000303 (CBS 2499 scaffold 24) and two on AHIQ01000316 (CBS 2499 scaffold 20). Interestingly, on the second deletion region, these genes encoded proteins involved in the uptake of sugars or the efflux of drugs, or in several carbon metabolic processes, encoding a galactose-1-phosphate uridylyltransferase, a galactokinase, a GAL10 bifunctional protein, a dtdp-glucose-dehydratase, a maltase and a β-glucosidase. A parologue of this β-glucosidase (sharing 68% and 67% sequence identity with AWRI 1499 (EIF45415) and CBS 2499 (JGI transcript number 51487) on the nucleotide and protein level, respectively) was found elsewhere in the ST05.12/22 genome (scaffold 8; gene 2952, β-glucosidase), for which AWRI 1499 and CBS 2499 were also found to contain a homologue (EIF48743, contig AHIQ0100078; JGI transcript number 26490, scaffold 17; 97% nucleotide identity with gene 2952 in the beer strain). For each β-glucosidase, both wine strains had identical homologues. Additionally, apart from another gene involved in carbon metabolism (β-galactosidase), a cluster of three genes involved in nitrogen metabolism (nitrate reductase, nitrite reductase and nitrate transporter) was found to be present in AWRI 1499 and CBS 2499 but missing in ST05.12/22. Consistent with these findings, AWRI 1499 and CBS 2499, both containing the nitrate assimilation gene cluster, were
found to grow on nitrate as sole nitrogen source (tested as mentioned in (1)), whereas ST05.12/22 lacking this gene cluster was not (Fig. 6).

**Distribution of genes uniquely found in AWRI 1499 and CBS 2499 but missing in ST05.12/22.** The existence of strain-specific genes suggests that these genes may have been lost in a particular strain, or may be acquired from another strain or species. Blastx analysis of the genes uniquely found in both wine strains revealed high homology (E<1e-19) with genes from other yeasts such as Candida, Debaryomyces Kluyveromyces, Meyerozyma, Ogataea, Saccharomyces, Scheffersomyces, Spathaspora, Schwanniomyces and Wickerhamomyces (Table 3). However, highest homology to the gene encoding a heavy metal binding protein in B. bruxellensis was found with a completely unrelated taxon to B. bruxellensis, a Streptomyces species (E-value of 1e-34) (Table 3), which could have act as a donor species for this gene. A PCR screen (for primers see Table S1) on the gene cluster targeting the beta-galactosidase, the nitrate reductase, the nitrite reductase and the adjacent nitrate transporter genes revealed that strains scoring positive or negative for one of these four genes generally also scored similarly for the remaining genes (Table 4). This suggests that this gene cluster has been completely lost in certain isolates. All but one isolate from the Cantillon brewery (ST05.12/48, subcluster II-B) (subcluster II-D) had lost this gene cluster. In addition, isolates ST05.12/12.21, ST05.12/28 and ST05.12/40 displayed this genomic deletion. Additionally, the three phylogenetically closely related beer isolates ST05.12/25, ST05.12/26 and ST05.12/27 (>99% fingerprint similarity (Fig. 2)) were found to contain the nitrate transporter gene, while they had lost the genes encoding the beta-galactosidase, the nitrate reductase and the nitrite reductase. Consistent with these findings, all isolates containing the complete nitrate
assimilation gene cluster displayed robust growth on nitrate as sole nitrogen source, whereas the isolates missing (part of) this gene cluster did not (Table 4). Further analysis of the three genes involved in the assimilation of nitrate revealed that all three genes were heterozygous in ST05.12/18, ST05.12/59, AWRI 1499 (ST05.12/62) and CBS 2499 (ST05.12/56), whereas they were homozygous in the other strains. Additionally, isolates from the genotype cluster II-B had highly conserved sequences for the three genes (for the nitrate reductase and nitrate transporter genes identical for all isolates; for the nitrite reductase gene identical for almost all isolates) (Fig. 7), illustrating their high genetic relatedness. Sequences obtained for these genes involved in *B. bruxellensis* nitrate assimilation were deposited in GenBank under the accession numbers KJ3590-KJ35643.

For the galactokinase, dtdp-glucose-dehydratase, maltase and β-glucosidase genes, positive and negative PCR results were found to be scattered over the different isolates tested (Table 4). For example, of the 26 isolates tested 17 scored positive for the maltase gene, while only seven scored positive for the β-glucosidase gene (Table 4). The random distribution of these genes over the *B. bruxellensis* clade could be potentially explained by a common ancestor having all these genes, which were lost by some strains in the course of evolution. Alternatively, this may be explained by the fact that the primers developed (based on the AWRI 1499 genome sequence) may have had one or more mismatches in comparison to the tested strains by which amplification failed.

**DISCUSSION**

Despite its economic importance, either for its spoilage or appreciated activity in specific beers, the ecology and genetic relationships between and within
Brettanomyces yeasts is still poorly understood. Here, we studied the genetic relationships between different Brettanomyces strains from all recognized Brettanomyces species, isolated from several food-related sources and geographic areas, and compared the genome sequence of a beer strain and wine strains.

First, all isolates were subjected to phylogenetic analysis based on LSU rRNA gene sequences and a number of established DNA fingerprinting techniques. Our results support earlier findings that Brettanomyces yeasts form a genetically diverse clade, even within a species, and are represented by several subgroupings (1, 16, 17, 32). Interestingly, expansion of our phylogenetic tree with all B. bruxellensis LSU rRNA gene sequences available in GenBank (55 additional sequences; August 2013) revealed no additional subclade within our B. bruxellensis clade (displaying a total of 13 polymorphic sites between the different B. bruxellensis isolates on a fragment of about 400 bp) (Fig. S4). Noteworthy, in this analysis all spoilage isolates collected from soft drinks such as cola and ginger ale or bantu beer grouped together, apart from all wine and most beer isolates, suggesting a link between the genotype and origin of the strains. Indeed, cluster analysis of the B. bruxellensis fingerprints obtained in this study indicates a strong correlation between the genetic profiles and the isolation source, rather than with geographic origin or year of isolation, thus suggesting niche adaptation. These results are in agreement with previously reported findings on Brettanomyces (for example, see references (1) and (16)). Also for other microorganisms, clustering of isolates according to the niche where they were isolated has been reported (62, 63). Conversely, our results are in disagreement to what has been found for, for example, S. paradoxus, a Saccharomyces species not related to industrial processes, for which geography seems more important than ecology in shaping the yeast’s population structure (68). For S. cerevisiae, a mixed population
structure was found, with lineages corresponding to geographic origin and others corresponding to niche (64). Interestingly, almost all wild isolates collected in this work (from the Cantillon brewery) clustered separately. This also resembles the findings of Vigentini et al. (16) who found that almost all their wild (wine) *B. bruxellensis* isolates were clearly separated from the CBS reference strains, representing isolates from different niches, also including wine isolates. It remains to be investigated whether these differences have a biological meaning or can be explained by the fact that the reference strains have become adapted to laboratory conditions, accompanied by changes in their genetic backbone (65).

In order to further investigate the genetic differences between strains originating from a different niche, a comparative genome analysis was carried out between a beneficial *B. bruxellensis* strain isolated from lambic fermentation (ST05.12/22) (sequenced in this study) and two wine spoilage strains, including one triploid (AWRI 1499 (ST05.12/62)) and one diploid strain (CBS 2499 (ST05.12/56)) (used as a reference). In this study ST05.12/22 was determined to be diploid, possessing a pair of closely related chromosomes with moderate levels of heterozygosity. Interestingly, triploid *B. bruxellensis* strains have been found to represent the vast majority of isolates from the wine industry (34), suggesting that the additional chromosome may confer a selective advantage for these strains in wineries. Also in *Saccharomyces* interspecific hybrids have been found that are allotriploid. These hybrids have been isolated from cold winemaking and brewing environments, where it is suggested that the allotriploid hybrids have a selective advantage over their parents (66; 67). So far, no other data is available on the ploidy level of *B. bruxellensis* strains isolated from other niches such as beer. Further research should therefore be
performed to find out whether a correlation exists between the level of ploidy and the
niche the strains occur.

In addition to strain-specific SNPs or InDels structural genome variation was
found between our strain and both wine strains, with some genomic regions duplicated
and others deleted in ST05.12/22. Further examination of the functional annotation of
the genes duplicated in the beer strain compared to the wine strains revealed no
indications that our beer strain would contain duplicated genes favoring its survival in
beer. Of the genomic loci that were absent in the beer strain, two regions were of
particular interest. These involved either the *B. bruxellensis* nitrate assimilation cluster
or a cluster of genes involved in carbon metabolism, two phenotypic features that have
been shown to vary considerably between *B. bruxellensis* strains (2). For example,
nearly one third of *B. bruxellensis* wine isolates failed to grow on nitrate as sole
nitrogen source (2). Additionally, while most isolates could grow on the hexose
monosaccharides glucose and fructose and the disaccharides sucrose, maltose,
cellobiose and trehalose, about one fifth of the tested isolates was unable to grow on
galactose. Further, sugars such as arabinose, lactose and raffinose did not support
growth of most isolates (2). Woolfit *et al.* (22) reported the presence of five genes
involved in nitrate assimilation in *B. bruxellensis* CBS 2499, including genes encoding
a nitrate reductase, a nitrite reductase, a nitrate transporter as well as two regulatory
genes, encoding a Zn(II)$_2$Cys$_6$ transcriptional factor for nitrate induction. Strains of
*Hansenula polymorpha* in which any of these genes were disrupted lost their ability to
grow on nitrate (68-70), showing their necessity in the assimilation of nitrate.
Recently, it has been shown that the ability to assimilate nitrate can render *B.
bruxellensis* able to out-compete *S. cerevisiae* in industrial fermentations, as *S.
cerevisiae* cannot use nitrate (71). Furthermore, nitrate assimilation has been shown to
give *B. bruxellensis* an improved ability to grow under anaerobic conditions and improve its fermentative metabolism (72). Along with an adjacent β-galactosidase gene, this cluster is specifically missing in ST05.12/22. Based on these findings it may be speculated that this gene cluster is less important for *B. bruxellensis* in certain fermentation systems such as brewing, thereby providing a selective pressure for its loss. On the other hand, whereas many beer isolates indeed represented nitrate negative phenotypes, PCR screening and phenotypic testing of our *B. bruxellensis* isolates did not reveal a clear correlation between possibility to assimilate nitrate and niche. Nonetheless, out of the five isolates from soft drinks tested (including one from bantu beer), four isolates were nitrate positive. Further research with more isolates from different origins is needed to elucidate whether a correlation exists between (non-)nitrate utilization phenotypes and niche. As soft drinks are often nitrogen poor (73), it may be expected that the ability to use nitrate may give nitrate utilization phenotypes an advantage to cause spoilage over microbes that cannot utilize nitrate.

Interestingly, in many strains (10 out of 14 containing the whole nitrate assimilation gene cluster), the nitrate assimilation gene cluster was shown to have undergone loss of heterozygosity resulting in identical alleles (haplotyped sequences). In contrast to Borneman *et al.* (34), our results do not support the suggestion that loss of heterozygosity in these genes is correlated with the inability to utilize nitrate, as all isolates which contained the complete gene cluster and showed growth on ammonium also displayed robust growth on nitrate as a sole nitrogen source. Further research should elucidate the impact of this loss of heterozygosity for nitrate assimilation in different ecosystems. In contrast, isolates that had lost part of the gene cluster or the complete gene cluster were unable to utilize nitrate. The sequences of the homozygous strain AWRI 1608, unable to grow on nitrate despite containing the complete nitrate...
assimilation locus (34), revealed its nitrite reductase and nitrate transporter gene sequences were identical to those of other strains able to grow on nitrate. For the
nitrate reductase gene, however, a difference of at least one amino acid was found
(aspartic acid in AWR 1608 versus serine in our strains) between AWRI 1608 and our
strains, which could have led to a less efficient enzyme, and thus to less efficient
nitrate assimilation. However, further research, e.g. by subjecting this strain to the
nitrate assimilation test performed in our study, is needed to confirm this.

Further, in comparison with both wine strains our beer strain was found to lack
a cluster of 12 genes, among which the majority involved in carbon metabolism,
encoding a galactose-1-phosphate uridylyltransferase, a galactokinase, a GAL10
bifunctional protein, a dtdp-glucose-dehydratase, a maltase and a β-glucosidase.
Together with the β-galactosidase mentioned above, the first three enzymes are
involved in the metabolism of galactose. β-glucosidases are well-known for their role
in flavor development in beer and wine (74, 75). Additionally, β-glucosidase has been
shown to play a role in the fermentation of cellobiose by B. bruxellensis (76-79)
Interestingly, we found that ST05.12/22 did contain another β-glucosidase gene, which
is also present in AWRI 1499 and CBS 2499. Further research is needed to investigate
whether the presence of this second β-glucosidase results in differences in glucosidase
activity and flavoring capability of B. bruxellensis strains. Furthermore, further studies
to determine how substrates and growth conditions affect the production of flavor
compounds is needed, which may help explain why specific strains are for example
not associated with any off-flavor.

The phenomenon of loss of nutrient utilization reminds of the concerted loss of
the galactose catabolism cluster in Japanese S. cerevisiae isolates compared to
European isolates, probably due to the fact that particular functions in the pathway
have fitness costs (80). A PCR screen performed on a selection of genes involved in carbon or nitrogen assimilation revealed a different distribution of the genes across the B. bruxellensis clade, corroborating the earlier observed phenotypic diversity between different B. bruxellensis strains (1). Further study of nitrate and carbon assimilation will reveal more insights in driving the phenotypes towards, or away from, utilization of specific nitrogen or carbon sources. Further, by comparing sufficiently large sets of whole genomes, coupled with functional and phenotypic analyses, we hope to be able to answer the question whether there are distinct groups of B. bruxellensis which have a distinct impact on the production of beer and wine or other beverages, or, more generally, to further understand the behavior of this economically important yeast.

ACKNOWLEDGEMENTS

We are grateful to Prof. G. Derdelinckx (KU Leuven) and F. Bajoulkakahi for assisting in the isolation of the B. bruxellensis strains. In addition, we thank everybody providing us with Brettanomyces strains and, especially the Cantillon brewery for giving us the opportunity to sample within their brewery. We appreciate the help of Dr. A. Justé for performing the molecular analyses and C. Verreth for technical assistance. Finally, we are grateful to Prof. C.D. Curtin for discussing our results.
REFERENCES


FIGURE LEGENDS

Fig. 1. Maximum likelihood tree (Tamura-Nei model) of all Brettanomyces strains investigated in this study based on partial large subunit ribosomal RNA gene sequences (457-471 bp). Bootstrap values >80% (based on 1,000 replicates) are given at the nodes of the tree. Saccharomyces cerevisiae was used as an outgroup. The origin of the different strains, i.e. beer, Cantillon brewery, soft drink, others and wine are highlighted in green, dark green, orange, blue and red, respectively.

Fig. 2. Dendrogram derived from the UPGMA linkage of Pearson correlation coefficients of combined fingerprinting data sets for all Brettanomyces strains investigated in this study. Isolates from B. anomalus (Dekkera anomala) (yellow), B. (Dekkera) bruxellensis (green), B. custersianus (red), B. naardenensis (blue) and B. nanus (pink) are grouped in the clusters I, II, III, IV and V, respectively (defined at a similarity percentage of 66%; marked by solid red line). At 80% similarity, 13 clusters can be distinguished (marked by dotted red line), among which the B. bruxellensis subclusters generally represent strains from a similar environment. “Blank” refers to the negative control (sterile distilled water). B. bruxellensis strains marked with a circle were shown to have the complete nitrate assimilation gene cluster, consisting of genes encoding a nitrate reductase, a nitrite reductase and a nitrate transporter. Isolates marked with a square lost the genes encoding the nitrate reductase and nitrite reductase. Isolates marked with a triangle lost the complete nitrate assimilation gene cluster. B. bruxellensis strains that are able or unable to utilize nitrate as a nitrogen source are indicated with a green or red colored mark, respectively. Isolates ST05.12/30 and ST05.12/54 (marked in orange) were both negative on ammonium and nitrate in our assay.
Fig. 3. Non-metric multidimensional scaling plot based on Pearson coefficient similarities of the combined fingerprinting data sets for (A) all Brettanomyces and (B) all *B. (Dekkera) bruxellensis* isolates investigated in this study. In panel A, isolates from *B. anomalus* (*D. anomala*), *B. bruxellensis*, *B. custersianus*, *B. naardenensis*, and *B. nanus*, are represented by circles, squares, triangles, plus signs, and reverse triangles, respectively. The diamant symbol corresponds to *Saccharomyces cerevisiae* and the X sign represents the negative control (sterile distilled water) (stress of plot = 0.15). The origin of the different strains, i.e. beer, Cantillon brewery, soft drink, others and wine are highlighted in green, dark green, orange, blue and red, respectively. In panel B, open dark green squares, closed green squares, red circles and orange triangles represent *B. bruxellensis* isolates from the Cantillon brewery, beer, wine and soft drinks, respectively (stress of plot = 0.08).

Fig. 4. Allele frequency distribution histogram, suggesting *Brettanomyces (Dekkera) bruxellensis* ST05.12/22 is a diploid strain. Calculations are based on the triploid reference strain AWRI 1499 (A) and the diploid strain CBS 2499 (B).

Fig. 5. Haplotype analysis of the three investigated *Brettanomyces (Dekkera) bruxellensis* isolates ST05.12/22, AWRI 1491 and CBS 2499. Distinct haplotypes were assembled for five conserved open reading frames and subjected to maximum likelihood phylogenies. These five loci represented genes encoding a nuclear protein required for actin cytoskeleton (*g1134.t1*) (A), a DNA primase small subunit (*g1822.t1*) (B), a protein component of the H/ACA snoRNP pseudouridylase complex
(g1851.t1) (C), and two hypothetical proteins (g2561.t1; g3222.t1) (D and E). Bootstrap values (based on 1,000 replicates) are given at the nodes of the tree.

Fig. 6. Phenotypic analysis of *Brettanomyces (Dekkera) bruxellensis* strains growing on either ammonium (NH$_4^+$) or nitrate (NO$_3^-$) (incubated for 7 days at 25°C). Both AWRI 1499 (ST05.12/56 (C)) and CBS 2499 (ST05.12/62 (D)) show growth, whereas ST05.12/22 (B) is not able to grow in medium with nitrate as sole nitrogen source. Tube A represents the negative control (non-inoculated medium).

Fig. 7. Phylogenetic analysis of the nitrate assimilation cluster in *B. bruxellensis*. Maximum-likelihood phylogenies were prepared from the haplotype-resolved ORFs for the predicted nitrate reductase (A), nitrite reductase (B), and nitrate transporter (C) proteins. Bootstrap values (based on 1,000 replicates) are given at the nodes of the tree. The origin of the different strains, i.e. beer, soft drink and wine are highlighted in green, orange and red, respectively. Circles represent the ability of the parent strain to utilize nitrate as a nitrogen source, while squares mark the inability of the parent strain to utilize nitrate as a nitrogen source. A triangle indicates that no conclusions could be made regarding nitrate assimilation (also found negative on ammonium in our assay).

All three genes were heterozygous in ST05.12/18, ST05.12/56 (CBS 2499), ST05.12/59 and ST05.12/62 (AWRI 1499), but were homozygous in the other strains. Strains ST05.12/25, ST05.12/26 and ST05.12/27 lost the nitrate and nitrite reductase genes. In addition to the *B. bruxellensis* strains investigated in this study, strain AWRI 1608, which is unable to utilize nitrate (34), has been included in the analysis.
<table>
<thead>
<tr>
<th>Species</th>
<th>Isolatea</th>
<th>Isolateb</th>
<th>LSU rRNA gene Genbank sequence accession number</th>
<th>Geographic origin</th>
<th>Year of isolation</th>
<th>Niche</th>
<th>Cluster</th>
<th>Subcluster</th>
</tr>
</thead>
<tbody>
<tr>
<td>B. anomalus</td>
<td>ST05.12/06</td>
<td>CBS 5111</td>
<td>KF790764</td>
<td>Ireland</td>
<td>-</td>
<td>Beer</td>
<td>I</td>
<td>I-A</td>
</tr>
<tr>
<td></td>
<td>ST05.12/15</td>
<td>MUCL 27703</td>
<td>KF790805</td>
<td>-</td>
<td>1909</td>
<td>Beer</td>
<td>I</td>
<td>I-B</td>
</tr>
<tr>
<td></td>
<td>ST05.12/13</td>
<td>MUCL 27702</td>
<td>KF790807</td>
<td>United Kingdom</td>
<td>1940</td>
<td>Bottled beer</td>
<td>I</td>
<td>I-B</td>
</tr>
<tr>
<td></td>
<td>ST05.12/19</td>
<td>MUCL 31218</td>
<td>KF790801</td>
<td>United Kingdom</td>
<td>-</td>
<td>Bottled beer</td>
<td>I</td>
<td>I-B</td>
</tr>
<tr>
<td></td>
<td>ST05.12/09</td>
<td>CBS 4460</td>
<td>KF790763</td>
<td>-</td>
<td>-</td>
<td>Cider</td>
<td>I</td>
<td>I-B</td>
</tr>
<tr>
<td></td>
<td>ST05.12/35</td>
<td>NRRL Y17520</td>
<td>KF790784</td>
<td>-</td>
<td>-</td>
<td>Cider</td>
<td>I</td>
<td>I-B</td>
</tr>
<tr>
<td></td>
<td>ST05.12/61</td>
<td>CBS 4210</td>
<td>KF790767</td>
<td>-</td>
<td>-</td>
<td>Cider</td>
<td>I</td>
<td>I-B</td>
</tr>
<tr>
<td></td>
<td>ST05.12/17</td>
<td>MUCL 49367</td>
<td>KF790803</td>
<td>The Netherlands</td>
<td>-</td>
<td>Soft drink</td>
<td>I</td>
<td>I-B</td>
</tr>
<tr>
<td></td>
<td>ST05.12/39</td>
<td>NRRL Y17522</td>
<td>KF790781</td>
<td>The Netherlands</td>
<td>-</td>
<td>Soft drink</td>
<td>I</td>
<td>I-B</td>
</tr>
<tr>
<td></td>
<td>ST05.12/26</td>
<td>MUCL 49865</td>
<td>KF790794</td>
<td>Belgium</td>
<td>-</td>
<td>Brewery</td>
<td>II</td>
<td>II-B</td>
</tr>
<tr>
<td></td>
<td>ST05.12/18</td>
<td>MUCL 27705</td>
<td>KF790802</td>
<td>South-Africa</td>
<td>-</td>
<td>Bantu-beer</td>
<td>II</td>
<td>II-A</td>
</tr>
<tr>
<td>B. bruxellensis</td>
<td>ST05.12/26</td>
<td>MUCL 49865</td>
<td>KF790794</td>
<td>Belgium</td>
<td>2012</td>
<td>Cantillon brewery</td>
<td>II</td>
<td>II-B</td>
</tr>
<tr>
<td></td>
<td>ST05.12/49</td>
<td>CBS 6055</td>
<td>KF790777</td>
<td>Belgium</td>
<td>2012</td>
<td>Cantillon brewery</td>
<td>II</td>
<td>II-D</td>
</tr>
<tr>
<td></td>
<td>ST05.12/48</td>
<td>CBS 6055</td>
<td>KF790778</td>
<td>Belgium</td>
<td>2012</td>
<td>Cantillon brewery</td>
<td>II</td>
<td>II-D</td>
</tr>
<tr>
<td></td>
<td>ST05.12/50</td>
<td>CBS 6055</td>
<td>KF790775</td>
<td>Belgium</td>
<td>2012</td>
<td>Cantillon brewery</td>
<td>II</td>
<td>II-D</td>
</tr>
<tr>
<td></td>
<td>ST05.12/51</td>
<td>CBS 6055</td>
<td>KF790774</td>
<td>Belgium</td>
<td>2012</td>
<td>Cantillon brewery</td>
<td>II</td>
<td>II-D</td>
</tr>
<tr>
<td></td>
<td>ST05.12/52</td>
<td>CBS 6055</td>
<td>KF790773</td>
<td>Belgium</td>
<td>2012</td>
<td>Cantillon brewery</td>
<td>II</td>
<td>II-D</td>
</tr>
<tr>
<td></td>
<td>ST05.12/61</td>
<td>CBS 6055</td>
<td>KF790804</td>
<td>United States of America</td>
<td>-</td>
<td>Dry ginger ale</td>
<td>II</td>
<td>II-A</td>
</tr>
<tr>
<td></td>
<td>ST05.12/59</td>
<td>CBS 6055</td>
<td>KF790768</td>
<td>United States of America</td>
<td>-</td>
<td>Dry ginger ale</td>
<td>II</td>
<td>II-A</td>
</tr>
<tr>
<td>Code</td>
<td>Strain</td>
<td>Species/Type</td>
<td>Country</td>
<td>Source</td>
<td>Role</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>-----------</td>
<td>--------</td>
<td>--------------</td>
<td>-------------</td>
<td>-----------------------------</td>
<td>----------------------------------------------------------------------</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>ST05.12/54</td>
<td>CBS 73</td>
<td>KF790772</td>
<td>France</td>
<td>-</td>
<td>Grape must</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>ST05.12/22</td>
<td>VIB X9085</td>
<td>KF790798</td>
<td>Belgium</td>
<td>-</td>
<td>Lambic beer</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>ST05.12/24</td>
<td>MUCL 27707</td>
<td>KF790796</td>
<td>Belgium</td>
<td>-</td>
<td>Lambic beer</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>ST05.12/25</td>
<td>MUCL 27700</td>
<td>KF790795</td>
<td>Belgium</td>
<td>-</td>
<td>Lambic beer</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>ST05.12/27</td>
<td>MUCL 30490</td>
<td>KF790793</td>
<td>Belgium</td>
<td>1989</td>
<td>Lambic beer</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>ST05.12/40</td>
<td>MUCL 30489</td>
<td>KF790797</td>
<td>Belgium</td>
<td>-</td>
<td>Lambic beer</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>ST05.12/36</td>
<td>NRRL Y1413</td>
<td>KF790792</td>
<td>Belgium</td>
<td>-</td>
<td>Lambic beer</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>ST05.12/21</td>
<td>MUCL 27706</td>
<td>KF790789</td>
<td>The Netherlands</td>
<td>-</td>
<td>Soft drink</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>ST05.12/30</td>
<td>CBS 8027</td>
<td>KF790789</td>
<td>The Netherlands</td>
<td>-</td>
<td>Soft drink</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>ST05.12/33</td>
<td>CBS 98</td>
<td>KF790786</td>
<td>The Netherlands</td>
<td>1939</td>
<td>Stout</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>ST05.12/34</td>
<td>CBS 97</td>
<td>KF790785</td>
<td>United Kingdom</td>
<td>1939</td>
<td>Stout</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>ST05.12/56</td>
<td>CBS 2499</td>
<td>KF790770</td>
<td>France</td>
<td>-</td>
<td>Wine</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>ST05.12/62</td>
<td>AWRI 1499</td>
<td>KF790766</td>
<td>Australia</td>
<td>-</td>
<td>Wine</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>B. custersiatus</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>ST05.12/11</td>
<td>CBS 4806</td>
<td>KF790809</td>
<td>South Africa</td>
<td>-</td>
<td>Bantu-beer</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>ST05.12/12</td>
<td>MUCL 27704</td>
<td>KF790808</td>
<td>South Africa</td>
<td>1960</td>
<td>Bantu-beer</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>ST05.12/29</td>
<td>CBS 8347</td>
<td>KF790791</td>
<td>The Netherlands</td>
<td>1996</td>
<td>Olives</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>B. naardenensis</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>ST05.12/10</td>
<td>CBS 6043</td>
<td>KF790810</td>
<td>The Netherlands</td>
<td>-</td>
<td>Carbonated tonic water</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>ST05.12/14</td>
<td>MUCL 27708</td>
<td>KF790806</td>
<td>The Netherlands</td>
<td>-</td>
<td>Soft drink</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>ST05.12/37</td>
<td>NRRL Y5740</td>
<td>KF790782</td>
<td>United States</td>
<td>-</td>
<td>Soft drink</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Strain Number</td>
<td>CBS Code</td>
<td>KF Number</td>
<td>Country/Region</td>
<td>Product Type</td>
<td>Clustering Results (combined data sets DNA fingerprint results) at a similarity percentage of 66%, perfectly corresponding to species delination;</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>---------------</td>
<td>----------</td>
<td>-----------</td>
<td>----------------</td>
<td>--------------</td>
<td>-------------------------------------------------------------</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>ST05.12/57</td>
<td>CBS 6040</td>
<td>KF790769</td>
<td>United States of America</td>
<td>Soda water</td>
<td>IV IV-A</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>B. nanus ST05.12/31</td>
<td>CBS 1956</td>
<td>KF790788</td>
<td>Sweden</td>
<td>Bottled beer</td>
<td>V V-A</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>ST05.12/32</td>
<td>CBS 1955</td>
<td>KF790787</td>
<td>Sweden</td>
<td>Bottled beer</td>
<td>V V-A</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

- Own isolate numbering;

b AWRI, Australian Wine Research Institute, Glen Osmond, Australia; CBS, Centraalbureau voor Schimmelcultures, Utrecht, The Netherlands; NRRL, Agriculture Research Service Culture Collection, National Center for Agricultural Utilization Research, Peoria, IL, USA; KaHoSL, Katholieke Hogeschool Sint-Lieven, Gent, Belgium; MUCL, Mycothèque de l’Université Catholique de Louvain, Louvain-la-Neuve, Belgium; VIB, Vlaams Instituut voor Biotechnologie, Leuven, Belgium;

c - unknown;

d Clustering results (combined data sets DNA fingerprint results) at a similarity percentage of 66%, perfectly corresponding to species delination;

e Clustering results (combined data sets DNA fingerprint results) at a similarity percentage of 80%;

f Sampled (February, 2012) from 50 year old cask; beer brewed on the 13th of November 2011;

g Sampled (February, 2012) from 6 to 7 year old cask; beer brewed on the 12th of December 2011;

h Sampled (February, 2012) from 8 year old cask; beer brewed on the 12th of December 2011;

i Isolated using Wallerstein Laboratory Nutrient agar (WLN);

j Isolated using Universal Beer Agar (UBA);
Isolated using Dekkera/Brettanomyces Differentiation Medium (DBDM).
<table>
<thead>
<tr>
<th>Library type/name</th>
<th>Obtained sequence data</th>
<th>Obtained sequence data after quality filtering</th>
</tr>
</thead>
<tbody>
<tr>
<td>500bp _2 x 100</td>
<td>1.6668Mb</td>
<td>1.350Mb</td>
</tr>
<tr>
<td>2kb _2 x 100</td>
<td>615Mb</td>
<td>502Mb</td>
</tr>
<tr>
<td>5kb _2 x 100</td>
<td>651Mb</td>
<td>502Mb</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Assembly</td>
<td></td>
</tr>
<tr>
<td>Main genome scaffold total</td>
<td>85</td>
</tr>
<tr>
<td>Main genome contig total</td>
<td>576</td>
</tr>
<tr>
<td>Main genome scaffold sequence total (Mb)</td>
<td>13.0</td>
</tr>
<tr>
<td>Main genome contig sequence total (Mb)</td>
<td>12.8</td>
</tr>
<tr>
<td>Max scaffold size (Mb)</td>
<td>1.4</td>
</tr>
<tr>
<td>Min scaffold size (kb)</td>
<td>1</td>
</tr>
<tr>
<td>Main genome scaffold N50 (Mb)</td>
<td>0.7</td>
</tr>
<tr>
<td>% main genome in scaffolds &gt; 50 Kb</td>
<td>97.6</td>
</tr>
</tbody>
</table>

Predicted gene models
- Average gene length (bp) | 1,569 |
- Average Protein length (amino acids) | 510 |
- Average exon frequency per gene | 1.15 |
- Average exon length (bp) | 1,335 |
- Average intron length (bp) | 248 |

Predicted gene models and supporting lines of evidence
- No. of gene models | 5,255 |
% complete (with start and stop codons) 99.9
% genes with homology support 95.0

Functional annotation of proteins
Proteins assigned a GO term 4,348
<table>
<thead>
<tr>
<th>AWRI 1499 Accession N°</th>
<th>CBS 2499 IGI transcript ID</th>
<th>AWRI 1499 Contig (GenBank Accession N°)</th>
<th>CBS 2499 scaffold (JGI name)</th>
<th>Function</th>
<th>GenBank homology (Blastx)</th>
<th>Organism (GenBank Accession N°)</th>
<th>ID%</th>
<th>E-value</th>
<th>S-score</th>
</tr>
</thead>
<tbody>
<tr>
<td>EIF46399</td>
<td>8711</td>
<td>AHIQ01000211</td>
<td>scaffold 18</td>
<td>pantothenate transporter fen2</td>
<td><em>Mesorocysta guilliermondii</em> ATCC 6260 (XP_001482616)</td>
<td>63.34</td>
<td>0.0</td>
<td>1538</td>
<td></td>
</tr>
<tr>
<td>EIF46400</td>
<td>23063</td>
<td>AHIQ01000211</td>
<td>scaffold 18</td>
<td>upf0145 protein</td>
<td><em>Streptomyces sp. R1-NS-10</em> (WP_013070227)</td>
<td>55.66</td>
<td>4e-34</td>
<td>316</td>
<td></td>
</tr>
<tr>
<td>EIF45404</td>
<td>26687</td>
<td>AHIQ01000280</td>
<td>scaffold 17</td>
<td>mfs drug transporter</td>
<td><em>Mesorocysta guilliermondii</em> ATCC 6260 (XP_001482160)</td>
<td>55.11</td>
<td>2e-177</td>
<td>1342</td>
<td></td>
</tr>
<tr>
<td>EIF45405</td>
<td>51428</td>
<td>AHIQ01000280</td>
<td>scaffold 17</td>
<td>putative mfs-mdt transporter</td>
<td><em>Ogataea parapolymorpha DL-1</em> (EFW97434)</td>
<td>45.75</td>
<td>6e-137</td>
<td>1073</td>
<td></td>
</tr>
<tr>
<td>EIF45407</td>
<td>62814</td>
<td>AHIQ01000280</td>
<td>scaffold 17</td>
<td>high-affinity glucose transporter</td>
<td><em>Schefteromyces stipitis</em> CBS 6054 (XP_013082755)</td>
<td>71.38</td>
<td>0.0</td>
<td>2037</td>
<td></td>
</tr>
<tr>
<td>EIF45408</td>
<td>26690</td>
<td>AHIQ01000280</td>
<td>scaffold 17</td>
<td>galactose-1-phosphate uridylyltransferase</td>
<td><em>Schefteromyces stipitis</em> CBS 6054 (XP_001383359)</td>
<td>65.94</td>
<td>8e-179</td>
<td>1322</td>
<td></td>
</tr>
<tr>
<td>EIF45409</td>
<td>8686</td>
<td>AHIQ01000280</td>
<td>scaffold 17</td>
<td>galactokinase</td>
<td><em>Mesorocysta guilliermondii</em> ATCC 6260 (EDK41764)</td>
<td>54.07</td>
<td>3e-152</td>
<td>1164</td>
<td></td>
</tr>
<tr>
<td>EIF45410</td>
<td>26691</td>
<td>AHIQ01000280</td>
<td>scaffold 17</td>
<td>gal10 bifunctional protein</td>
<td><em>Candida tenuis</em> ATCC 10573 (EGV61616)</td>
<td>58.08</td>
<td>0.0</td>
<td>2080</td>
<td></td>
</tr>
<tr>
<td>EIF45411</td>
<td>26692</td>
<td>AHIQ01000280</td>
<td>scaffold 17</td>
<td>dldp-glucose -dehydratase</td>
<td><em>Debaryomyces hansenii</em> CBS767 (XP_457784)</td>
<td>49.02</td>
<td>8e-99</td>
<td>787</td>
<td></td>
</tr>
<tr>
<td>Accession</td>
<td>Start</td>
<td>Stop</td>
<td>Scaffold</td>
<td>Description</td>
<td>Species</td>
<td>Location</td>
<td>Score</td>
<td>E-value</td>
<td>Bit Score</td>
</tr>
<tr>
<td>-----------</td>
<td>-------</td>
<td>------</td>
<td>----------</td>
<td>-------------</td>
<td>---------</td>
<td>----------</td>
<td>-------</td>
<td>----------</td>
<td>-----------</td>
</tr>
<tr>
<td>EIF45412</td>
<td>78562</td>
<td>AHIQ01000280</td>
<td>scaffold 17</td>
<td>hexose transporter</td>
<td>Candida intermedia (CAO79523)</td>
<td>47.73</td>
<td>6e-144</td>
<td>1118</td>
<td></td>
</tr>
<tr>
<td>EIF45413</td>
<td>8690</td>
<td>AHIQ01000280</td>
<td>scaffold 17</td>
<td>maltase</td>
<td>Saccharomyces kudriavzevii IFO 1802 (EJT44539)</td>
<td>63.4</td>
<td>0.0</td>
<td>1954</td>
<td></td>
</tr>
<tr>
<td>EIF45414</td>
<td>36189</td>
<td>AHIQ01000280</td>
<td>scaffold 17</td>
<td>multidrug resistance regulator 1</td>
<td>Ogataea parapolymorpha DL-1 (EFW97551)</td>
<td>34.62</td>
<td>4e-61</td>
<td>584</td>
<td></td>
</tr>
<tr>
<td>EIF45415</td>
<td>51487</td>
<td>AHIQ01000280</td>
<td>scaffold 17</td>
<td>beta-glucosidase</td>
<td>Schwanniomyces etchellsi (ACF93471)</td>
<td>59.26</td>
<td>0.0</td>
<td>2706</td>
<td></td>
</tr>
<tr>
<td>EIF45416</td>
<td>51392</td>
<td>AHIQ01000280</td>
<td>scaffold 17</td>
<td>hexose transporter</td>
<td>Wickerhamomyces ciferrii (CCH41021)</td>
<td>50.36</td>
<td>0.0</td>
<td>1439</td>
<td></td>
</tr>
<tr>
<td>EIF45248</td>
<td>26813</td>
<td>AHIQ01000303</td>
<td>scaffold 24</td>
<td>beta-galactosidase</td>
<td>Kluyveromyces lactis NRRL Y-1140 (XP_452194)</td>
<td>58.39</td>
<td>0.0</td>
<td>1575</td>
<td></td>
</tr>
<tr>
<td>EIF45249</td>
<td>51850</td>
<td>AHIQ01000303</td>
<td>scaffold 24</td>
<td>nitrate reductase</td>
<td>Ogataea parapolymorpha DL-1 (EFW95688)</td>
<td>59.32</td>
<td>0.0</td>
<td>2067</td>
<td></td>
</tr>
<tr>
<td>EIF45250</td>
<td>26815</td>
<td>AHIQ01000303</td>
<td>scaffold 24</td>
<td>nitrite reductase</td>
<td>Ogataea parapolymorpha DL-1 (EFW95689)</td>
<td>62</td>
<td>0.0</td>
<td>1159</td>
<td></td>
</tr>
<tr>
<td>EIF45251</td>
<td>145655</td>
<td>AHIQ01000303</td>
<td>scaffold 24</td>
<td>nitrate transporter</td>
<td>Ogataea parapolymorpha DL-1 (EFW95690)</td>
<td>59</td>
<td>0.0</td>
<td>604</td>
<td></td>
</tr>
<tr>
<td>EIF45193</td>
<td>8788</td>
<td>AHIQ01000316</td>
<td>scaffold 20</td>
<td>cytochrome mitochondrial precursor</td>
<td>Spathaspora passalidarum NRRL Y-27907 (EGW32589)</td>
<td>50.68</td>
<td>2e-113</td>
<td>904</td>
<td></td>
</tr>
<tr>
<td>EIF45194</td>
<td>31164</td>
<td>AHIQ01000316</td>
<td>scaffold 20</td>
<td>fungal specific transcription factor domain protein</td>
<td>Kluyveromyces lactis NRRL Y-1140 (XP_453742)</td>
<td>45.55</td>
<td>7e-97</td>
<td>876</td>
<td></td>
</tr>
</tbody>
</table>

* As determined by read depth analysis (CNVnator) and/or Blast analysis of the AWRI 1499 and CBS 2499 genome sequence against the ST05.12/22 assembly, and confirmed by a PCR screen (see also Data sets S5 and S6);
Based on best *B. bruxellensis* GenBank blast hit;

*B. bruxellensis* hits excluded;

ID%, identity percentage;

* Expected value.
### TABLE 4 Distribution of genes\(^a\) uniquely found in *Brettanomyces bruxellensis* AWRI 1499 (ST05.12/62) and CBS 2499 (ST05.12/56) over different *Brettanomyces* strains\(^b\)

<table>
<thead>
<tr>
<th>CDS (GenBank Accession N(^\circ))</th>
<th>Function</th>
<th>Brettanomyces bruxellensis ST05.12/62</th>
</tr>
</thead>
<tbody>
<tr>
<td>Contig AHIQ01000280</td>
<td></td>
<td></td>
</tr>
<tr>
<td>EIF45409</td>
<td>Galactokinase</td>
<td>+ + + + + + + + + + + + + + + + + + +</td>
</tr>
<tr>
<td>EIF45411</td>
<td>dTdp-glucose-dehydratase</td>
<td>+ + + + + + + + + + + + + + + + + + +</td>
</tr>
<tr>
<td>EIF45413</td>
<td>Maltase</td>
<td>+ + + + + + + + + + + + + + + + + + +</td>
</tr>
<tr>
<td>EIF45415</td>
<td>β-glucosidase</td>
<td>+ + + + + + + + + + + + + + + + + + +</td>
</tr>
<tr>
<td>Contig AHIQ01000303</td>
<td></td>
<td></td>
</tr>
<tr>
<td>EIF45248</td>
<td>β-galactosidase</td>
<td>+ + + + + + + + + + + + + + + + + + +</td>
</tr>
<tr>
<td>EIF45249</td>
<td>nitrate reductase</td>
<td>+ + + + + + + + + + + + + + + + + + +</td>
</tr>
<tr>
<td>EIF45250</td>
<td>nitrite reductase</td>
<td>+ + + + + + + + + + + + + + + + + + +</td>
</tr>
<tr>
<td>EIF45251</td>
<td>nitrate transporter</td>
<td>+ + + + + + + + + + + + + + + + + + +</td>
</tr>
</tbody>
</table>

\(^a\) As determined by PCR amplification using primers targeting the almost complete ORF (for primers see Table S1); PCR amplification was performed with 10 ng genomic DNA. All bands were of the expected size;

\(^b\) Strains marked with an asterisk were able to utilize nitrate as sole nitrogen source. All isolates, with exception of ST05.12/30 and ST05.12/54, were able to utilize ammonium in our assay;

\(^c\) +, band; blank, no band.
FIGURE 1
B. naardenensis
B. anomalus (D. anomala)
B. custersianus
B./D. bruxellensis

FIGURE 2
FIGURE 3
FIGURE 6

\[
\begin{array}{cccc}
\text{NH}_4^+ & & \text{NO}_3^- \\
A & B & C & D \\
\end{array}
\]