F1000Research

Check for updates

DATA NOTE

REVISED BeerDeCoded: the open beer metagenome project [version 2; referees: 2 approved, 1 approved with reservations]

Jonathan Sobel iD , Luc Henry iD , Nicolas Rotman, Gianpaolo Rando

Hackuarium Association, Renens, Switzerland

Abstract

Next generation sequencing has radically changed research in the life sciences, in both academic and corporate laboratories. The potential impact is tremendous, yet a majority of citizens have little or no understanding of the technological and ethical aspects of this widespread adoption. We designed BeerDeCoded as a pretext to discuss the societal issues related to genomic and metagenomic data with fellow citizens, while advancing scientific knowledge of the most popular beverage of all. In the spirit of citizen science, sample collection and DNA extraction were carried out with the participation of non-scientists in the community laboratory of Hackuarium, a not-for-profit organisation that supports unconventional research and promotes the public understanding of science. The dataset presented herein contains the targeted metagenomic profile of 39 bottled beers from 5 countries, based on internal transcribed spacer (ITS) sequencing of fungal species. A preliminary analysis reveals the presence of a large diversity of wild yeast species in commercial brews. With this project, we demonstrate that coupling simple laboratory procedures that can be carried out in a non-professional environment with state-of-the-art sequencing technologies and targeted metagenomic analyses, can lead to the detection and identification of the microbial content in bottled beer.

Open Peer Review

Referee Status: ✔ ✔ ?

	Invited Referees		
	1	**2**	**3**
REVISED version 2 published 20 Oct 2017	✔ report	✔ report	? report
version 1 published 11 Sep 2017	? report	? report	? report

1 **Bastian Greshake** iD , Goethe University, Germany

2 **Matthew L. Bochman** iD , Indiana University Bloomington, USA

3 **Kristoffer Krogerus** iD , VTT Technical Research Centre of Finland Ltd., Finland

Discuss this article

Comments (0)

Corresponding authors: Jonathan Sobel (jonathan.sobel@unil.ch), Gianpaolo Rando (gianpaolo.rando@swissdecode.com)

Author roles: Sobel J: Conceptualization, Data Curation, Formal Analysis, Funding Acquisition, Investigation, Methodology, Project Administration, Resources, Software, Supervision, Validation, Visualization, Writing – Original Draft Preparation, Writing – Review & Editing; **Henry L**: Conceptualization, Funding Acquisition, Project Administration, Supervision, Validation, Writing – Original Draft Preparation, Writing – Review & Editing; **Rotman N**: Formal Analysis, Investigation, Methodology, Resources; **Rando G**: Conceptualization, Funding Acquisition, Investigation, Methodology, Project Administration, Resources, Supervision, Validation, Writing – Original Draft Preparation, Writing – Review & Editing

Introduction

Beer is probably the world's oldest and most widely consumed alcoholic beverage on the planet, with a worldwide production of nearly 2 billion hectolitres (2·10E11 litres) annually [The Barth Report, Hops 2015/2016], and, as DNA sequencing becomes increasingly cheap, whole genome sequencing and metagenomic analyses are being explored as tools to better understand brewing in particular, and food fermentation in general[1]. Complex microbial communities influence the wine- and cheesemaking process throughout[2,3]. Indeed, microbial communities contribute to nutritional and aromatic properties, as well as shelf life of the products. In the case of wine, microorganisms are present in the soil, on the grapes, and in the fermenter, being carried over from the vine to the must to the wine, and there is increasing evidence for the existence of an important microbial contribution to the notion of "terroir" (i.e regional environmental factors that affect the properties of the final product)[4–7]. One question that remains unanswered is whether there is such a thing as a "terroir" for beer.

Of particular interest is sour beers, such as lambic and gueuze, beverages produced without the controlled addition of known yeast cultivates. Instead, the wort is exposed to ambient air, allowing naturally occurring bacteria and yeasts to start the fermentation and leading to a production that is difficult to standardize. To our knowledge, three initiatives are currently exploring the role of the beer microbiome in the brewing process and how it shapes the characteristics of the final product. Using metagenomic analyses, Kevin Verstrepen and colleagues at KU Leuven, Belgium, study the production of lambic, a traditional Belgian beer produced by spontaneous fermentation [VIB project 35]. Similarly, Matthew Bochman and colleagues at Indiana University, USA, have recently published preliminary results showing how the microbial community evolved over the fermentation process, together with the relative abundance of the organic acids that give sour beer its characteristic taste[8,9]. Similarly, researchers at the University of Washington, USA, have studied open-fermentation beer and discovered a novel interspecific hybrid yeast[10].

To investigate the microbial composition of a collection of commercial beers, we initiated BeerDeCoded in the context of Hackuarium, a Swiss not-for-profit organisation that supports unconventional research projects and promotes the public understanding of science. Members of the Hackuarium community are interested in participatory biology and want to promote interdisciplinary citizen research and innovation outside traditional institutions, using low-cost, simple and accessible technologies. The goal of the BeerDeCoded project is not only to broaden the scientific knowledge about beer, but also to improve the public understanding of issues related to personal genomics, food technology, and their role in society. With the release of this first data set, we built the proof of concept for a targeted metagenome analysis pipeline for beer samples that can be used in high schools, citizen science laboratories, craft breweries or industrial plants.

Methods

Beer sample preparation

The content of each beer sample was mixed to homogeneity by inversing the bottle several times. 50 mL were transferred into a conical tube and centrifuged (5000 rpm, 20 min, 4°C) to collect cells and other precipitable material. Pellets were resuspended with 1 mL TE buffer (Tris 10 mM, EDTA 1 mM,pH 8.0) and transferred into 1.5 mL tubes. The samples were centrifuged (10000 rpm, 10 min, 4°C), the supernatant was removed and the pellet stored frozen (-20°C) until future analyses. The ZR Fecal DNA MiniPrep kit (Zymo Research) was used for DNA extraction with minor modifications to the original protocol[11]. Sludge pellets were used instead of the 50-100 mg of fecal material suggested by the manufacturer.

Quality control for DNA extraction

To ensure the DNA was free from proteins and other contaminants, the absorbance of DNA samples was measured at 230, 260 and 280 nm using a NanoDrop 2000 spectrophotometer (Thermo Fisher Scientific).

ITS amplification

Yeast genomic DNA was amplified using the fungal hypervariable region ITS1 (internal transcribed spacer 1) as previously described[11] using the following primers: BITS (5'–CTACCTGCGGARG-GATCA–3') and B58S3 (5'– GAGATCCRTTGYTRAAAGTT–3'). Typical PCR reactions contained 5–100ng of DNA template. Amplicon size (500nt) was verified using gel electrophoresis and with a fragment analyser. ITS amplicons were purified using AMPure XP beads following the manufacturer's instructions (Beckman Coulter). Dual indices and Illumina sequencing adapters were attached using the Nextera XT Index Kit following manufacturer's instructions (Illumina).

Sequencing

MiSeq sequencing was performed using the MiSeq v3 reagent kit protocol (Illumina). Briefly, the amplified DNA was quantified using a fluorimetric method based on ds-DNA binding dyes (Qubit). Each DNA sample was diluted to 4 nM using 10 mM Tris pH 8.5 and 5 uL of diluted DNA from each library were pooled. In preparation for cluster generation and sequencing, 5 uL of the pooled final library was denatured with 5 uL of freshly diluted

0.2 N NaOH and combined with 30% PhiX control library to serve as an internal control for low-diversity libraries. After loading the samples on the MiSeq, paired 2x 300bp reads were generated and exported as FASTq files.

Bioinformatics analysis

The curated set of ITS sequences from the Refseq database (Targeted Loci) was used to build an ITS index for the Burrows-Wheeler Aligner (BWA, version 0.7.13)[12]. The BWA was used with standards parameters to map the paired-end reads of each beer from the fastq files to our ITS index. The BAM files were sorted and indexed using samtools[13]. A quality control of the BAM files was performed using SAMstat (version 1.5)[14]. A read quality threshold above 3 (MAPQ score) was applied in order to remove low quality and non-unique mapping reads. Subsequently, the number of ITS per beer and per species were counted and only species with over 10 reads were taken into consideration. Visualization of the results were performed with R (version 3.4.0).

Results

Over the month of June 2015, a total of 124 individuals contributed over 10,000 Euros to a crowdfunding campaign that provided financial resources for the first stage of the BeerDeCoded project. Reaching out to the public through this campaign also enabled crowdsourcing a collection of 120 beer samples from 20 countries. We have subsequently demonstrated that it is possible to extract DNA directly from bottled beer using low cost methodologies, typically available to citizen scientists (see Methods).

The internal transcribed spacer regions (ITS) of fungal species[15] were then amplified and, after quality control, 39 samples were sent for DNA sequencing. These 39 commercial beers originated from 5 different countries: 30 were from Switzerland, five from Belgium, two from Italy, one from France and one from Austria. We obtained an average library size of 600K reads (min 350K, max 2400K see Table 1) with more than 99% of reads mapping to the ITS database per sample.

A total of 42 fungal species were identified, 24 of which were present only in a single brew. This high variety of wild yeasts in commercial beers was unexpected (Figure 1 A), with some brews containing traces of up to more than 10 different fungal species (Figure 1 B). The beer in which we measured the highest ITS diversity (19 fungal species) was Waldbier 2014 Schwarzkiefer, an
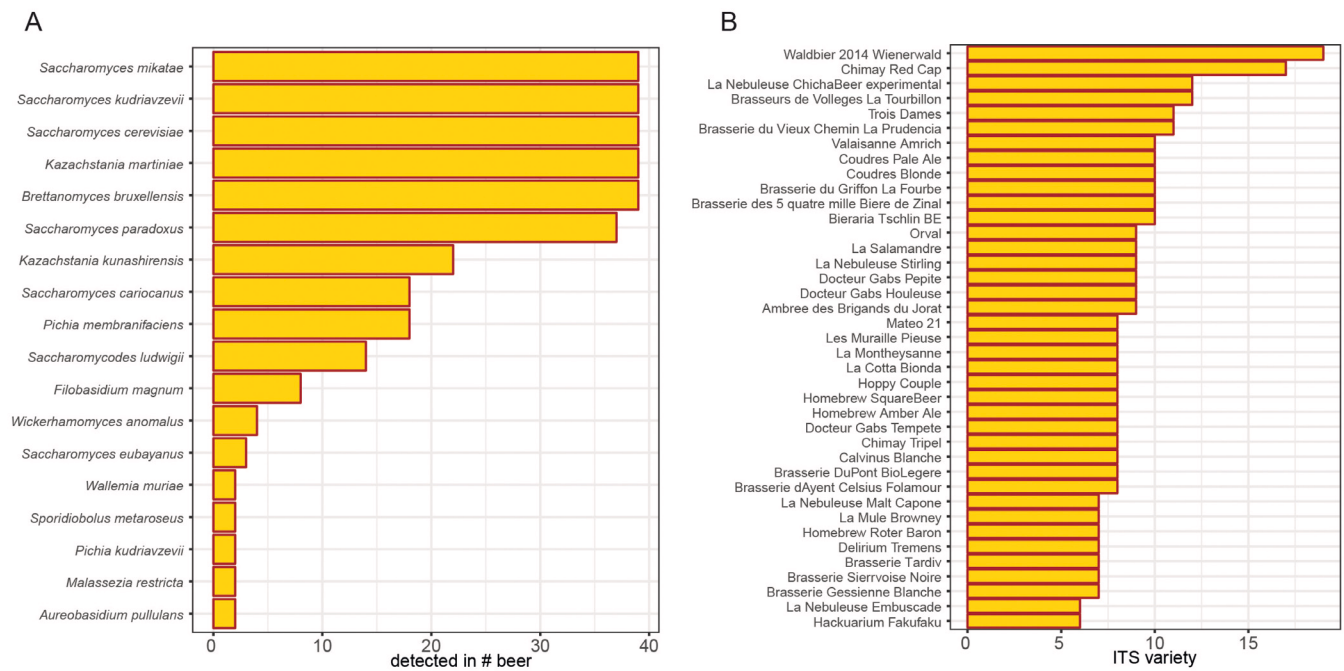
Figure 1. Barplot graph representing. (**A**) the number of beers containing the species (n=36) occurring in at least two samples. Species (n=52) present in only one sample were excluded for clarity. (**B**) represents the number of fungal species identified in each of the 39 bottled beers.

Table 1. Sequencing libraries statistics.

Beer library	total read count	unmapped read count	mapping percentage [%]
Ambree des Brigands du Jorat	645239	20674	99,97
Bieraria Tschlin BE	640291	15700	99,98
Brasserie dAyent Celsius Folamour	634121	17162	99,97
Brasserie des 5 quatre mille Biere de Zinal	600066	16194	99,97
Brasserie du Griffon La Fourbe	377774	9592	99,97
Brasserie du Vieux Chemin La Prudencia	454462	12490	99,97
Brasserie DuPont BioLegere	483889	12486	99,97
Brasserie Gessienne Blanche	379492	6164	99,98
Brasserie Sierrvoise Noire	353605	14524	99,96
Brasserie Tardiv	585357	18061	99,97
Brasseurs de Volleges La Tourbillon	418709	11948	99,97
Calvinus Blanche	473262	10944	99,98
Chimay Red Cap	552594	7806	99,99
Chimay Tripel	587167	11089	99,98
Coudres Blonde	652259	36080	99,94
Coudres Pale Ale	431653	16170	99,96
Delirium Tremens	627271	10432	99,98
Docteur Gabs Houleuse	681220	25303	99,96
Docteur Gabs Pepite	597756	10987	99,98
Docteur Gabs Tempete	644890	8640	99,99
Hackuarium Fakufaku	489232	9443	99,98
Homebrew Amber Ale	578714	12711	99,98
Homebrew Roter Baron	350211	17288	99,95
Homebrew SquareBeer	421463	8670	99,98
Hoppy Couple	612653	14486	99,98
La Cotta Bionda	681507	16781	99,98
La Montheysanne	381861	11297	99,97
La Mule Browney	402023	6068	99,98
La Nebuleuse ChichaBeer experimental	670287	18244	99,97
La Nebuleuse Embuscade	591798	22089	99,96
La Nebuleuse Malt Capone	637362	15770	99,98
La Nebuleuse Stirling	681512	29075	99,96
La Salamandre	643011	18059	99,97
Les Muraille Pieuse	391583	11006	99,97
Mateo 21	392582	7935	99,98
Orval	524342	21694	99,96
Trois Dames	2600874	78441	99,97
Valaisanne Amrich	368145	8278	99,98
Waldbier 2014 Schwarzkiefer	350005	16224	99,95

Austrian beer brewed using pine cones collected in local forests. Two other beers contained more than 12 fungal species: La Nébuleuse Cumbres Rijkrallpa (a sour/wild ale beer made with cranberries and the fermented corn "Chicha") and Chimay Red Cap, a Belgian trappist beer. Using hierarchical clustering, we built a proximity tree of the different beers (Figure 2).

Consistent with its widespread use for fermentation, brewer's yeast (*Saccharomyces cerevisiae*) was detected in all beer samples, accounting for between 11% (Orval, an ale beer by Belgian Brasserie d'Orval) and 99% (Tempête, an ale from the Swiss brewery Docteur Gab's) of all sequencing reads. In most samples, *S. cerevisiae* was present at very high levels (typically 90–97% of reads, Figure 3). More surprisingly, *Saccharomyces mikatae*, a species used in winemaking[16] was also relatively abundant in all samples (0.5–5%). Interestingly, most brews were found to contain low to medium abundance of multiple other yeast species, including *Saccharomyces kudriavzevii* and *Saccharomyces eubayanus* (a probable parent of *Saccharomyces pastorianus*) and *Brettanomyces bruxellensis* (typically used for the production of the Belgian beers). Non-conventional, as well as wild yeast, such as *Saccharomyces cariocanus* and *Saccharomyces paradoxus*, two species closely related to *Saccharomyces cerevisiae* were also found. Another example is *Kazachstania sp.*, a wild yeast of commonly found in brines[17]. The presence of this species may be of interest, as it was previously reported that adding the parent *Kazachstania servazzi* to the brewing process 24 hours before the ale yeast contributed to the production of high level of esters, producing a strong fruity and floral aroma[18].

Discussion and future perspectives
While a continuous process of market consolidation has lead to 5 companies controlling more than half of global beer production, there has been an explosion of craft industries over the past years, especially in Europe and North America. In 1978 there were 89 large industrial breweries in the USA. In 2016, there were 5,301, among them 3,132 small, independent microbreweries (American Brewers Association). There is a parallel with Hackuarium, an independent "craft" science initiative that has branched out from large institutional research institutes and provides an environment that allows scientists to explore topics that are rarely found in academia or industry. What is truly unique is the participation of individuals with no formal science training, and therefore the strong focus on citizen science and communication. With the BeerDeCoded project, we explored the potential of crowdfunding and crowdsourcing in engaging members of the general public in the production of scientific knowledge. We demonstrated that it is possible to execute complex molecular analyses on everyday products using limited resources and technical support from research institutions, and no financial support from traditional funding sources. The resulting dataset contains the ITS profile of 39 bottled beers from five different countries, revealing the low abundance but widespread presence of wild fungal species. It is a proof of concept that sequencing beer metagenomic information can be done, at least partly, with the help of the public. For the current analysis, we relied on high-throughput sequencing technology available to us through a partnership, a technology that may be out of reach for individuals working in non-traditional research environments. In the future, we would like to overcome this limitation, for example by providing a pipeline based on portable sequencing technologies, such as Oxford Nanopore's minION instrument. Further analyses could also go as far as shedding light on the so-called biological "dark matter" of the beer ecosystem[19,20].

With the costs of DNA sequencing falling dramatically, and with the emergence of portable and user-friendly instrumentation, we believe that it is a favorable time to expand the application of DNA analysis to novel fields, including food and beverage. This industry is starting to explore the potential of genome sequencing to understand the contribution of various species to product characteristics. The sequencing of the full genome of 157 brewing yeast strains was, for example, recently reported[21]. Metagenomic analyses could also have important implications for the optimization and batch-to-batch reproducibility of the various fermentation processes, as well as quality control, traceability and authentication of the products. One hypothesis that could be investigated further in the future is whether the presence of a specific fungal species can be diagnostic for a unique geographic area. In our data set, the non *Saccharomyces* yeast that contributes to wine aroma through the production of volatile compounds, *Wickerhamomyces anomalus*, was found exclusively in five of the brews manufactured in Switzerland. The limited sample size, however, does not allow us to draw a statistically significant conclusion, and it remains to be seen if *W. anomalus* is present in beers from other locations as well. Due to inherent limitations of DNA sequencing, it is difficult to anticipate whether the microbes identified are likely to be having an impact on the fermentation process. However, based on the identification of strains present in brews with desired characteristics, controlled experiments in which the microbial composition of the brew is altered could allow us to investigate if the presence of specific microorganisms affects flavour[22]. The origin of each yeast species could also be investigated; i.e. whether they come with the ingredients or from the environment at the production site. Techniques to sample airborne DNA exist[23]. Furthermore, other protocols could also be used to catalogue plant DNA[24], such as malt and hop varieties, and to map the bacterial diversity.
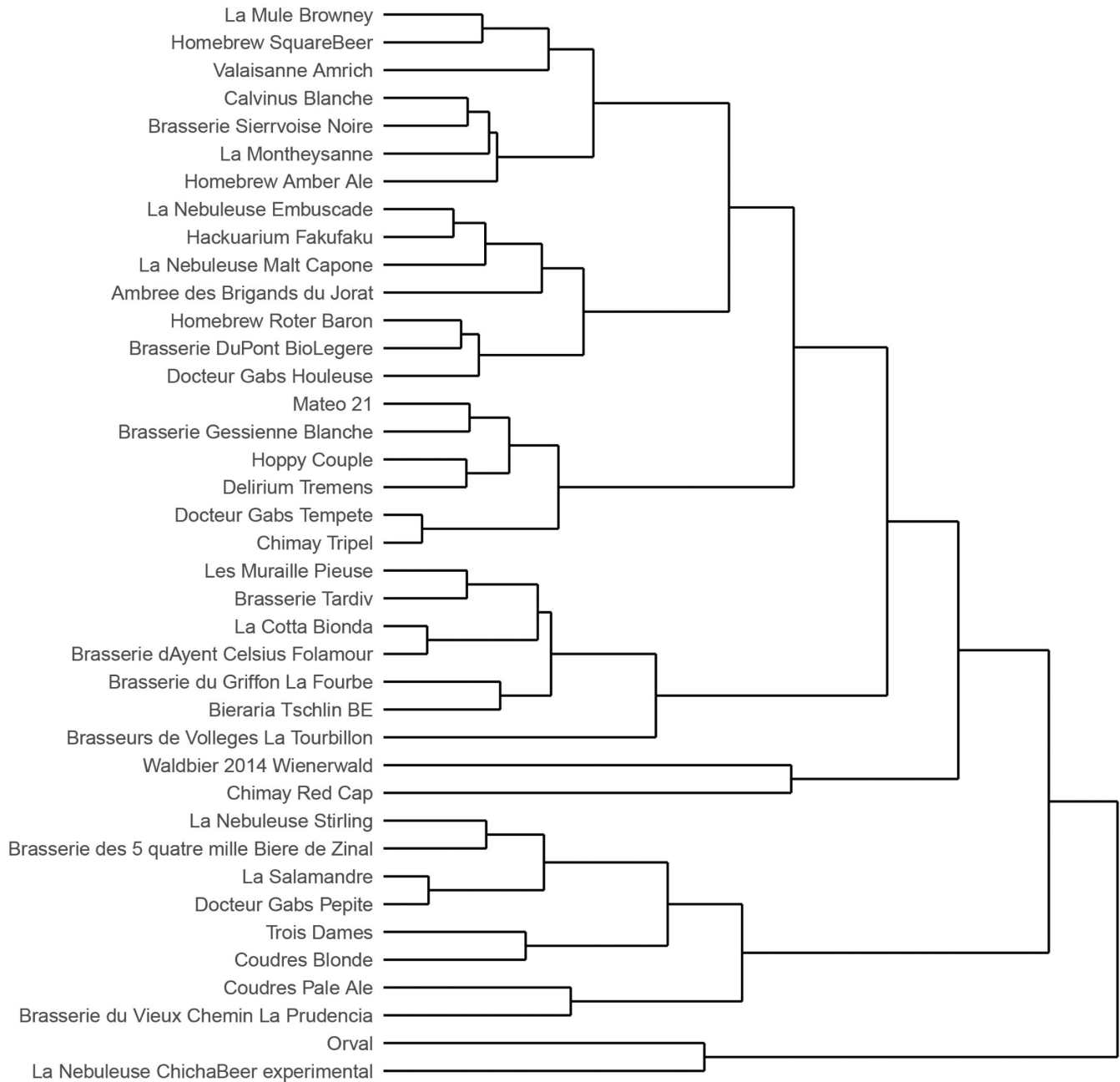
Figure 2. Hierarchical clustering of the 39 beers included in this study, based on their fungal content. We applied the Ward's method on the Euclidean distance computed on the log10 counts matrix.
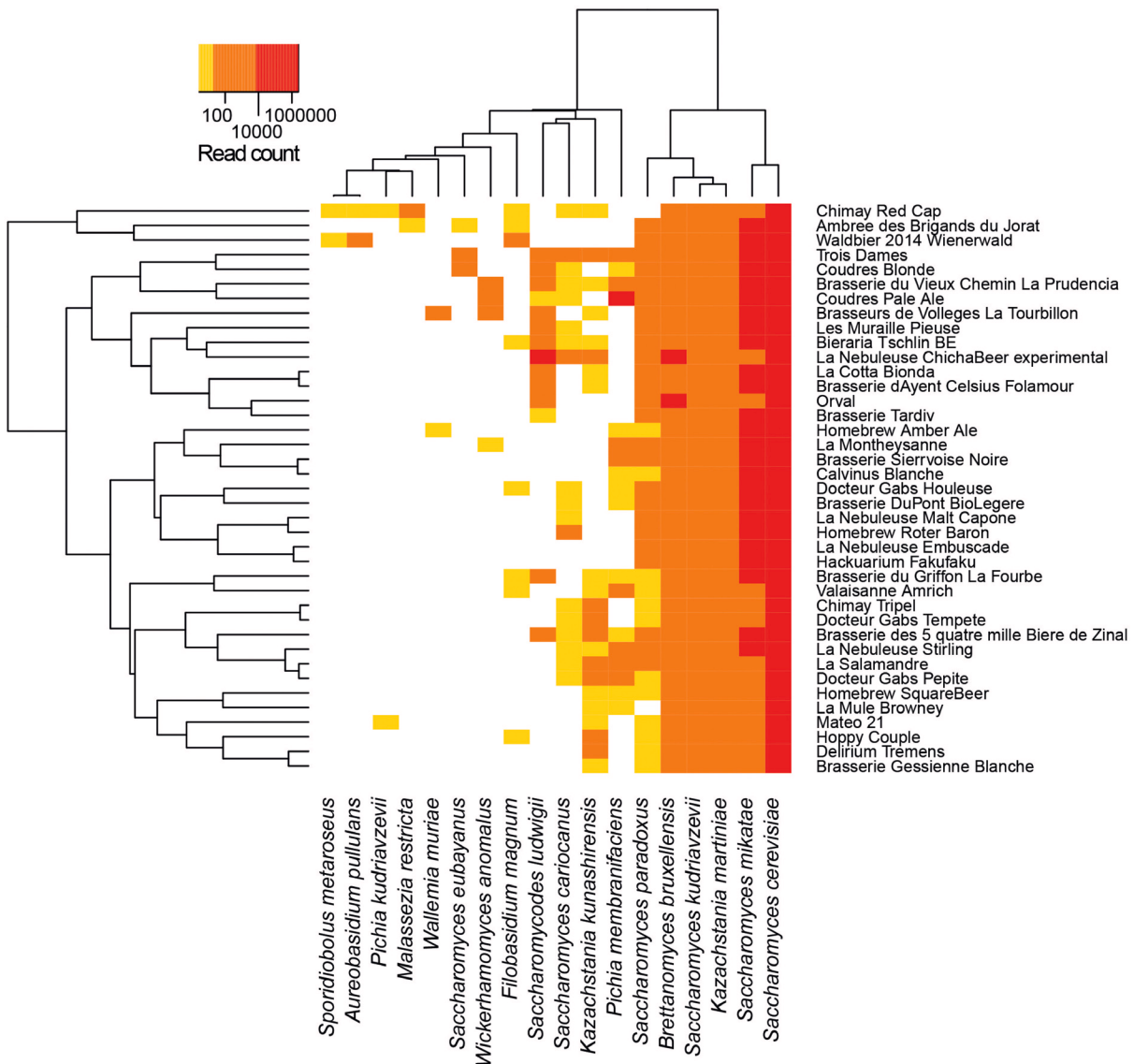
Figure 3. Heatmap of the number of reads per ITS per beer. Only ITS with more than 10 reads and present in at least two beers are shown.

In order to standardize and simplify our pipeline, and facilitate the contribution of new data and their further analysis by individuals not involved in this initial study, we are in the process of developing a BeerDeCoded repository and a Galaxy instance[25]. This tool will enable any citizen scientist to carry out beer metagenomics and reproduce our analysis. In the meantime, we encourage researchers from other laboratories, microbreweries and citizen laboratories to further explore our data set, and invite them to consider contributing additional data in the future.

Data availability

The dataset contains the metagenomic profiles for 39 beers. The data was obtained using a targeted approach based on the phylogenetic typing with internal transcribed spacers (ITS) of ribosomal sequences. All methods, quality control, processed tables, metadata and code are accessible at: https://github.com/beerdecoded/Beer_ITS_analysis. The raw data are stored in the SRA database in the bio project PRJNA388541

References

1. De Filippis F, Parente E, Ercolini D: **Metagenomics insights into food fermentations.** *Microb Biotechnol.* 2017; **10**(1): 91–102.
   PubMed Abstract | **Publisher Full Text** | **Free Full Text**

2. Liu Y, Rousseaux S, Tourdot-Maréchal R, *et al.*: **Wine microbiome: A dynamic world of microbial interactions.** *Crit Rev Food Sci Nutr.* 2017; **57**(4): 856–873.
   PubMed Abstract | **Publisher Full Text**

3. Delcenserie V, Taminiau B, Delhalle L, *et al.*: **Microbiota characterization of a Belgian protected designation of origin cheese, herve cheese, using metagenomic analysis.** *J Dairy Sci.* 2014; **97**(10): 6046–56.
   PubMed Abstract | **Publisher Full Text**

4. Bokulich NA, Collins TS, Masarweh C, *et al.*: **Associations among Wine Grape Microbiome, Metabolome, and Fermentation Behavior Suggest Microbial Contribution to Regional Wine Characteristics.** *mBio.* 2016; **7**(3): pii: e00631-16.
   PubMed Abstract | **Publisher Full Text** | **Free Full Text**

5. Bokulich NA, Thorngate JH, Richardson PM, *et al.*: **Microbial biogeography of wine grapes is conditioned by cultivar, vintage, and climate.** *Proc Natl Acad Sci U S A.* 2014; **111**(1): E139–E148.
   PubMed Abstract | **Publisher Full Text** | **Free Full Text**

6. Pinto C, Pinho D, Cardoso R, *et al.*: **Wine fermentation microbiome: a landscape from different Portuguese wine appellations.** *Front Microbiol.* 2015; **6**: 905.
   PubMed Abstract | **Publisher Full Text** | **Free Full Text**

7. Belda I, Zarraonaindia I, Perisin M, *et al.*: **From Vineyard Soil to Wine Fermentation: Microbiome Approximations to Explain the "*terroir*" Concept.** *Front Microbiol.* 2017; **8**: 821.
   PubMed Abstract | **Publisher Full Text** | **Free Full Text**

8. Bochman ML, Young J: **Mapping the sour beer microbiome.** Experiment.com. Technical report, Indiana University; Wild Pitch Yeast, LLC and Blue Owl Brewing, 2017.
   Reference Source

9. Osburn K, Amaral J, Metcalf SR, *et al.*: **Primary souring: a novel bacteria-free method for sour beer production.** *bioRxiv.* 2017.
   Publisher Full Text

10. Heil CS, Burton JN, Liachko I, *et al.*: **Identification of a novel interspecific hybrid yeast from a metagenomic open fermentation sample using Hi-C.** *bioRxiv.* 2017.
    Publisher Full Text

11. Bokulich NA, Bergsveinson J, Ziola B, *et al.*: **Mapping microbial ecosystems and spoilage-gene flow in breweries highlights patterns of contamination and resistance.** *eLife.* 2015; **4**: e04634.
    PubMed Abstract | **Publisher Full Text** | **Free Full Text**

12. Li H, Durbin R: **Fast and accurate short read alignment with Burrows-Wheeler transform.** *Bioinformatics.* 2009; **25**(14): 1754–1760.
    PubMed Abstract | **Publisher Full Text** | **Free Full Text**

13. Li H, Handsaker B, Wysoker A, *et al.*: **The Sequence Alignment/Map format and SAMtools.** *Bioinformatics.* 2009; **25**(16): 2078–2079.
    PubMed Abstract | **Publisher Full Text** | **Free Full Text**

14. Lassmann T, Hayashizaki Y, Daub CO: **Samstat: monitoring biases in next generation sequencing data.** *Bioinformatics.* 2011; **27**(1): 130–131.
    PubMed Abstract | **Publisher Full Text** | **Free Full Text**

15. Schoch CL, Seifert KA, Huhndorf S, *et al.*: **Nuclear ribosomal internal transcribed spacer (its) region as a universal dna barcode marker for fungi.** *Proc Natl Acad Sci U S A.* 2012; **109**(16): 6241–6246.
    PubMed Abstract | **Publisher Full Text** | **Free Full Text**

16. Bellon JR, Schmid F, Capone DL, *et al.*: **Introducing a new breed of wine yeast: interspecific hybridisation between a commercial *Saccharomyces cerevisiae* wine yeast and *Saccharomyces mikatae*.** *PLoS One.* 2013; **8**(4): e62053.
    PubMed Abstract | **Publisher Full Text** | **Free Full Text**

17. Hill AE: **Brewing Microbiology: Managing Microbes, Ensuring Quality and Valorising Waste.** *Number 289 in Woodhead Publishing Series in Food Science, Technology and Nutrition.* Woodhead Publishing Ltd., 1 edition, 2015; **6**.
    Reference Source

18. Gibson B, Krogerus K, Ekberg J, *et al.*: **Non-conventional yeast as a new tool for beer flavour modification.** Technical report, VTT Technical Research Centre of Finland, 2015.
    Reference Source

19. Robbins RJ, Krishtalka L, Wooley FC: **Advances in biodiversity: metagenomics and the unveiling of biological dark matter.** *Stand Genomic Sci.* 2016; **11**(1): 69.
    PubMed Abstract | **Publisher Full Text** | **Free Full Text**

20. Rinke C, Schwientek P, Sczyrba A, *et al.*: **Insights into the phylogeny and coding potential of microbial dark matter.** *Nature.* 2013; **499**(7459): 431–7.
    PubMed Abstract | **Publisher Full Text**

21. Gallone B, Steensels J, Prahl T, *et al.*: **Domestication and Divergence of *Saccharomyces cerevisiae* Beer Yeasts.** *Cell.* 2016; **166**(6): 1397–1410.e16.
    PubMed Abstract | **Publisher Full Text** | **Free Full Text**

22. Hong X, Chen J, Liu L, *et al.*: **Metagenomic sequencing reveals the relationship between microbiota composition and quality of Chinese Rice Wine.** *Sci Rep.* 2016; **6**(1): 26621.
    PubMed Abstract | **Publisher Full Text** | **Free Full Text**

23. Jiang W, Liang P, Wang B, *et al.*: **Optimized DNA extraction and metagenomic sequencing of airborne microbial communities.** *Nat Protoc.* 2015; **10**(5): 768–779.
    PubMed Abstract | **Publisher Full Text**

24. Nakamura S, Tsushima R, Ohtsubo K: **A novel method for the preparation of template DNA for PCR from beer to detect materials and to develop DNA markers to evaluate the quality of beer.** *Biosci Biotechnol Biochem.* 2013; **77**(4): 820–831.
    PubMed Abstract | **Publisher Full Text**

25. Afgan E, Baker D, van den Beek M, *et al.*: **The galaxy platform for accessible, reproducible and collaborative biomedical analyses: 2016 update.** *Nucleic Acids Res.* 2016; **44**(W1): W3–W10.
    PubMed Abstract | **Publisher Full Text** | **Free Full Text**

F1000Research

Open Peer Review

Current Referee Status: ✔ ✔ ❓

Version 2

Referee Report 02 November 2017

doi:10.5256/f1000research.13990.r27165

✔ **Bastian Greshake** 🆔
Department for Applied Bioinformatics, Institute for Cell Biology and Neuroscience, Goethe University, Frankfurt am Main, Germany

The authors have satisfactory addressed the issues mentioned in the review of the first article version.

Competing Interests: No competing interests were disclosed.

Referee Expertise: Fungal metagenomics & bioinformatics

I have read this submission. I believe that I have an appropriate level of expertise to confirm that it is of an acceptable scientific standard.

Referee Report 02 November 2017

doi:10.5256/f1000research.13990.r27163

❓ **Kristoffer Krogerus** 🆔
VTT Technical Research Centre of Finland Ltd., Espoo, Finland

I thank the authors for addressing all the concerns I brought up. However, while the post-alignment QC has reduced the amount of false positives, I'm still concerned about the lack of pre-alignment read trimming (which should be standard practice). The methods section or material on the project's github page make no mention of any trimming of the reads prior to alignment. As demonstrated in my first referee report, by trimming potential adaptor and primer sequences and removing poor quality bases at the ends of reads, the fungal diversity of the beers is reduced further (i.e. the amount of false positives are decreased), even with lenient post-alignment filtering (MAPQ < 3). I would recommend that the authors either redo the analysis once more with proper pre-alignment filtering as well, or clearly state in the results and discussion section that no such step was carried out and that the results may therefore contain a large number of false positives. For example, it is highly unlikely that all beers would contain traces of *Saccharomyces mikatae*, a yeast that to my knowledge never before has been isolated from fermentation environments and only ever from forest samples in Asia (1, 2). As demonstrated by the results in my previous referee report, *S. mikatae* is no longer detected in the samples I analyzed when pre-alignment quality trimming was performed.

References

1. Naumov GI, James SA, Naumova ES, Louis EJ, Roberts IN: Three new species in the Saccharomyces sensu stricto complex: Saccharomyces cariocanus, Saccharomyces kudriavzevii and Saccharomyces mikatae.*Int J Syst Evol Microbiol*. 2000; **50 Pt 5**: 1931-42 PubMed Abstract | Publisher Full Text

2. Bellon JR, Schmid F, Capone DL, Dunn BL, Chambers PJ: Introducing a new breed of wine yeast: interspecific hybridisation between a commercial Saccharomyces cerevisiae wine yeast and Saccharomyces mikatae.*PLoS One*. 2013; **8** (4): e62053 PubMed Abstract | Publisher Full Text

Competing Interests: No competing interests were disclosed.

I have read this submission. I believe that I have an appropriate level of expertise to confirm that it is of an acceptable scientific standard, however I have significant reservations, as outlined above.

Referee Report 20 October 2017

doi:10.5256/f1000research.13990.r27164

Matthew L. Bochman (iD)

Molecular and Cellular Biochemistry Department, Indiana University Bloomington, Bloomington, IN, USA

The authors have adequately addressed all of my concerns.

Competing Interests: I also participated in a crowdfunding campaign to use next-gen enunciating to analyze beer samples (https://experiment.com/projects/mapping-the-sour-beer-microbiome).

I have read this submission. I believe that I have an appropriate level of expertise to confirm that it is of an acceptable scientific standard.

Version 1

Referee Report 28 September 2017

doi:10.5256/f1000research.13607.r26415

Kristoffer Krogerus (iD)

VTT Technical Research Centre of Finland Ltd., Espoo, Finland

This data note describes the fungal microbiome of 39 (commercial and homebrewed) beers as determined by next generation sequencing of ITS amplicons. The project was crowdfunded and many of the individual funders were also involved in providing beer samples and assistance during DNA extraction. While the results will be of interest, particularly for the brewing industry, I have some concerns with the analysis methods and the results presented in this first version of the manuscript.

Major comments:

To determine the microbiome of the beers, the authors align the raw sequencing reads to a concatenated dataset containing fungal ITS sequences. To my understanding, no quality control or filtering was

performed prior to and after the alignment. This will cause a large number of false positives, as many of the intragenic ITS sequences are very similar. To demonstrate, I repeated the analysis on six samples I retrieved from the NCBI SRA:

SRR5740352: Chimay Red Cap
SRR5740353: Chimay Tripel
SRR5740362: Waldbier 2014 Wienerwald
SRR5740364: La Nebuleuse Chicha
SRR5740374: Orval
SRR5740375: Trois Dames

According to the results presented in the manuscript, each of the samples contained traces of at least 11 different species (see Figure 1 and Figure 3).

What I did to the sequencing reads was
1. Trim them using 'cutadapt' as follows (any similar tool would do the same job)
- Remove 20 first bases of each read
- Remove bases from end of read when quality score is less than 15
- Remove any reads shorter than 200 base pairs
- Approximately 80% of the bases were retained from each set of reads after these steps
2. After this the reads were aligned to the concatenated ITS sequence dataset using bwa mem with default settings as the authors had done. Reads mapping to the different sequences were then counted with the script used by the authors (obtainable from github). The results with no post alignment filtering: https://www.dropbox.com/s/llg94fgk23ag264/Beer_results_nofilter.txt

3. After this I removed all reads that did not map to a unique location (i.e. could be mapped to the ITS sequences of several species) and reads where the two paired reads mapped to different sequences. This was done by removing all alignments with a MAPQ score below 4 and 'awk': https://www.dropbox.com/s/0iimh5fbb40qeh0/Beer_results_mapq4.txt

As can be seen, the diversity is reduced considerably, and if all hits where the read count is less than 10 are also removed (as the authors had done), most samples now contain only *S. cerevisiae* and/or *B. bruxellensis*.

4. The amount of false positives can be further reduced by filtering by a higher MAPQ score (e.g. 30): https://www.dropbox.com/s/3x4gyylsiykyu7o/Beer_results_mapq30.txt

I therefore suggest that the authors redo the analysis with proper filtering to remove poor quality alignments and false positives in the results. The results and conclusions will subsequently have to be rewritten accordingly.

Minor comments:

In the Methods section, under Beer sample preparation: I assume the DNA was extracted from the frozen yeast pellet? Any reason why it was not attempted to extract DNA from the beer itself, e.g. using the method described in reference (23)? This would allow analysis of filtered beers as well.

Is Figure 2 necessary, since Figure 3 shows the clustering as well? Also, why does the clustering in Figures 2 and 3 differ? Were these generated using different clustering methodologies?

It is mentioned that some of the beers contained speciality ingredients, such as pine cones. Do the authors know at what point in the production process these were added (i.e. pre- or post-boil)? This would have a large impact on how these ingredients affect the beer microbiome.

Is the rationale for creating the dataset(s) clearly described?
Yes

Are the protocols appropriate and is the work technically sound?
Partly

Are sufficient details of methods and materials provided to allow replication by others?
Partly

Are the datasets clearly presented in a useable and accessible format?
Yes

Competing Interests: No competing interests were disclosed.

I have read this submission. I believe that I have an appropriate level of expertise to confirm that it is of an acceptable scientific standard, however I have significant reservations, as outlined above.

Author Response 10 Oct 2017
Luc Henry, Hackuarium, Switzerland

Regarding the major comments:
- We thank the referee for his thorough analysis of our results and for his comments. In our analysis, we initially choose to favor sensitivity over specificity and we did not filtered the reads based on their quality. We considered that, with 300 bp paired-end reads on ITS amplicons, we had enough specificity. But as referee 3 pointed out, we should take care of multiple mapping reads to for instance discriminate between the Saccharomyces sp. that are quite close to each other. Using a filter on MAPQ score > 30 is quite stringent, but the referee's point on non-unique reads is critical. We have now re-performed the analysis using a series of MAPQ threshold. As expected, the higher the threshold, the more we lose some of the ITS detected, and some sensitivity. In our update of the article, we report results after using a filter with a MAPQ score of 3. This led to a reduction of the total fungal species identified, from 88 to 42, as well as of the unique occurrences, from 52 to 24.
In addition we have performed quality control (QC) with samstat of our BAM files in order to see the quality score distribution of the reads of the alignment to the ITS database. The result of samstat has been added to our github repository. Samstat results show that about 3.2 % of reads are not aligned when reads with MAPQ score > 3 are taken into account. In addition we have checked the distribution of fragments size in the BAM files. We have an average fragment size of 383 bp with a sd of 10 bp. We have a very low fraction of fragments below 200 bp, or discordant pairs. Thanks to the comments of the Referee 3, we improved the specificity of the analysis and we may have excluded artifacts. Due to this re-analysis, we have updated Figure 1, Figure 2 and Figure 3, the code in our github repository as well as the main text. We have also added the QC informations.

Regarding the minor comments:

- With the goal to maximise our chances to obtain some preliminary results with the budget of the kickstarter campaign, we decided to use the pelleted material with the educated assumption that pelleted cells may protect DNA better than DNA left in solution. Reference (23) intrigued us: their DNA preparation method requires additional lyophilization, pulverization and digestion with amylase and we decided to start with an easier protocol to facilitate participation from the general public.
- We included both Figure 2 and Figure 3 to provide two different representations of the same results. The differences in the tree are due to the different display of the dendrogram. The same three is underling based on the same distance (euclidian distance on log10 counts matrix, with ward clustering method).
- The last comment is an interesting work hypothesis. We however do not know when these ingredients are added to the brewing process, and whether they were pre-treated before addition. If these ingredients contained living microorganisms at the moment of addition, they may indeed affect the brewing process and the taste of the final product. We believe collecting this metadata is outside the scope of this dataset description.

Competing Interests: No competing interests were disclosed.

? **Matthew L. Bochman** iD

Molecular and Cellular Biochemistry Department, Indiana University Bloomington, Bloomington, IN, USA

In this manuscript, Sobel *et al*. present fungal microbiome data from 39 different beers as the culmination of a crowdfunded citizen science campaign. These data will be of interest to citizen scientists and financial backers of the project, as well as those in the fermentation (especially beer) industry. Overall, the data seem sound, but I have some concerns:

Major comments
Were any controls for contamination used, *i.e.*, are all of the fungi identified actually from the beer samples? The sequencing of a non-beer sample such as water that had been handled in the same way as the beer samples would help to determine if fungal DNA contamination occurred during sample processing.

In line with the comment above, the manuscript states that "…microorganisms, or their DNA, could be carried over from the ingredients to the final product." Can the authors comment on whether they know if they are detecting the fungi themselves or DNA remnants from fungi that came from the raw ingredients of the beer? Again, one could add purified control DNA to a mash, brew and bottle a beer, and then try to detect that DNA by PCR in the end product (or even at various stages along the brewing and fermentation process). Attempting this with various concentrations of DNA would also yield information on how many cells of a particular species would be necessary on malted barley, for instance, to be detected in the final beer.

Minor comments

In the introduction, the authors state that "…sour beer…[is] produced without the controlled addition of known yeast cultivates." Although this may be true for some types of sour beer like lambic and gueuze, many sour beers made in the U.S. are inoculated with known strains of yeast. In those cases, the souring bacteria are usually the unknowns.

Why was a fecal DNA prep kit used for DNA extraction?

The authors collected 120 beers from 20 countries but only sequenced the fungi from 39 (mostly from Switzerland). Is there an explanation for this attrition?

Is the rationale for creating the dataset(s) clearly described?
Yes

Are the protocols appropriate and is the work technically sound?
Partly

Are sufficient details of methods and materials provided to allow replication by others?
Partly

Are the datasets clearly presented in a useable and accessible format?
Yes

Competing Interests: I also participated in a crowdfunding campaign to use next-gen enunciating to analyze beer samples (https://experiment.com/projects/mapping-the-sour-beer-microbiome).

I have read this submission. I believe that I have an appropriate level of expertise to confirm that it is of an acceptable scientific standard, however I have significant reservations, as outlined above.

Author Response 10 Oct 2017
Luc Henry, Hackuarium, Switzerland

Regarding major comments:
- We did not include a water sample to account for any contaminant during the DNA extraction process. Processing such a control at this point would not reflect the original experimental conditions. We will include this control in the future when we will process new beers.
- We did think of experiments to identify DNA remnants from raw ingredients but did not have the means to perform them. Indeed, the beer samples were sent from all over the world (Europe and Switzerland for the 39 reported in this data set) and we had no possibility to collect other related samples (raw ingredients, brewing environment, etc). We think that it is out of the scope of the current study. In this setup, it is not possible to know afterwards from which ingredient the DNA comes from and we comment on this in the text.

Regarding minor comments:
- We rephrased the sentence about sour beers accordingly.
- Similar to soil and fecal material, beer samples were found to contain PCR inhibitors (see Juvonen and Haikara, J. Inst. Brew. 115(3), 167–176, 2009) that can interfere with the preparation of samples for sequencing analyses. The ZR Fecal DNA MiniPrep kit can overcome this because it

provides filter columns designed to remove PCR inhibitors. This methodological choice was previously described by Bukolich and coworkers (eLife 2015;4:e04634 DOI: 10.7554/eLife.04634). - A majority of the samples we received were from industrial/filtered beers. Unfortunately the volumes we had at hand (typically a 330 mL bottle or below) did not yield enough material (DNA of good purity) to obtain sequencing results, as judged by QC of PCR products. In order to detect DNA in these beers, we would probably need to process a much larger volume of beer. We therefore did not include these samples in the final data set.

Competing Interests: No competing interests were disclosed.

Referee Report 13 September 2017

Bastian Greshake (iD)

Department for Applied Bioinformatics, Institute for Cell Biology and Neuroscience, Goethe University, Frankfurt am Main, Germany

This article describes how innovative, participant-driven research projects can create an interesting data set outside the traditional Academy. This is an extremely laudable goal and the resulting data will be of interest to a broad audience. In its current state the article is a hybrid between data note, methods article and research article that delivers preliminary results. Regardless of the form of the article, the methods section would benefit a lot from a more detailed description of the methodology (see details below). Similarly, the analyses and results are a bit lacking at this stage, especially with respect to the basic metrics of the data sets (again, see below).

Major comments

Methods
- the methods section should be significantly improved/extended for a better understanding. I'm aware that most/all the things are on *GitHub*, but having to crosslink these makes it hard to follow. Specifically the following things should be improved upon:
  - What modifications were done to the protocol of the *ZR Fecal DNA MiniPrep* kit? (suggestion: putting the modified protocol to https://www.protocols.io/ if deemed useful)
  - Which parameters were used to perform the *bwa* alignment?
  - What was the size of the reference database that was used for the mapping?
  - How was the hierarchical clustering done that is described in the methods section? Which clustering method was used? Which distance measure was used?

Results
- *"We obtained an average library size of 600K reads (min 350K, max 2400K)"* This is a rather large difference between the different libraries. Does the number of species found correlate with the sequencing depth? I.e. would you have found more species if you had sequenced more data for the smaller libraries? A rarefaction analysis would be useful to understand the impact of sequencing depth on species recovery. A minor, related suggestion: Having a table of sequencing statistics so that the reader can compare the samples.
- A major thing that is not mentioned in the results/discussion is the number of reads which could not be mapped against the reference database. How many reads of each library did not belong to any of the reference ITS sequences? And are the non-mapping reads similar to each other or can be

clustered into OTUs? This would be needed to understand how many species/OTUs are in a given sample but could not be classified due to a lack of reference database. Without this the ITS diversity in a sample cannot be correctly estimated.

Minor comments:
- *"we built the proof of concept for a targeted metagenome analysis pipeline for beer samples that can be used in high schools, citizen science laboratories, craft breweries or industrial plants"* It would be good to at least briefly discuss how this is currently limited by the need to have access to a high-throughput sequencer.
- It would be great if "*terroir*" could be defined in the introduction for those not too familiar with oenology
- *"a total of 88 fungal species were identified, including 52 unique occurrences"* are unique occurrences those species which are only found in a single beer? I'd suggest rephrasing it for a better understanding.
- *"Interestingly, most brews were found to contain low to medium presence of multiple other yeast species, including Saccharomyces bayanus (used in winemaking and cider fermentation), Saccharomyces kudriavzevii and Saccharomyces pastorianus (used in lager manufacturing), Saccharomyces eubayanus (a probable parent of Saccharomyces pastorianus) and Brettanomyces bruxellensis (typically used for the production of the Belgian beer styles)"* please include citations for these explanations of the different taxa.
- It's a matter of taste, but I recommend rethinking the use of *"microbial dark matter"*, c.f. http://merenlab.org/2017/06/22/microbial-dark-matter/ for an explanation of why.

Is the rationale for creating the dataset(s) clearly described?
Yes

Are the protocols appropriate and is the work technically sound?
Partly

Are sufficient details of methods and materials provided to allow replication by others?
Partly

Are the datasets clearly presented in a useable and accessible format?
Yes

Competing Interests: No competing interests were disclosed.

Referee Expertise: Fungal metagenomics & bioinformatics

I have read this submission. I believe that I have an appropriate level of expertise to confirm that it is of an acceptable scientific standard, however I have significant reservations, as outlined above.

Author Response 10 Oct 2017

Luc Henry, Hackuarium, Switzerland

We thank the referee for his constructive remarks, and in particular for his comments on our methods section. We have added the details he requested (see below).

Regarding major comments:

- The minor modifications to the ZR Fecal DNA MiniPrep kit instructions are already described in the methods section of the article, as well as on the methods description available in the GitHub repository. These consist of, instead of starting with 50-100 mg of fecal material as suggested by the manufacturer, using a sludge pellet obtained through centrifuging 50 mL of beer (as described in the methods). Similar to soil and fecal material, beer samples were found to contain PCR inhibitors (see Juvonen and Haikara, J. Inst. Brew. 115(3), 167–176, 2009). The ZR Fecal DNA MiniPrep kit can overcome this because it provides filter columns designed to remove PCR inhibitors. This choice was previously described by Bukolich and coworkers (eLife 2015;4:e04634 DOI: 10.7554/eLife.04634).
- In order to do the alignment we have used BWA (version 0.7.13) with standards parameters with 300 bp paired-end reads.
- The ITS reference database contains 5361 ITS sequences. The average size of ITS is 585 bp with a standard deviation of 90 bp.
- Hierarchical clustering was done by applying the basic Ward clustering algorithm with the euclidian distance computed on the log10 read count. We modified the legend of the Figure 2 accordingly.
- We have to acknowledge the relatively large size variability between the different libraries. The amount of ITS detected can be affected by the sequencing depth and by the richness of the beer ecosystem. Accordingly, if the sequencing is not deep enough, we will clearly miss some low abundance species. If the beer sample contains a low variety of fungal species, sequencing deeper will not however provide additional information. In our analysis, it does not seem to be any correlation between the library size and the variety of ITS detected. The sample producing the largest library (2,4 mio reads, "Les Trois Dames") is not the one with the largest detected ITS variety (11 species). Also, the beer sample containing the largest ITS diversity ("Waldbier 2014 Schwarzkiefer", with 38 fungal species) had a library size of only about 0.35 mio reads. A rarefaction analysis is beyond the scope of this dataset description. As suggested, we added a table (Table 1) with the different informations regarding the libraries such as the mapping percentage and the total number of reads. In this table we observe that more than 99% of reads map to the ITS database. Consequently, there is a limited interest to try to find missing species or Operational Taxonomic Units (OTUs), and we can reasonably conclude that the ITS database used is comprehensive enough.

Regarding minor comments:
- The current data-set is a proof of concept that sequencing beer metagenomic information can be done, at least partly, with the help of the public. For the current analysis, we indeed had to rely on high-throughput sequencing technology available to us through a partnership with the genomic facility at the University of Lausanne. In the future, we would like to overcome this limitation, e.g. by using a minION sequencer. A remark was added to the discussion.
- The text was modified to clarify the notion of "terroir".
- The text was modified to clarify the notion of "unique occurrences".
- The text was modified (species were removed) to reflect changes due to an updated sensitivity of the analysis (based on another referee's comment), and a reference was added.
- The so-called "microbial dark matter" concept is regularly used by microbiologist and we think that it points to an interesting hypothesis worth mentioning, although it has nothing to do with the dark matter in physics, as explained in the blog post mentioned.

Competing Interests: No competing interests were disclosed.

The benefits of publishing with F1000Research:

- Your article is published within days, with no editorial bias

- You can publish traditional articles, null/negative results, case reports, data notes and more

- The peer review process is transparent and collaborative

- Your article is indexed in PubMed after passing peer review

- Dedicated customer support at every stage

For pre-submission enquiries, contact research@f1000.com

F1000Research