



**HAL**  
open science

## Whole-genome sequencing and intraspecific analysis of the yeast species *Lachancea quebecensis*.

Kelle C Freel, Anne Friedrich, Véronique Sarilar, Hugo Devillers, Cécile Neuvéglise, Joseph Schacherer

### ► To cite this version:

Kelle C Freel, Anne Friedrich, Véronique Sarilar, Hugo Devillers, Cécile Neuvéglise, et al.. Whole-genome sequencing and intraspecific analysis of the yeast species *Lachancea quebecensis*.. *Genome Biology and Evolution*, 2016, 8 (3), pp.733-741. 10.1093/gbe/evv262 . hal-02635461

**HAL Id: hal-02635461**

**<https://hal.inrae.fr/hal-02635461>**

Submitted on 27 May 2020

**HAL** is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

**Whole-genome sequencing and intraspecific analysis  
of the yeast species *Lachancea quebecensis***

Kelle C. Freel<sup>1†</sup>, Anne Friedrich<sup>1†</sup>, Véronique Sarilar<sup>2,3</sup>, Hugo Devillers<sup>2,3</sup>, Cécile Neuvéglise<sup>2,3</sup>  
and Joseph Schacherer<sup>1\*</sup>

1. Department of Genetics, Genomics and Microbiology, University of Strasbourg - CNRS, UMR7156, Strasbourg, France
2. INRA, UMR1319 Micalis, Jouy-en-Josas, France
3. AgroParisTech, UMR1319 Micalis, Jouy-en-Josas, France

†These authors contributed equally to this work

\*Author for Correspondence: Joseph Schacherer, Department of Genetics, Genomics and Microbiology, University of Strasbourg, Strasbourg, France, +33 03 68 85 19 61, schacherer@unistra.fr

**Short title:** Population genomics of *L. quebecensis*

**Abstract**

The gold standard in yeast population genomics has been the model organism *Saccharomyces cerevisiae*. However, the exploration of yeast species outside the *Saccharomyces* genus is essential to broaden the understanding of genome evolution. Here, we report the analyses of whole genome sequences of 9 isolates from the recently described yeast species *Lachancea quebecensis*. The genome of one isolate was assembled and annotated, and the intraspecific variability within *L. quebecensis* was surveyed by comparing the sequences from the 8 other isolates to this reference sequence. Our study revealed that these strains harbor genomes with an average nucleotide diversity of  $\pi = 2 \times 10^{-3}$  which is slightly lower, although on the same order of magnitude, as that previously determined for *S. cerevisiae* ( $\pi = 4 \times 10^{-3}$ ). Our results show that even though these isolates were all obtained from a relatively isolated geographic location, the same ecological source, and represent a smaller sample size than is available for *S. cerevisiae*, the levels of divergence are similar to those observed in this model species. This divergence is essentially linked to the presence of two distinct clusters delineated according to geographic location. However, even with relatively similar ranges of genome divergence, *L. quebecensis* has an extremely low global phenotypic variance of 0.062 compared to 0.59 previously determined in *S. cerevisiae*.

**Keywords:** genome evolution, *Lachancea quebecensis*, phenotype, intraspecific diversity

## Introduction

The recent advent of affordable genome sequencing has opened an unprecedented door to understanding the molecular mechanisms driving genome evolution among closely related organisms. While sequence data provides a method to obtain a thorough understanding of diversification in a variety of eukaryotes, yeasts are the best models for in-depth genomic studies. Indeed, due to the small and compact nature of the genomes of the Saccharomycotina yeasts, the relative ease in obtaining high quality sequence data, and the vast number of natural isolates available, this group of organisms is well suited for intraspecific analysis and population genomic research (Dujon 2010; Liti and Schacherer 2011).

While the most studied groups among the Saccharomycotina are within the *Saccharomyces* genus, (Hittinger et al. 2010; Almeida et al. 2014), with the best known being *Saccharomyces cerevisiae* (Liti et al. 2009; Schacherer et al. 2009; Wang et al. 2012; Cromie et al. 2013; Skelly et al. 2013; Strobe et al. 2015), it is essential to broaden the view of genomic diversity among yeasts. To ensure a thorough understanding of different forces driving diversification, a range of species should be investigated. The genus *Lachancea* harbors multiple species, characterized by genomic histories that comprise a continuum of rearrangements, ranging from mostly collinear to significantly reshuffled genomes. Isolates are found all over the world and inhabit many niches including soil, plants, and insects as well as processed food and beverages (Naumova et al. 2007).

Recent studies have paved the way for the genus *Lachancea* to be a model system for the exploration of intraspecific diversity in both mitochondrial (mt) and nuclear genomes in non-*Saccharomyces* yeasts (Friedrich et al. 2012; Jung et al. 2012; Freel et al. 2014a; Friedrich et al. 2015). An analysis of the intraspecific diversity among the mitochondrial genomes of *Lachancea kluyveri* (Jung et al. 2012) suggested that there has been recent rapid evolution resulting in variable genomes, while the largest intraspecific survey of mt genomes in yeast to date, revealed that *Lachancea thermotolerans* harbors highly conserved sequences (Freel et al. 2014a). These studies demonstrated that different levels of purifying selection could influence genomes of closely related species. Furthermore, the first interspecific study of nuclear genome evolution outside of the genus *Saccharomyces*, focused on *Lachancea kluyveri*,

was recently reported (Friedrich et al. 2015) and revealed that even within a single genome, different evolutionary patterns can exist.

In the genus *Lachancea*, the most recently described species is *Lachancea quebecensis* (Freel et al., 2015). Thus far, the majority of isolates from this species have been obtained from Québec, Canada while two additional strains have also been reported from Japan (Naumova et al. 2007). First identified as a population related to *L. thermotolerans*, these isolates reportedly have demonstrated an ability to grow on melibiose, distinguishing them from their sister species (Lachance and Kurtzman 2011). While there are limited isolates available for this species, this lineage represents an opportunity to explore the genomes of a species very closely related to *L. thermotolerans*, ultimately enhancing the utility of the genus *Lachancea* as a model for genome evolution, population genomics, and mechanisms driving speciation. Additionally, the currently available *L. quebecensis* strains mainly originated from the same general geographic location and were all isolated from tree bark, allowing the possibility for fine scale comparison of isolates from the same ecological niche.

Here, we analyzed the genomes of 9 *L. quebecensis* isolates that support the recent delineation of this species and confirm that this lineage is most closely related to *L. thermotolerans*. The genome sequences of the 9 strains provided an opportunity for the analysis of a novel species and investigation of the divergence among isolates from a novel *Lachancea* species. Interestingly, we found that these genomes from isolates inhabiting a single environmental niche and relatively restricted geographic location harbor levels of genome diversity equal to that found in *S. cerevisiae*, while the phenotypic diversity is not as variable based on the value of the global phenotypic variance across all conditions tested. Our results highlight the importance of investigating various natural populations of yeast in order to obtain a global view of genome evolution in the Saccharomycotina.

## Results and Discussion

Nine *Lachancea quebecensis* isolates were used in our study. These strains were determined to be haploid, based on flow cytometry analysis, and share very similar electrophoretic karyotypes that demonstrate there is variability in the size of the largest chromosome (Figure S1). All strains belong to the recently described species *L. quebecensis* and are most closely related to *Lachancea thermotolerans* (Freel et al. 2015). They were all obtained from tree bark (either maple, oak, or birch) found in essentially two locations: either the Gaspé Peninsula or Saint Lawrence Valley, in the province of Québec, Canada (Table 1, Figure 1A) (Charron et al. 2014).

### Genome sequencing, assembly, and annotation

The 9 *L. quebecensis* isolates were sequenced with Illumina HiSeq2000 technology, with an average mean theoretical coverage of 75X. The reads of isolate LL2013\_164 (CBS 14088) were used for complete *de novo* assembly. We performed annotation on the 51 scaffolds larger than 5 kb, which had a cumulative size of 10,225,810 bp and a GC content of 46.6%. This GC content is comparable to what is found in *L. thermotolerans* (47.3%) (Souciet et al. 2009), however, it is higher than the average from other *Lachancea* species, including *L. kluyveri* (41.5%) (Payen et al. 2009; Souciet et al. 2009; Friedrich et al. 2015) and *L. lanzarotensis* (44.3%) (Sarilar et al. 2015).

Genome annotation was performed using the Amadea Annotation transfer tool (Isoft, France) with *Lachancea kluyveri* (Souciet et al. 2009) as the reference genome, followed by manual curation. This strategy allowed for the identification of 5,075 putative protein-coding genes and 71 pseudogenes. Spliceosomal introns were predicted in 323 genes (6.4% of the genes), only 11 of which were interrupted by two introns. Whenever possible, functional annotation was transferred from *S. cerevisiae*, RefSeq protein sequences, or experimentally validated proteins from other *Lachancea* species. During this process, 3,892 proteins showed at least 50% sequence similarity with *S. cerevisiae*. Finally, all but 70 putative proteins have a homolog in Saccharomycotina yeasts. A single case of inter-kingdom horizontal

transfer specific to *L. quebecensis* was predicted (LAQU0S32e00122g), which provides this species with a putative FMN-dependent NADH-azoreductase.

### **Gene content variation compared to the closest related species: *L. thermotolerans***

*L. thermotolerans* is the species most closely related to *L. quebecensis*. The genome of *L. thermotolerans* CBS 6340<sup>T</sup> is slightly larger than that of CBS 14088 by approximately 165 kb, and harbors 102 additional genes. Structural comparison between the CBS 6340<sup>T</sup> and CBS 14088 genomes revealed that their chromosomes are highly collinear. Only three major synteny breakpoints were detected, two of which corresponded to a reciprocal translocation involving scaffolds LAQU0S24 and LAQU0S07 (Figure 2). We identified the genes flanking these breakpoints (Figure 2). The third synteny breakpoint involved scaffold LAQU0S10, however it is difficult to determine if it is a true reciprocal translocation as the breakpoints in *L. thermotolerans* correspond to scaffold ends in *L. quebecensis*. Further comparison with the *L. waltii* strain CBS 6430<sup>T</sup> revealed that both rearrangements occurred in the *L. quebecensis* branch. At least 4,827 genes were found to be syntenic homologs with a mean sequence similarity of 92.8%. The majority of the remaining 248 genes have a non-syntenic homolog in *L. thermotolerans*. Among the 73 exceptions, some are noteworthy: for instance, LAQU0S45e00144g, a *MELI* homolog encoding  $\alpha$ -galactosidase which enables the fermentation of melibiose, LAQU0S10e01904g, a triphenylmethane reductase like protein, mostly found in bacteria but also with a *Kluyveromyces lactis* homolog (KLLA0C19547g), and LAQU0S33e00144g, a putative gamma-glutamyltranspeptidase involved in amino acid transport.

### **Intraspecific genetic diversity across the *L. quebecensis* isolates**

To survey the intraspecific variability within *L. quebecensis*, we compared the sequence of the 8 other *L. quebecensis* isolates to the newly assembled reference genome. A total of 243,358 SNPs distributed across 69,346 polymorphic positions were identified, with the number of SNPs ranging from 41 to 59,109 per isolate (Table 2). There were only 41 and 127 SNPs identified between the reference genome



(LL2013\_164) and two other isolates, LL2013\_162 and LL2012\_118, respectively. This demonstrates that these strains are highly similar. However, it is interesting to note that while LL2013\_164 and LL2012\_118 were both isolated from maple tree bark, they were isolated from different samples collected in different years from two distinct collection sites (Table 1). Additionally, LL2013\_164 and LL2013\_162, which are incredibly similar genetically, were collected from distinct sites and isolated from maple and birch tree bark, respectively (Table 1). This demonstrates that nearly clonal isolates of *L. quebecensis* are present in the population sampled in Québec. Among the SNPs identified, 9,407 correspond to singletons *i.e.* are found only in one genome. The phylogenetic relationship was constructed based on these polymorphic positions (Figure 1B) and revealed the presence of 2 main clusters that are in accordance with the geographical origins of the strains. Indeed, one cluster is composed of isolates originating from the Gaspé Peninsula, while the other harbors strains from the St Lawrence Valley (Figure 1A, B). Links between phylogenetic cluster and geographic origin have been observed in the closely related *L. kluyveri* (Friedrich et al. 2015) while the ecological origin was linked to some clusters in *S. cerevisiae* (Liti et al. 2009; Schacherer et al. 2009). Between these two main clusters, the genetic divergence ranges from 0.5% to 0.6% (Figure 1B), which is approximately what was reported in *S. cerevisiae* (Schacherer et al. 2009). Within each of the two sub-clusters, the divergence is less than 0.1%. However, among the 5 isolates from the St Lawrence Valley, varying levels of divergence are observed among 3 that share less than 0.005% (LL2013\_164, LL2013\_162 and LL2012\_118), while further including LL2011\_022<sup>T</sup> leads to a level of 0.03%, and among all 5 isolates there is a maximum of 0.1% divergence observed (when LL2012\_095 is also considered). While this phylogeny was constructed with the genomes from only 9 isolates, which were all obtained from the same ecological source and general geographic location, we observed the same level of diversity seen in *S. cerevisiae* (Liti et al. 2009; Schacherer et al. 2009; Wang et al. 2012; Cromie et al. 2013), even though the samples included form two clear clusters in contrast to the continuum revealed by the *S. cerevisiae* phylogeny.

Global *L. quebecensis* genetic diversity, estimated by the average pairwise divergence ( $\pi = 2 \times 10^{-3}$ ) demonstrated that this species harbors similar values to those estimated for *S. cerevisiae* ( $\pi = 4 \times 10^{-3}$ )



(Schacherer et al. 2009) and *S. pombe* ( $\pi = 3 \times 10^{-3}$ ) (Jeffares et al. 2015). However, it must be noted that these strains represent a very limited data set composed of only 9 genome sequences from isolates obtained from a single geographic region (Québec, Canada) and ecological niche (tree bark). However, it is an order of magnitude lower than both the most closely related species, *Lachancea kluyveri* ( $\pi = 17 \times 10^{-3}$ ) (Friedrich et al. 2015) as well as the more distantly related *S. uvarum* ( $\pi = 12 \times 10^{-3}$ ) (Almeida et al. 2014). These differences could be attributed in part to the fact that *L. quebecensis* is represented currently by a dramatically smaller set of isolates sequenced (9) in comparison to 28 and 54 in the *L. kluyveri* and *S. uvarum* studies, respectively. The dichotomy observed between the two *Lachancea* species is probably due to the fact that all *L. quebecensis* isolates are from the same geographical region and ecological niche, essentially representing a subpopulation level sampling of this species, while the *L. kluyveri* isolates were derived from diverse global geographical locations and were obtained from various environmental sources including tree exudate, soil, and *Drosophila*. Indeed, analysis of individual subpopulations from Europe and Asia were characterized by dramatically lower values of  $\pi$  equal to  $0.6 \times 10^{-3}$  and  $7.5 \times 10^{-3}$ , respectively (Friedrich et al. 2015). The following four *L. quebecensis* strains were previously identified: two originated from Japan and were isolated from leaves (NBRC 10066) and flowers (NBRC 10067) (Naumova et al. 2007), one was obtained from black knot on a cherry tree in Québec (UWOPS 79-139) and one from *Drosophila* sp. in Ontario (UWOPS 82-231). It would be interesting to obtain additional whole genome sequences to determine the full extent of the diversity within the currently known isolates of this species.

### **Phenotypic diversity observed among *L. quebecensis* isolates**

In order to survey the overall phenotypic diversity of these 9 isolates, they were grown in a total of 33 conditions (Figure 3A), which tested the ability of the strains to survive in a range of conditions. The media assessed growth on a variety of carbon and nitrogen sources as well as compounds that affect different cellular functions including: protein stability (formamide, KCl and SDS), osmotic stress (NaCl), sterol biosynthesis (nystatin), transcription (6-azauracil), and translation (anisomycin and cycloheximide).

When all conditions were considered, the isolates formed three phenotypic clusters (Figure 3A, B). Each cluster was a mix of representatives from both phylogenetic subgroups shown in Figure 2B and strains among each group share levels of genome divergence of approximately 0.6%. The first group (I) is composed of three isolates, and demonstrated an enhanced growth rate on galactose (2%), ethanol (2%) xylose (2%) and also tolerated a higher level of osmotic stress (NaCl 1.5 M) in comparison to the other groups. The second (II) and third groups (III) share more similar profiles, with group II distinguished in particular by poor growth on anisomycin at 20  $\mu\text{g} / \text{mL}$ . The global phenotypic variance among all of the 33 conditions tested was 0.067 (Figure 3C), which is an order of magnitude lower than what was observed for *S. cerevisiae* (0.59) as well as *S. paradoxus* (0.35) (Warringer et al. 2011). Although these global values for *S. cerevisiae* and *S. paradoxus* were determined across larger sample sizes of 38 and 35 isolates, respectively, even in small subpopulations phenotypic variance is high. For example, analysis of three *S. cerevisiae* isolates from Malaysia, yielded a value of approximately 0.5 while four strains of *S. paradoxus* from the Far East harbor a value of 0.35 (Warringer et al. 2011). It is difficult to compare phenotypic variation amongst different laboratories and to previous studies due to changes in methodology and available tools as well as media conditions tested. Furthermore, it might be expected that divergent species differentially utilize various media types as in nature they likely inhabit various ecological niches. Assessing the phenotypic variance for species across multiple conditions is an attempt to broadly summarize the phenotypic diversity in a population. The values obtained across 33 media conditions for the 9 *L. quebecensis* isolates studied here are consistently low with the highest levels in variation observed on galactose (Figure 3C). All strains exhibited reduced growth in the presence of a suite of compounds associated with various cellular stresses. However, an additional analysis with these 12 conditions removed resulted in the same clustering pattern (Figure S2A, B). The lack of correlation between genotype and phenotype was surprising considering the extremely low level of nucleotide divergence (<0.005%) observed between some of the isolates in the St Lawrence Valley clade (LL2012\_162 and LL2012\_164). However, it must be noted that although three groups can be distinguished, overall the 9 strains share relatively similar phenotypic profiles. Using the global

phenotypic variance to compare species as was previously done by Warringer and colleagues is not a definitive method to define phenotypic variation, however it is an initial attempt to broadly summarize trait variation in different populations (Warringer et al. 2011). As divergent populations and phenotypes of yeasts are explored, better tools to assess and compare multiple groups of isolates and datasets will be developed, continually elucidating a better understanding of yeast diversity and genome evolution.

### Conclusion

The genomes of the 9 *L. quebecensis* isolates obtained in this study confirm the relationship of this species to the closely related sister species *L. thermotolerans*. These strains were obtained from approximately the same geographic location and environmental niche and while they were isolated from different samples, the close proximity of their origins likely biases the view of genome diversity in *L. quebecensis*. It could prove interesting to continue the exploration of this species by sequencing the other few isolates currently known. A sample size of 9 genomes from strains isolated from tree bark all generally located in the same geographical area harbor the same level of genetic diversity as has been observed in *S. cerevisiae* isolated from global locations and a wide range of ecological niches. The *L. quebecensis* diversity is essentially explained by the presence of two phylogenetic groups that reflect a slight difference in geographic distribution. While there is a relatively similar range of genome divergence within the *L. quebecensis* and *S. cerevisiae* populations the global phenotypic variance is dramatically lower in *L. quebecensis*.

## Materials and Methods

### Yeast strains, flow cytometry and electrophoretic karyotyping

All strains were generously provided by Christian Landry and were originally isolated using previously described methods (Charron et al., 2014). Samples were prepared for DNA content analysis using flow cytometry and pulse field gel electrophoresis using previously described protocols (Freel et al., 2015).

### Genomic DNA extraction and Illumina high-throughput sequencing

The 9 *L. quebecensis* isolates were grown overnight at 30°C in 20 ml of YPD medium to early stationary phase before cells were harvested by centrifugation. Total genomic DNA was subsequently extracted using the MasterPure Yeast DNA purification kit (Cat No MPY80200) according to the manufacturer's instructions (epicentre, Germany). For all genomes, 480-bp insert libraries were produced and sequenced on an Illumina HiSeq2000 platform. 100-bp paired-end reads were obtained and trimmed according to quality criteria with Cutadapt (Martin 2011).

### Reference genome annotation of strain LL2013\_164

Since there was no reference genome for *L. quebecensis* available, SOAPdenovo2, version r240 (Luo et al. 2012) was used to assemble the LL2013\_164 reads, with a k-mer size of 63. The assembly was annotated with the Amadea Annotation transfer tool (Isoft, France), and considered as the reference genome for the species. Functional annotation was performed as in *L. lanzarotensis* (Sarilar et al. 2015) and *Cyberlindnera fabianii* (Freel et al. 2014b).

The final assembly of this reference genome consisted of 119 scaffolds larger than 1 kb, revealing a genome size of 10,535,401 bp that contains 47,783 ambiguous sites (maximum scaffold size: 1,366,161 bp, mean scaffold size: 8,750 bp, N50: 453 kb (7 scaffolds), N80: 222 kb (18 scaffolds))

### Population genomic analysis

The 8 other sequences were compared to this reference genome to obtain insights into genomic intraspecific variability. The clean reads were mapped with BWA (version 0.7, option -I -q 25 -n 8 -o 2) (Li and Durbin 2009) and SNPs and short indels were identified on the basis of the mpileup files generated by SAMtools (version 0.1.18) (Li et al. 2009).

To obtain a view of the phylogenetic relationships among the different isolates, a neighbor-joining tree of the 9 strains using the 69,346 polymorphic positions detected in the whole population was constructed with the software package SplitsTree (Huson and Bryant 2006). Branch lengths are proportional to the number of segregating sites that differentiate each node. An estimation of  $\pi$ , the average pairwise nucleotide diversity, was calculated with Variscan (Hutter et al. 2006).

Synteny conservation analysis of *L. quebecensis* LL2013\_164 (CBS 14088), *L. thermotolerans* CBS 6340<sup>T</sup>, and *L. waltii* CBS 6430<sup>T</sup> was performed with SynChro (Drillon et al., PLoS ONE 2014).

### Phenotyping of *L. quebecensis*

All *L. quebecensis* isolates in this study were subjected to growth on a total of 33 different media types, the specific composition of each medium can be found in Figure 3A. The conditions tested were in standard YPD with the addition of the relevant stressor. In media testing carbon utilization, 2% glucose was substituted with 2% of the following carbon sources: sorbitol, mannose, EtOH, acetate, galactose, xylose, ribose, glycerol, glucose, fructose, and sucrose. To prepare the master matrices, the isolates were grown in a 96 well format in liquid YPD overnight. After overnight growth, the colonies were pinned on Singer PlusPlates of YPD media using the Singer RoToR HDA Microbial Pinning Robot (Singer Instruments, Somerset, UK) and allowed to grow approximately 30 hours. The plate with 96 colonies was then used to construct 384 colony matrices on YPD, where each original colony was plated in 4 replicates. After an additional 30 hours, the 384 matrices were pinned onto the various media types (Figure 3A) as well as control plates with YPD. All plates were scanned after 24, 40, and 48 hours of growth at 30°C using an 8-bit gray scale and resolution of 300 dpi. The colony size from each condition

was used as a measure of growth and compared to that on the standard YPD control plates. Images were analyzed using the Gitter package in R in order to obtain growth ratios for all isolates on all media types. Data from the 40 hour time point was used to construct the heat maps and calculate global phenotypic variance. Heatmaps and PCA analyses were generated with R also using the growth ratio data, which was also used to calculate the global and individual condition phenotypic variance (RCoreTeam 2013) using the ggplot2 and FactorMineR packages.

### Acknowledgements

We thank Christian Landry for generously providing the *Lachancea quebecensis* isolates. This work was supported by an ANR grant (2010-BLAN-1606) (J.S. and C.N.) and ANR Young Investigator grant (2011-JSV6-004-01) (J.S.). We also thank the Université de Strasbourg (IdEx 2012 Attractivité) for their financial support and the BioImage platform (IBMP-CNRS, Strasbourg) for their support.

[1] The sequence of the *L. quebecensis* (CBS 14088) genome has been deposited at European Nucleotide Archive under the accession number ERS870159 and study ID PRJEB8847.

[2] All Illumina sequencing reads generated in this study have been deposited in the European Nucleotide Archive under the study accession number PRJEB10592.

## References

- Almeida P, Gonçalves C, Teixeira S, Libkind D, Bontrager M, Masneuf-Pomarède I, Albertin W, Durrens P, Sherman DJ, Marullo P et al. 2014. A Gondwanan imprint on global diversity and domestication of wine and cider yeast *Saccharomyces uvarum*. *Nat Commun* **5**:4044.
- Charron G, Leducq JB, Bertin C, Dube AK, Landry CR. 2014. Exploring the northern limit of the distribution of *Saccharomyces cerevisiae* and *Saccharomyces paradoxus* in North America. *FEMS Yeast Res* **14**(2): 281-288.
- Cromie GA, Hyma KE, Ludlow CL, Garmendia-Torres C, Gilbert TL, May P, Huang AA, Dudley AM, Fay JC. 2013. Genomic Sequence Diversity and Population Structure of *Saccharomyces cerevisiae* Assessed by RAD-seq. *G3: Genes|Genomes|Genetics* **3**(12): 2163-2171.
- Drillon G, Carbone A, Fischer G. 2014. SynChro: a fast and easy tool to reconstruct and visualize synteny blocks along eukaryotic chromosomes. *PLoS One* **9**(3):e92621.
- Dujon B. 2010. Yeast evolutionary genomics. *Nature reviews Genetics* **11**(7): 512-524.
- Freel KC, Charron G, Leduc LG, Landry CR, Schacherer J. 2015. *Lachancea quebecensis* sp. nov., a yeast species consistently isolated from tree bark in the Canadian province of Québec. *International Journal of Systematic and Evolutionary Microbiology*.
- Freel KC, Friedrich A, Hou J, Schacherer J. 2014a. Population Genomic Analysis Reveals Highly Conserved Mitochondrial Genomes in the Yeast Species *Lachancea thermotolerans*. *Genome biology and evolution* **6**(10): 2586-2594.
- Freel KC, Sarilar V, Neuvéglise C, Devillers H, Friedrich A, Schacherer J. 2014b. Genome Sequence of the Yeast *Cyberlindnera fabianii* (*Hansenula fabianii*). *Genome Announcements* **2**(4): e00638-00614.
- Friedrich A, Jung P, Reisser C, Fischer G, Schacherer J. 2015. Population genomics reveals chromosome-scale heterogeneous evolution in a protoploid yeast. *Mol Biol Evol* **32**(1): 184-192.
- Friedrich A, Jung PP, Hou J, Neuvéglise C, Schacherer J. 2012. Comparative Mitochondrial Genomics within and among Yeast Species of the *Lachancea* Genus. *PloS one* **7**(10).
- Hittinger CT, Goncalves P, Sampaio JP, Dover J, Johnston M, Rokas A. 2010. Remarkably ancient balanced polymorphisms in a multi-locus gene network. *Nature* **464**(7285): 54-58.
- Huson DH, Bryant D. 2006. Application of phylogenetic networks in evolutionary studies. *Molecular Biology and Evolution* **23**(2): 254-267.
- Hutter S, Vilella AJ, Rozas J. 2006. Genome-wide DNA polymorphism analyses using VariScan. *Bmc Bioinformatics* **7**: 409.
- Jeffares DC, Raillis C, Rieux A, Speed D, Prevorovsky M, Mourier T, Marsellach F, Iqbal Z, Lau W, Cheng TMK et al. 2015. The genomic and phenotypic diversity of *Schizosaccharomyces pombe*. *Nature genetics* **47**: 235-241.
- Jung PP, Friedrich A, Reisser C, Hou J, Schacherer J. 2012. Mitochondrial Genome Evolution in a Single Protoploid Yeast Species. *G3 (Bethesda)* **2**(9): 1103-1111.



- Lachance MA, Kurtzman CP. 2011. *Lachancea Kurtzman (2003)*. In: *The Yeasts, a Taxonomic Study, 5th Edition*. Edited by Kurtzman CP, Fell JW, Boekhout T. Elsevier, Amsterdam pp 511-519.
- Li H, Durbin R. 2009. Fast and accurate short read alignment with Burrows-Wheeler transform. *Bioinformatics* **25**(14): 1754-1760.
- Li H, Handsaker B, Wysoker A, Fennell T, Ruan J, Homer N, Marth G, Abecasis G, Durbin R. 2009. The Sequence Alignment/Map format and SAMtools. *Bioinformatics* **25**(16): 2078-2079.
- Liti G, Carter DM, Moses AM, Warringer J, Parts L, James SA, Davey RP, Roberts IN, Burt A, Koufopanou V et al. 2009. Population genomics of domestic and wild yeasts. *Nature* **458**(7236): 337-341.
- Liti G, Schacherer J. 2011. The rise of yeast population genomics. *Comptes rendus biologiques* **334**(8-9): 612-619.
- Luo R, Liu B, Xie Y, Li Z, Huang W, Yuan J, He G, Chen Y, Pan Q, Liu Y et al. 2012. SOAPdenovo2: an empirically improved memory-efficient short-read *de novo* assembler. *GigaScience*: 1-18.
- Martin M. 2011. Cutadapt removes adapter sequences from high-throughput sequencing reads. *EMBnet journal* **17**(1): 10-12.
- Naumova ES, Serpova EV, Naumov GI. 2007. Molecular Systematics of *Lachancea* Yeasts. *Biochemistry (Moscow)* **72**(12): 1356-1362.
- Payen C, Fischer G, Marck C, Proux C, Sherman DJ, Coppee JY, Johnston M, Dujon B, Neuvéglise C. 2009. Unusual composition of a yeast chromosome arm is associated with its delayed replication. *Genome Res* **19**(10): 1710-1721.
- RCoreTeam. 2013. R: A Language and Environment for Statistical Computing R Foundation for Statistical Computing, Vienna, Austria.
- Sarilar V, Devillers H, Freel KC, Schacherer J, Neuvéglise C. 2015. Draft Genome Sequence of *Lachancea lanzarotensis* CBS 12615(T), an Ascomycetous Yeast Isolated from Grapes. *Genome Announcements* **3**(2): e00292-00215.
- Schacherer J, Shapiro JA, Ruderfer DM, Kruglyak L. 2009. Comprehensive polymorphism survey elucidates population structure of *Saccharomyces cerevisiae*. *Nature* **458**(7236): 342-345.
- Skelly DA, Merrihew GE, Riffle M, Connelly CF, Kerr EO, Johansson M, Jaschob D, Graczyk B, Shulman NJ, Wakefield J et al. 2013. Integrative phenomics reveals insight into the structure of phenotypic diversity in budding yeast. *Genome Research* **23**(9): 1496-1504.
- Soucié JL, Dujon B, Gaillardin C, Johnston M, Baret PV, Cliften P, Sherman DJ, Weissenbach J, Westhof E, Wincker P et al. 2009. Comparative genomics of protoploid Saccharomycetaceae. *Genome Res* **19**(10): 1696-1709.
- Strope PK, Skelly DA, Kozmin SG, Mahadevan G, Stone EA, Magwene PM, Dietrich FS, McCusker JH. 2015. The 100-genomes strains, an *S. cerevisiae* resource that illuminates its natural phenotypic and genotypic variation and emergence as an opportunistic pathogen. (1549-5469 (Electronic)).

Wang Q-M, Liu W-Q, Liti G, Wang S-A, Bai F-Y. 2012. Surprisingly diverged populations of *Saccharomyces cerevisiae* in natural environments remote from human activity. *Molecular ecology* **21**(22): 5404-5417.

Warringer J, Zörgö E, Cubillos FA, Zia A, Gjuvsland A, Simpson JT, Forsmark A, Durbin R, Omholt SW, Louis EJ et al. 2011. Trait Variation in Yeast Is Defined by Population History. *PLoS Genet* **7**(6): e1002111.

**Table 1.** Description of the 9 *Lachancea quebecensis* strains studied.

Strain ID	Year isolated	Geographical Origin	Ecological Origin	Phylogenetic group*	FACS
LL2011_022 <sup>T</sup> (CBS 14138)	2011	St Lawrence Valley, Québec, Canada	Maple tree bark	Group 1 (<0.005%)	Haploid
LL2012_068	2012	Gaspé Peninsula, Québec, Canada	Oak tree bark	Group 2 (<0.1%)	Haploid
LL2012_073	2012	Gaspé Peninsula, Québec, Canada	Maple tree bark	Group 2 (<0.1%)	Haploid
LL2012_078	2012	Gaspé Peninsula, Québec, Canada	Maple tree bark	Group 2 (<0.1%)	Haploid
LL2012_083	2012	Gaspé Peninsula, Québec, Canada	Maple tree bark	Group 2 (<0.1%)	Haploid
LL2012_095	2012	St Lawrence Valley, Québec, Canada	Maple tree bark	Group 1.A (<0.1%)	Haploid
LL2012_118	2012	St Lawrence Valley, Québec, Canada	Maple tree bark	Group 1 (<0.005%)	Haploid
LL2013_162	2013	Mont-St-Anne, Québec, Canada	Birch tree bark	Group 1 (<0.005%)	Haploid
LL2013_164 (CBS 14088)	2013	Mont-St-Anne, Québec, Canada	Maple tree bark	Group 1 (<0.005%)	Haploid

\*See Figure 1 for the delineation of the phylogenetic groups.

**Table 2.** SNPs and indels per strain in comparison to the reference (LL2013\_164) genome.

Strain ID	# of SNPs	SNPs/kb in Genes	SNPs/kb in intergenic regions	# of small indels
LL2011_022 <sup>T</sup> (CBS 14138)	2,335	0.2	0.6	32
LL2012_068	59,109	4.4	10.1	360
LL2012_073	58,528	4.4	10.0	447
LL2012_078	57,726	4.3	9.9	624
LL2012_083	57,069	4.3	9.7	3,726
LL2012_095	8,423	0.7	1.6	601
LL2012_118	127	0.03	0.2	61
LL2013_162	41	0.02	0.1	83

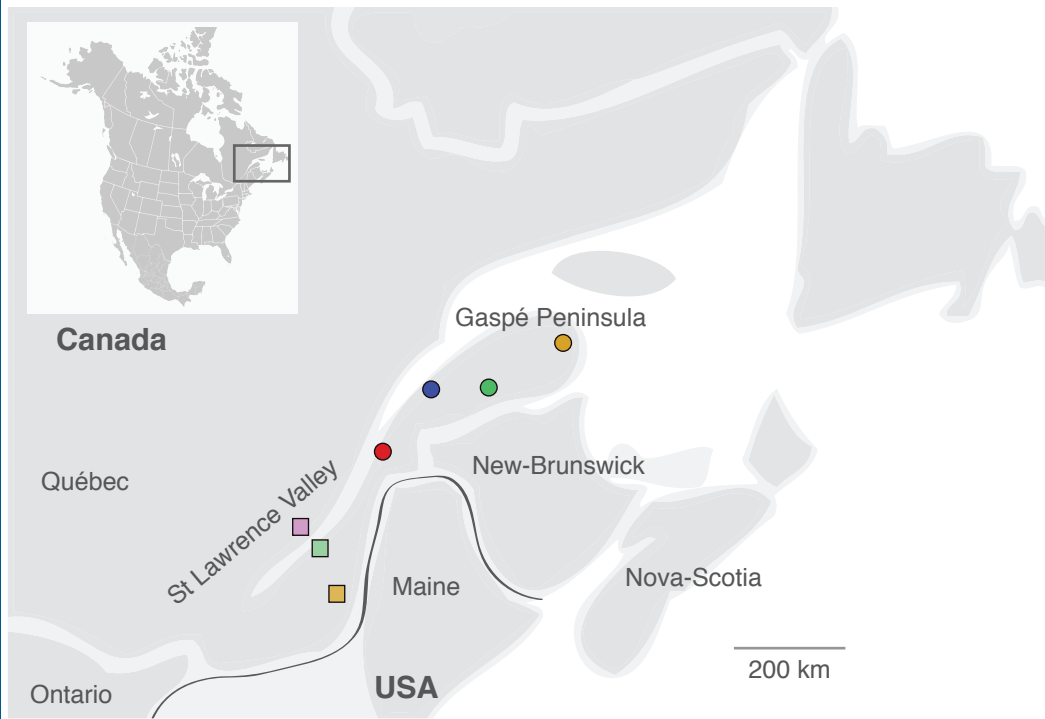
## Figure legends

**Figure 1.** (A) Map of strain origins. Five strains were obtained from sites in the St. Lawrence Valley area (three squares) while the remaining four strains were isolated from the Gaspé Peninsula (four circles). (B) Strain relationships based on the 69,346 polymorphic positions identified in 9 genomes of *L. quebecensis*. The divergence among isolates in each group is less than 0.1%.

**Figure 2.** Whole genome dot-plot comparison between *L. quebecensis* CBS 14088 and *L. thermotolerans* CBS 6340<sup>T</sup>. Only scaffolds with at least 5 orthologous gene pairs were considered. The six-frame translations of both genomes were compared via MUMmer 3.0. Homologous regions are plotted as squares, which are color coded according to their relative orientation: red for parallel orientation and blue for anti-parallel orientation. Synteny breakpoints and chromosomal rearrangements are highlighted with circles and arrows, respectively. The green circles indicate the reciprocal translocation that occurred in *L. quebecensis* between homologues of *KLTH0A01804g* and *KLTH0A01826g*, as well as *KLTH0E10010g* and *KLTH0E10054g*. The purple circle highlights the third major synteny breakpoint, which involved the homologous regions *KLTH0G04730g–KLTH0G04752g* and *KLTH0E02200g–KLTH0E02310g*. (\*) indicate that the scaffolds were reverse-complemented, gene order is well conserved however the strand sequence was inverted.

**Figure 3.** Phenotypic diversity among the 9 *L. quebecensis* strains on 33 different media after 40 hours of growth at 30°C. The three phenotypic groups discussed are indicated by Roman numerals (I, II, and III). Growth ratios calculated from colony size on the media of interested and that on YPD media were used for all comparative analyses. (A) Heatmap of the 9 strains that were grown in 33 different media including multiple carbon sources (which replaced the 2% glucose in standard YPD media) as well as other compounds that induced stress on various cellular functions. (B) A PCA analysis of the 9 *L. quebecensis* isolates analyzed. (C) Phenotypic variance in the 9 *L. quebecensis* strains across all 33 media conditions tested.

Figure 1.



B

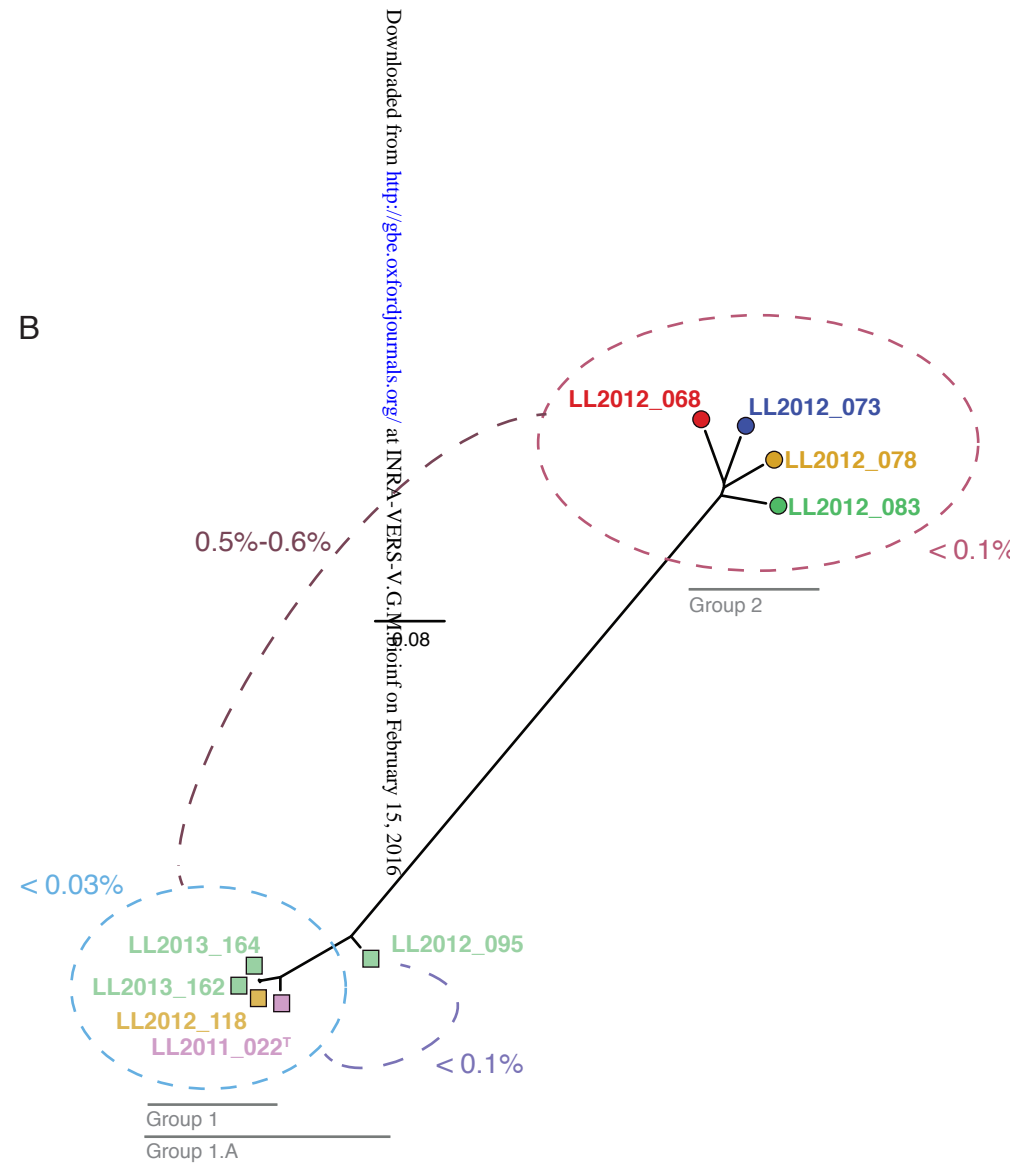


Figure 2.

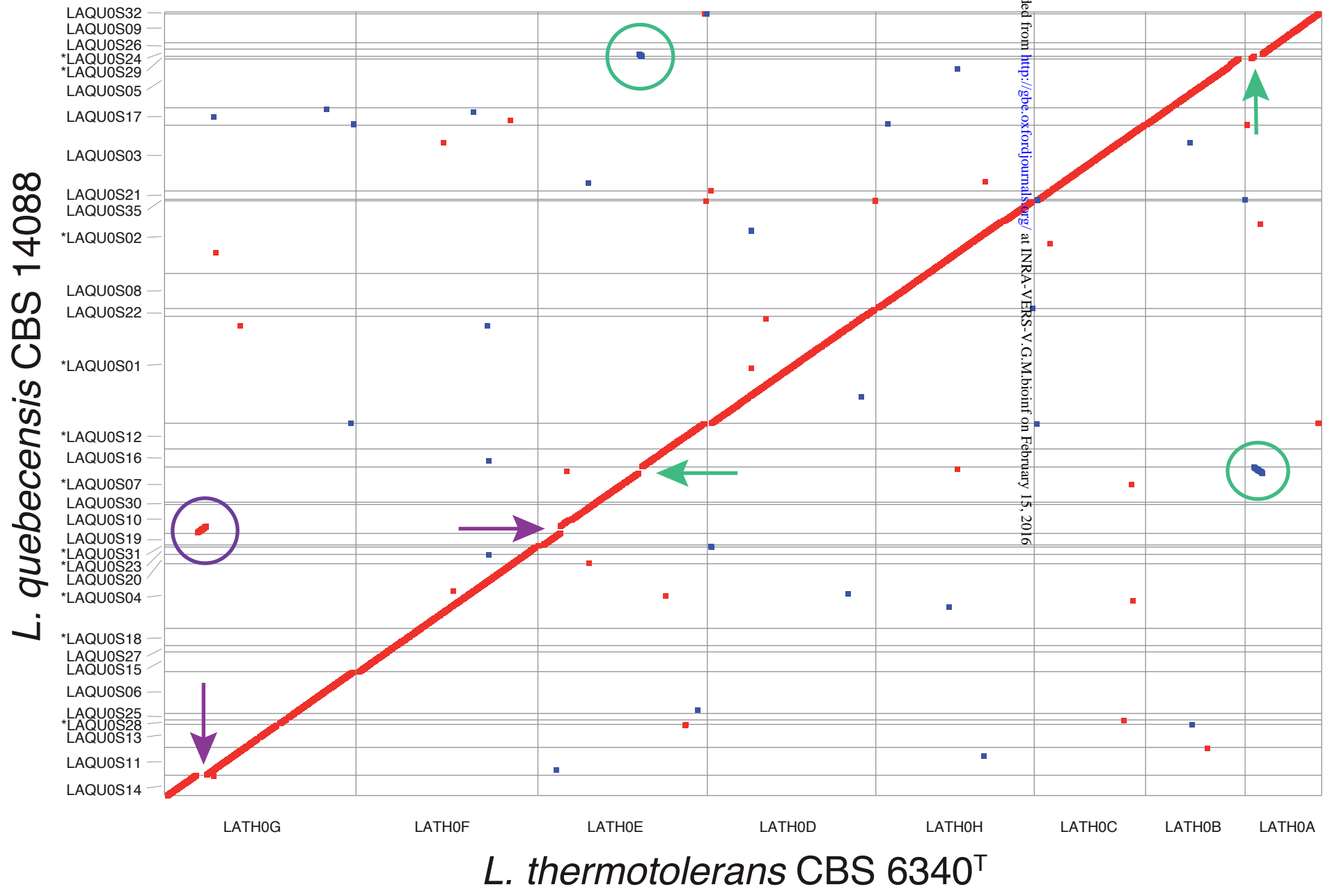


Figure 3.

