De la diversité à la domestication chez Saccharomyces cerevisiae

Jean-Luc Legras

Université de Montpellier

HABILITATION A DIRIGER DES RECHERCHES

De la diversité à la domestication chez la levure *Saccharomyces cerevisiae*

par

Jean-Luc LEGRAS
Jean-luc.legras@inra.fr

Session : 2018

Table des matières

Parcours Professionnel et Scientifique

1. Curriculum vitae

LEGRAS Jean-Luc
Né le 29 Octobre 1963

Ingénieur de Recherche
UMR 1083 INRA, Montpellier-SupAgro, Université Montpellier 1
2 place Viala
34061 Montpellier Cedex 1
Email : jean-luc.legras@inra.fr

Formation

1992 - 93 Diplôme National d'œnologue, Université de Bourgogne, Dijon, Juillet 1993.
1986 - 90 Doctorat d'ingénieur en Sciences des aliments (Microbiologie), ENSA Montpellier Juillet 1990. Mention très honorable.
1985 - 86 DEA Sciences des Aliments, Nutrition, Fermentation, USTL Montpellier.
1983 – 86 Ingénieur Agronome, ENSA de RENNES, 1983 – 1986
1980 Baccalauréat C

Stages et expérience professionnelle

1986-1989 : Stage de thèse au Laboratoire de Microbiologie Industrielle et Génétique des microorganismes, ENSA (Supagro) Montpellier
1989-1990 : Stage de recherche: Dept of Environmental Engineering, Biotechnology division, TECHNION, Haifa, ISRAEL . Screening de champignon pour leur activité amylolytique
1990-1991 : ATER Laboratoire d'Œnologie , Université de Bourgogne, DIJON.

Séjour professionnel à l'INRA

1992- 2010 Ingénieur de recherche à l'unité de recherches Vigne et Vin puis UMR SVQV INRA Colmar. Responsable du laboratoire d'analyse des vins accrédité COFRAC de 1994 à 2001.
2001-2010 : Etude de la diversité des levures œnologique s
2010- actuellement : Ingénieur de recherche à l'UMR Sciences pour l'œnologie INRA SupAgro Montpellier

2. Communications scientifiques

Article dans des revues à comité de lecture (les articles les plus importants sont grisés)

A1. Legras JL, Galeote V, Bigey F, Camarasa C, Marsit S, Nidelet T, Sanchez I, Couloux A, Guy J, Franco-Duarte R, Schuller D, Sampaio JP , Dequin S. (2018) Adaptation of *S. cerevisiae* to fermented food environments reveals remarkable genome plasticity and the footprints of domestication. (submitted)

A2. Börlin, M., Claisse, O., Albertin, W., Salin, F., Legras, J.L. Masneuf-Pomarede Quantifying anthropogenic effect on vineyard *S. cerevisiae* metapopulation diversity (in prep)

A3. Eder, M., Sanchez, I., Brice, C., Camarasa, C., Legras, J.L., Dequin, S. (2018) QTL mapping of volatile compound production in *Saccharomyces cerevisiae* during alcoholic fermentation. BMC Genomics. In press.

A4. Stefanini I, Albanese D, Sordo M, Legras J-L, De Filippo C, Cavalieri D, Donati C. (2017). SaccharomycesIDentifier, SID: strain-level analysis of *Saccharomyces cerevisiae* populations by using microsatellite meta-patterns. Scientific Report 7:15343. DOI : 10.1038/s41598-017-15729-3

A5. Coi, A. L., Bigey, F., Mallet, S., Marsit, S., Zara, G., Gladieux, P., Galeote, V., Budroni, M., Dequin, S., Legras, J. L. (2017). Genomic signatures of adaptation to wine biological aging conditions in biofilm-forming flor yeasts. Molecular Ecology, 26 (7), 2150-2166. DOI : 10.1111/mec.14053

A6. Ferreira, D., Galeote, V., Sanchez, I., Legras, J. L., Julien Ortiz, A., Dequin, S. (2017). Yeast multi-stress resistance and lag phase characterization during wine fermentation. FEMS Yeast Research, 17 (6), 37-9

A7. Viel, A., Legras, J. L. Nadai, C. Carlot, M. , Lombardi, A. , Crespan, M. , Migliaro, D., Giacomini, A. , Corich, V. (2017). The geographic distribution of *Saccharomyces cerevisiae* isolates within three Italian neighboring winemaking regions reveals strong differences in yeast abundance, genetic diversity and industrial strain dissemination. Frontiers in Microbiology. 8. 1595 DOI : 10.3389/fmicb.2017.01595

A8. Börlin, M., Venet, P., Claisse, O., Salin, F., Legras, J. L., Masneuf-Pomarede, I. (2016). Cellar-associated *Saccharomyces cerevisiae* population structure revealed high diversity and perennial persistence in Sauternes wine estates. Applied and Environmental Microbiology, 82 (10), 2909-2918. DOI : 10.1128/AEM.03627-15

A9. Coi, A. L., Legras, J. L., Zara, G., Dequin, S., Budroni, M. (2016). A set of haploid strains available for genetic studies of *Saccharomyces cerevisiae* flor yeasts. FEMS Yeast Research, 16 (6), 9 p. DOI : 10.1093/femsyr/fow066

A10. Dulermo, R., Legras, J. L., Brunel, F., Devillers, H., Sarilar, V., Neuveglise-Degouy, C., Nguyen, H.-V., Daran-Lapujade, P. (2016). Truncation of Gal4p explains the inactivation of the GAL/MEL regulon in both Saccharomyces bayanus and some *Saccharomyces cerevisiae* wine strains. FEMS Yeast Research, 16 (6), 11 p. DOI :10.1093/femsyr/fow070

A11. Legras, J. L., Moreno-Garcia, J., Zara, S., Zara, G., Garcia-Martinez, T., Mauricio, J. C., Mannazzu, I., Coi, A. L., Bou Zeidan, M., Dequin, S., Moreno, J., Budroni, M. (2016). Flor yeast: new perspectives beyond wine aging. Frontiers in Microbiology, 7, 11 p. DOI : 10.3389/fmicb.2016.00503

A12. Masneuf-Pomarede, I., Salin, F., Börlin, M., Coton, E., Coton, M., Jeune, C. L., Legras, J. L. (2016). Microsatellite analysis of Saccharomyces uvarum diversity. FEMS Yeast Research, 16 (2), 12 p. DOI : 10.1093/femsyr/fow002

A13. Tapsoba, F., Savadogo, A., Legras, J. L., Zongo, C., Traore, A. S. (2016). Microbial diversity and biochemical characteristics of Borassus akeassii wine. Letters in Applied Microbiology, 63 (4), 297-306. DOI : 10.1111/lam.12619

A14. Vallverdu Queralt, A., Biler, M., Meudec, E., Le Guerneve, C., Vernhet, A., Mazauric, J. P., Legras, J. L., Loonis, M., Trouillas, P., Cheynier, V., Dangles, O. (2016). p-Hydroxyphenyl-pyranoanthocyanins: An Experimental and Theoretical Investigation of Their Acid-Base Properties and Molecular Interactions. International Journal of Molecular Sciences, 17 (11). DOI : 10.3390/ijms17111842

A15. Almeida, P., Barbosa, R., Zalar, P., Imanishi, Y., Shimizu, K., Turchetti, B., Legras, J. L., Serra, M., Dequin, S., Couloux, A., Guy, J., Bensasson, D., Gonçalves, P., Sampaio, J. P. (2015). A population genomics insight into the Mediterranean origins of wine yeast domestication. Molecular Ecology, 24 (21), 5412-5427. DOI : 10.1111/mec.13341

A16. Marongiu, A., Zara, G., Legras, J. L., Del Caro, A., Mascia, I., Fadda, C., Budroni, M. (2015). Novel starters for old processes: use of *Saccharomyces cerevisiae* strains isolated from artisanal sourdough for craft

beer production at a brewery scale. Journal of Industrial Microbiology and Biotechnology. DOI : 10.1007/s10295-014-1525-1

A17. Marsit, S., Mena, A., Bigey, F., Sauvage, F. X., Couloux, A., Guy, J., Legras, J. L., Barrio, E., Dequin, S., Galeote, V. (2015). Evolutionary advantage conferred by an eukaryote-to-eukaryote gene transfer event in wine yeasts. Molecular Biology and Evolution, 32 (7), 1695-1707. DOI : 10.1093/molbev/msv057

A18. Tapsoba, F., Legras, J. L., Savadogo, A., Dequin, S., Traore, A. S. (2015). Diversity of *Saccharomyces cerevisiae* strains isolated from Borassus akeassii palm wines from Burkina Faso in comparison to other African beverages. International Journal of Food Microbiology, 211, 128-133. DOI : 10.1016/j.ijfoodmicro.2015.07.010

A19. Brice C, Sanchez I, Bigey F, Legras J-L, Blondin B. 2014. A genetic approach of wine yeast fermentation capacity in nitrogen-starvation reveals the key role of nitrogen signaling. BMC Genomics 15:495

A20. Legras J-L, Erny C, Charpentier C. 2014. Population Structure and Comparative Genome Hybridization of European Flor Yeast Reveal a Unique Group of *Saccharomyces cerevisiae* Strains with Few Gene Duplications in Their Genome. PLoS One 9:e108089.

A21. Brion C, Ambroset C, Sanchez I, Legras J-L, Blondin B. (2013). Differential adaptation to multi-stressed conditions of wine fermentation revealed by variations in yeast regulatory networks. BMC Genomics 14:681.

A22. Steyer D., Erny C., Claudel P., Riveil G, Karst F., & Legras J-L (2013). Genetic analysis of geraniol metabolism during fermentation. Food Microbiology, 33(2), 228–34.

A23. Steyer D, Ambroset C, Brion C, Erny C, Claudel P, Sanchez I, Delobel P, Blondin B, Karst F, Legras JL(2012) QTL mapping of the production of wine aroma compounds by yeast.. *BMC Genomics*, 13, 573.

A24. Stefanini, I., Dapporto, L., Legras, J.-L., Calabretta, A., Di Paola, M., De Filippo, C., Viola, R., et al. (2012). Role of social wasps in *Saccharomyces cerevisiae* ecology and evolution. *Proceedings of the National Academy of Sciences*. 109(33), 13398–403 doi:10.1073/pnas.1208362109

A25. Erny, C., Raoult, P., Alais, A., Butterlin, G., Delobel, P., Matei-Radoi, F., Casaregola, S., et al. (2012). Ecological success of a group of *Saccharomyces cerevisiae*/Saccharomyces kudriavzevii hybrids in the Northern European wine making environment. Applied and environmental microbiology, 78(9), 3256-3265 DOI : 10.1128/AEM.06752–11.

A26. Nguyen, H.-V., Legras, J.-L., Neuvéglise, C., & Gaillardin, C. (2011). Deciphering the Hybridisation History Leading to the Lager Lineage Based on the Mosaic Genomes of Saccharomyces bayanus Strains NBRC1948 and CBS380T. (L. N. Rusche, Ed.)PLoS ONE, 6(10), e25821.

A27. Galeote, V., Bigey, F., Beyne, E., Novo, M., Legras, J.-L., Casaregola, S., & Dequin, S. (2011). Amplification of a Zygosaccharomyces bailii DNA Segment in Wine Yeast Genomes by Extrachromosomal Circular DNA Formation. (N. Nikolaidis, Ed.)PLoS ONE, 6(3), e17872.

A28. Sicard, D., & Legras, J.-L. (2011). Bread, beer and wine: Yeast domestication in the Saccharomyces sensu stricto complex. *Comptes Rendus Biologies*. Academie des sciences. doi: 10.1016/j.crvi.2010.12.016.

A29. Legras, J. L., Erny, C., Le Jeune, C., Lollier, M., Adolphe, Y., Demuyter, C., et al. (2010). *Saccharomyces cerevisiae* activates two different resistance mechanisms when exposed to octanoic and decanoic acids. *Applied and environmental microbiology*, *76*(22), 7526-7535.

A30. Duchene, E., Legras, J.-L., Karst, F., Merdinoglu, D., Claudel, P., Jaegli, N., et al. (2009). Variation of linalool and geraniol content within two pairs of aromatic and non-aromatic grapevine clones. *Australian Journal of Grape and Wine Research*, *15*(2), 120-130.

A31. Novo, F. M., Bigey, F., Beyne, E., Galeote, V., Gavory F, Mallet, S., Cambon, B., Legras JL, Winckler, P, Casaregola S., Dequin S. 2009. Eukaryote-to-eukaryote gene transfer events revealed by the genome sequence of the wine yeast *Saccharomyces cerevisiae* EC1118. *Proceedings of the National Academy of Sciences of the USA.* 106, 16333-8

A32. Charpentier C, Colin A, Alais A, Legras JL 2009 French Jura Flor Yeast: genotypical and technological diversity. *Antonie van Leewenhoek*. *95*, 263-73

A33. Ezeronye O, Legras J.L., 2009 Genetic analysis of *Saccharomyces cerevisiae* strains isolated from Palm Wine in Eastern Nigeria. Comparison to other African strains. *Journal of Applied Microbiology* 106, 1569-78

A34. Legras JL, Merdinoglu D, Cornuet JM, Karst F 2007. Bread, Beer and Wine: *Saccharomyces cerevisiae* diversity reflects human history. *Molecular Ecology* 16, 2091-2102

A35. Le Jeune C, Lollier M, Demuyter C, Erny C, Legras JL, Aigle M & Masneuf-Pomarède I (2007) Characterization of natural hybrids of *Saccharomyces cerevisiae* and *Saccharomyces bayanus var. uvarum.* FEMS Yeast Research, 7, 540-549

A36. Ayoub M.J., Legras J.L., Saliba R., Gaillardin C. (2006) Application of Multi Locus Sequence Typing to the analysis of the biodiversity of indigenous *Saccharomyces cerevisiae* wine yeasts from Lebanon. Journal of Applied Microbiology, 100, 699-711

A37. Legras JL, Ruh O, Merdinoglu D and Francis Karst, 2005 Selection of hypervariable microsatellite loci for the characterization of *Saccharomyces cerevisiae* strains. International Journal Food Microbiology 102, 73-83

A38. Demuyter, C., Lollier, M., Legras, J.-L., & LeJeune, C. (2004). Predominance of Saccharomyces uvarum during spontaneous alcoholic fermentation, for three consecutive years, in an Alsatian winery. Journal of Applied Microbiology, 97(6), 1140–8.

A39. Legras, J.-L., & Karst, F. (2003). Optimisation of interdelta analysis *for Saccharomyces cerevisiae* strains characterisation. FEMS Microbiology Letters, 221(2), 249–255.

A40. Kaakeh R., Legras JL, Arnaud, A. Galzy, P. (1991) : Purification and properties of the nitrile hydratase of a new strain of *Rhodococcus* sp. Zentralbl. Mikrobiol. (1991), 146(2), 89-98

A41. Legras JL, Kaakeh R., Arnaud, A. Galzy, P. (1989) : Purification and properties of a b-glucosidase from a nitrile hydratase producing *Brevibacterium* strain R312. J. Basic. Microbiol. 29, 655-699

A42. Legras JL, Kaakeh R., Arnaud, A. Galzy, P. (1989) : Degradation of cyanoglucosides by *Brevibacterium* sp R312 strain. J. Gen. Appl. Microbiol., 35, 451-461.

A43. Legras JL, Jory M., Arnaud, A. Galzy, P. (1989) : Detoxication of cassava pulp using *Brevibacterium* sp R312. Applied Microbiology and Biotechnology 33, 529-533

A44. Legras JL, Chuzel G., Arnaud, A. Galzy, P. (1989) : Natural nitriles and their metabolism. World J. Microbiol. Biotechnol., 6, 83-108.

Parmi ces 42 articles scientifiques publiés dans des revues à comité de lecture, je suis en position de premier auteur dans 10, et en position de dernier auteur dans 7. Ils totalisent 1404 citations soit une moyenne de 36,95 citations par article cité (WoS). Mon h index est de 18.

Chapitres d'ouvrage

B1. Legras, J. L., Galeote, V., Camarasa, C., Blondin, B., Dequin, S. (2017). Ecology, Diversity and Applications of Saccharomyces Yeasts in Food and Beverages. In: Yeast Diversity in Human Welfare (p. 283-321). SGP : Springer Science + Business Media Singapore.

B2. Blondin, B. Dequin, S., Querol, A. Legras, J.L. (2009) Genome of *Saccharomyces cerevisiae* and related yeast. In "Biology of Microorganisms on Grapes, in Must and in Wine" König H, Unden G. Fröhlich J (Eds.). Springer

B3. Karst, F. et Legras, JL. (2005). Méthodes de caractérisation moléculaire des levures *S. cerevisiae*. In « Les fermentations au service des produits de terroir ». Editors : Montel MC Beranger C ; Bonnemaire J., Quae editions, INRA-Paris :, 2005. – p 280-6

Communications orale dans un congrès

C1. Legras, J. L., Galeote, V., Coi, A. L., Bigey, F., Marsit, S., Camarasa, C., Sanchez, I., Nidelet, T., Couloux, A., Guy, J., Franco-Douarte, R., Schuller, D., Sampaio, J. P., Budroni, M., Dequin, S. (2017). Adaptation of Yeast to Anthropogenic Environments Using Comparative Genomics. Presented at 27. International Conference on Yeast Genetics and Molecular Biology (ICYGMB), Prague, CZE (2017-08-27 - 2017-09-01).

C2. Sicard, D. , Michel, E. , Nidelet, T. , Farrera, L. , Legras, J. L. , Robert, Y. , Casaregola, S. , Dequin, S. , Howell, K. (2017). Ecology and domestication of yeasts in wine and bread ecosystems . Presented at MicrobiOccitanie 2017 - 1. Rencontre des Microbiologistes Région Occitanie, Toulouse, FRA (2017-04-24 - 2017-04-26).

C3. Legras, J. L. (2017). Origins of yeast domestication, as revealed from wine. Presented at American Association of Physical Anthropologists 86th Annual Meeting, New Orleans, USA (2017-04-19 - 2017-04-22).

C4. Legras, J. L. , Galeote, V. (2016). Application of Saccharomyces yeasts for Industrial Fermentation. Presented at Summer School of the YEASTCELL ITN - Application of Saccharomyces yeasts for Industrial Fermentation, Montpellier, FRA. (6-06-2016 – 9-06-2016)

C5. Marsit, S. , Mena, A., Bigey, F., Sauvage, F. X., Couloux, A., Guy, J., Legras, J. L., Barrio, E., Dequin, S., Galeote, V. (2015). Adaptive Advantage Conferred by an Eukaryote-to-Eukaryote Gene Transfer Event in Wine Yeasts. Presented at ESF-EMBO Symposium, Exploring the genomic complexity and diversity of eukaryotes, Sant Feliu de Guixols, ESP. (17-10-2015 - 22-10-2015)

C6. Eder, M., Sanchez, I., Rigou, P., Nidelet, T., Camarasa, C., Legras, J. L., Dequin, S. (2015). Deciphering the Genetic and Metabolic Bases of Yeast Aroma Properties. Presented at ISSY 32, Perugia, ITA (2015-09-13 - 2015-09-17).

C7. Legras, J. L. (2014). La biodiversité chez Saccharomyces : origine, filiation. Presented at Journée technique des Œnologues "La microbiologie du vin", Taissy, FRA (2014-11-06 - 2014-11-06).

C8. Marsit, S., Bigey, F., Legras, J. L., Dequin, S., Galeote, V. (2015). Adaptive advantage of oligopetide transporters acquired by horizontal transfer during wine fermentation. Presented at 10. Symposium International d'Œnologie de Bordeaux, Œno2015 , Bordeaux, FRA.

C9. Marsit, S. , Bigey, F., Legras, J. L., Galeote, V., Dequin, S. (2014). Adaptive role of horizontally transferred oligopeptide transporters in wine yeasts. Presented at 31. International Specialised Symposium on Yeast, Nova-Gorica/Vipava, SVN.

C10. Marsit, S. , Bigey, F., Legras, J. L., Dequin, S., Galeote, V. (2014). Avantage adaptatif des transporteurs d'oligopeptides Fot1/2 acquis par transfert horizontal de gènes durant la fermentation alcoolique. Presented at Levures, Modèles & Outils – 11ème Rencontre des Levuristes Francophones, Bordeaux, FRA.

C11. Legras, J. L., Coi, A. L., Bigey, F., Galeote, V., Marsit, S., Couloux, A., Guy, J., Franco-Duarte, R., Schuller, D., Sampaio, J. P., Budroni, M., Dequin, S. (2015). New insights into the adaptation of yeast to wine making using comparative genomics. Presented at 10. Symposium International d'Œnologie de Bordeaux, Œno2015 , Bordeaux, FRA.

C12. Coi, A. L. , Legras, J. L., Bigey, F., Galeote, V., Mannazzu, I., Budroni, M., Dequin, S. (2014). Characterization of genes involved in the adaptation of flor strains to biological ageing conditions of sherry wines. Presented at 31. International Specialised Symposium on Yeast, Nova-Gorica/Vipava, SVN.

C13. Legras, J. L. , Coi, A. L., Bigey, F., Galeote, V., Marsit, S., Couloux, A., Guy, J., Franco-Duarte, R., Schuller, D., Sampaio, J. P., Budroni, M., Dequin, S. (2015). New insights into adaptation of yeast to anthropic environment using comparative genomics. Presented at 32. International Specialized Symposium on Yeasts-Yeast Biodiversity and biotechnology in the twenty-first century, Perugia, ITA.

C14. [Présentation orale] Legras, J. L. , Coi, A. L., Bigey, F., Galeote, V., Marsit, S., Couloux, A., Guy, J., Sampaio, J. P., Budroni, M., Dequin, S. (2013). New insights into the adaptation of yeast to anthropic niches using comparative genomics. Focus on flor yeast genome. Presented at ESF-EMBO Comparative Genomics of Eukaryotic Microorganisms: Complexity Patterns in Eukaryotic Genomes, Sant Feliu de Guixols, ESP.

C15. [Présentation orale] Legras, J. L. , Galeote, V., Bigey, F., Coi, A. L., Marsit, S., Cambon, B., Budroni, M., Dequin, S. (2013). First insight into the adaptation of yeast to anthropic niches from genome sequence. Presented at 5. Conference on Physiology of Yeast and Filamentous Fungi, Montpellier, FRA.

C16. Legras, J. L., Bigey, F., COI, A. L., Galeote, V., Charpentier, C., Dequin, S. (2012). Diversity and adaptation of flor yeast : new data for an old question. Presented at 13. International Congress on Yeasts (ICY 2012), Madison, USA.

C17. Sicard, D., Legras, J. L. (2010). Diversité des levures *Saccharomyces cerevisiae* et domestication. Presented at Element transposables, Virus et domestication des espèces , Lyon, FRA (2010-05-11 - 2010-05-12).

C18. Legras J.L. 2008 Analyse phénotypique et transcriptomique de la réponse de *Saccharomyces cerevisiae* aux acides gras à moyenne chaîne, Levure Modèle Outil 8, La colle sur Loup, 26-29 Octobre 2008.

C19. Legras, J. L., Blondin, B., Casaregola, S., Dequin, S. (2008). Diversité moléculaire et structure du génome chez les levures *S. cerevisiae* œnologique s. In: Les ressources génétiques à l'heure des génomes (p. 523-523). Actes du BRG (7). Presented at 7. Colloque National du BRG - Les Ressources Génétiques à l'heure des Génomes, Strasbourg, FRA (2008-10-13 - 2008-10-15). Paris, FRA : Bureau des Ressources Génétiques.

C20. Legras, J. L., Cornuet, J.-M., Karst, F. (2007). *Saccharomyces cerevisiae* diversity reflects human history. Presented at 26. International Specialised Symposium on Yeasts, Sorrento, ITA (2007-06-03 - 2007-06-07).

C21. Legras, J. L., Merdinoglu, D., Karst, F. (2005). Microsatellite characterization of *Saccharomyces cerevisiae* strains reveals population structure linked to geographic and technological origins. Presented at Biomicroworld, Badajoz, ESP (2005-03-15 - 2005-03-18).

C22. Legras J.L. ; Ruh O. ; Merdinoglu D. and Karst F (2004). Caractérisation de populations de *S. cerevisiae* par microsatellites : Les levures se sont-elles propagées avec l'extension de la culture de la vigne ? Vinitech, 3 Décembre 2004 présentation orale

Communications sous forme d'affiches dans un congrès

P1. Eder, M., Sanchez, I., Rigou, P., Nidelet, T., Camarasa, C., Legras, J. L., Dequin, S. (2017). Deciphering the Genetic and Metabolic Bases of Yeast Aroma Properties. Presented at ISSY 33, Cork, IRL (2017-06-26 - 2017-06-29).

P2. Nidelet, T., Carbonetto, M. B., Dugat-Bony, E., Callon, C., Delbes, C., Farrera, L., Legras, J. L., Dequin, S., Sicard, D., Howell, K. (2016). Ecology and diversity of yeasts in fermented food ecosystems. Presented at 16. International Symposium on Microbial Ecology - ISME16, Montréal, CAN (2016-08-21 - 2016-08-26).

P3. Legras, J. L., Coi, A. L., Galeote, V. , Bigey, F., Marsit, S., Duarte, F., Schuller, D., Sampaio, J. P., Budroni, M., Dequin, S. (2015). Deciphering the adaptation of yeast to anthropic environment using comparative genomics. Presented at ESF-EMBO Symposium, Exploring the genomic complexity and diversity of eukaryotes, Sant Feliu de Guixols, ESP.

P4. Legras, J. L., Coi, A. L., Bigey, F., Galeote, V., Marsit, S., Couloux, A., Guy, J., Duarde, R., Schuller, D., Sampaio, J. P., Budroni, M., Dequin, S. (2014). New insights into the adaptation of yeast to anthropic niches using comparative genomics. Focus on flor yeast.. Presented at EMBO Conference Series: Experimental Approaches to Evolution and Ecology, Heidelberg, DEU (2014-10-12 - 2014-10-15).

P5. Börlin, M., Salin, F.; Legras J. L.; Masneuf-Pomarede, I. (2015) Wine grape S.cerevisiae diversity and population structure reveal possible genetic admixture to commercial wine strains. Presented at ISSY 32, Perugia, ITA (2015-09-13 - 2015-09-17)

P6. Coi, A. L. , Legras, J. L., Zara, G., Bigey, F., Galeote, V., Dequin, S., Budroni, M. (2015). The genetic basis of adaptation of flor strains : a comparative genome and genetic analysis. Presented at 32. International Specialised Symposium on Yeasts-Yeast biodiversity and biotechnology in the twenty-first century, Perugia, ITA.

P7. Coi, A. L., Legras, J. L., Bigey, F., Galeote, V., Budroni, M., Dequin, S. (2013). Characterization of genes involved in yeast adaptation to Sherry-like wines' biological ageing identified by a genome based approach. Presented at ESFEMBO Comparative Genomics of Eukaryotic Microorganisms: Complexity Patterns in Eukaryotic Genomes, Sant Feliu de Guixols, ESP.

P8. Marsit, S. , Bigey, F., Legras, J. L., Dequin, S., Galeote, V. (2013). Role of horizontally transfered genes in the adaptation of wine yeasts to their environnment. Presented at 5. Conference on Physiology of Yeast and Filamentous Fungi- PYFF5, Montpellier, FRA.

P9. Marsit, S., Bigey, F., Legras, J. L., Dequin, S., Galeote, V. (2013). Adaptive role of horizontally transferred genes during wine fermentation.. Presented at ESF-EMBO Comparative Genomics of Eukaryotic Microorganisms: Complexity Patterns in Eukaryotic Genomes, Sant Feliu de Guixols, ESP.

P10. Legras, J. L. , Coi, A. L., Galeote, V., Bigey, F., Marsit, S., Cambon, B., Budroni, M., Dequin, S. (2014). Adaptation des levures *S. cerevisiae* à des niches anthropogéniques à partir d'une approche de génomique des populations. Exemple de l'élevage en voile. Presented at Levures, Modèles & Outils – 11ème Rencontre des Levuristes Francophones, Bordeaux, FRA.

P11. Erny, C., Alais, A., Legras, J. L. (2009). Ecological success of a group of *Saccharomyces cerevisiae* *Saccharomyces kudriavzevii hybrids. . Presented at 27. International specialised Symposium on Yeasts Issy-27, Paris, FRA .

P12. Legras, J. L., Schuller, D., Sampaio, J. P., Enjalbert, J. (2010). Deciphering yeast life cycle from population data. EMBO Conference Series Experimental Approaches to Evolution and Ecology in Yeast, Heidelberg, DEU .

P13. Legras JL Enjalbert JM & Karst F (2008). How did Man history affect *Saccharomyces cerevisiae* diversity? EMBO. Comparative Genomics of Eukaryotic microorganisms, Heideberg ,1-5 October 2008

P14. Legras, J. L., Cornuet, J.-M., Karst, F. (2008). De la diversité des levures *Saccharomyces cerevisiae* à l'histoire de l'homme.. Presented at 7. Colloque national "Ressources génétiques". Les Ressources Génétiques à l'heure des Génomes, Strasbourg, FRA (2008-10-13 - 2008-10-15).

P15. Charpentier, C., Legras, J. L. (2007). Les levures de voile du Jura : diversité génotypique et technologique. Presented at 8. International Symposium of Enology, Bordeaux, FRA (2007-06-25 - 2007-06-27).

P16. Legras, J. L., Cornuet, J.-M., Karst, F. (2007). How did Man history affect *Saccharomyces cerevisiae* diversity?. Presented at EFS-EMBO. Comparative Genomics of Eukaryotic microorganisms, Sant Feliu de Guixols, ESP (2007-10-20 - 2007-10-25).

P17. Legras, J. L., Erny, C., Adolphe, Y., Le Jeune, C., Lollier, M., Blondin, B., Karst, F. (2007). Phenotypic analysis and transcriptome profiling of Saccharomyces response to medium chain fatty acids. Presented at 26. International Specialised Symposium on Yeasts, Sorrente, ITA (2007-06-03 - 2007-06-07).

P18. Erny C, Legras J. L., Adolphe Y., Le Jeune C., Lollier M., Delobel P., Blondin B. et Karst F. (2006) Analyse transcriptionnelle et phénotypique de la réponse de la levure *Saccharomyces cerevisiae* aux acides gras à moyenne chaîne. Congrès de la Société Française de Biochimie et Biologie Moléculaire, Clermont Ferrand, 7-8 Septembre.

P19. Ezeronye O. U. and Legras J.L.; 2005, Molecular analysis of *Saccharomyces cerevisiae* strains of Palm wine fermentations. Biomicroworld 15-18 Mars 2005, Badajoz, Spain poster

P20. Legras J.L. ; Merdinoglu D. and Karst Francis 2003 : Microsatellite characterization of *Saccharomyces cerevisiae* highly suggests that wine yeast has spread all over the world with Vitis vinifera extension . Colmar 4th Biology Symposium. 16-17 Octobre 2003 - Poster Primé

P21. Legras, J. L., Ruh, O., Merdinoglu, D., Karst, F. (2003). Sélection de loci microsatellites hypervariables pour la caractérisation des souches de *Saccharomyces cerevisiae*. In: Rencontres 2003 des microbiologistes de l'INRA (p. 147). Presented at 3. Rencontres des microbiologistes de l'INRA, Dourdan, FRA (2003-05-05 - 2003-05-07). Paris, FRA : INRA Editions.

P22. Karst, F., Legras, J. L. (2003). Rôle des phytostérols dans la fermentation œnologique . In: Rencontres 2003 des microbiologistes de l'INRA (p. 106). Presented at 3. Rencontres des microbiologistes de l'INRA, Dourdan, FRA (2003-05-05 - 2003-05-07). Paris, FRA : INRA Editions.

P23. Legras, J.L., Karst, F. (2002). Optimization and modelization of *Saccharomyces cerevisiae* interdelta caractérisation. IUMS, The World of Microbes, Paris, 27 Juillet-1 Aout 2002

P24. Legras, J. L., Meyer, J.P., Legname, E., Schaeffer, A. (1996). Etude de la flore levurienne de différents terroirs alsaciens. In: Les terroirs viticoles. Concept. Produit. Valorisation (p. 469-471). Presented at 1. Colloque international, Angers, FRA (1996-07-17 - 1996-07-18).

Publications techniques

T1. Brice, C., Legras, J.L., Tesnière, C. et Blondin B. (2015). Pourquoi les levures ont-elles des besoins en azote différents ? Revue des Œnologues, 171, 14-17

T2. Brice, C, Sanchez I, TEsnière, C., Bigey F, Legras JL, et Blondin B. (2015). Une étude du besoin en azote des levures œnologiques révèle le rôle clé de la signalisation azote dans le contrôle de la vitesse de fermentation. Revue des Œnologues, 154, 21-24

T3. Galeote V, Bigey F, Beyne E, Novo M, Legras J-L, Casaregola S, & Dequin, S (2011). Ce que nous apprend le séquençage des levures œnologique s. Revue Française d'Œnologie , 248, 2-3.
T4. Legras JL (2011) Les levures du vin : leur biologie – En quoi la diversité nous éclaire-t-elle sur leurs propriétés et leurs origines. Actes du colloque de la percée du Vin Jaune 4/11/2012.
T5. Legras J.L. ; Meyer J.P. (2000) Levures œnologiques, un large choix pour s'adapter aux vins. Viti, 256, 13-16
T6. Gresser C.; Legras, J.L.; Gerland, C. (2000) Sélection de levures de cépages : Vers une meilleure expression de la typicité. Les vins d'Alsace, Août , 46-48
T7. Legras J.L., Meyer J.P., Schaeffer A., Meistermann E., Brechtbuhler Ch, (1997). Evolution récente de la viticulture alsacienne. Revue Française d'Œnologie 162, 9-10

Sélection de souches industrielles

L1. Sélection de deux souches pour la vinification (1998-2001): GE7, levuline B201. A venant à licence 81014SF
L2. Sélection d'un souche pour la vinification des Gewurztraminer (2001): Vitilevure 58W3. licence
L3. Sélection d'un mutant de la souche Eg8 faiblement producteur d'acide acétique (nom commercial ALS) 2007

Communications à des journées techniques

T8. Legras J.L. (2014) « Que savons-nous de plus aujourd'hui sur la diversité des levures *Saccharomyces cerevisiae* en œnologie. Journée technique Institut des Sciences de la Vigne et du Vin, Université de Bordeaux « Microorganismes indigènes. Diversité, intérêts et mise en œuvre » , Bordeaux, 10 Avril 2014.
T9. Legras JL (2014) La biodiversité chez Saccharomyces : origine, filiation. Journée technique des Œnologie s, Epernay, 22 Octobre 2014.
T10. Legras J.L., Karst, F., Demuyter, C, Lollier M, Le jeune C., Walter B. (2004) Caractéristiques et physiologiques des levures alsaciennes Réseau RVVS, Colmar, 1-2 Avril 2004.
T11. Legras J.L., Merdinoglu, D., Karst, F. (2004) La caractérisation des levures *Saccharomyces cerevisiae* de vin par les microsatellites suggère qu'elles se sont répandues à travers le monde avec l'extension de la vigne
T12. Legras J.L. (2003) « Caractérisation génétique des levures alsaciennes : des levures des parcelles à leur implantation dans les chais. ».– Journées de formation l'ITV Ostheim, 19 Janvier 2003.

Réseau International Collaborations

D. Cavalieri , Université de Florence, Italie (A4, A24)
V. Corich, University of Padoue, Italie (A7)
M. Budroni, Università di Sassari, Sassari, Italy . (Thèse AL Coi en cotutelle, A16,A11,A9,A5)
J.P. Sampaio, Universidade Nova, Lisbonne, Portugal (A15)
D. Schuller, Universidade do Minho, Braga, Portugal (A1 sousmis)
N. Dominy Dartmouth College USA.

3. Participation à l'animation de la recherche

Encadrement d'étudiants niveau master 2

Pauline Raoult (2006) Stage de 3ème année d'ingénieur ESITPA, Rouen (6mois).
Caractérisation d'une souche mutante de la levure Eg8 produisant moins d'acide acétique », 6 mois

Laetitia Muller (2003) Stage d'ingénieur, Polytech'Lille (6 mois).
Effet du levurage sur la diversité microbienne des fermentations viniques.
Evaluation des flores après étalement et à partir de l'amplification de l'ADNr de la biomasse.

Olivier Ruh (2003) : Stage de DESS Analyses Biologiques et Chimiques. Strasbourg ; (10 mois).
Développement d'une méthode de genotypage des levures *S.cerevisiae* par de l'ADN microsatellite.

Agathe Bursin (1998) Stage de DNO IUVV Université de Bourgogne, Dijon (4 mois).
Etude et caractérisation de la population levurienne de vinifications en flore indigène

Séverine Mornand (1996) Stage DNO Université de Reims (8 mois).
Influence du débourbage sur la production d'acides organiques

Clarisse Meyer (1994) Stage de DNO (4mois) Université de Bourgogne, Dijon (4 mois).
Influence du débourbage sur la production d'acides organiques.

Anne Valérie Péneau-Rousseau (1994) Mémoire d'ingénieur ENITA Bordeaux (6 mois).
Etude de la flore levurienne de différents terroirs, identification par la technique electrophorèse en champ pulsé

Fabrice Masson (1991) : DEA Ampélologie - Œnologie. Faculté des Science de Dijon, Université de Bourgogne (6 mois).
Incidence des β-glucosidases de *Saccharomyces cerevisiae* et de préparations enzymatiques commerciales sur les anthocyanes de la baie de raisin

Encadrement d'étudiants en stage de Doctorat

Coresponsabilité directe d'encadrement doctoral

Damien Steyer 2011 : Ecole doctorale Université de Strasbourg. Directeur de Thèse Francis Karst,
Etude génétique du métabolisme des acides gras et des terpènes aromatiques chez la levure en conditions fermentaires

Ana-Lisa Coi 2014 : Ecole doctorale : Sciences des Procédés, Sciences des Aliments. Montpellier Supagro. Directrices de thèse : Sylvie Dequin et Marilena Budroni
A genome based approach to characterize genes involved in yeast adaptation to Sherry-like wines' biological ageing

Marine Borlin 2015 : École doctorale : Sciences de la Vie et de la Santé, Université de Bordeaux. Directrice de thèse : Isabelle Masneuf
Diversité et structure de population des levures *Saccharomyces cerevisiae* à l'échelle du vignoble bordelais

Matthias Eder 2017 : Ecole doctorale : Sciences des Procédés, Sciences des Aliments. Montpellier Supagro. Directrice de thèse : Sylvie Dequin
Deciphering the genetic and metabolic bases of yeast aroma properties

Amandine Deroite en cours: Ecole doctorale : Sciences des Procédés, Sciences des Aliments. Montpellier Supagro. Directrice de thèse : Sylvie Dequin
Bases génétiques et réduction de la production d'acide acétique chez des hybrides *Saccharomyces*

Participation à l'encadrement doctoral

Claire Brice 2014 : Ecole doctorale : Sciences des Procédés, Sciences des Aliments. Montpellier Supagro. Directeur de thèse : B. Blondin
Étude des bases génétiques et physiologiques du besoin en azote des levures *Saccharomyces cerevisiae* en fermentation alcoolique

Participation à des Jury de thèse en temps qu'examinateur

Marie José Ayoub 2006 : Directeur de thèse : C. GAILLARDIN, Ecole doctorale : Ecole Doctorale ABIES, Paris
Mohand Sadoudi 2014 : Directeur de thèse : H. Alexandre, Ecole doctorale : Environnement santé, Dijon
Cédric Grangeteau 2016 : Directeur de thèse : M. Guilloux Benatier, Ecole doctorale : Environnement santé, Dijon

Autres encadrements de stage :

Stagiaires de DUT ou BTS : 18 étudiants.

Participation à des comité éditoriaux de revue

Je participe depuis 2017 à au comité éditorial de la revue « Yeast » en temps que section editor.

4- Participation à l'enseignement

Cours et travaux dirigés: « Introduction to Population Genomics » Wroclaw, Pologne, 15-11-2017 (2h cours + 2h TD)

Travaux pratiques de microbiologie et contrôle des vins SupAgro Montpellier – spécialisation Viticulture-Œnologie (2011-2017) 16 heures

Invitation comme « Visiting Professor » (10 jours), Sassari, Italie Juillet 2013

Initiation à l'Assurance Qualité : DESS Valorisation Substances Naturelles Végétales, Université de Strasbourg (2003-6). 3 heures

5. Participation à l'animation collective

J'ai été plusieurs fois membre des conseils d'unité à Colmar (SVQV), puis à Montpellier (SPO), membre des conseils de gestion du département INRA Transformation des Produits Végétaux puis Caractérisation Elaboration des Produits issus de l'Agriculture entre 1996 et 2000, membre du conseil de gestion du centre de Colmar (1996-2000), représentant du personnel élu pour le corps des ingénieurs en CAPN (3 mandats entre 2003 et 2015), et je suis actuellement membre du conseil scientifique du département MICA depuis 2017. Mes participations aux instances collectives ou paritaires traduisent simplement mon souhait de contribuer à la vie de l'institut, soit en représentant le personnel auprès de l'administration, soit au niveau de l'organisation de l'unité. J'ai animé l'organisation d'une journée scientifique de l'UMR pour qu'elles se connaissent mieux et ainsi favoriser les échanges entre équipes : une collaboration ne se décrète pas, elle peut s'obtenir à partir de la prise de conscience de l'intérêt mutuel qu'elle représente. J'ai aussi participé à la réflexion sur la réorganisation de l'unité et marqué ma préférence pour une scission de notre ancienne équipe pluridisciplinaire mais trop importante vers deux équipes sœurs plus spécialisées du point de vue thématique : ADEL et FLAM.

6. Projets/contrats de recherche

Projet où j'apparais en temps que porteur:

ECO-NET 2008-2009 (ministère affaires étrangères) (17Keuros) « Diversité et capacité technologique des flores de levures de vinification d'Europe Centrale » Collaboration avec F

Matei-Radoi, Université des sciences agronomiques et vétérinaires de Bucarest (Roumanie).
Viniflhor 2008-2009 (7 keuros) « Préservation du pool terpénique des moûts en fermentation - Etude du catabolisme des terpénols par les levures » (financement partiel de la thèse de D. Steyer.
ANR PEAKYEAST : 2016-01-01 (- 2019) (410Keuros) « Moving the wine yeast *Saccharomyces cerevisiae* up its adaptive peak in grape must »

Projet où j'apparais en temps que responsable de tâche:
ANR VINICULTURE 2017-01-01 (-2020) (52 Keuros) « Vignes et vins en France du Néolithique au Moyen Age. Approche intégrée en archéosciences », porteur L Bouby ISEM, Montpellier.

7. Résumé de mon parcours professionnel

Le bilan de mon parcours scientifique, doit être évalué en fonction de mon parcours professionnel qui inclut des activités scientifiques, d'enseignement et l'encadrement d'un plateau analytique de service en œnologie.

En 1980, j'ai choisi une filière d'ingénieur agronome avec l'espoir de me spécialiser en biotechnologie. J'ai intégré l'ENSA de Rennes, mais en l'absence de cette spécialité à Rennes, j'ai choisi l'option de « Microbiologie Industrielle » proposée à l'ENSA de Montpellier (SupAgro), au Laboratoire de Microbiologie Industrielle et de Génétique des Micro-organismes sous la direction du Pr. P. GALZY. Au cours de mon stage de DEA, j'ai isolé et caractérisé des bactéries capables de métaboliser l'acrylonitrile. La nitrile hydratase de la meilleure de ces souches a été ensuite étudiée de manière plus approfondie (Kaahkeh et al. 1991). J'ai poursuivi ce travail dans une direction un peu différente par un travail de thèse sur le thème «Etude des voies de dégradation des cyanoglucosides par une souche de *Brevibacterium* sp. Application à la détoxication du manioc », qui m'a amené à purifier la β-glucosidase de cette bactérie, d'en étudier l'action conjointe avec sa nitrile-hydratase et son amidase sur les cyanoglucosides des amandes ou du manioc. A la fin de cette thèse, j'étais l'auteur de 5 articles (4 en premier auteur) associés à ce travail. Mon domaine de spécialité était donc le screening de souches de bactéries pour des activités enzymatiques, l'enzymologie, l'identification de bactéries, et je m'étais ouvert à la fermentation lactique du manioc.

A la fin de cette thèse, j'ai eu la chance de réaliser mon service militaire au titre de la coopération (VSNA) dans le laboratoire du Prof J. Ziffer, au Technion à Haifa, Israel. Au

cours de ce séjour j'ai réalisé un screening de souches de champignons filamenteux capables d'hydrolyser l'amidon non gélatinisé que j'ai isolés à partir de différentes origines. Ce séjour m'a permis de découvrir le fonctionnement d'un laboratoire de type anglo-saxon pour son financement et révélé l'importance de l'activité de publication.

A mon retour en France, après avoir soutenu ma thèse, j'ai été recruté en temps qu'ATER à l'Université de Bourgogne dans le laboratoire d'œnologie, où j'ai cette fois encadré pour la première fois un étudiant de DEA, dont le sujet était l'étude de « l'impact des β-glucosidases de *S. cerevisiae* sur la couleur des vins ».

Durant cette année j'ai enseigné les travaux pratiques d'analyse des vins, découvert l'œnologie, réalisé mes premières vinifications, et découvert l'importance des levures dans l'élaboration de l'arome du vin. Ce séjour a décidé de mon orientation professionnelle vers l'œnologie car il m'a décidé à me présenter au concours ouvert pour un poste d'ingénieur de Recherche à l'INRA à la station de Recherche Vigne et Vin de Colmar.

Le profil affiché de ce poste d'ingénieur comprenait à la fois une activité de recherche et l'animation d'un plateau technique analytique dédié aux analyses de routine des vins pour la profession viticole. En ce qui concerne l'activité recherche, le thème général de recherche de l'équipe étant « la caractérisation de l'impact des terroirs alsaciens sur la qualité des vins », il m'avait été proposé de développer un projet de recherche de caractérisation des flores de levures des différents terroirs. En ce qui concerne le volet analytique, après avoir obtenu le diplôme d'œnologue requis pour assurer ce poste, il m'a été confié la mise en place de l' accréditation COFRAC pour l'analyse des vins et la responsabilité de ce plateau analytique. Cette mission a soulevé un grand nombre de difficultés que j'ai résolues peu à peu : nous avons obtenu l'accréditation COFRAC du laboratoire en 1994, mais achevé de résoudre nos difficultés seulement vers 1997. Ceci c'est fait au détriment de mon activité de recherche, mais j'ai réussi néanmoins à maintenir une activité minimale, qui a abouti à la sélection de 3 souches de levures pour la vinification, et à la publication de 2 articles techniques et un poster dans un colloque avec actes.

L'arrêt de l'activité analytique de service en 2001 m'a permis de reprendre de plein pied une activité de recherche et le projet d'analyse de la diversité des levures de sous la direction de F. Karst (Université de Strasbourg), en lui redonnant une nouvelle dynamique. Dans un premier temps, nous avons cherché à développer de nouvelles méthodes de

caractérisation moléculaire des levures : PCR interdelta, ou typage par microsatellites que nous avons appliqué alors à différentes populations de levure. De plus l'expérience des années 1992-96, m'a fait prendre conscience que la diversité des levures devait être analysée avec des outils de génétique de population, ce qui ne se faisait pas à cette époque pour *S. cerevisiae*. Le développement de marqueurs microsatellites pour *S. cerevisiae* a marqué une rupture dans la capacité de répondre aux questions abordées jusqu'ici à partir du typage moléculaire chez la levure. Elle aurait permis de revisiter de nombreuses questions laissées en suspend comme les dynamiques de populations entre régions, de la parcelle au chai, ou d'étudier de nouvelles populations de levures en utilisant ma base de données de génotypes.

En réponse à la forte pression de mon département scientifique (CEPIA), j'ai infléchi mon projet de recherche vers des thèmes plus en rapport avec la technologie, comme la résistance des levures aux acides gras inhibiteurs, ou l'analyse des bases génétique de la composition aromatique des vins. Ces thèmes m'ont ouvert la voie vers les « Omics » en m'amenant à me former à la transcriptomique et la génétique quantitative chez la levure.

A la suite d'une évaluation insuffisante de notre équipe en 2009 par l'AERES, conduisant à sa réorientation vers l'amélioration de la vigne (et au changement de département scientifique INRA), j'ai demandé une mobilité sur Montpellier à l'UMR SPO, dans l'équipe de S. Dequin, pour proposer comme projet la recherche des déterminants géniques expliquant l'adaptation des levures de voile à partir d'une approche génomique. Cette dernière phase de mon parcours professionnel m'a ainsi amené à me former aux outils de la génomique et à la bioinformatique associée, et glisser vers la génomique des populations.

Synthèse de l’activité de recherche

Introduction

La fermentation alcoolique est une étape décisive de la transformation du mout de raisin en vin. Depuis la découverte du rôle des levures par Pasteur à la fin du 19ème siècle, la nature de ce microorganisme, sa spécificité en fonction des régions, des crus, des vins de l'ampleur de sa contribution à l'arôme des vins ont été le sujet de débats récurrents pour les viticulteurs. Chaque vigneron pense souvent posséder une flore unique dans ses parcelles, adaptée au terroir local et à la transformation du raisin. De plus, l'image d'une flore régionale de levure est de plus en plus véhiculée par les médias, avec une connotation de respect de la nature. En parallèle, la technique du levurage qui s'est largement développée dans les années 1980-90 pour remédier en premier lieu aux problèmes d'arrêts des fermentations puis pour l'amélioration de la qualité organoleptique, est aussi très critiquée par certain vinificateurs au nom de la perte d'authenticité.

En 1990, la vision des viticulteurs d'une flore de terroir s'opposait à celle de certains scientifiques qui montraient la rareté de ce microorganisme sur les fruits et raisins sains (Davenport, 1974; Rosini et al., 1982; Poulard et al., 1985) qui posaient la question de l'origine des fermentations non ensemencées. L'italien A. Martini résume cette situation en affirmant que la microbiologie du vin a été dominée par un certain nombre de dogmes inexacts dont le premier était:

"La levure de vin par excellence, *Saccharomyces cerevisiae,* (et synonymes) est ubiquiste dans la nature, avec une préférence forte pour le sol des vignobles et des vergers… ce premier dogme a un corollaire reposant sur la croyance que chaque microclimat est caractérisé par une flore de *Saccharomyces* spécifique, adaptée par sélection pour la fermentation du cépage local dans cet environnement..." (Martini and Vaughan-Martini 1990; Martini 1993).

Pour ces mêmes auteurs *S. cerevisiae* n'étant présente que dans les chais, elle était donc une levure domestiquée par excellence. Cependant dans le même temps, Cuinier et al. (1981) montraient qu'une souche de levure pouvait devenir majoritaire dans une cuve tout en étant absente du chai montrant que la position extrême des levures absentes du raisin ne pouvait tenir. Aussi, les questions de spécificité des levures de terroir, et des pratiques susceptibles de les modifier comme les itinéraires phytosanitaires ne pouvaient être séparées de l'implantation dans les chais.

Ces années 1990 ont été également marquées par l'avènement des techniques de

biologie moléculaire permettant la différenciation des levures, et offrant donc les outils pour essayer de résoudre ces questionnements d'écologie. On peut citer successivement le développement des techniques de RFLP de l'ADN mitochondrial permettant de différentier les levures de bière, (Aigle et Moll 1984), le profil RFLP de l'ADN nucléaire (Degré et al. 1989), l'électrophorèse en champ pulsé (Vezinhet et al. 1990) et enfin 1992, une technique simplifiée de RFLP de l'ADN mitochondrial (Querol et al. 1990). La première méthode basée sur l'utilisation de la PCR est apparue un peu plus tard (Ness et al. 1993) mais était moins discriminante que les méthodes précédentes.

Partant de ce questionnement, mon parcours scientifique, a évolué vers l'amélioration des méthodes de typage, l'exploration de la structure de population chez *S. cerevisiae*, l'adaptation de *S. cerevisiae* à ses niches écologiques en lien avec la domestication et l'exploration des bases génétiques des propriétés des souches.

1. Diversité des levures et terroirs Alsaciens

A mon arrivée à l'INRA, à Colmar, le laboratoire d'œnologie de l'unité Vigne et Vin, était engagé dans un programme de caractérisation des terroirs viticoles Alsaciens en collaboration avec une équipe d'agronome de l'INRA de Colmar, et l'Unité de recherche Vigne et Vin d'Angers. Les agronomes proposaient une définition des terroirs viticoles reposant sur la notion de séquence écogéopédologique mise au point par R Morlat (INRA) dans la vallée de la Loire, et incluant la nature de la roche sous-jacente, le faciès pédologique, et le paysage associé. Cette approche a été complétée pour les terroirs Alsaciens par l'orientation des versants (Lebon et al. 1993). Au sein de cette thématique, il m'a été proposé d'évaluer la diversité des levures en lien avec les terroirs alsaciens. A ce moment, les deux méthodes disponibles et accessibles pour différencier les levures étaient le profil RFLP de l'ADN mitochondrial (Aigle et al. 1984), et l'électrophorèse en champ pulsé (ECP)(Vezinhet et al. 1990) . Cette dernière méthode étant reconnue comme une méthode à la fois performante et assez rapide de différenciation des levures venait d'être appliquée avec succès à l'étude de la dynamique des populations de levures présentes durant des vinification en rouge à Bordeaux (Frezier and Dubourdieu 1992), nous l'avons donc retenue pour l'étude des populations des levures de terroirs alsaciens.

Notre objectif était d'évaluer la spécificité des levures de terroir, leur rémanence dans la parcelle et leur implantation dans les cuves. Pour cela nous avions choisi d'analyser la flore de quatre terroirs de caractéristiques pédoclimatiques très contrastées, situés sur les versants de deux vallées Alsaciennes, en minivinification en asepsie contrôlée dans la cuverie expérimentale. Pour les conditions aseptiques le jus de 20 kg de raisin a été laissé fermenter, et 40 isolats ont été caractérisés en début milieu et fin de fermentation par ECP. Pour les minivinifications traditionnelles un seul prélèvement réalisé en fin de fermentation a été analysé de la même manière. La première présentation des résultats obtenus pour les campagnes 1992, 1993 et 1994 (représentant l'analyse de 1920 souches en ECP) a été faite en 1996.
Nous avions observé que pour les vinifications de raisins en conditions aseptiques, un profil était présent sur trois parcelles et majoritaire dans deux. Il avait été rencontré sur les deux parcelles en 1992 1993 et 1994. Mais ce maintient pluriannuel n'était pas observé sur la troisième parcelle. La quatrième parcelle avait montré quant à elle une flore différente des trois autres et variable dans le temps. Par contre, dans la cuverie expérimentale avec une quantité de raisin plus grande nous avions observé un très grand polymorphisme en fin de fermentation, l'absence de souche dominante de parcelle, et l'absence des profils précédemment relevés. De plus les souches C19 et Eg8, largement utilisées dans le chai et en Alsace, étaient retrouvées dans quelques fermentations. Pour publier cela, il me paraissait essentiel de confirmer l'identité (ou tout au moins un lien de parenté) des souches de terroirs différents, par une seconde méthode de caractérisation moléculaire, ce qui n'a pas été fait.
Ce travail a été poursuivi en collaboration avec une équipe de l'IUT de Colmar, en nous focalisant sur la dynamique des populations du raisin au chai. Nous avons réalisé des prélèvements chez 8 viticulteurs sur deux parcelles de raisins et dans les 2 cuves correspondantes du chai. Les caryotypes moyens obtenus à partir de la biomasse des prélèvements de raisins et des cuves étaient semblables dans 3 cas sur 16. Pour les comparaisons dans lesquelles les caryotypes de raisin et de chai différaient, l'analyse de clones isolés, nous a fourni un cas supplémentaire de correspondance entre profils majoritaires des raisins et du chai, et enfin pour deux autres parcelles, un caryotype minoritaire du raisin s'était implanté dans la cuve (rapport DNO A. Bursin 1998). En conclusion, le transfert de raisin semble apporter des levures au chai mais le devenir de ces souches semble globalement incertain (6 cas sur 16). Enfin, durant ces essais,

nous avons observé le maintient d'une population de souches de *S. uvarum* dans la cuverie (Demuyter et al. 2004).

2. Développement de méthodes pour explorer la diversité des levures au niveau intra spécifique

Les analyses réalisées durant la période 1992 - 2000 m'ont persuadé de la relativité des performances des méthodes de typage par ECP chez *S. cerevisiae*. En effet, lors de comparaison de nombreuses souches par ECP, bien que les profils soient réalignés grâce à une référence, certains profils caryotypes simples (16 bandes) étaient très fréquents et présentaient des correspondances imparfaites suggérant que malgré leur grande similarité les souches pouvaient être différentes. De plus, les deux autres méthodes populaires à ce moment : le profil de restriction de l'ADN mitochondrial et l'amplification des séquences localisées entre les éléments delta (LTR des rétrotransposons TY1 et TY2 de levure) aussi appelée PCR interdelta, présentaient des limites : la première posait des soucis de spécificité et de reproductibilité (Nguyen , communication personnelle) alors que la seconde, plus aisée à utiliser et plus rapide, était par contre moins performante que les deux autres méthodes (Vezinhet et al. 1993).

Aussi, avec l'arrêt de l'activité analytique, et mon retour vers une activité de recherche à temps plein, il m'a paru prioritaire de disposer de meilleures méthodes. Comme celles ci avaient été mises au point avant l'achèvement du séquençage du génome de la levure *S. cerevisiae* S288C en 1996 (Goffeau et al. 1996), il était possible d'utiliser cette information pour les améliorer et évaluer leur capacité.

2.1. Optimisation de la méthode de typage par PCR interdelta (A39)

La première version des PCR interdelta étant peu performante, il nous a paru intéressant d'essayer d'utiliser la séquence du génome de S288C pour évaluer et améliorer cette méthode et essayer de modéliser le profil obtenu. Nous avons observé en premier que le nombre de sites de fixation des deux amorces était très différent et avons donc recherché de nouvelles amorces dans des régions conservées des éléments delta du génome de S288C. Ces nouvelles amorces (delta 12 et delta 21) révèlent un polymorphisme supérieur lorsqu'elles sont utilisées pour le typage des levures (figure 1)(Schuller et al. 2004). Par ailleurs, la recherche des sites d'hybridation de ces amorces par BLAST m' a permis de prédire 24 bandes potentielles. Parmi ces 24

bandes, celles qui étaient prédites à partir de primers identiques n'étaient pas retrouvées, comme cela est décrit pour l'AFLP. Nous avons ainsi pu remarquer également que cette méthode ne puisait son information qu'à partir de 7 sur 16 chromosomes dans le cas de S288C. Enfin, il est à noter que les profils de levures de vin, ne produisent souvent qu'entre 5 et 12 bandes.

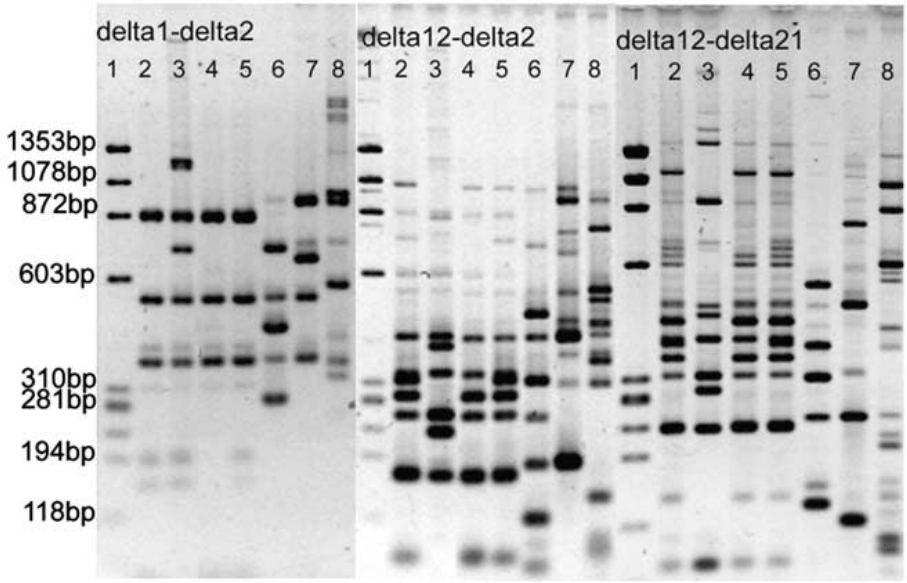


Figure 1 : Comparaison de gels d'électrophorèse obtenus en utilisant les paires d'amorces delta1–delta2, delta12–delta2 and delta12–delta21 sur différentes levures. Pistes 1–10 de chaque gel: 1 PM φX174 digéré par HaeIII; 2, S288C; 3, FL100; 4, W303; 5, FY1689; 6, AWRI350; 7, AWRI750; 8, Clib319. (tiré de A39).

Cette version optimisée du typage delta fournissant des profils plus informatifs que l'ancienne version, est devenue aujourd'hui une des méthodes de typage les plus populaires (citée par 161 articles) après le typage de l'ADNm (Querol and Barrio 1990).

2.2. Loci microsatellites polymorphes pour la différenciation des souches *S. cerevisiae*

Le typage interdelta utilise pour ancrage les éléments delta qui sont répartis assez aléatoirement dans le génome. Cependant, il est possible que des bandes de même taille soient obtenues à des endroits différents du génome. Ce risque d'homoplasie est d'autant plus élevé que la résolution utilisée pour les électrophorèses est basse. De plus cette méthode est sensible à la qualité et la quantité de l'ADN, ou à la polymérase utilisée pour l'amplification, défauts qu'elle partage avec les typages RAPD (Gil-

Lamaignere et al. 2003). Aussi, nous nous sommes tournés vers les marqueurs microsatellites pour la différentiation des levures. L'utilisation de marqueurs microsatellites s'était largement répandue pour la caractérisation moléculaire de plantes, d'animaux ou même l'homme pour l'identité judiciaire, et les premiers travaux publiés chez la levure (Hennequin et al. 2001, Perez et al. 2001) montraient à la fois le potentiel de cette méthode, mais ne semblaient pas suffisamment performants pour différencier les levures œnologiques. Aussi nous avons recherché de nouveaux loci en utilisant la séquence du génome de S288C et en nous focalisant sur les motifs de grande taille. Nous avons ainsi identifié les 6 loci les plus polymorphes qu'il était possible de combiner pour le typage de *S. cerevisiae* (A37)(Figure 2)

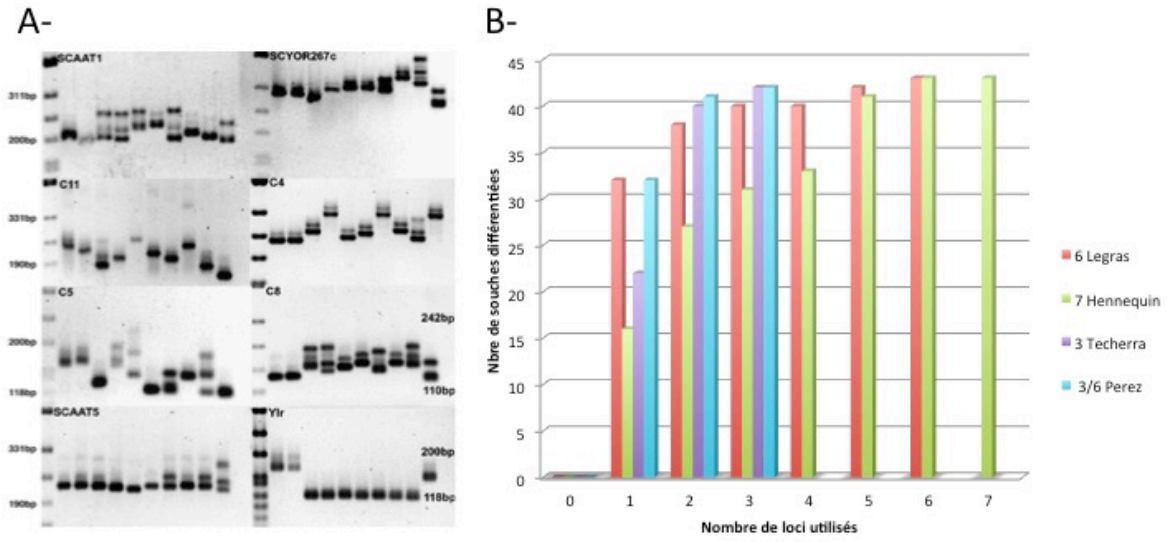


Figure 2 : A- Polymorphisme de 10 souches de levures révélé par 8 loci microsatellites. B- Evolution du pouvoir discriminatoire obtenu pour différentes combinaison de loci: 6 marqueurs (Legras et al. 2005), 7 marqueurs (Hennequin et al. 2001), 3 loci (González Techera et al. 2001), 3 parmi 6 (Pérez et al. 2001). Les 3 loci de Gonzalez et de Perez et al, sont les mêmes et inclus dans les 6 que j'ai proposés.

2.3. Loci microsatellites pour la différenciation d'autres espèces du genre Saccharomyces

Dans le prolongement du travail précédent, nous avons utilisé cette stratégie pour d'autres espèces de levures pour lesquelles peu ou pas d'informations étaient disponibles. Cela nous a permis de caractériser des souches de l'espèce *S. kudriavzevii* isolées à partir du chêne et la partie *S. kudriavzevii* du génome d'hybrides interspécifiques d'une levure industrielle Eg8 (A25). De même nous avons sélectionné plusieurs loci complémentaires pour la caractérisation de souches de l'espèce *S.*

uvarum (Masneuf et al. 2016) afin d'explorer sa structure de population. J'ai aussi proposé un volet similaire pour la caractérisation de l'espèce *Dekkera bruxellensis* qui pose de grands problèmes de vinification (dans le cadre d'un projet non retenu par l'ANR en 2011). Depuis, plusieurs jeux de loci microsatellites ont été développés pour la caractérisation de différentes espèces œnologiques (Albertin et al. 2014; Albertin et al. 2016; Hranilovic et al. 2017) démontrant ainsi la qualité de cette technique pour explorer la diversité des levures du phylum des Ascomycètes a coût modéré.

3. *Structure de population chez S. cerevisiae*

La méthode de typage interdelta permet de comparer les souches à peu de coût et est intéressante pour décrire la dynamique des souches de levure durant la fermentation. En comparaison, les marqueurs microsatellites sont d'un coût légèrement plus élevé mais révèlent un polymorphisme plus élevé. Ces marqueurs co-dominants ouvrent de plus un champ analytique plus large en permettant les analyses de ploïdie, de diversité génétique, de différentiation de population, de phylogénie (Lowe et al. 2004) ainsi que les inférences de flux de gènes ou de scénarii démographiques. Au moment où j'ai entrepris ce travail, l'utilisation de ces marqueurs chez *Coccidioides immitis* (Fisher et al. 2001) était remarquable car elle avait montré en Amérique l'existence de 2 populations de ce champignon pathogène ayant divergé lors de la colonisation Nord-Sud de ce continent par les humains. De la même manière ces marqueurs avaient été utilisés chez *Candida albicans* pour comparer des populations de différents pays entre elles (Fundyga et al. 2002), ou chez des champignons pathogènes (Enjalbert et al. 2005).

A ce moment, bien que l'on sache que des souches de *S. cerevisiae* avait été isolées de nombreux produits fermentés dont l'histoire remonte à l'antiquité (vin, bière, pain, fromage…), et que ces souches avaient été isolées sur les différents continents, personne n'avait identifié de structure de population chez ce champignon. De plus la question de l'origine de cette levure était toujours débattue par les partisans de la domestication de l'espèce car uniquement rencontrée dans les produits fermentés et non retrouvée dans l'environnement à moins d'utiliser des techniques d'enrichissement (Martini 1993).

Pour ces raisons j'ai entrepris de collecter des souches de *S. cerevisiae* isolées dans des origines variées dans les différentes régions du globe afin de rechercher quelle structure de population pouvait exister chez *S. cerevisiae.*

3.1. Structure de population et influence de l'homme chez *S. cerevisiae* (A34)

Pour cette partie j'ai sollicité l'appui de JM Cornuet, généticien des populations, (INRA, CBGP, Montpellier). Cette collecte de souches m'a permis d'obtenir le génotype de 651 souches de levures à 12 loci microsatellites. Un cladogramme obtenu à l'aide de la distance d'arc DC de Cavalli Sforza (Cavalli-Sforza and Edwards 1967) ou ici avec la distance de Bruvo (Bruvo et al. 2004) montre clairement le regroupement des souches en fonction de leur origine. En particulier, on peut facilement reconnaître les groupes de levures isolées à partir de vin, sake, pain, fromage, vin de palme, de chêne etc… (Figure 3). Afin, d'évaluer si ces populations étaient régionales, nous avons comparé les matrices de distances génétiques entre populations avec les distances géographiques. Dans la mesure, ou l'analyse de ces populations indique que 28% de la variabilité génétique est expliquée par l'isolement géographique (sauf pour les levures de vin) ceci suggère que chacun de ces clusters possède une origine locale. En ce qui concerne les levures de vin, l'analyse de différentiation entre populations de régions viticoles mesurée à l'aide de la distance Fst montre une position basale de la population provenant du Liban et un regroupement entre régions qui suggère d'une part une origine des levures de vin située en Mésopotamie, et d'autre part la présence de route de migrations au travers de la Méditerranée, la vallée du Rhône, ou le Danube. Enfin l'inférence bayésienne (ABC) d'un modèle démographique de différenciation entre les populations les plus divergentes, utilisant comme date d'arrivée en France 500 av. JC, nous a permis d'obtenir pour les populations les plus divergentes une date compatible avec l'avènement de l'agriculture.

La présence de tels groupes de souches associés à des produits fermentés a été décrite en 2005 (Fay and Benavides 2005) pour les levures de vin et sake, simultanément à la présentation de nos premiers résultats (2003, 2004 et 2005) et avant la publication de nos résultats. Ce regroupement des isolats en fonction de leur origine technologique a amené J. Fay à proposer la domestication des levures de vin et de sake en s'appuyant sur leur plus faible diversité. Néanmoins la définition de ce concept chez les microorganismes pose question.

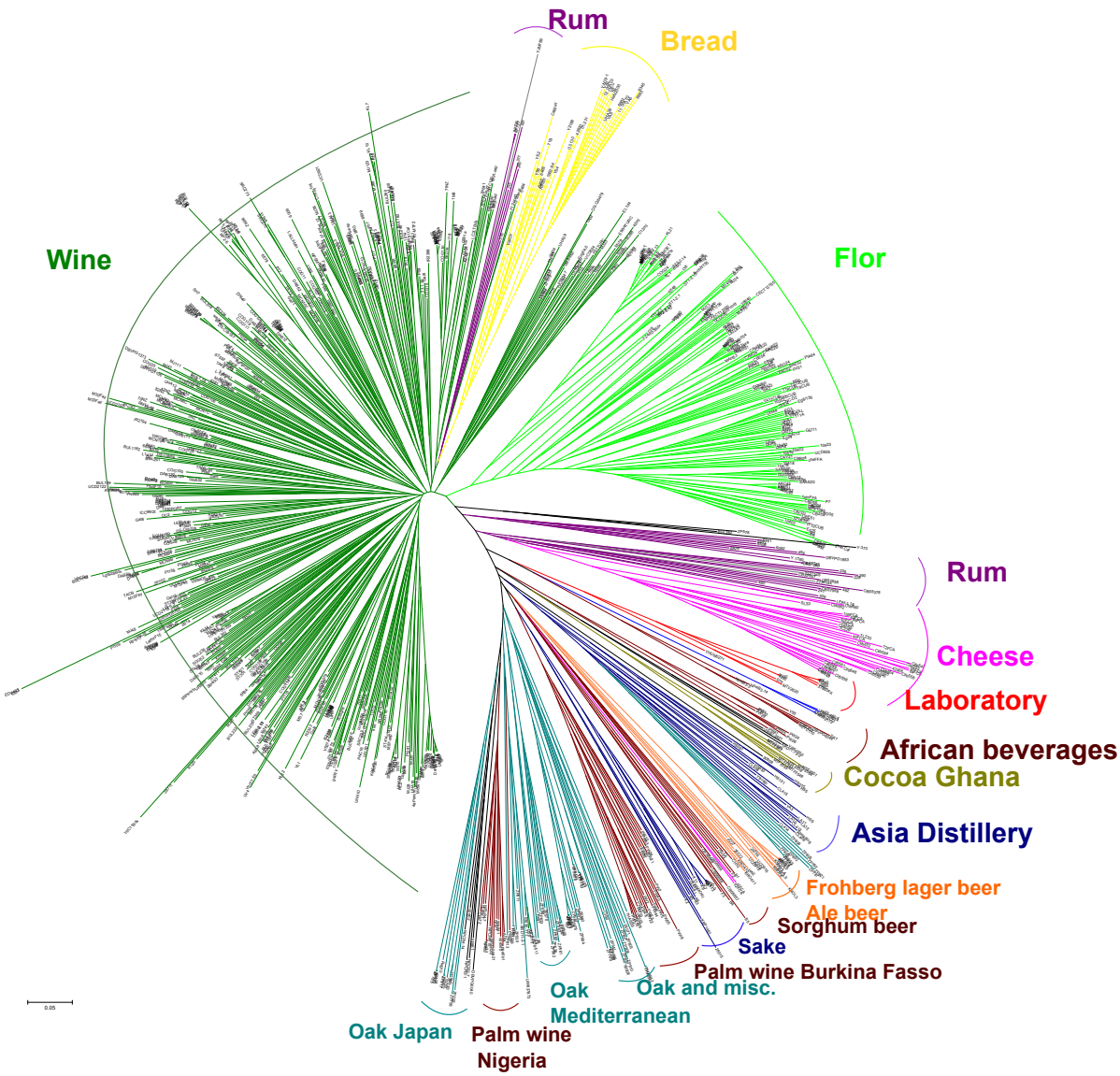


Figure 3: Arbre non enraciné représentant les relations obtenues entre 1000 levures de différentes origines, mis à jour à partir des données de A34. Cette version est construite à partir d'une matrice de distance de Bruvo (Bruvo et al. 2004) et un regroupement selon la méthode du « neighbor joining ».

Un des spécialistes de la domestication chez les animaux propose la définition suivante « Espèce élevée en captivité et par conséquent modifiée de son ancêtre sauvage de manière à la rendre plus utile aux humains qui contrôlent sa reproduction (pour les animaux) et son alimentation» (Diamond 2002). Pour les isolats *S. cerevisiae* des levains de pain, de bière, de voile, de sake, la propagation très ancienne par l'homme donne des arguments solides pour la domestication. Le cas des levures du vin est plus discutable car le cycle des levures semble comprendre à la fois une partie au vignoble et une partie au chai. En effet, les résultats des expériences de dissémination menées à

l'aide d'électrophorèse en champ pulsé semblaient indiquer un faible retour des levures utilisées du chai vers le vignoble (Valero et al. 2005), suggérant une compartimentation entre chais et vignobles. Cependant nous n'avons pas observé de notre coté de différentiation entre souches de raisins et de chais, et nous avons remarqué la présence de plusieurs isolats de raisins aux génotypes très semblables à ceux de préparations commerciales.

3.2. Insectes et vection chez *S. cerevisiae* (A24)

Alors que les levures *S. cerevisiae* sont quasi absentes des raisins durant leur premières phases de développement, leur fréquence augmente de manière très importante en fin de maturité, à l'automne, en particulier sur des raisins altérés (Mortimer 2000). Ceci amène nécessairement à poser la question du mécanisme permettant l'arrivée de ces levures sur les raisins et donc envisager un mécanisme de vection. Les mouches *Drosophila melanogaster* sont attirées par les produits fermentés mais pas par les fruits (Becher et al. 2012) suggérant qu'elles ne sont pas le bon vecteur, ou tout au moins celui fournissant la première inoculation. La présence élevée de levures *S. cerevisiae* sur les raisins altérés (Mortimer 2000) suggère qu'une blessure est à l'origine de la première inoculation des baies. Sachant que ni les drosophiles du genre *D. melanogaster,* ni les abeilles ne disposent de pièces buccales capables de blesser la pellicule des baies, D. Cavalieri (Université de Florence) s'est tourné vers les guêpes et les frelons. Son étudiante en doctorat I. Stefanini, avait collecté des polistes et frelons, et isolé la flore levurienne dont *S. cerevisiae*. D. Cavalieri m'a alors contacté pour analyser les populations de levures *S. cerevisiae* recueillies et inférer l'origine de ces souches. Le génotype de ces isolats à des loci microsatellites a indiqué comme origine le vin ou le raisin pour la majorité puis le pain, suggérant en effet un rôle vraisemblable de vecteur (figure 4). L'association entre levures et guêpes a été démontrée finalement en alimentant des femelles prêtes à hiberner avec des levures marquées qui ont été retrouvées dans le tube digestif des guêpes de la génération suivante. L'ensemble de ce projet a contribué à montrer l'importance des insectes dans la vection des levures.

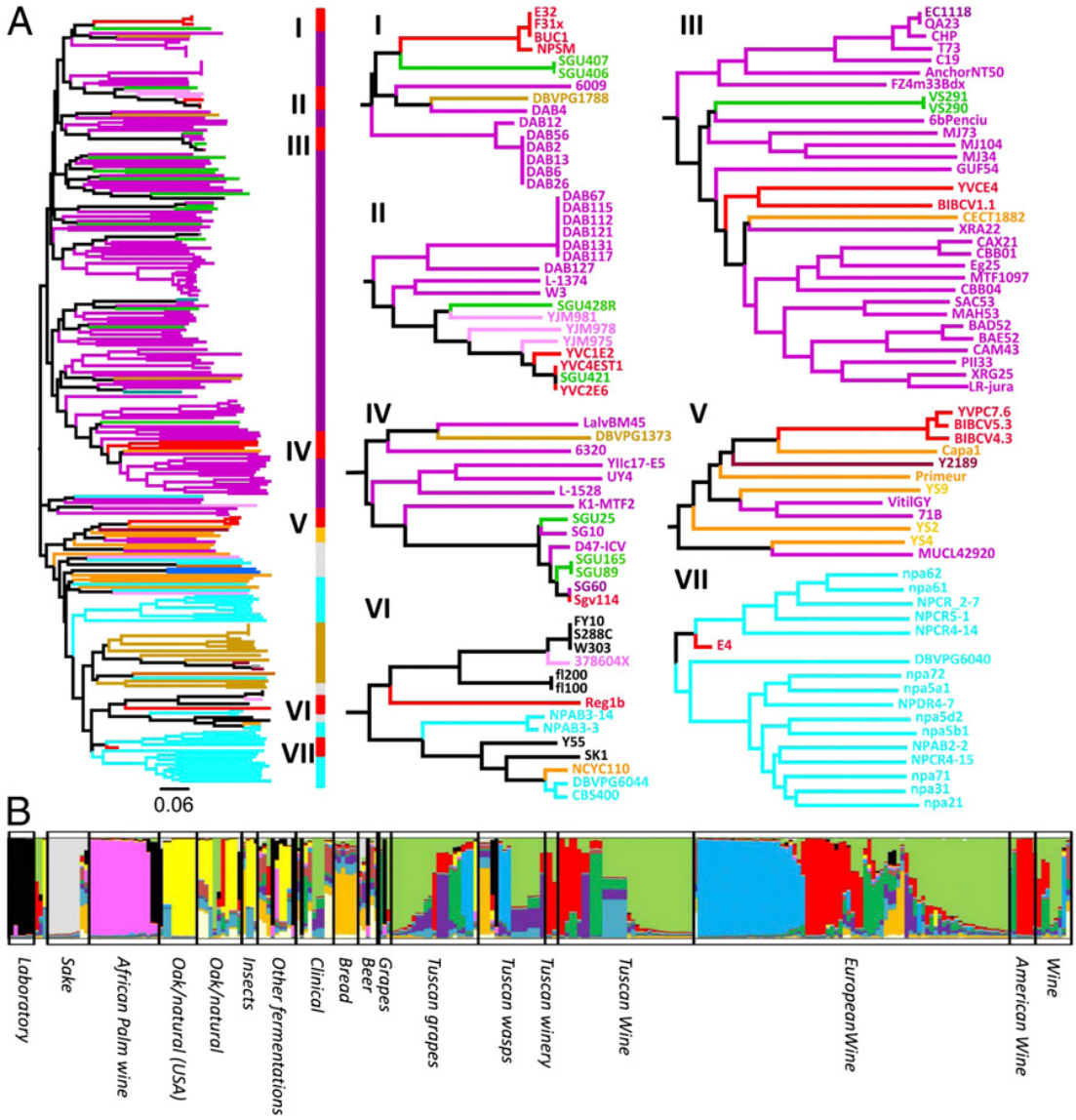


Figure 4 : Analyse des populations de levures isolées de guêpes à l'aide de microsatellites (A24). Arbre montrant le regroupement de 17 souches de *S. cerevisiae* isolées de guêpes parmi 256 souches de différentes origines. L'arbre a été construit selon Legras et al. 2007. Les branches sont colorées en fonction de l'origine d'isolement : code couleur: rouge, insecte; violet, vin; vert, raisins; orange, pain et bière; rose, patient; bleu ciel, autres fermentations ; brun clair, autres sources naturelles. Les positions des isolats de guêpes dans l'arbre sont indiquées par une barre rouge. I, II, III, IV, V, VI, and VII: détails de la structure des clusters contenant les isolats de guêpes. (B) Inférence des clusters ancestraux réalisés avec Instruct pour les 256 souches de *S. cerevisiae* avec les proportions de lignées ancestrales pour chaque cluster.

3.3. Facteurs influençant la diversité des levures *S. cerevisiae* au vignoble

Le développement d'une méthode de typage des levures par microsatellites offrait la possibilité de revisiter la dynamique des levures, d'évaluer les différenciations régionales suggérées par notre étude de 2007, d'évaluer l'impact de pratiques phytosanitaires, d'analyser les relations entre vignobles, ou du vignoble au chai. Je n'ai pu aborder ces aspects en 2006- 2007, alors que j'avais des outils mieux adaptés à ces questions. En effet, la demande de mon département scientifique pour un recentrage vers des thèmes plus technologiques, m'a amené à décliner une première proposition de collaboration faite par D. Schuller (Braga, Portugal) en 2007 sur ce thème. Avec mon arrivée dans le département MICA, ces thèmes devenaient davantage compatibles avec les objectifs du CT3 du département (actuellement CT2). En 2011, j'ai alors accepté de collaborer avec V. Corrich (Université de Padova) en encadrant le travail d'une étudiante en thèse (A. Viel) qui souhaitait comparer les flores levuriennes de régions viticoles Italiennes. En octobre 2012, I. Masneuf (Bordeaux) m'a proposé de construire ensemble et de co-encadrer le travail de thèse de M. Borlin (2012-2015) qui avait pour objectifs d'évaluer les effets de différents facteurs sur la diversité des levures des vignobles Bordelais.

Dans le cadre de la thèse d'A. Viel (Padova, Italie) une analyse d'une collection d'isolats obtenus dans trois régions viticoles du nord est de l'Italie et différenciées par leur profil de restriction de l'ADNm ont été ensuite caractérisés à des loci microsatellites. La première particularité de ce travail a été de révéler la parenté d'un grand nombre d'isolats aux LSA utilisées dans ces régions : entre 9,5% et 51,5% suivant la région. Le génotypage de 219 souches déjà différentiées (183 génotypes différents en ADNm) a conforté ces résultats mais apparaître de petites variations entre souches industrielles et les isolats. La cartographie de 23 populations de levures (sur 31) présentant au moins 3 levures différentes non apparentées aux LSA a permis de mettre en évidence une population de levure particulière dans la région la plus à l'est (LPAO) que l'on peut observer sur cette analyse canonique discriminante des populations ancestrales (Figure 5), ainsi que par les Fst interpopulations.

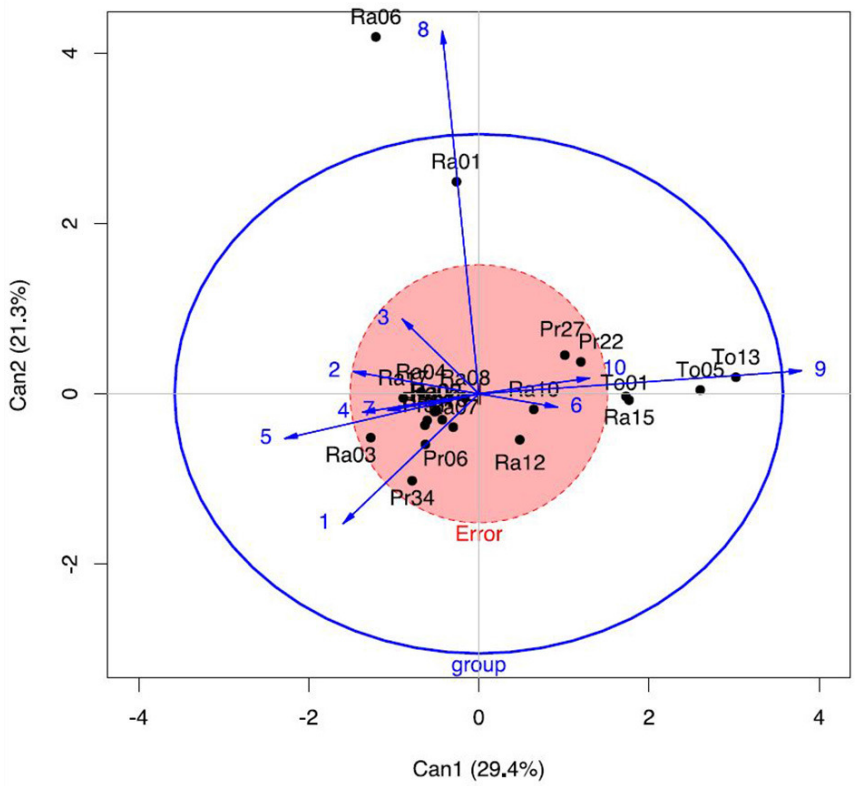


Figure 5 : Différentiation de populations isolées dans 3 régions viticoles Italiennes par analyse discriminante de leur population ancestrale inférées à l'aide d'InStruct. L'ellipse extérieure indique la variation des groupes autour de la moyenne, alors que le cercle rouge central indique la dispersion associée à la variation intragroupe globale. Les populations indiquées TO Ra et Pr correspondent respectivement aux régions LPAO, CVPAO, et PAO (A7)

Cette étude m'a offert également la possibilité de constater que le typage par microsatellites permet de différencier plus de génotypes que par RFLP de l'ADNm (206 génotypes différents sur 209 souches, contre 183 avec la RFLP de l'ADNm) tout en faisant apparaître plus clairement la structure de population (i.e. ici avec DAPC, figure 6). Cela pose donc la question de l'adéquation des différentes méthodes couramment utilisées pour différencier les levures pour des études biogéographiques, en particulier de la RFLP de l'ADNm mais aussi du typage interdelta par rapport aux deux précédentes méthodes.

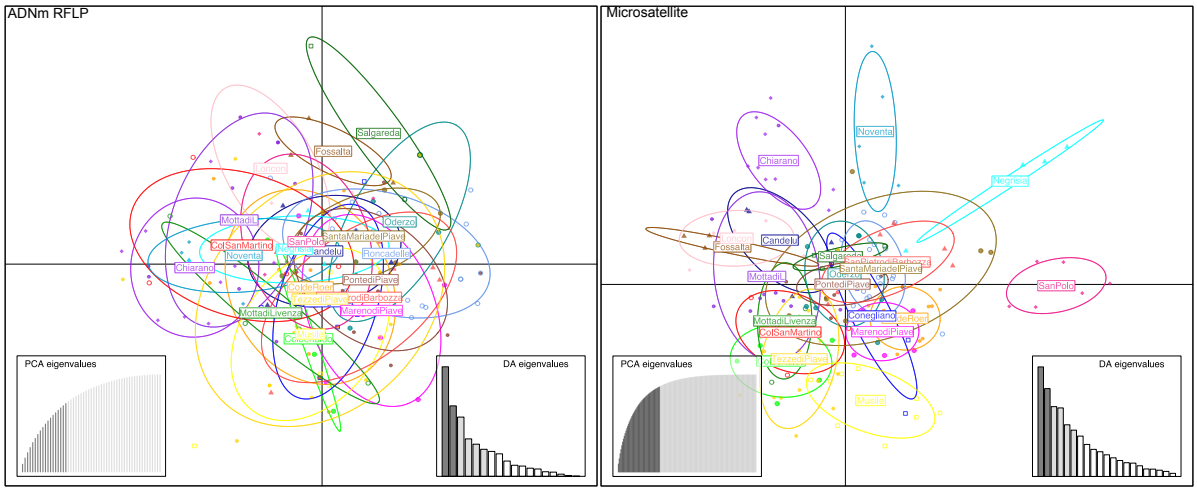


Figure 6: Comparaison des structures de populations obtenues avec DAPC à partir des profils de restriction de l'ADNm (gauche) et des génotypes en microsatellites (droite) de 171 isolats de raisins données de A7.

Toutefois la thèse d'A Viel posait un problème dans l'échantillonnage des populations de levures puisqu'il avait été « filtré » par la première sélection de profils de restriction d'ADNm différents, ce qui pouvait amener à limiter la diversité trouvée. De plus un grand nombre de vignobles avaient été analysés mais le plus souvent un faible nombre de génotypes différents avaient été isolés par vignoble ce qui laisse penser que l'exploration n'avait pas été assez profonde. La thèse de M. Borlin co-encadrée avec I Masneuf (Bordeaux) a été l'occasion de compléter ce travail en essayant de corrigeant ces aspects : en appliquant directement un typage par microsatellites, et en menant plusieurs échantillonnages par région. Plusieurs questions ont été abordées : les relations entre les populations de levures de différents vignobles, les relations historiques entre les populations de ces vignobles, l'impact de pratiques culturales phytosanitaires sur la diversité levurienne, et les relations entre les populations de raisin et du chai.

Nous avons ainsi mis en évidence des différenciations variables entre appellations du Bordelais, particulièrement élevées pour la population de Médoc. Des inférences bayésienne ABC des scénarii démographiques les plus probables propose que ces populations se sont différentiées à partir de celle de Saint Emilion (Figure 7). Un travail similaire avait été entrepris pour l'analyse des populations de différentes régions de Nouvelle Zélande, et des flux interrégionaux de levures très différents entre régions (Knight and Goddard 2015) avaient été estimés. La même méthode d'inférence

a montré des flux plus similaires entre les vignobles du Bordelais. Ceci s'explique aussi sans doute par la plus grande proximité géographique entre les régions viticoles.
Par ailleurs, en ce qui concerne les populations des raisins et des chais, nous avons en premier retrouvé 25% d'isolats apparentés aux LSA utilisées dans la région, confirmant donc les résultats d'A. Viel. La comparaison globale des flores de levures de 4 chais et de parcelles de raisin proches des chais ne fait apparaître en moyenne aucune différentiation, et l'inférence des flux de gènes entre les compartiments chai et raisin indique que des échanges élevés, plus importants dans le sens raisin chai que l'inverse. Aussi, contrairement aux résultats de Valero et al (2005), les deux compartiments chai et raisins des parcelles apparaissent comme fortement liés.
L'itinéraire phytosanitaire est un second facteur susceptible d'impacter les diversités fongiques de chaque parcelle. Dans les conditions de notre essai, qui présentait un dispositif plurirégional appariant domaines en agriculture traditionnelle et en agriculture biologique, nous n'avons observé globalement aucune différentiation entre flore *S. cerevisisiae* des parcelles de vigne en fonction de l'itinéraire technique (A2).
En complément de ce travail, le suivi de population dans des chais de l'appellation Sauternes fait à la fois apparaître la similarité des populations entre domaines, et le maintient de certaines souches sur 23 ans. Enfin, durant la vinification nous avons observé un glissement des populations au fil de l'avancement de la récolte et des différents tris de pourriture noble (A8).

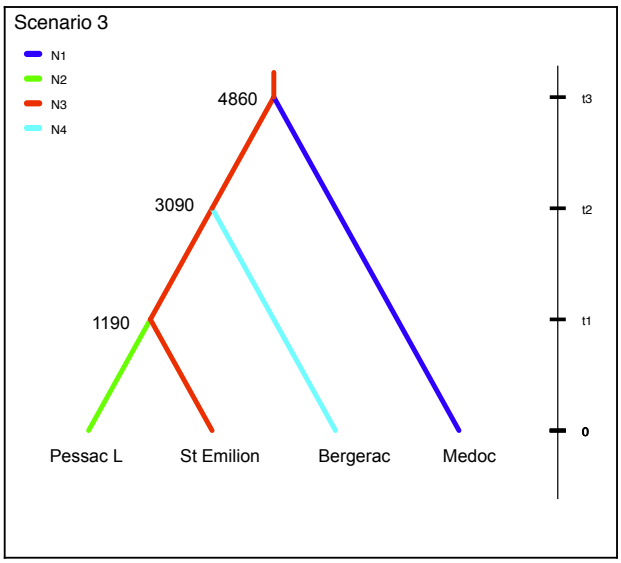


Figure 7 : Inférence de l'histoire des populations du vignoble Bordelais à partir du profil microsatellites des populations par « Approximate Bayesian Computation ». Les divergences entre régions ont été inférées à partir du profil de microsatellites à 17 loci. Les périodes de temps sont données en générations à partir du présent (méioses) (A2)

3.4. Une population de levure Méditerranéenne comme origine des levures de vin

Le profil de diversité des levures de vin obtenu en 2007 suggère leur développement à partir d'une ressource locale associée à un environnement non modifié par l'homme. C'est pour cela que je me suis rapproché de JP Sampaio (Universidade Nova, Lisbonne) qui avait isolé des souches *S. cerevisiae* à partir d'écorces de chêne en Europe pour aborder cette question avec lui. Dans un premier temps le génotypage par microsatellite a permis d'identifier un groupe particulier de levures de chêne isolées autour de la Méditerranée, puis le séquençage des génomes par JP Sampaio et son groupe, a permis de caractériser cette population de levures de chêne proche des levures de vin. Cette population est aujourd'hui la ressource naturelle de levure la plus proche de la population des levures de vin actuelle (A15). A partir des données génomiques, un modèle démographique a été construit par la méthode d'inférence démographique basée sur une approximation de la diffusion (δaδi) (Gutenkunst et al. 2009), qui a indiqué une divergence récente des deux populations, (postérieure à la dernière glaciation). Ces deux populations sont chacune en expansion et ont échangé entre elles des individus de manière asymétrique, plus nombreux dans le sens chêne/vin.

3.5. Hybridation dans le genre *Saccharomyces*

La formation d'hybrides a été par contre décrite de nombreuses fois chez les plantes et les animaux et peut être perçue comme un moyen d'atteindre de nouveaux optimum de fitness dans un paysage adaptatif (Mallet 2005; Mallet 2007). Des cas d'hybridations interspécifiques dans le genre *Saccharomyces* ont été progressivement décrits au cours des 30 dernières années, en premier pour les souches de bière de type lager (Pedersen 1985) puis pour des souches de vin et le cidre (Masneuf et al. 1998) jusqu'à ce qu'il apparaisse que leur présence était élevée dans les boissons fermentées à basse température (González et al. 2008; Belloch et al. 2009; Peris et al. 2011).
En 2006, nous avons obtenu un mutant faiblement producteur d'acide acétique à partir de la souche Eg8 isolée en 1979 à l'INRA de Colmar, initialement sélectionnée comme starter pour les vinifications de vins aromatiques contenant des thiols variétaux (Gewurztraminer, sauvignon blanc, colombard…) mais connue également pour sa cryophilie et sa capacité à achever les fermentations. En séquençant les gènes *ALD6* et *ALD5*, nous avons observé que cette souche était un hybride *S. cerevisiae x S.*

kudriavzevii. Nous nous sommes alors interrogés sur la spécificité de cette souche par rapport aux autres hybrides similaires déjà décrits, et sur l'origine des parties *S. cerevisiae* et *S. kudriavzevii* du génome. Connaître l'origine de la souche aurait pu nous offrir la possibilité de reconstruire un hybride similaire. Le génotype de 12 loci microsatellites de la partie *S. cerevisiae* du genome, nous a permis de rapprocher cette souche des levures de voile et plus spécialement d'un groupe de souches isolées en Alsace, Allemagne, Autriche et en Hongrie, et qui étaient toutes issues de la même hybridation (A25). Par contre ni la population de *S. kudriavzevii* que nous avons collectée à partir de chênes en Ardèche, ni les souches isolées au Portugal ni les autres hybrides n'étaient proches de la fraction *S. kudriavzevii* du génome de cet hybride. En comparant la distance moyenne δm^2 (basée sur les différences du nombre de répétitions des motifs microsatellites) pour les levures de ce groupe d'hybride en comparaison du groupe des levures de vin, nous avons estimé de manière très grossière la divergence de ce groupe à la moitié de la dispersion des levures de vin, ce qui indique que cette migration est ancienne et indique un réel succès écologique. Ce succès pourrait provenir de l'association de la résistance à l'alcool provenant de son génome de levure de voile, et de la cryophilie offerte par *S. kudriavzevii*.

Un autre exemple d'hybridation complexe d'une souche adaptée au mout de bière est celui des souches de l'espèce *S. bayanus*. Lors du séquençage des gènes MAL32 et 31, chez la souche CBS380, HV Nguyen (Micalis, INRA Jouy) avait identifié des allèles de *S. cerevisiae* et donc supposé une origine hybride de cette souche entre *S. uvarum* et *S. cerevisiae*. Nous avons réalisé une hybridation de l'ADN génomique de NBRC1948 et CBS380 sur une puce à ADN Affymetrix yeast genome 2.0 pour *S. cerevisiae* pour connaître l'ampleur du contenu *S. cerevisiae*. La répartition des fragments entre les deux souches, nous a fait proposer la possibilité d'un croisement entre NBRC 1948 avec *S. uvarum* pour donner la souche CBS380. La recherche de la trace d'un tel croisement sur les gènes de la partie *S. uvarum* a permis d'observer deux types d'allèles, suggérant la présence d'un ancêtre d'une autre espèce inconnue que nous avons appelée *S. lagerae* (A26). Ceci nous a amené à proposer le schéma suivant des relations possibles entre les souches du genre *S. bayanus* et *S. pastorianus* (Figure 8).

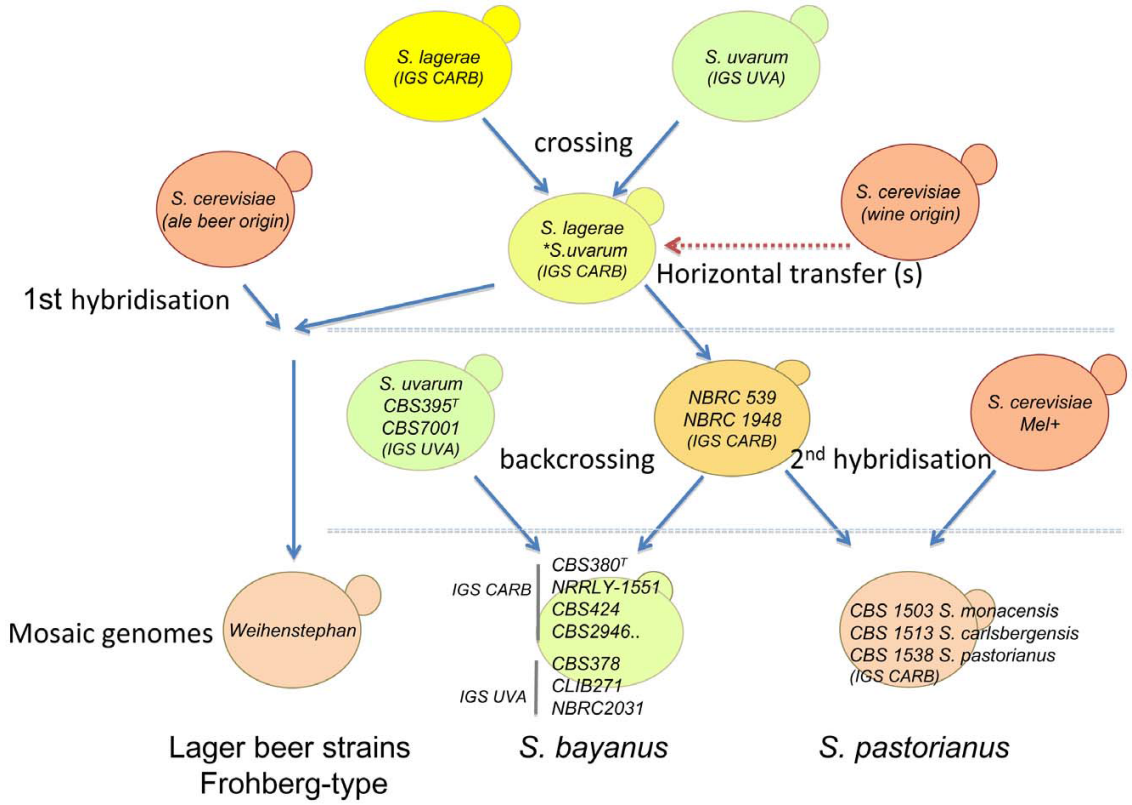


Figure 8 : Schéma hypothétique présentant les croisements entre les espèces *S. uvarum*, *S. lagerae* et *S. cerevisiae* (A26).

Au delà de la clarification des relations entre espèces du complexe *S. bayanus*, qui avaient amené à de nombreuses controverses, les souches CBS380 et NBRC1948 présentent une amplification notables des locus MAL et MTY1 impliqués dans l'assimilation et le transport du maltotriose en teneur importante dans les mouts de bière. Cette amplification des locus MAL vraisemblablement en lien avec la composition du milieu semble représenter également un exemple d'adaptation associé à un environnement anthropique.

3.6. Les levures de voiles un groupe spécifique proche des levures de vin

L'élevage sous voile des vins, aussi appelé vieillissement biologique est pratiqué en Espagne, dans (la région de Jerez), en Italie (Sardaigne), en Hongrie (Région de Tokaij) et en France (Jura). Durant cet élevage du vin, les levures ont un métabolisme respiratoire et utilisent le glycérol et l'éthanol comme source de carbone à la différence des levures durant la fermentation (Figure 9). Seules certaines levures sont capables de produire un voile, et celles-ci ont été classées en 4 races différentes (Martínez et al. 1995). Cependant, les analyses moléculaires ont montré que ces souches

appartiennent toutes à l'espèce *S. cerevisiae* et présentent un profil d'ITS particulier (Esteve-Zarzoso et al. 2001).

Figure 9 : Biofilm de levure sur du vin obtenu à l'aide de levures de voile.

Une collection de souches rassemblées à partir d'élevages de vin jaune du Jura, par le Laboratoire Départemental d'Analyse de Poligny dans les années 1990-2000 était étudiée pour ses caractéristiques technologiques par C. Charpentier de l'IUVV à Dijon. Dans la mesure où ces souches donnaient naissance à des voiles fin ou épais, nous nous demandions si il était possible de différencier ainsi les souches donnant l'un ou l'autre type de voile. Aussi, j'ai caractérisé ces souches à l'aides de techniques de typage classiques : ECP, et typage interdelta (A32).
De manière intéressante, le regroupement obtenu avec le typage interdelta nous a permis en premier d'observer une corrélation statistiquement significative avec la capacité à produire des voiles de types différents, alors qu'aucune corrélation n'était détectée par ECP. De plus, la répartition des levures par cluster n'était pas aléatoire du Nord au Sud du Jura. Cependant, la caractérisation obtenue dans un second temps à l'aide des marqueurs microsatellites a fourni une plus forte corrélation entre les génotypes et les types de voiles (A20) suggérant un meilleur potentiel a révéler des aspect biogéographiques. De plus, cette confrontation des génotypes en microsatellites des levures de voiles à ceux des levures des autres origines montre assez clairement la proximité des levures de voile avec les levures de vin. Ainsi, le groupe contenant la souche EC1118 et les levures apparentées sont issues d'un croisement entre une levure de vin et une de voile. Cette particularité de la souche EC1118, dont le génome a

été le premier séquencé d'une levure commerciale majeure (Novo et al. 2009) (projet auquel j'ai participé) indique que cette souche présente un caractère hybride entre deux familles de levures. Elle n'est donc pas le meilleur modèle pour prédire le comportement de l'une ou l'autre classe de levure.
Enfin, du fait du contraste entre le style de vie de ces levures de voile par rapport à celui des levures de vin « fermentaires », et de leur proximité génétique, ces deux groupes représentent un modèle intéressant pour étudier l'adaptation des levures à leur niche écologique.

En conclusion de cette partie, le développement d'une méthode de caractérisation moléculaire, nous a ouvert la possibilité d'explorer la diversité des levures dans les différentes niches fermentaires de vin, de pain, de bière, de vin de palme, de dolo (A16, A18, A25, A26, A33, A35, A36 etc…), de montrer la structure de population chez *S. cerevisiae* et d'aider à en comprendre l'écologie. Elle montre l'importance du choix des méthodes, et on ne peut que regretter que les marqueurs microsatellites n'aient été davantage utilisés dans les multiples études écologie pour obtenir de meilleures inférences historiques, sur le fonctionnement des écosystèmes, ou sur le style de vie des levures. Avec J. Enjalbert (INRA, Moulon) j'ai en particulier développé un modèle permettant d'estimer le taux d'homothalisme, d'outcrossing et d'approcher le taux de méiose à partir du profil d'hetérozygotie (article en préparation).
Il est probable que la chute du cout des méthodes de séquençage amène à séquencer des génomes entiers pour quelques euros, rendant alors cette méthode obsolète. Enfin cette partie m'a ouvert à la question de la domestication des levures, et m'amené à collaborer avec D. Sicard, pour la rédaction d'une revue sur cette question (A28).

4. *Adaptation de S. cerevisiae aux niches anthropogéniques*

La différentiation de clusters de souches en fonctions des niches écologiques telles que nous les avons observées en 2007 et la meilleure capacité des levures isolées de fruits et de raisin à achever la fermentation alcoolique œnologique en comparaison des souches d'autres origines et en particulier de chêne (Camarasa et al. 2011) amène immédiatement à s'interroger sur les mécanismes mis en place chez la levure pour affronter les stress spécifiques de ces niches. Pourtant, trois exemples d'adaptation chez la levure de vin étaient décrits: celui de la translocation entre *ECM32* porté par le

chromosome VIII et *SSU1* porté par le chromosome XVI, permettant une induction d'un transporteur expulsant les sulfites hors de la cellule (Pérez-Ortín et al. 2002), celui de l'inactivation des aquaporines *AQY1* et *AQY2* (Will et al. 2010) procurant aux cellules une meilleure vigueur en fermentation et enfin celui de l'amplification des gènes *CUP1* chez les levures de vin en relation avec les traitements phytosanitaires de la vigne au cuivre (Fay et al. 2004; Warringer et al. 2011). Ceci contrastait avec les multiples exemples de loci sous sélection déjà reportés chez l'homme (Akey et al. 2002; Voight et al. 2006; Sabeti et al. 2007) les plantes (Tenaillon et al. 2004) et les animaux (Gu et al. 2009; Akey et al. 2010; Rubin et al. 2010). Il y avait aussi un intérêt logique pour mener des approches de génomique des populations afin de faire apparaitre les contraintes auxquels ont été soumis les génomes des levures dans des niches anthropogéniques. Sur ces aspects, depuis mon arrivée à Montpellier dans l'équipe de S. Dequin (UMR SPO) je me suis focalisé sur l'adaptation obtenue à partir de la variation génétique existante, alors que ma collègue V. Galeote s'est penchée avec S. Dequin sur les « nouvelles régions » issues de transferts horizontaux comme celles détectées dans le génome de EC1118 (Novo et al. 2009). Le couple de populations de levures de vin et levures de voile me paraissait un bon modèle pour aborder la question de la domestication des levures car ces deux populations sont très contrastées dans leur style de vie et très proches génétiquement. Nous avons ensuite élargi cette comparaison à des groupes de plusieurs origines technologiques, en incluant les levures de voile et en les comparant aux levures isolées de ressources naturelles : de chêne.

4.1 Adaptation des levures de voiles.

Les spécificités phénotypiques des levures de voile ont attiré l'attention depuis longtemps. Un des premiers essais destinés à en comprendre les bases a été mené par Infante en Espagne (Infante et al. 2003). Cette étude a mis en évidence de nombreuses aneuploïdies chez 2 souches de voile et suggérait que les changements du nombre de copies étaient un des facteurs à l'origine de leurs spécificités. Aussi, nous avions entrepris, à Colmar, de comparer les profils d'hybridation sur puce de différentes levures de voile (de France, Italie, Hongrie et Espagne) en comparaison des levures de vin, pour faire apparaître des régions dont l'amplification pourrait expliquer l'adaptation. A notre surprise, nous n'avons trouvé que 3 gènes amplifiés (*MCH2,*

YKL222W et *FRE2*) pour une partie des souches de voile suggérant que contrairement aux résultats précédents l'amplification génique n'expliquait pas les spécificités des levures de voile (A20) (Figure 10).

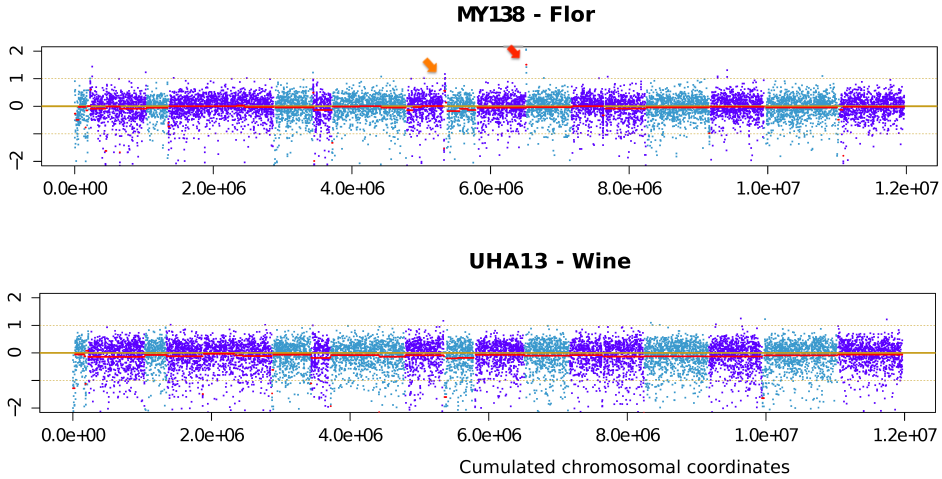


Figure 10 : Karyoscope présentant les variations de signal d'hybridation au long du chromosome pour la levure de voile My181 et la levure de vin UHA13(A20). La flèche rouge indique la région MCH2, YKL222W, FRE2 et la flèche orange la région PHO12 et IMD2 qui sont amplifiées chez la levure de voile My138.

Aussi, dans un second temps nous recherché les spécificités des levures de voile grâce à un projet de re-séquençage. En préalable, nous avons construit des dérivés haploïdes de souches de vin et de voile afin de disposer de souches haploides pour valider la fonction de gènes particuliers ou pour mener ultérieurement des approches de génétique quantitative. Après avoir délété le gène HO, nous avons construit des paires de souches de signes sexuels opposés sans marqueurs de sélection. Ses souches (2 souches de vin et 3 souches de voile) sont disponibles auprès du CIRM levure (INRA – Jouy). Nous avons également modifié un milieu minimum pour disposer d'un milieu de voile de composition plus proche du vin que le milieu YNB – éthanol souvent utilisé (A9).

Le séquençage de 9 souches de voile et 9 souches isolées de vin a en premier lieu confirmé la proximité phylogénétique des levures de voile et de vin bien que ces deux populations soient clairement séparées (Fst=0.43). (Figure 11). La levure EC1118 dont le génome a été séquencé apparaît bien comme un hybride entre une levure de vin et une levure de voile. La divergence entre levures de voile étant 1/3 de celle observée entre les levures de vin, ces souches sont apparues après les levures de vin, peut être avec le début de la maitrise de la vinification et de la conservation des vins. Grace à la collaboration avec P. Gladieux (BGPI, INRA-CIRAD, Montpellier), nous avons pu inférer

un modèle démographique selon la méthode δaδi pour les populations de levures de voile méditerranéenne ou du Jura. Ces modèles montrent une population de levures de vin en expansion, alors que les levures de voile apparaissent en contraction. Toutefois ces scénarii sont assez imprécis du fait du petit nombre de souches pris en compte dans l'analyse.

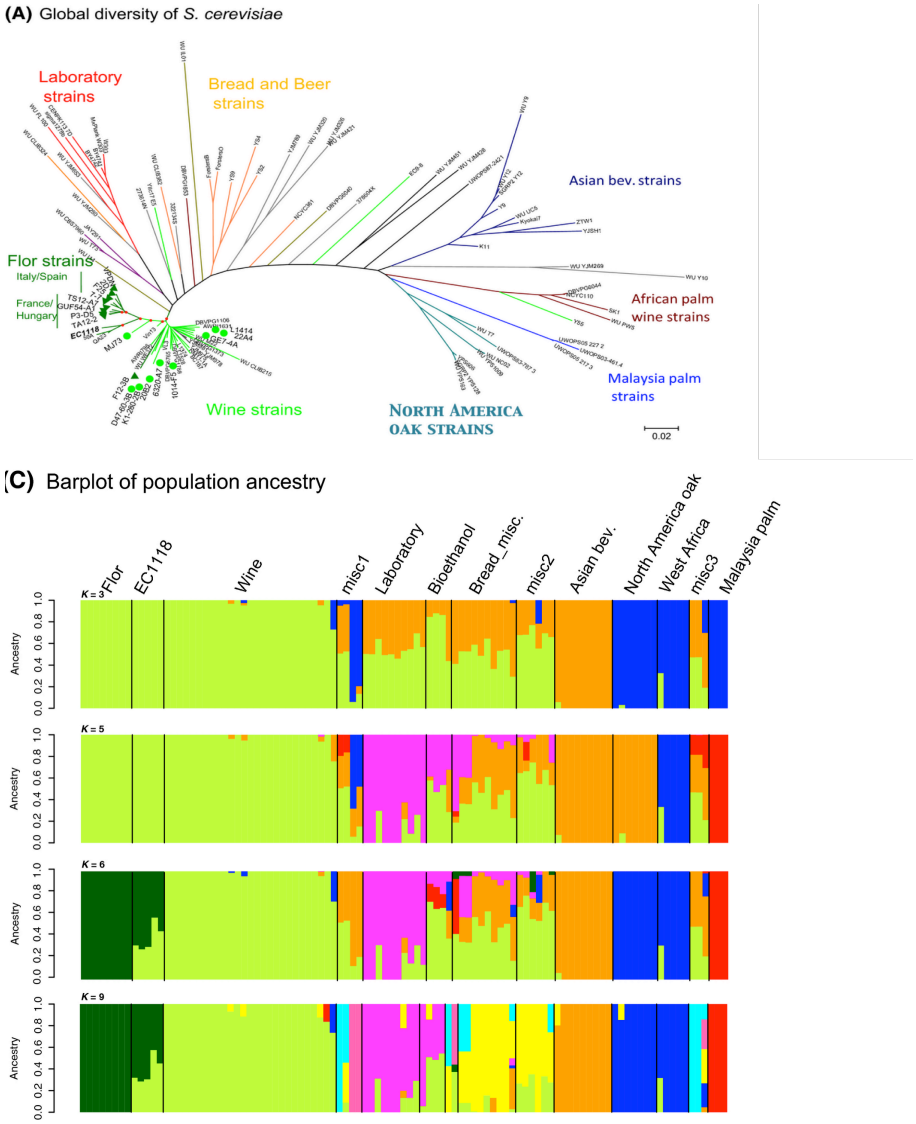


Figure 11 : Relations entre levures de vin et voile établies à partir des données génomiques (A5) Généalogie obtenue par la méthode du maximum de vraisemblance, entre les 9 levures de voile (triangles) et 9 levures de vin (cercles) et en comparaison d'autres levures. La généalogie a été inférée à partir de 427 587 positions variantes présentant moins de 20% de données manquantes. (C) Inférence des groupes ancestraux avec admixture (Alexander et al. 2009) en supposant de K = 3 à 9 groupes ancestraux.

L'analyse comparée des génomes fait apparaître des régions très différentiées entre levures de vin et levures de voile, comme par exemple la région du transporteur de zinc *ZRT1* ainsi que d'autres transporteurs de métaux divalents *(SMF1, ZRT3, ALR1*), ou la région promotrice de *FLO11* ainsi que de bien nombreux autres gènes. En accord avec cela, le sous-ensemble des gènes présentant des positions variantes prédites pour impacter la fonction des protéines (et différentiant levures de vin et de voile) semble être enrichi pour de nombreuses fonctions importantes pour la croissance du voile. Nous avons pu détecter une sélection positive sur le transporteur de zinc *ZRT1*, à l'aide du « branch-site test» qui teste la présence d'un taux de mutations non synonymes plus élevé dans une branche par rapport au reste d'une phylogénie. Une seconde méthode de détection d'un balayage sélectif , XP-CLR (qui estime la vraisemblance d'un balayage sélectif en modélisant le spectre des fréquences alléliques associé à la différentiation de deux populations), appliquée entre ces 2 populations indique une sélection positive pour plusieurs autres régions. Enfin, nous avons pu vérifier la spécialisation phénotypique de certains allèles très divergents entre levures de vin et de voile pour les gènes *RGA2* et *SFL1*.

Cette partie mérite d'être complétée par une approche de génétique quantitative, utilisant un croisement entre une levure de vin et de voile pour identifier des QTL expliquant les différences phénotypiques entre levures de vin et de voile. De plus les composants de la matrice exocellulaire des voiles devraient être caractérisés. Une analyse comparative des transcriptomes de levures de vin et de voile sur milieu synthétique de voile mais avec agitation (pour éviter sa formation), a été également réalisée. Mais du fait de difficultés expérimentales (que nous avons finalement réglées) nous n'avons obtenu qu'une seule comparaison entre levures de vin et de voile. Ce point devrait aussi être prochainement complété, afin de pouvoir éclairer sur les fonctionnements respectifs des levures de voiles et de vin sur les milieux de voile, indépendamment de l'état lié à la vie en biofilm.

4.2 Adaptation de levures de différentes niches anthropogéniques (A1)

En complément de la comparaison de ces 2 niches, plusieurs autres niches présentent des caractéristiques très différentes. Le jus de raisin est caractérisé par sa richesse en sucres, son bas pH (3-4) et des teneurs en azote et en lipides limitées, le jus de canne est assez similaire avec une teneur en sucre proche (10-18%) et un pH moins acide (5 -

5.5) alors que les laits fermentés contiennent des sucres issus de l'hydrolyse du lactose. Enfin, les levures de chêne utilisent les sucres produits par les plantes (Sampaio and Gonçalves 2008) qui ne semblent pas en quantités importantes.

Partant des résultats précédents montrant les différences de propriétés entre les souches de vin et les autres origines (Spor et al. 2008; Camarasa et al. 2011), nous avons construit ce projet de manière à obtenir des groupes avec des effectifs plus équilibrés (vin, voile, fromage, pain, rhum, chêne), avec les objectifs de faire apparaître les différences phénotypiques entre ces groupes, d'essayer de les relier aux différences génotypiques, et par ailleurs de détecter dans les génomes des souches les indices génomiques de l'adaptation à ces différentes niches écologiques. Nous avons pour cela séquencé 40 autres souches et pu bénéficier de 26 autres génomes qui ont donné plus de poids à la population de levures de vin principalement. L'ensemble de ces données nous permet de couvrir les populations de vin (28 individus), de voile (8 individus), de rhum (8 individus), de fromage (7 individus), de chêne méditerranéen (4 individus), et de chêne américain (4 individus).

Du point de vue phénotypique, les levures de vins se sont avérées les souches ayant la meilleure capacité à fermenter, alors que les souches de fromages, ou les souches de chêne étaient incapables d'achever la fermentation. En revanche, les souches de fromages avaient la meilleure capacité à assimiler le galactose, indiquant bien une spécialisation phénotypique des souches de ces groupes.

Le séquençage nous a permis de rechercher et compléter la structure de population connue chez *S. cerevisiae* (Liti et al. 2009; Schacherer et al. 2009; Cromie et al. 2013) en caractérisant deux groupes supplémentaires (voile, chêne méditerranéen, et fromage). Pour tous ces groupes, nous avons pu observer une chute rapide du déséquilibre de liaison (DL) indiquant la présence d'évènements méiotiques. Cependant, les groupes issus des produits fermentés présentent des fréquences d'outcrossing (croisement entre individus non apparentés) au moins dix fois plus élevées. Nous avons aussi estimé le taux de méiose et d'outcrossing dans chaque groupe qui auraient plus évoluer en fonction des niches. Le taux de méiose ne varie pas de manière très claire avec la niche, mais par contre le taux d'outcrossing varie très fortement, expliquant ainsi la forte hétérozygotie des populations de produits fermentés.

Du point de vue génomique, chacun de ces groupes présente des spécificités: des variations du nombre de copies, parfois connue pour être associées aux conditions

environnementale: région de *CUP1* pour les levures de vin (mais aussi pour les levures de rhum), gènes *MAL* pour les levures de pain mais aussi *DLD3* pour les levures de fromages ce qui n'était pas connu. Par ailleurs la présence de différentes régions issues de transferts horizontaux, ou d'introgressions ont été identifiées, et plusieurs semblent remarquablement spécifiques à certains groupes.

Enfin, pour faire apparaître des signes d'adaptation dans ces génomes, nous avons utilisé des méthodes visant à faire apparaitre des différenciations intergroupes locales (hapFLK) (Fariello et al. 2013), ainsi que des méthodes intragroupes reposant sur l'augmentation du DL (iHS, SVD, omega) (Voight et al. 2006; Hussin et al. 2010; Alachiotis et al. 2012) ou encore sur des modifications du spectre de fréquences associées à la fixation de mutations adaptatives : H de Fay et Wu (Fay and Wu 2000). Cette batterie de tests a fourni un nombre assez élevé de cibles intéressantes. La comparaison des populations de levures de vin de rhum, et de chêne Méditerranéen a permis par exemple de faire apparaître le transporteur de stérols *AUS1* au rôle essentiel pour les levures de vin en vinification, ou encore le transporteur de thiamine *THI72* pour les levures de rhum, ainsi que plusieurs autres régions. Le rôle des allèles de ces gènes doit être validé du point de vue fonctionnel.

Pour l'analyse intragroupe nous avons mis l'accent sur le groupe des levures vin pour lequel nous avions le plus d'individus. Mais nous avons exploré également le groupe de levures de fromage, de rhum et de chêne Méditerranéen avec les mêmes méthodes.

A part iHS qui présente une distribution normale, les autres statistiques n'ont pas de distribution connue. Il était nécessaire de comparer la distribution observée à celle obtenue en simulant des populations ayant la même démographie sous une évolution neutre. Pour cela nous avons utilisé le modèle démographique proposé pour les levures de vin et de chêne Méditerranéen proposé par Almeida (A15).

Bien que nous n'ayons pas de loci commun à toutes les méthodes et peu de loci sont partagés, les régions pointées pour les levures de vin, présentent un enrichissement pour les transports d'acides aminés, ce qui est cohérent avec l'importance de l'azote souvent limitant en fermentation œnologique (Figure 12). La présence du cluster *ARR1* impliqué dans la résistance à l'arsenate appliqué sur les vignes comme fongicide, ou du gène *RIM15* déjà pointé dans les analyses de QTL de fermentation donne une forte plausibilité à ces candidats. Néanmoins l'effet des allèles de ces gènes devra être validé.

Enfin, chez les levures de fromage, nous avons trouvé un signal de sélection pour la région *GAL7-GAL1*- introgressée d'une espèce basale du complexe *Saccharomyces* et qui fournit vraisemblablement un avantage sélectif à ces souches sur milieu contenant du galactose et du glucose (validation en cours). Après sa première évaluation, nous complétons ces résultats pour soumettre une version corrigée du manuscrit (A1).

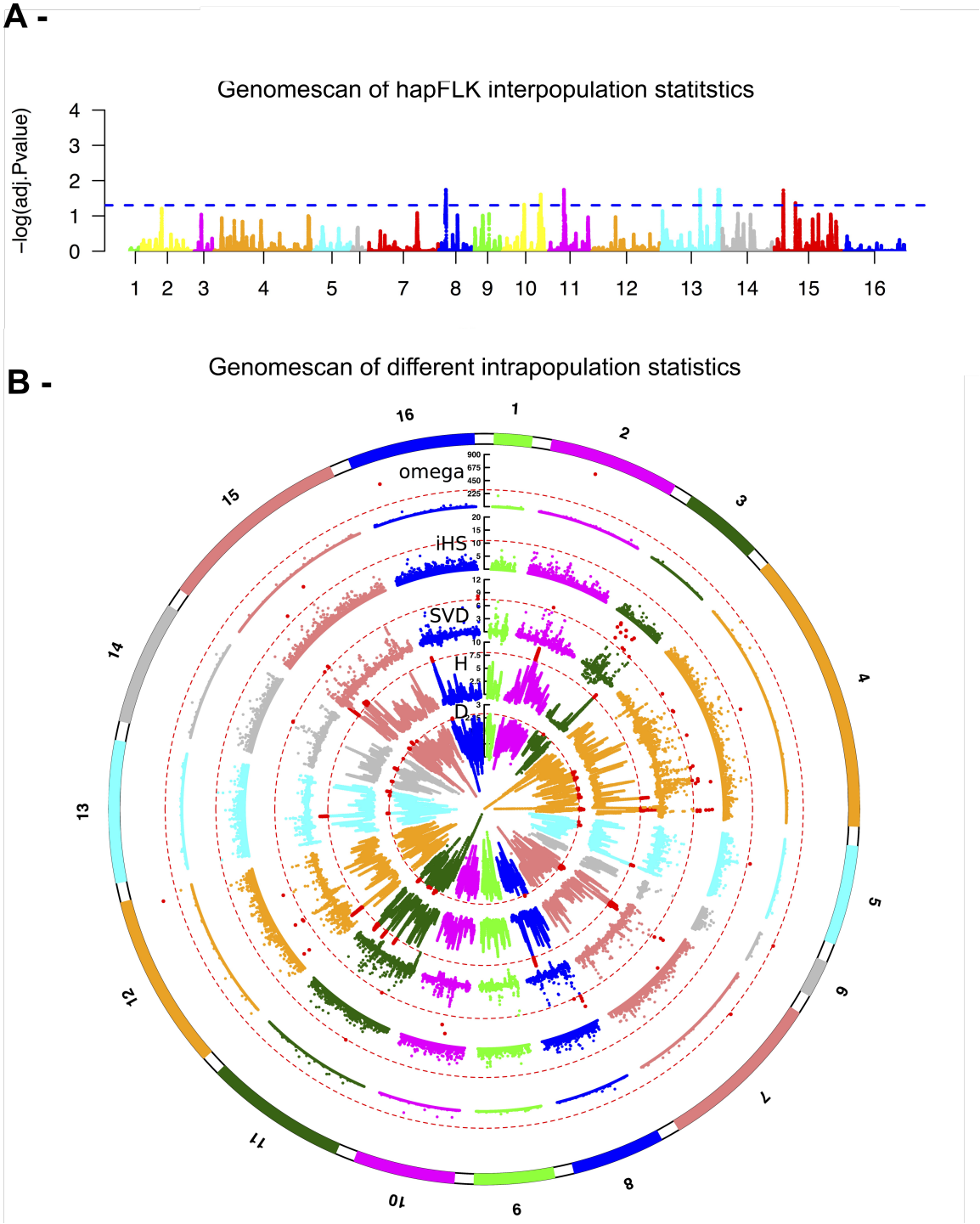


Figure 12 : régions du génome potentiellement sous sélection A- identifiées par la méthode différentielle hapFLK (log des probabilités corrigées pour la multiplicité des

tests) B- identifiées par des méthodes internes aux populations pour les levures de vin : Omega, D, et SVD obtenu à partir de 57987 SNPs phasés. Les points en rouge au dessus du cercle pointillé rouge sont des cibles potentiellement sous sélection positive.

Globalement ces résultats complètent la proposition de Goddard et Greig présentant *S. cerevisiae* comme un généraliste (Goddard and Greig 2015) et suggèrent que du fait de la remarquable plasticité de son génome, *S. cerevisiae* est un complexe de populations spécialisées avec des génomes « sur-mesure ».

5. *Bases moléculaires de la diversité physiologique des levures S. cerevisiae œnologiques*

La diversité physiologique des souches de levures devrait en toute logique faire pendant à la diversité génétique. Pourtant, l'impact de cette diversité sur les propriétés des souches était peu étudié jusqu'en 2009 et l'avènement des approches combinant génomique et phénotype haut débit (Liti et al. 2009).

5.1. Résistance aux acides gras à moyenne chaine inhibiteurs de fermentation

L'achèvement de la fermentation alcoolique est une des priorités pour la plupart des œnologues. Les carences azotée ou lipidique peuvent conduire à des arrêts de fermentation mais aussi la présence dans le mout d'acides gras à moyennes chaines qui sont inhibiteurs des fermentations. Alors que l'implication d'un transporteur *PDR12* était connue pour les acides gras en C6, et C8 rien n'était décrit pour l'acide décanoïque. Plusieurs mécanismes étaient envisageables : export de ces acides gras en dehors de la cellule, utilisation par la cellule pour la production d'acides gras plus longs, ou enfin détoxification au travers de la production d'éthyl-esters. Pour y répondre, j'ai construit cette étude avec ma collègue de l'IUT Génie Biologique de Colmar C. Erny, en combinant l'analyse d'un groupe de 78 souches, l'analyse de la réponse du transcriptome à l'exposition vis à vis de ces acides, et le screening d'une collection de souches avec les transporteurs candidats inactivés (A29).
Au niveau expérimental, le phénotype de 78 souches œnologique a montré que la résistance à l'acide octanoïque et l'acide décanoïque était variable et impliquait au moins deux mécanismes différents. Les modifications du transcriptome (en collaboration avec B Blondin, SupAgro, Montpellier) en réponse à l'exposition à ces molécules font également apparaître des mécanismes de réponse différents : induction

importante du transporteur *PDR12* pour l'acide octanoïque, et action conjointe de deux transporteurs *TPO1*, et *PDR12* (mais à un moindre niveau) pour l'acide décanoïque. Ceci a été vérifié ensuite à l'aide de souches délétées pour chacun de ces deux gènes. Enfin pour ce deuxième acide, nous avons noté une plus forte induction de l'éthyl ester synthase *EEB1* à l'origine de la synthèse de décanoate d'éthyl, un composé important de l'arôme des vins (Figure 13).

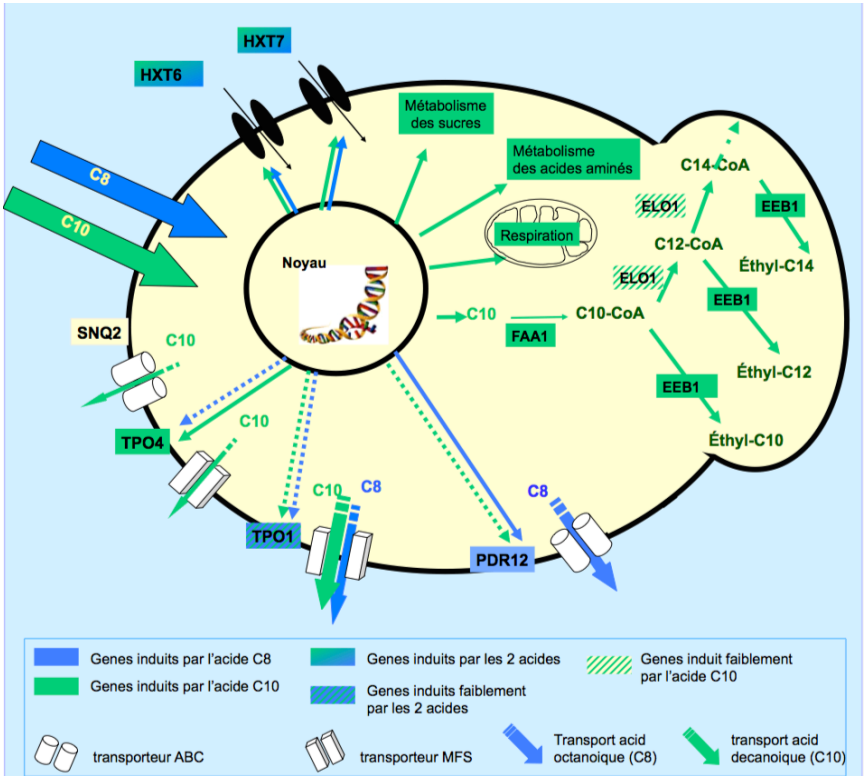


Figure 13 : Représentation schématique de la réponse de la levure *S. cerevisiae* U13 aux acides octanoïque et décanoïque

5.2. Bases génétiques de la production de composés volatils durant la fermentation

La synthèse de composés volatils est une des propriétés essentielles du rôle des levures œnologiques. Alors que les mécanismes de biosynthèse de nombreux composés sont connus comme celui des acétates (Verstrepen et al. 2003), des éthyl esters (Saerens et al. 2006), (pour revue voir Swiegers et al. 2005), nous n'avions que peu d'informations sur les mécanismes expliquant les variations de biosynthèse des différents composés. De plus, les mécanismes d'action des levures sur certains composés aromatiques du raisin sont mal connus : 60% des terpénols libres sont perdus durant la fermentation alcoolique (King and Dickinson 2000), et aucune explication n'était établie en 2008. Le développement de l'analyse génétique quantitative chez la levure, ouvrait la possibilité d'explorer les bases moléculaires des

différences de production ou de transformation des composés aromatiques, sans passer par le filtre subjectif des gènes candidats. J'ai donc proposé au Prof F. Karst (Université de Strasbourg, UMR SVQV Colmar) un projet de thèse explorant les possibilités de l'analyse de QTL sur la production de composés aromatiques par la levure durant la fermentation ainsi que sur la dégradation du géraniol. Nous avons pour cela utilisé une population de ségrégants issus d'un croisement entre la levure de laboratoire S288C et la levure de vin 59A, cartographiée à l'INRA de Montpellier par B. Blondin (SupAgro, Montpellier) à l'aide de puces Affymetrix S98. Ce travail aura été pour moi également l'occasion de co-diriger mon premier étudiant en doctorat, D. Steyer (2008-2011), sous la direction de F. Karst. Nous avons analysé l'évolution de 33 composés dans deux dispositifs expérimentaux et identifiés au total 8 régions expliquant de 39 à 72 % des variations observées (Figure 14). L'une de ces régions contenait le facteur de transcription *PDR8*, qui régulait différemment chez les deux souches parentales 59A et S288C le transporteur *QDR2* responsable de l'export du nérolidol hors de la cellule. Un second QTL majeur observé pour la production du phenyl-2-éthanol est lié à une des particularités de S288C : un allèle non fonctionnel de *ABZ1* chez S288C. Ce travail avait pour nous vocation à être une preuve de concept de l'intérêt d'une stratégie QTL pour des métabolites. Cependant, le faible nombre de QTL obtenus malgré la forte héritabilité a pointé la faible puissance permise par une population expérimentale de 30 souches (A23).

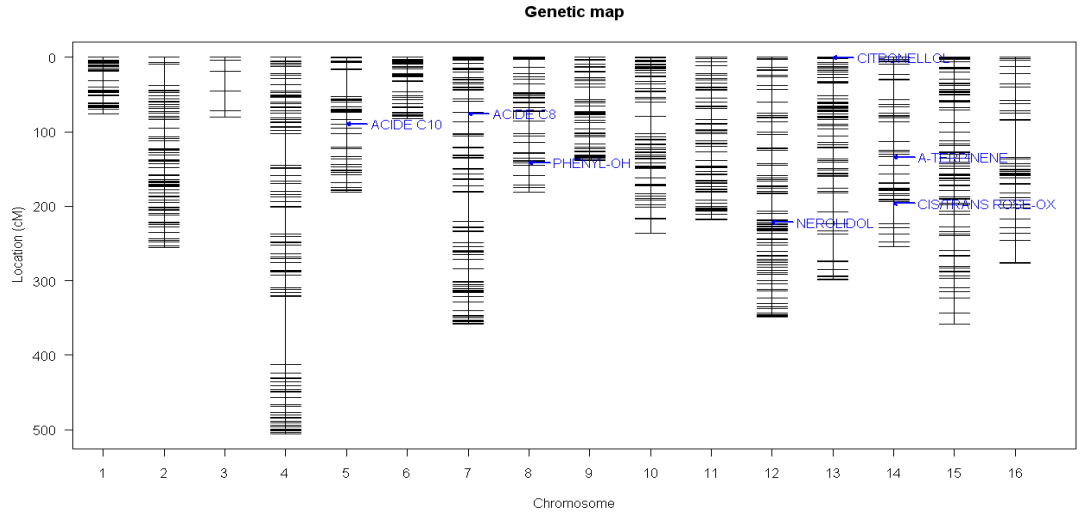


Figure 14: Localisation des mQTL impliqués dans la production de composés volatils par *S. cerevisiae* sur la carte génétique. PHENYL-OH= 2-phényléthanol, CIS/TRANS ROSE OX= rapport cis rose oxyde/trans rose oxyde) (thèse D Steyer, 2011)

Aussi il me paraissait nécessaire de poursuivre ce travail, à l'aide d'une population de taille plus large, et d'un croisement entre deux parents à la fois différents mais obtenus à partir du vin. Ceci a pu être mis en place durant la thèse de M. Eder, que j'ai co – encadrée avec S. Dequin. Nous avons utilisé une population de 130 ségégants issus d'un croisement entre une levure de vin et une levure « champenoise », déjà en partie produite pour la recherche de QTL associé aux besoins d'azote dans le cadre de la thèse de C. Brice dont j'avais participé à l'encadrement (2012-13, thèse soutenue en 2013), (A19). Grace à cette population de ségregants nous avons recherché des QTL expliquant l'origine des différences de teneurs en alcool supérieurs, esters, en géraniol et ses dérivés, en sulfure de dimethyl (DMS), ainsi qu'au flux du métabolisme central carboné.

En utilisant différentes stratégies de recherche nous avons identifié 65 QTL, dont 55 qui influencent la formation de 30 composés volatils, 14 avec un effet sur la consommation des sucres, et le métabolisme carboné central. En plus du nombre de QTL détectés, cette population de ségrégants nous a aussi permis d'identifier des QTL en interaction pour le propanol, l'ethyl lactate, et l'éthyl octanoate. Enfin, parmi 10 des régions détectées, nous avons validé les contributions de 14 gènes impliqués dans le transport et le métabolisme de l'azote, le métabolisme carboné central, la synthèse des acides gras et sa régulation, et expliqué une fraction des différences de concentration de différents composés volatils (A3). La multiplicité des QTL obtenus, montre clairement l'utilité de cet outil pour rechercher les bases génétiques des différences de production de composés volatils durant la fermentation. Il permet également d'envisager d'améliorer des souches par croisement par sélection assisté par marqueurs pour les allèles ayant le plus d'impact.

6. *Projets en cours et à venir*

La multitude de produits fermentés par la levure *S. cerevisiae* et consommés par l'homme au travers du monde démontre d'une étonnante attirance pour ces produits. Cependant, cet attrait pour les produits fermentés ne semble pas l'apanage de l'homme : des musaraignes (Wiens et al. 2008), les singes (Hockings et al. 2015), les aye-ayes (Gochman et al. 2016) sont attirés par les jus de fruits fermentés. De plus une mutation dans *ADH4* (A294V) permettant aux hominidés d'assimiler l'alcool est

apparue il y a 10 millions d'années et pourrait être associée à un changement de mode vie plus terrestre survenu lors d'un changement climatique au milieu du Miocène (Carrigan et al. 2014). Il est donc logique de s'interroger sur les origines des populations de levures fermentant les fruits, et sur l'origine de ce savoir faire apparu de si nombreuses fois : pourquoi, quand et comment ? La domestication des levures pour les fermentations de fruits est elle réellement récente ?
Les aspects liés à la domestication de *S. cerevisiae* peuvent être abordés à partir du modèle des levures *S. cerevisiae* de vin ou de voile. En complément de l'étude de la spécialisation de ces levures à leur environnement à l'aide d'une analyse génétique décrite plus haut, j'ai souhaité faire évoluer mon projet sur quatre points.

6.1. Comment : Reproduire l'adaptation des levures *S. cerevisiae* de vin dans leur environnement (PeakYeast – 2016-2019)

Nous avons détecté dans les parties précédentes plusieurs régions du génome des levures de vin qui semblent fournir aux levures de vin une meilleure capacité à se reproduire dans leur milieu. Ces différents évènements (transferts horizontaux, variabilité de gènes de transport de l'azote ou des stérols, perte de fonctionnalité des aquaporines) sont des indicateurs que les levures *S. cerevisiae* de vin se sont adaptées à leur habitat. Mais cet environnement ne se limite pas à la cuve du chai, car il comprend aussi la vigne où les insectes (guêpes, les drosophiles…) peuvent jouer un rôle de vecteur du chai au raisin (ou l'inverse). Les souches *S. cerevisiae* de vin peuvent se retrouver sur les raisins avec les souches de chêne, mais pas dans les forêts avoisinantes (Hyma and Fay 2013) et en revanche les souches de chêne présentent une moins bonne capacité à fermenter indiquant bien une spécialisation, sans que les deux compartiments soient totalement isolés. Ces arguments militent bien en faveur de la domestication des levures *S. cerevisiae* de vigne.
Cependant la domestication d'espèces c'est souvent accompagnée de l'accumulation de mutations délétères (Glémin and Bataillon 2009; Marsden et al. 2016). Il a été montré chez la levure, que la dérive génétique expliquait davantage les différences phénotypiques que l'adaptation (Warringer et al. 2011) alors qu'on ne connaît pas le niveau d'adaptation général des levures au moût de raisin. Aussi, il est probable qu'une fraction d'allèles délétères fixés dans la population des levures de vin lors du « bottleneck » initial, ou accumulés depuis par relâchement de la sélection, limite le potentiel d'adaptation des levures *S. cerevisiae* de vin. La présence d'allèles avec des

effets opposés dans un QTL pour un croisement vin/voile argue en cette faveur (A17). Il est possible de tester cette hypothèse en utilisant un pool génétique incluant une diversité large fortement recombinée dans une approche d'évolution expérimentale dans des moûts de raisin. Nous avons construit pour cela des populations recombinantes présentant différents niveaux de diversité que nous adaptons au mout de raisin par des fermentations répétées intercalées de cycles méiotiques. Notre objectif est en premier d'évaluer l'évolution des fréquences des allèles au cours de la sélection, puis d'identifier les allèles favorables pour l'environnement « jus de raisin », seul ou avec sa composante microflore. En second nous souhaitons également rechercher si l'on peut observer les même mécanismes (HGT, structure de population et profil de sélection) dans d'autres populations de levures du mout de raisin. En effet, la pratique du levurage est extrêmement récente (1978-80), et plusieurs autres espèces de levures se développent dans un mout de raisin non ensemencé avant *S. cerevisiae.* Ces espèces de levures sont des témoins potentiels de la domestication des levures de vin, et peut être même de l'écosystème entier : vigne/vin/microflore.

6.2. Quand : Retrouver l'histoire de la vinification à partir de l'ADN ancien (Viniculture 2017-2020)

On sait aujourd'hui que le vin a joué un rôle important dès le Néolithique, et en particulier en France on connaît de mieux en mieux les circuits d'échange de vin et les sites de production. L'objectif de l'ANR Viniculture qui est portée par l'archéobotaniste L Bouby (ISEM, Montpellier) comprend deux volets, le premier de mieux caractériser la diversité des cépages et leur répartition du Néolithique au moyen-âge en France, et le second de préciser les techniques de vinification : fruits retenus, modalités de conservation, et flore de fermentations, en s'appuyant sur une approche d'archéogénomique basés sur l'ADN ancien extrait à partir de céramiques. Dans ce projet, je souhaite retrouver l'environnement microbien des fermentations et de conservation des vins dans l'antiquité. Ces fermentations étaient conduites dans des céramiques de manière similaire aux vinifications de la région de Kakhétie en Georgie. Récemment, des populations ont été isolées de telles vinifications (Capece et al. 2013) que j'ai identifiées comme levures de voile à partir de leur génotype. Ceci paraît logique du fait de l'élevage sans sulfites et d'une fermeture imparfaite des cuves. Aussi, il est probable, que les vins antiques présentaient également des populations de levures de voile, ou subissaient un élevage sous voile durant leur conservation. La

caractérisation des flores de ces vins anciens est un moyen de remonter aux conditions de vinification, et éventuellement nous fournir des informations sur l'évolution des génomes de levures de vin ou de voile, à condition, qu'il soit possible d'extraire suffisamment d'ADN d'artefacts anciens.

6.3. Quand : Inférer l'histoire démographique des levures de vins avec des bases réalistes

Pouvoir dater un événement historique au travers de leur signature génétique a été une tentation à laquelle peu d'auteurs ont résisté. Deux paramètres importants sont utilisés: un taux de mutation et un nombre de générations par an. La plupart des études s'appuient sur des taux de mutation établis en conditions de laboratoire sur milieu de culture riche, pour des souches de laboratoires (Drake 1991; Lynch et al. 2008; Zhu et al. 2014). En revanche, des expérimentations récentes d'adaptation de levures en milieu contenant une teneur élevée en éthanol (Voordeckers et al. 2015) ou des inférences reposant sur des données historiques pour la bière (Gallone et al. 2016) suggèrent que ce taux de mutation pourrait être 100 fois plus élevé dans un milieu de fermentation. Il n'y a pas d'estimation faite à ce jour pour la fermentation du mout de raisin avec une expérimentation dédiée d'analyse d'accumulation de mutations, et il me paraît donc nécessaire d'estimer spécifiquement le taux de mutation des levures de vin en fermentation pour le confronter à des modèles démographiques.
Le second paramètre utilisé pour dater ces évènements historiques est le nombre annuel de générations. On peut à priori diviser le cycle de vie de la levure *S. cerevisiae* de vin en trois phases : la phase récolte-fermentations (la mieux connue), sans doute une phase d'hibernation, et une phase de multiplication durant la végétation (figure 15). La fourchette utilisée la plus courante est de 1 à 8 générations par jours soit 2920 à 365 générations par an (Fay and Benavides 2005), mais il s'agit d'une estimation très grossière sans doute irréaliste :

- admettons une baie (1g) altérée contenant 1 million de cellules (Mortimer and Polsinelli 1999) pour une tonne de baies, cela représente potentiellement un mout contenant environ 1 cellule par ml au total
- partir d'une cellule pour en produire 134 millions/ml (population atteinte en fermentation) requiert 27 générations, qui vont assurer la fermentation avant d'être disséminées au vignoble (cycle Raisin – cuve simple)

- La réinoculation d'autres cuves à partir de la première, soit au travers des populations présentes sur le matériel vinaire, soit par les mouches devrait amener un inoculum très supérieur (1000 fois ?) et donc le réensemencement à une cellule par ml d'une seconde cuve devrait produire 27 générations de plus et augmenter le nombre de générations d'autant par cycle annuel

Cette phase devrait à priori produire largement moins d'une centaine de générations par an, alors qu'elle est la phase où à priori l'expansion des populations sera la plus large. Dans le cas d'une population de levures se maintenant dans le chai, à la fin de la vendange après les fermentations, et l'exception des levures de voile, la croissance des levures sera minimale du fait de l'anaérobiose et de l'absence de sucres, ou de l'absence de liquide dans le cas de biofilm restant sur les parois des cuves. Le nombre de générations entre cycles devrait donc être faible ou nul.

Après une vection, le plus probablement réalisée par les insectes ou par l'air (Vigentini et al. 2014), la phase hibernation à l'extérieur du chai devrait à priori être menée à partir sous la forme de spores qui sont des formes de survie. L'expérimentation de Stefanini, (A22) montre que l'hibernation et la dissémination des levures *S. cerevisiae* de vin est compatible avec une hibernation avec les insectes. Nous n'avons pas d'indication sur sa durée, mais une phase d'hibernation ne devrait pas conduire à un nombre élevé de générations. Enfin, nous n'avons pas d'idée précise du nombre de générations subies au vignoble durant la phase végétative, aussi bien sur la vigne qu'avec les insectes. Au total, les chiffres indiquant plus que quelques centaines de générations ne peuvent être qu'irréalistes.

Aussi une expérimentation réalisée en environnement contrôlé à l'aide de souches marquées à l'aide de résistance aux antibiotiques (i.e. résistance à l'érythromycine acquise par mutation) dispersées sur des vignes à l'automne et recueillies à l'automne suivant pourrait permettre d'estimer le nombre de mutations accumulées (mutations x nombre de générations) durant cette phase.

Enfin, les données génomiques de projets larges (1002 yeast genomes), ou le re-séquençage de populations en pool pourront nous permettre d'inférer des scénarios démographiques probables pour les populations de vigne et vin par rapport à celles de voile et chêne Méditerranéen par exemple. Ainsi ce projet nous permettrait d'obtenir des horloges moléculaires plus justes et nous éclairer sur l'échelle de temps qui a

amené à la spécialisation des levures de vin, et à la différentiation de populations de vignobles.

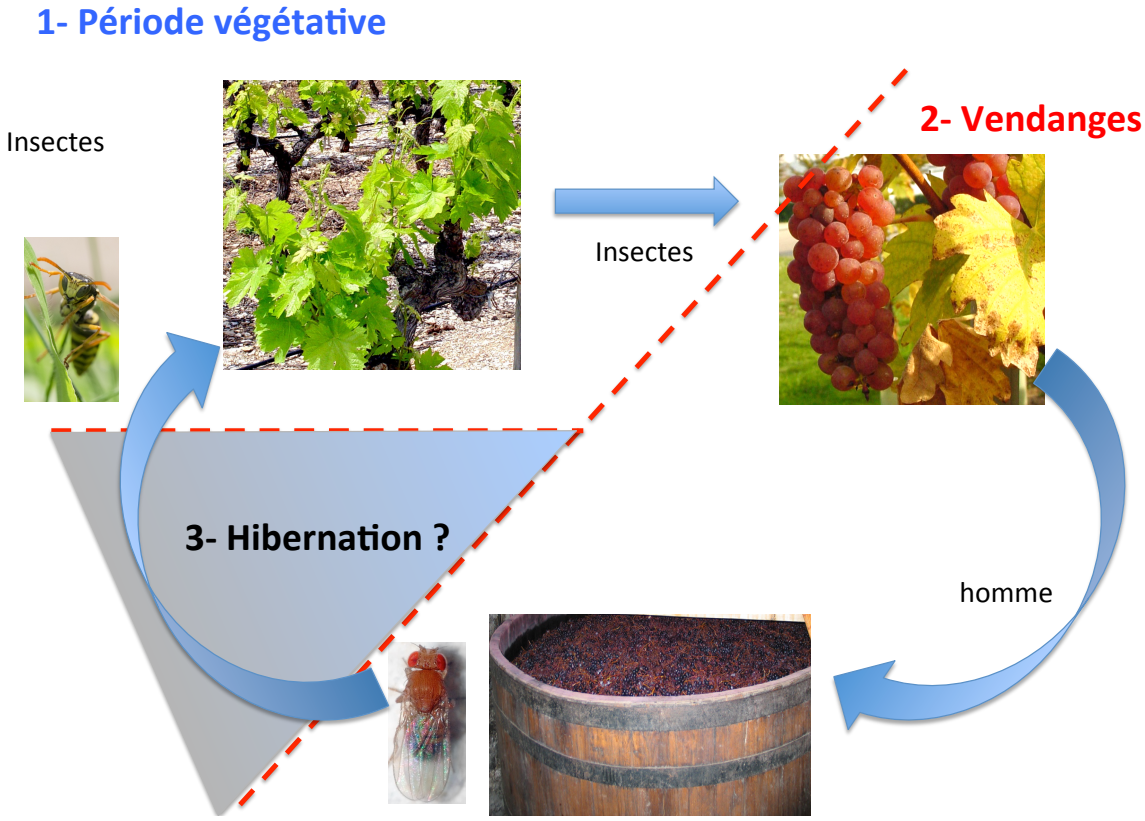


Figure 15 : Cycle écologique schématique de la levure *S. cerevisiae* de vigne.

6.4. Pourquoi *S. cerevisiae*

Alors que cette espèce de levures est minoritaire dans les jus de raisins, la manière dont *S. cerevisiae* domine les fermentations œnologiques classiques (hormis les fermentations à basse température) est remarquable. La résistance à l'éthanol a été longtemps considérée comme la raison principale conduisant à la dominance de *S. cerevisiae*, mais une première entaille a ce dogme a été proposé par Goddard (Goddard 2008) qui a montré que la chaleur produite durant la fermentation contribuait avec la production d'alcool à la dominance de *S. cerevisiae*. Plus récemment, J Fay a montré à partir d'expériences de compétitions, que cette capacité de *S. cerevisiae* à dominer les autres espèces était une acquisition récente dans l'évolution, car seul *S. paradoxus* était capable d'entrer en compétition avec *S. cerevisiae* dans les conditions testées (Williams et al. 2015). Cette capacité de *S. cerevisiae* à dominer les autres levures dans le mout de

raisin est bien liée à la résistance à l'éthanol, mais aussi à la capacité à puiser les acides aminés et les vitamines (Williams et al. 2015). Toutefois, cette étude menée en cultures agitées et donc aérobies ne tient pas compte d'un facteur essentiel en vinification : la forte anaérobiose. Nous avons observé au laboratoire que des fermentations sous agitation pouvaient réduire ou supprimer l'avantage sélectif de *S. cerevisiae*. Les levures ont besoin de stérols pour leur croissance, et en absence d'oxygène, les levures ne peuvent les synthétiser. En vinification, *S. cerevisiae* est capable de puiser les stérols du mout de raisin grâce à des transporteurs spécifiques *AUS1* et *PDR11* (Kohut et al. 2011), et en présence d'oxygène cet import est bloqué sous le contrôle de *SUT1* (Bourot and Karst 1995) et du sensor *UPC2.* Nous avons enfin trouvé *AUS1* comme cible de sélection chez les levures de vin, aussi, il paraît essentiel d'évaluer comment la capacité à utiliser les stérols exogènes a évolué chez les ascomycètes, en particulier pour les gènes *AUS1*, *PDR11*, et *SUT1*, en lien avec la croissance en anaérobiose.

7. Positionnement de ce projet

7.1 Positionnement du projet

Mon projet s'inscrit en premier lieu dans la thématique de notre équipe « Amélioration Diversité et Ecologie des Levures » de l'UMR SPO. Ce que je propose ici, est complémentaire des approches domestication levures menées par D. Sicard DR2 en boulangerie sur la même espèce, mais avec les spécificités du vin, et sera naturellement construit en collaboration avec elle ainsi aussi que de V. Galeote CR1 qui s'est tournée sur les avantages sélectifs offerts par les transferts horizontaux chez les levures en particulier en contexte œnologique. Nous avons tous trois des spécialités différentes : D. Sicard en tant que spécialiste de l'écologie évolutive, V Galeote comme biologiste moléculaire, et moi plutôt positionné sur des aspects de diversité et génomique des populations, et participons tous les trois au projet PeakYeast en cours. De plus le volet transporteur de stérols sera bien sur mené en collaboration avec C. Tesnière qui s'intéresse à l'impact des stérols du mout en vinification et leur rôle physiologique chez les levures *Saccharomyces* ou non-*Saccharomyces*. Enfin, je peux également compter sur l'appui du reste de l'équipe S. Dequin, B. Blondin pour des les aspects de physiologie/amélioration des souches, F. Bigey pour la bioinformatique, et aussi l'équipe FLAM pour le volet métabolisme. La partie la plus novatrice du projet : la dissémination de levure devra être mené avec mes collègues travaillant sur la vigne d'AGAP (P This, JP Peros) ou du LEPS présents sur le site de Montpellier. Il faut aussi

rappeler l'importance de la communauté Montpelliéraine d'écologie et d'évolution, ou de génétique des populations, auprès de qui j'ai noué des contacts pour les aspects de recherche de sélection (P Gladieux BGPI, M Gautier et R Vitalis CBGP…), ou de l'ISEM (S Glémin). Aussi, il me semble que nous disposons au niveau local de tous les outils pour mettre en œuvre ce projet de manière satisfaisante.
Ce projet axé sur la domestication s'intègre complètement dans les axes du département MICA de l'INRA, en particulier du champ thématique CT2. Enfin, nous faisons partie du GDR CNRS iGenolevure dans lequel nous échangeons avec nos collègues Français, ou espagnols (T Gabaldon, Barcelone).
La difficulté de ce thème « domestication » reste par contre la concurrence internationale importante. Je crois que notre spécificité reste d'avoir une meilleure connaissance des problèmes de vinification que certains collègues étrangers ayant de meilleures compétences en génomique. De plus nous sommes une équipe reconnue sur ce thème, puisque j'ai été contacté par N Dominy un anthropologue Américains, pour présenter nos résultats sur la domestication des levures à partir de mon article de Molecular Ecology de 2007, et de la revue faite avec D Sicard dans CR Acad Science en 2011 dans une cession du congrès de l' « American Physical Anthropologist Association ». Il m'a proposé de collaborer avec lui à l'étude de la domestication de *S. cerevisiae* en analysant des populations de levures africaines.

Une seconde question est celle de l'intérêt sociétal de ces projets. Ce projet peut paraître fondamental. Cependant, je crois qu'ancrer notre connaissance scientifique dans un contexte historique (si on le peut) est un moyen de rapprocher science et citoyen. L'art de la vinification est issu du patrimoine d'un savoir faire collectif empirique. L'art, puis la science ont permis de développer de nouvelles technologies et fait progresser le savoir faire des vinificateurs, en particulier en ce qui concerne la sécurité des vinifications. Aujourd'hui, l'excès d'interventions œnologique s a pu amener certains viticulteurs à rejeter certaines innovations, mais il n'est pas non plus réaliste d'idéaliser les vins anciens.
Mon projet à pour but de chercher à comprendre la domestication de la levure *S. cerevisiae* avec l'évolution de la vinification. Les populations de levures ont du vraisemblablement s'adapter aux changements technologiques dont nous ne connaissons que les derniers épisodes, comme l'utilisation des sulfites ou du cuivre.

Connaître les changements passés peut aussi nous éclairer sur les évolutions inéluctables qui nous attendent avec le changement climatique.
Cette évolution des populations de levures fait partie du patrimoine historique des vignobles, et concerne donc les viticulteurs. Je pense qu'il est possible d'aborder l'analyse de populations régionales à l'aide d'approches participatives impliquant les viticulteurs en particulier pour recueillir les flores de vignobles, ce qui est d'ailleurs dans l'esprit des accords de Nagoya.

7.2 Pourquoi une HDR

C'est pour moi à la fois une reconnaissance et le moyen de contribuer à l'animation scientifique et collective de l'équipe, qui je crois est de la responsabilité de chacun, scientifiques ou ingénieurs. Mon objectif est aussi de pouvoir proposer et encadrer des projets qui me sont chers tout en étant dans le cadre des thèmes de l 'équipe. J'ai émis ce souhait il y a déjà 8 ans, avant ma mobilité, mais dû le reporter avec le chevauchement entre les projets issus de mon travail à Colmar et mes nouveaux projets Montpelliérains. Enfin disposer d'un encadrement suffisant dans l'unité est important, car nous avons aussi été en situation de refuser un financement de thèse faute d'HDR dans l'équipe.

8. *Bibliographie*

Aigle M, Erbs D, Moll M. 1984. Some molecular structures in the genome of lager brewing yeasts. J. Am. Soc. Brew. Chem. 42:1–7.

Akey JM, Ruhe AL, Akey DT, Wong AK, Connelly CF, Madeoy J, Nicholas TJ, Neff MW. 2010. Tracking footprints of artificial selection in the dog genome. Proc. Natl. Acad. Sci. U. S. A. 107:1160–1165.

Akey JM, Zhang G, Zhang K, Jin L, Shriver MD. 2002. Interrogating a high-density SNP map for signatures of natural selection. Genome Res. 12:1805–1814.

Alachiotis N, Stamatakis a., Pavlidis P. 2012. OmegaPlus: A scalable tool for rapid detection of selective sweeps in whole-genome datasets. Bioinformatics 28:2274–2275.

Albertin W, Chasseriaud L, Comte G, Panfili A, Delcamp A, Salin F, Marullo P, Bely M. 2014. Winemaking and Bioprocesses Strongly Shaped the Genetic Diversity of the Ubiquitous Yeast Torulaspora delbrueckii. PLoS One 9:e94246.

Albertin W, Setati ME, Miot-Sertier C, Mostert TT, Colonna-Ceccaldi B, Coulon J, Girard P, Moine V, Pillet M, Salin F, et al. 2016. Hanseniaspora uvarum from winemaking environments show

spatial and temporal genetic clustering. Front. Microbiol. 6.

Almeida P, Barbosa R, Zalar P, Imanishi Y, Shimizu K, Turchetti B, Legras J-L, Serra M, Dequin S, Couloux A, et al. 2015. A population genomics insight into the Mediterranean origins of wine yeast domestication. Mol. Ecol. 24:5412–5427.

Becher PG, Flick G, Rozpedowska E, Schmidt A, Hagman A, Lebreton S, Larsson MC, Hansson BS, Piškur J, Witzgall P, et al. 2012. Yeast, not fruit volatiles mediate Drosophila melanogaster attraction, oviposition and development. Funct. Ecol. 26:822–828.

Belloch C, Pérez-Torrado R, González SS, Pérez-Ortín JE, García-Martínez J, Querol A, Barrio E. 2009. Chimeric genomes of natural hybrids of *Saccharomyces cerevisiae* and Saccharomyces kudriavzevii. Appl. Environ. Microbiol. 75:2534–2544.

Börlin M, Venet P, Claisse O, Salin F, Legras J-L, Masneuf-Pomarède I. 2016. Cellar-associated *Saccharomyces cerevisiae* population structure revealed high diversity and perennial persistence in Sauternes wine estates. Appl. Environ. Microbiol. [Internet]:AEM.03627-15.

Bourot S, Karst F. 1995. Isolation and characterization of the *Saccharomyces cerevisiae* SUT1 gene involved in sterol uptake. Gene 165:97–102.

Brice C, Sanchez I, Bigey F, Legras J-L, Blondin B. 2014. A genetic approach of wine yeast fermentation capacity in nitrogen-starvation reveals the key role of nitrogen signaling. BMC Genomics 15:495.

Bruvo R, Michiels NK, D'Souza TG, Schulenburg H. 2004. A simple method for the calculation of microsatellite genotype distances irrespective of ploidy level. Mol. Ecol. 13:2101–2106.

Camarasa C, Sanchez I, Brial P, Bigey F, Dequin S. 2011. Phenotypic Landscape of *Saccharomyces cerevisiae* during Wine Fermentation: Evidence for Origin-Dependent Metabolic Traits.Harris S, editor. PLoS One 6:e25147.

Capece A, Siesto G, Poeta C, Pietrafesa R, Romano P. 2013. Indigenous yeast population from Georgian aged wines produced by traditional 'Kakhetian' method. Food Microbiol. 36:447–455.

Carrigan M a, Uryasev O, Frye CB, Eckman BL, Myers CR, Hurley TD, Benner S a. 2014. Hominids adapted to metabolize ethanol long before human-directed fermentation. Proc. Natl. Acad. Sci.:6–11.

Cavalli-Sforza LL, Edwards AWF. 1967. PHYLOGENETIC ANALYSIS: MODELS AND ESTIMATION PROCEDURES. Evolution 21:550–570.

Charpentier C, Colin A, Alais A, Legras J-L. 2009. French Jura flor yeasts: genotype and technological diversity. Antonie Van Leeuwenhoek 95:263–273.

Coi AL, Legras J-L, Zara G, Dequin S, Budroni M. 2016. A set of haploid strains available for genetic studies of *Saccharomyces cerevisiae* flor yeasts. FEMS Yeast Res. 16:1–26.

Cromie GA, Hyma KE, Ludlow CL, Garmendia-Torres C, Gilbert TL, May P, Huang A a, Dudley AM, Fay JC. 2013. Genomic sequence diversity and population structure of *Saccharomyces*

cerevisiae assessed by RAD-seq. G3 (Bethesda). 3:2163–2171.

Cuinier, C., Bouix M. et Leveau, J.-Y., 1981. Méthode d'étude de l'origine des levures en œnologie . Identification des espèces et différenciation des clones. Vignes et Vins, 302, 3 - 7.

Degré R, Thomas DY, Ash J, Mailhiot K, Morin A, Dubord C. 1989. Wine Yeasts Strain Identification. Am. J. Enol. Vitic. 40:309–315.

Demuyter C, Lollier M, Legras J-L, Le Jeune C, LeJeune C. 2004. Predominance of Saccharomyces uvarum during spontaneous alcoholic fermentation, for three consecutive years, in an Alsatian winery. J. Appl. Microbiol. 97:1140–1148.

Diamond J. 2002. Evolution, consequences and future of plant and animal domestication. Nature 418:700–707.

Drake JW. 1991. A constant rate of spontaneous mutation in DNA-based microbes. Proc. Natl. Acad. Sci. 88:7160–7164.

Enjalbert J, Duan X, Leconte M, Hovmøller MS, De Vallavieille-Pope C. 2005. Genetic evidence of local adaptation of wheat yellow rust (Puccinia striiformis f. sp. tritici) within France. Mol. Ecol. 14:2065–2073.

Erny C, Raoult P, Alais a., Butterlin G, Delobel P, Matei-Radoi F, Casaregola S, Legras JL. 2012. Ecological Success of a Group of *Saccharomyces cerevisiae*/Saccharomyces kudriavzevii Hybrids in the Northern European Wine-Making Environment. Appl. Environ. Microbiol. 78:3256–3265.

Esteve-Zarzoso B, Peris-Torán MJ, García-Maiquez E, Uruburu F, Querol A. 2001. Yeast population dynamics during the fermentation and biological aging of sherry wines. Appl. Environ. Microbiol. 67:2056–2061.

Fariello MI, Boitard S, Naya H, SanCristobal M, Servin B. 2013. Detecting signatures of selection through haplotype differentiation among hierarchically structured populations. Genetics 193:929–941.

Fay JC, Benavides JA. 2005. Evidence for domesticated and wild populations of *Saccharomyces cerevisiae*. PLoS Genet. 1:66–71.

Fay JC, McCullough HL, Sniegowski PD, Eisen MB. 2004. Population genetic variation in gene expression is associated with phenotypic variation in *Saccharomyces cerevisiae*. Genome Biol. 5:R26.

Fay JC, Wu CI. 2000. Hitchhiking under positive Darwinian selection. Genetics 155:1405–1413.

Fisher MC, Koenig GL, White TJ, San-Blas G, Negroni R, Alvarez IG, Wanke B, Taylor JW. 2001. Biogeographic range expansion into South America by Coccidioides immitis mirrors New World patterns of human migration. Proc. Natl. Acad. Sci. U. S. A. 98:4558–4562.

Frezier V, Dubourdieu D. 1992. Ecology of yeast-strain *Saccharomyces cerevisiae* during spontaneous fermentation in a spontaneous fermentation in a Bordeaux winery. Am. J. Enol.

Vitic. 43:375–380.

Fundyga RE, Lott TJ, Arnold J. 2002. Population structure of Candida albicans, a member of the human flora, as determined by microsatellite loci. Infect. Genet. Evol. 2:57–68.

Gallone B, Steensels J, Prahl T, Soriaga L, Saels V, Herrera-Malaver B, Merlevede A, Roncoroni M, Voordeckers K, Miraglia L, et al. 2016. Domestication and Divergence of *Saccharomyces cerevisiae* Beer Yeasts. Cell 166:1397–1410.e16.

Gil-Lamaignere C, Roilides E, Hacker J, Müller F-MC. 2003. Molecular typing for fungi--a critical review of the possibilities and limitations of currently and future methods. Clin. Microbiol. Infect. 9:172–185.

Glémin S, Bataillon T. 2009. A comparative view of the evolution of grasses under domestication: Tansley review. New Phytol. 183:273–290.

Gochman SR, Brown MB, Dominy NJ. 2016. Alcohol discrimination and preferences in two species of nectar-feeding primate. R. Soc. Open Sci. 3:160217.

Goddard MR. 2008. Quantifying the complexities of *Saccharomyces cerevisiae*'s ecosystem engineering via fermentation. Ecology 89:2077–2082.

Goddard MR, Greig D. 2015. *Saccharomyces cerevisiae*: a nomadic yeast with no niche? FEMS Yeast Res. 15:1–6.

González SS, Barrio E, Querol A. 2008. Molecular characterization of new natural hybrids of *Saccharomyces cerevisiae* and S. kudriavzevii in brewing. Appl. Environ. Microbiol. 74:2314–2320.

González Techera a, Jubany S, Carrau FM, Gaggero C. 2001. Differentiation of industrial wine yeast strains using microsatellite markers. Lett. Appl. Microbiol. 33:71–75.

Gu J, Orr N, Park SD, Katz LM, Sulimova G, MacHugh DE, Hill EW. 2009. A genome scan for positive selection in thoroughbred horses. PLoS One 4:e5767.

Gutenkunst RN, Hernandez RD, Williamson SH, Bustamante CD. 2009. Inferring the joint demographic history of multiple populations from multidimensional SNP frequency data. PLoS Genet. 5:e1000695.

Hennequin C, Thierry A, Richard G-F, Lecointre G, Nguyen H-V, Gaillardin C, Dujon B. 2001. Microsatellite typing as a new tool for identification of *Saccharomyces cerevisiae* strains. J. Clin. Microbiol. 39:551–559.

Hockings KJ, Bryson-Morrison N, Carvalho S, Fujisawa M, Humle T, McGrew WC, Nakamura M, Ohashi G, Yamanashi Y, Yamakoshi G, et al. 2015. Tools to tipple: ethanol ingestion by wild chimpanzees using leaf-sponges. R. Soc. Open Sci. 2:150150.

Hranilovic A, Bely M, Masneuf-Pomarede I, Jiranek V, Albertin W. 2017. The evolution of Lachancea thermotolerans is driven by geographical determination, anthropisation and flux between different ecosystems. PLoS One 12:e0184652.

Hussin J, Nadeau P, Lefebvre J-F, Labuda D. 2010. Haplotype allelic classes for detecting ongoing

positive selection. BMC Bioinformatics 11:65.

Hyma KE, Fay JC. 2013. Mixing of vineyard and oak-tree ecotypes of *Saccharomyces cerevisiae* in North American vineyards. Mol. Ecol. [Internet]:1–14.

Infante JJ, Dombek KM, Rebordinos L, Cantoral JM, Young ET. 2003. Genome-wide amplifications caused by chromosomal rearrangements play a major role in the adaptive evolution of natural yeast. Genetics 165:1745–1759.

King A, Dickinson JR. 2000. Biotransformation of monoterpene alcohols by *Saccharomyces cerevisiae*, Torulaspora delbrueckii and Kluyveromyces lactis. Yeast 16:499–506.

Knight S, Goddard MR. 2015. Quantifying separation and similarity in a *Saccharomyces cerevisiae* metapopulation. ISME J. 9:361–370.

Kohut P, Wüstner D, Hronska L, Kuchler K, Hapala I, Valachovic M. 2011. The role of ABC proteins Aus1p and Pdr11p in the uptake of external sterols in yeast: Dehydroergosterol fluorescence study. Biochem. Biophys. Res. Commun. 404:233–238.

Lebon E, Dumas V, Mettauer H, Morlat R. 1993. Caractérisation intégrée du vignoble Alsacien: Aspects méthodologiques et application à l'étude des composantes naturelles des principaux terroirs. J. Int. des Sci. la Vigne du Vin 27:235–253.

Legras J-L, Erny C, Charpentier C. 2014. Population Structure and Comparative Genome Hybridization of European Flor Yeast Reveal a Unique Group of *Saccharomyces cerevisiae* Strains with Few Gene Duplications in Their Genome. PLoS One 9:e108089.

Legras J-L, Merdinoglu D, Cornuet J-M, Karst F. 2007. Bread, beer and wine: *Saccharomyces cerevisiae* diversity reflects human history. Mol. Ecol. 16:2091–2102.

Legras J-L, Ruh O, Merdinoglu D, Karst F. 2005. Selection of hypervariable microsatellite loci for the characterization of strains. Int. J. Food Microbiol. 102:73–83.

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 458:337–341.

Lowe A, Harris S, Ashton P. 2004. Ecological Genetics. Design, Analysis and Application.

Lynch M, Sung W, Morris K, Coffey N, Landry CR, Dopman EB, Dickinson WJ, Okamoto K, Kulkarni S, Hartl DL, et al. 2008. A genome-wide view of the spectrum of spontaneous mutations in yeast. Proc. Natl. Acad. Sci. U. S. A. 105:9272–9277.

Mallet J. 2005. Hybridization as an invasion of the genome. Trends Ecol. Evol. (Personal Ed. 20:229–237.

Mallet J. 2007. Hybrid speciation. Nature 446:279–283.

Marsden CD, Ortega-Del Vecchyo D, O'Brien DP, Taylor JF, Ramirez O, Vilà C, Marques-Bonet T, Schnabel RD, Wayne RK, Lohmueller KE. 2016. Bottlenecks and selective sweeps during domestication have increased deleterious genetic variation in dogs. Proc. Natl. Acad. Sci. 113:152–157.

Martínez P, Codón AC, Pérez L, Benítez T. 1995. Physiological and molecular characterization of flor yeasts: polymorphism of flor yeast populations. Yeast 11:1399–1411.

Martini A. 1993. Origin and domestication of the wine yeast *Saccharomyces cerevisiae*. J. Wine Res. 4:165–176.

Martini A, Vaughan-Martini A. 1990. Grape must fermentation: past and present. (Spencer JFT, Spencer DM, editors.). Springer Verlag

Masneuf I, Hansen J, Groth C, Piskur J, Dubourdieu D. 1998. New hybrids between Saccharomyces sensu stricto yeast species found among wine and cider production strains. Appl. Environ. Microbiol. 64:3887–3892.

Mortimer R, Polsinelli M. 1999. On the origins of wine yeast. Res. Microbiol. 150:199–204.

Mortimer RK. 2000. Evolution and Variation of the Yeast (Saccharomyces) Genome. Genome Res. 10:403–409.

Ness F, Lavallée F, Dubourdieu D, Aigle M, Dulau L. 1993. Identification of yeast strains using the polymerase chain reaction. J. Sci. Food Agric. 62:89–94.

Nguyen H-V, Legras J-L, Neuvéglise C, Gaillardin C. 2011. Deciphering the Hybridisation History Leading to the Lager Lineage Based on the Mosaic Genomes of Saccharomyces bayanus Strains NBRC1948 and CBS380T.Rusche LN, editor. PLoS One 6:e25821.

Novo M, Bigey F, Beyne E, Galeote V, Gavory F, Mallet S, Cambon B, Legras J-L, Wincker P, Casaregola S, et al. 2009. Eukaryote-to-eukaryote gene transfer events revealed by the genome sequence of the wine yeast *Saccharomyces cerevisiae* EC1118. Proc. Natl. Acad. Sci. 106:16333–16338.

Pedersen M. 1985. DNA sequence polymorphisms in the genus Saccharomyces. II. Analysis of the genes RDN1, HIS4, LEU2 and Ty transposable elements in Carlsberg, Tuborg, and 22 Bavarian brewing strains. Carlsb. Res. Commun. 50:262–272.

Pérez-Ortín JE, Querol A, Puig S, Barrio E. 2002. Molecular characterization of a chromosomal rearrangement involved in the adaptive evolution of yeast strains. Genome Res. 12:1533–1539.

Pérez M a, Gallego FJ, Martínez I, Hidalgo P. 2001. Detection, distribution and selection of microsatellites (SSRs) in the genome of the yeast *Saccharomyces cerevisiae* as molecular markers. Lett. Appl. Microbiol. 33:461–466.

Peris D, Belloch C, Lopandić K, Álvarez-Pérez JM, Querol A, Barrio E. 2012. The molecular characterization of new types of *S. cerevisiae* x S. kudriavzevii hybrid yeasts unveils a high genetic diversity. Yeast 29:81-91.

Querol A, Barrio E. 1990. A rapid and simple method for the preparation of yeast mitochondrial DNA. Nucleic Acids Res. 18:1657.

Rubin C-J, Zody MC, Eriksson J, Meadows JRS, Sherwood E, Webster MT, Jiang L, Ingman M, Sharpe T, Ka S, et al. 2010. Whole-genome resequencing reveals loci under selection during chicken domestication. Nature 464:587–591.

Sabeti PC, Varilly P, Fry B, Lohmueller J, Hostetter E, Cotsapas C, Xie X, Byrne EH, McCarroll SA, Gaudet R, et al. 2007. Genome-wide detection and characterization of positive selection in human populations. Nature 449:913–918.

Saerens SMG, Verstrepen KJ, Van Laere SDM, Voet ARD, Van Dijck P, Delvaux FR, Thevelein JM. 2006. The *Saccharomyces cerevisiae* EHT1 and EEB1 genes encode novel enzymes with medium-chain fatty acid ethyl ester synthesis and hydrolysis capacity. J. Biol. Chem. 281:4446–4456.

Sampaio JP, Gonçalves P. 2008. Natural populations of Saccharomyces kudriavzevii in Portugal are associated with oak bark and are sympatric with *S. cerevisiae* and S. paradoxus. Appl. Environ. Microbiol. 74:2144–2152.

Schacherer J, Shapiro JA, Ruderfer DM, Kruglyak L. 2009. Comprehensive polymorphism survey elucidates population structure of *Saccharomyces cerevisiae*. Nature 458:342–345.

Schuller D, Valero E, Dequin S, Casal M. 2004. Survey of molecular methods for the typing of wine yeast strains. FEMS Microbiol. Lett. 231:19–26.

Sicard D, Legras J-L. 2011. Bread, beer and wine: Yeast domestication in the Saccharomyces sensu stricto complex. C. R. Biol. 334:229–236.

Spor A, Wang S, Dillmann C, de Vienne D, Sicard D. 2008. 'Ant' and 'grasshopper' life-history strategies in *Saccharomyces cerevisiae*. PLoS One 3:e1579.

Stefanini I, Dapporto L, Legras J-L, Calabretta A, Di Paola M, De Filippo C, Viola R, Capretti P, Polsinelli M, Turillazzi S, et al. 2012. Role of social wasps in *Saccharomyces cerevisiae* ecology and evolution. Proc. Natl. Acad. Sci. U. S. A. 109:13398–13403.

Swiegers JH, Bartowsky EJ, Henschke P a., Pretorius IS. 2005. Yeast and bacterial modulation of wine aroma and flavour. Aust. J. Grape Wine Res. 11:139–173.

Tenaillon MI, U'Ren J, Tenaillon O, Gaut BS. 2004. Selection versus demography: a multilocus investigation of the domestication process in maize. Mol. Biol. Evol. 21:1214–1225.

Valero E, Schuller D, Cambon B, Casal M, Dequin S. 2005. Dissemination and survival of commercial wine yeast in the vineyard: a large-scale, three-years study. FEMS Yeast Res. 5:959–969.

Verstrepen KJ, Van Laere SDM, Vanderhaegen BMP, Derdelinckx G, Dufour J-P, Pretorius IS, Winderickx J, Thevelein JM, Delvaux FR. 2003. Expression Levels of the Yeast Alcohol Acetyltransferase Genes ATF1, Lg-ATF1, and ATF2 Control the Formation of a Broad Range of Volatile Esters. Appl. Environ. Microbiol. 69:5228–5237.

Vezinhet F, Blondin B, Hallet J-N. 1990. Chromosomal DNA patterns and mitochondrial DNA polymorphism as tools for identification of enological strains of *Saccharomyces cerevisiae*. Appl. Microbiol. Biotechnol. 32:568–571.

Vigentini I, De Lorenzis G, Fabrizio V, Valdetara F, Faccincani M, Panont C a., Picozzi C, Imazio S, Failla O, Foschino R. 2014. The vintage effect overcomes the terroir effect: a three year survey

on the wine yeast biodiversity in Franciacorta and Oltrepo Pavese, two northern Italian vine-growing areas. Microbiology 161:362–373.

Voight BF, Kudaravalli S, Wen X, Pritchard JK. 2006. A map of recent positive selection in the human genome. PLoS Biol. 4:0446–0458.

Voordeckers K, Kominek J, Das A, Espinosa-Cantú A, De Maeyer D, Arslan A, Van Pee M, van der Zande E, Meert W, Yang Y, et al. 2015. Adaptation to High Ethanol Reveals Complex Evolutionary Pathways. PLoS Genet. 11:1–31.

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:e1002111.

Wiens F, Zitzmann A, Lachance M-A, Yegles M, Pragst F, Wurst FM, von Holst D, Guan SL, Spanagel R. 2008. Chronic intake of fermented floral nectar by wild treeshrews. Proc. Natl. Acad. Sci. U. S. A. 105:10426–10431.

Will JL, Kim HS, Clarke J, Painter JC, Fay JC, Gasch AP. 2010. Incipient balancing selection through adaptive loss of aquaporins in natural *Saccharomyces cerevisiae* populations. PLoS Genet. 6:e1000893.

Williams KM, Liu P, Fay JC. 2015. Evolution of ecological dominance of yeast species in high-sugar environments. Evolution 69:2079–2093. A

Zhu YO, Siegal ML, Hall DW, Petrov D a. 2014. Precise estimates of mutation rate and spectrum in yeast. Proc. Natl. Acad. Sci. U. S. A. 111:E2310-8.

Annexes : Publications représentatives

ELSEVIER

FEMS Microbiology Letters 221 (2003) 249–255

FEMS MICROBIOLOGY Letters

www.fems-microbiology.org

Optimisation of interdelta analysis for *Saccharomyces cerevisiae* strain characterisation

Jean-Luc Legras *, Francis Karst

Institut National de la Recherche Agronomique (INRA), Unité Mixte de Recherche Vigne et Vins d'Alsace, 28 rue de Herrlisheim, BP 507, 68021 Colmar Cedex, France

Received 12 December 2002; received in revised form 27 February 2003; accepted 5 March 2003

First published online 26 March 2003

Abstract

A new primer pair (delta12–delta21) for polymerase chain reaction-based yeast typing was designed using the yeast genome sequence. The specificity of this primer pair was checked by the comparison of the electrophoresis pattern with a virtual profile calculated from Blast data. The analysis of 53 commercial and laboratory *Saccharomyces cerevisiae* yeast strains showed a clear improvement of interdelta analysis using the newly designed primers.

Keywords: Wine; Molecular typing; Genetic variability; Delta element; *Saccharomyces cerevisiae*

1. Introduction

Numerous studies have been conducted on wine yeast since the early work of Pasteur. Yeast strains have specific enzymatic activities necessary for the production of key aromatic compounds during alcoholic fermentation. Enzymes appear to be required for the release of bound compounds synthesised from grapes such as mercaptopentanone in Sauvignon wines [1] or volatile phenols in Gewurztraminer wines [2]. Such uncoverings led to the technological success of tested commercial dry-wine yeast strains in the process of wine making. However, these important enzymatic differences are laborious to analyse in different strains and a more rapid but just as definitive method is necessary. Recent molecular biological techniques have allowed the characterisation of yeast strains [3,4], leading to a new era for ecological surveys. These techniques have enabled the population dynamics of *Saccharomyces cerevisiae* strains in vineyards or wineries to be studied [5–10], as well as the control of industrially dried yeast production. They also proved extremely beneficial for yeast laboratories testing strains for their enological properties in order to optimise wild-strain isolates collections. Since the first application of mitochondrial DNA restriction profiling to brewing yeast [11], several techniques have been developed. Pulsed-field electrophoresis karyotyping applied to enological yeast strains [12] is often regarded as a time-consuming procedure although it is highly efficient [13]. Other polymerase chain reaction (PCR) techniques were developed with randomly amplified polymorphic DNA (RAPD) primers for wine yeast [14,15], but poor discrimination of strains was obtained with each primer set. De Barros Lopez et al. [16] developed an interesting technique using primers based on intron splicing sites.

More recently, a powerful microsatellite-based technique has been developed [17]. It differentiates yeast from various origins (French enological, medical isolates, Asian yeast, etc.) through the analysis of seven loci. Similar techniques were developed [18,19], but as they were based on one locus analysis only, these approaches were, however, not powerful enough for an accurate characterisation of numerous yeast strains. The potential of such a technique for enological strains has still to be evaluated. An AFLP-based analysis of yeast strains, clustering beer, whisky, bakery, and sake yeast, was also described [20]. However, these two methods require DNA sequence analysis for accurate determination.

* Corresponding author. Tel.: +33 (3) 89224993; Fax: +33 (3) 89224989.
E-mail address: legras@colmar.inra.fr (J.-L. Legras).

0378-1097/03/$22.00
doi:10.1016/S0378-1097(03)00205-2

The use of a rapid PCR-based protocol, relying on the amplification of interdelta regions, was initially proposed in 1993 [21]. Delta elements form the LTR flanking retrotransposons TY1 and TY2 in yeast, but can also be found separate from these retrotransposons and are called solo delta elements. About 300 such delta elements are described in the genome of S288C and are therefore good candidate targets for identification of polymorphisms. This interdelta method is now often used for routine analysis of yeast strains [22], but is less discriminatory than the previously mentioned pulsed-field electrophoresis [13].

In this report, the interdelta method has been improved utilising the now complete yeast genome sequence database. Analysis of the database allowed the design of the interdelta primers to be optimised, as well as the development of a new method compatible with numeric profiling.

2. Materials and methods

2.1. Yeast strains

Laboratory and enological yeast strains used in this work are listed in Table 1.

2.2. DNA extraction

DNA was prepared from 10 ml YPD cultures (Yeast Extract (Difco) 1% w/v, Bactopeptone (Difco) 1% w/v, glucose 2% w/v) agitated for 24 h at 28°C.

2 ml of the culture was centrifuged in an Eppendorf tube (5000 rpm, 5 min). 400 μl of lysis buffer (Tris 10 mM, pH 7.6, EDTA 1 mM, NaCl 100 mM, Triton X-100, 2% w/v, sodium dodecyl sulphate (SDS) 1% w/v), 400 μl of phenol/chloroform/iso-amyl alcohol (25/24/1 v/v), and 600 μl of glass beads were added to the pellet. The mixture was vortexed for 4 min. Then 200 μl of Tris EDTA (pH 7.6) buffer was added, and the mixture centrifuged for 5 min at 6000 rpm. 500 μl of chloroform/iso-amyl alcohol (98/2 v/v) was added to the upper phase and after gentle agitation, the mixture was submitted to centrifugation (14 000 rpm for 2 min). Two volumes of ethanol were added to the aqueous phase. After centrifugation (14 000 rpm, 5 min), the nucleic acid pellet was dissolved in 10 mM TE buffer pH 8.0.

DNA from commercial dried yeast was prepared in the same way after rehydrating 50 mg of yeast powder in 500 μl 50 mM EDTA, for 15 min.

2.3. DNA amplification

PCR amplifications were carried out in 25 μl reaction volumes containing 5–20 ng yeast DNA, 10 mM Tris pH 9.0, 50 mM KCl, 0.1% Triton X-100, 0.2 mg ml^{-1} gelatin, 200 mM of each dNTP, 2.5 mM of $MgCl_2$ and 1 μM for

Table 1
Saccharomyces strains used in the study

Strain	Origin – Commercial name
S. carlsbergensis Clib176	Clib laboratory strain
S. exiguus Clib179	Clib laboratory strain
S. bayanus Clib181	Clib laboratory strain
S. kluyveri Clib182	Clib laboratory strain
S. paradoxus Clib228	Clib laboratory strain
S. uvarum Clib533	Clib laboratory strain
S. uvarum D24	Wild-type isolate, INRA Colmar
S. bayanus Clib181	Clib laboratory strain
S. servazzii Clib187	Clib laboratory strain
Saccharomyces cerevisiae laboratory strains	
SC288C	Laboratory strain
FL100	Laboratory strain
FL200	Laboratory strain
SW303	Laboratory strain
FY1679	Laboratory strain
Clib176	Clib brewery strain
Clib227	Clib laboratory strain
Clib319	Bakery yeast
Saccharomyces cerevisiae enological yeast	
182007	IOEC 18-2007
1M8	Wild-type isolate, INRA Colmar
1N1d	Wild-type isolate, INRA Colmar
522D	Davis Montrachet
58W3	Alsaflore (INRA Colmar)
595 Davies	Vitilevure 'B'
67J	Enolevure OV
7013	Enolevure SL
70S1	Vitilevure Albaflore
71B	71B (INRA Narbonne)
7303	Enolevure CR
AWRI350	AWRI350
AWRI796	AWRI796
B10	GE7 (INRA Colmar)
B94/201	Ceres (INRA Colmar)
C19	C19 (INRA Colmar)
CH158	SIHA4
CIVC8130	CHP
CY3079	CY3079
D1	Wild-type isolate, ULP Strasbourg
D576	SIHA7
DV10	DV10
E51	SIHA2
Eg8/136	ALS (INRA Colmar)
K1	Enolevure K34 (INRA Montpellier)
L13	LAl13
L1414	L1414
L2056	Lalvin 2056
LW128-91	Hefix 1000
LW 185-25	Hefix 2000
LW317-29	Oenoferm Klosterneubourg
MBZ1	levuline MBZ1
MV94017	Vitilevure 'Sauvignon'
R2	KD
RC212	Lavin RC212
UP3OY5	BRG
RCA17	Lalvin RCA17
RHST	RHST
ST-Clib2026	Zymaflore ST
Vl1	Zymaflore VL1
Vl3	Zymaflore VL3
WET 136	SIHA3
xx	Spindal Aromatic
yy	Vitilevure Rubiflore
zz	Zymasil

each oligonucleotide primer of the delta1 family and delta2 family. Primer sequences for delta 1 and delta 2 are those described previously [21] and newly designed primers are delta12 (5′-TCAACAATGGAATCCCAAC-3′) and delta21 (5′-CATCTTAACACCGTATATGA-3′).

Amplification reactions were performed with a Stratagene thermal cycler using the following programme: 4 min at 95°C followed by 35 cycles of 30 s at 95°C, 30 s at 46°C and 90 s at 72°C and a finishing step of 10 min at 72°C. For the delta1/delta2 primer pair, the annealing temperature was 42°C for the first five cycles and 45°C for the following cycles as described in [21].

2.4. Electrophoresis

Amplification products were separated by electrophoresis on 10 or 15 cm 2% agarose gels submitted to 100 V for 1 h in 1×TBE buffer. Nusieve 3:1 agarose (BMA) was used for newly designed primers, since better resolution was required.

2.5. Numerical analysis

After staining with ethidium bromide (10 μg ml^{-1}), gels were scanned with a Gel Doc 1000 apparatus (Bio-Rad), and compared with Molecular Analyst Fingerprinting plus (Bio-Rad) after normalisation of the profiles. Clustering of profiles was done using the UGMPA calculation methods based on Dice coefficient. Gel images and dendrograms were processed with Adobe Photoshop and CorelDraw.

2.6. Sequence alignment tools

Blast searches were performed with the 'nearly exact match option' at NCBII. As differences were noticed for chromosome III, between NCBI and SGD data, only coordinates obtained from SGD were kept for that chromosome.

The statistical significance threshold for reporting matches against database sequences was 1000, and word size 7. Sequences and coordinates were treated in a 4D® database, and band-size calculation was established using an Excel® data sheet. Possible amplification sites were selected from nine base-length cohesive 3′ end sequences. Possible bands were determined from the coordinates given by the Blast procedure; one possible band corresponds to two primers on different strands with opposite coordinates (5′ 3′). Band weight was calculated from the distance between 3′ ends of each primer increased by the size of the respective primers. Melting temperature was calculated for the complementary sequences from the approximate model of Bolton and McCarthy [23]. As the low melting temperature was unlikely to give small bands under our amplification conditions only those possibilities corresponding to a calculated T_m greater than 25°C were retained.

```
5′ TGTTGGAATA AAAACCAACT ATCGTCTATC AACTAGTAGT CATACTATCA
   ATATATTATC ATATACGGTG TTAGATAGTG ACATAAGTTA TTATAGAAGC
   TGTCACCGAA GTTAGAGGAA GCTGAAATGC AAGGATCGAT AATGTAATAG
   GATAATGAAA CATATGAAAT GGAATGAGGA ATAATCGTAA TATTGGTACA
   TAGAAATATA GATTCCATTA TGGGGATTCC TATATCATCG AGGAGAACTT
   CTAGTACATT CTGTATACCT AATATTATAG CCTTTATCAA CGATGGAATC
   CCAACAATTA TTTCAAAATT CACCCATTTC TCA 3′
```

Fig. 1. Location of the different primers on YOLC delta3. Underlined letters correspond to conserved zones [25]. Primers designed by Ness et al. [21] are written with italics, delta12 and delta21 primers are written with bold letters.

3. Results and discussion

3.1. Optimisation of the interdelta primers

The location of the two primers delta1 and delta2 given previously [21] on the delta element no. YOLC delta3 is indicated in Fig. 1.

We compared the homology of both delta1 and delta2 primers to the whole *S. cerevisiae* genome. Resulting Blast searches showed that primer delta1 (78 hits) gave poor homology. However, primer delta2 (303 hits) displayed a high consensus with the delta elements dispersed within the genome with an uneven matching close to the 5′ end. The low T_m of these primers [21] could be responsible for less stable patterns after amplification [24]. Comparison, using the sequence of delta element YOLC delta3, enabled the design of two new primers: delta12 and delta21. Primer delta12 is close to the primer delta1 in the well-conserved ATG region in TYA/GAG open reading frame [25]. Primer delta21 was chosen 60 bp upstream of primer delta2 in a 22-bp well-conserved region [25]. It has a higher T_m.

Data from the Blast searches made for each primer were then compared with the location of the 305 delta elements that we located on the *S. cerevisiae* genome (SGD, 2002). The different delta elements that are potential amplification start points are indicated for each chromosome in

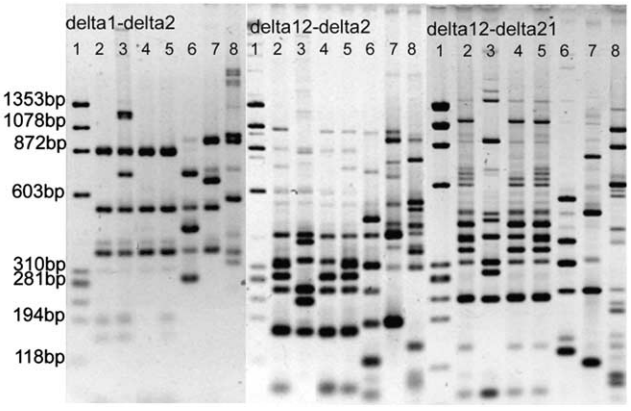


Fig. 2. Comparison of electrophoresis gels obtained for different yeast strains with delta1–delta2, delta12–delta2 and delta12–delta21. Lanes 1–10 of each gel: 1, molecular mass markers ϕX174 digested by HaeIII; 2, S288C; 3, FL100; 4, W303; 5, FY1689; 6, AWRI350; 7, AWRI750; 8, Clib319.

Table 2. The results showed the major improvement when primer delta12 was used instead of primer delta1. Indeed, primer delta1 gave only 34 hybridisation sites whereas delta12 gave 180 sites. However, primers delta2 and delta 21 showed similar efficiency.

Fig. 2 displays electrophoresis patterns obtained with different combinations of these four primers. As expected, profiles showing more bands were obtained with delta12–delta21 and delta12–delta2 primer pair combinations. Only three fragments were amplified with delta1–delta2 for strain S288C, whereas eight and 11 fragments were amplified with primer combinations delta12–delta2 and delta12–delta21. The whole gel including seven strains gave 15 or 27 or 28 different bands with primer pairs delta1–delta2, delta12–delta2, and delta12–delta21, respectively. Since a better association (T_m and amplification patterns) was achieved with primer pair delta12–delta21, we focused our study on this primer pair. However, the delta12–delta2 primer combination gave similar results.

3.2. Theoretical analysis of possible amplification products

The specificity of a method is a key point for its reproducibility, which is the drawback of RAPD techniques. We therefore checked if delta elements were the only source of amplification using primers delta12–delta21. We determined first the annealing sites of each primer by performing Blast searches on the S288C genome and then predicted the number of fragments that could be amplified.

Table 3 lists all the expected amplification fragments deduced from the Blast data. A great number of bands were predicted from 70 bp to 4 kb or higher; however, we chose to focus only on fragments <1.5 kb. T_m was determined for each homologous sequence found for each primer by Blast in order to assess the likelihood of amplification. Therefore, theoretical fragments obtained from sites with low homology (T_m lower than 25°C) were excluded. This reduced the number of virtual bands to 25 (Table 3); 24 of the virtual bands arose between two delta elements and the last additional virtual band was instead due to the presence of a partially homologous region adjacent to a bonafide delta element. This virtual profile also shows an uneven representation of chromosomes: two-thirds of the possible fragments are obtained from chromosomes 3, 4, 7 and 16; coincidentally, no amplification was forecasted for chromosomes 1, 2, 6, 8, 9, 11, 12, 13, 14 even if a high number of delta elements are possible amplification start points (i.e. chromosome 12, see Table 2). This result shows that only a limited number of delta elements is involved in the amplification pattern.

Experimental data show that among the 25 expected fragments, only 11 intense bands are obtained for S288C (Fig. 2). Furthermore, these bands were purified, cloned, sequenced and compared with databank sequences. Every single one of the 11 determined sequences matched one of the 25 predicted fragments (Table 3, in bold letters). Among the 14 virtual bands that were not amplified, 10 had the same primer at both ends. It has been described that this situation leads to a 'PCR-suppressive effect' [26,27]. According to this theory, the fragments with two complementary ends will form a 'panhandle-like structure' during the PCR, thus preventing further annealing and amplification. It should, however, be noticed that the fragment at 620 bp arose from a region located between two delta12 primers; it might have been obtained because of

Table 2
Theoretical number of delta elements giving amplification[a]

Chromosome no.	Delta elements[b]	Primer[c]			
		Delta1	Delta2	Delta12	Delta21
1	8	2	5	5	4
2	19	3	10	10	11
3	17	2	11	11	10
4	34	4	26	25	25
5	26	3	11	12	12
6	10	1	2	5	5
7	32	2	19	19	18
8	15	0	5	7	6
9	6	0	4	2	3
10	20	0	12	9	10
11	12	1	6	6	5
12	27	2	16	20	19
13	21	1	14	14	12
14	9	2	7	8	8
15	25	2	15	14	11
16	24	9	16	13	13
Total	305	34	179	180	172

[a]Theoretical number of delta elements was determined by Blast searches as described in Section 2.
[b]Number of delta elements identified on *S. cerevisiae* genome.
[c]PCR primers are described in Section 2.

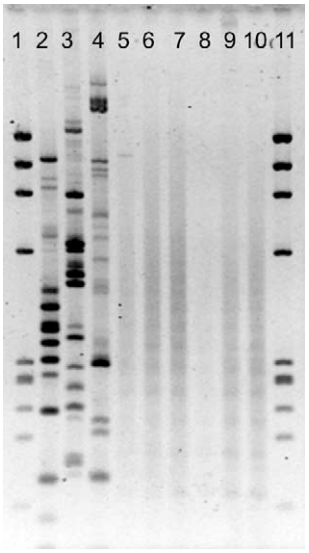


Fig. 3. Electrophoretical patterns obtained for different yeasts with delta12–delta21 primers. Lanes 1–11 of each gel: 1, molecular mass markers ϕX174 digested by HaeIII; 2,3, *S. cerevisiae* S288C and Clib227; 4, *S. carlsbergensis* Clib176; 5, *S. exiguus* Clib179; 6, *S. bayanus* Clib181; 7, *S. kluyveri* Clib182; 8, *S. servazzii* Clib187; 9, *S. paradoxus* Clib228; 10, *S. uvarum* D24; 11, molecular mass markers.

the simultaneous amplification of a 385-bp fragment at the same site that had primed the extension of the DNA.

Two forecasted fragments with asymmetrical primers at 1385 and 1159 were only weakly detected. Several attempts were made to increase the amplification level of these faint bands, without noticeable success. One explanation for this could be that amplification from multiple sites could reduce the level of individual signals. Nevertheless, this hypothesis is not valid for the three bands at 81, 140 and 426 bp obtained for chromosome 16 (Fig. 2).

In conclusion, all of the detected bands corresponded to one calculated possibility, and the differences between the virtual profile and the electrophoretical pattern can be explained mainly by a PCR-suppressive effect.

3.3. *Evaluation of the technique with strains of the* Saccharomyces *genera*

Sequences homologous to parts of TY retrotransposons have been found in various other yeast genomes [28], but the level of homology varies greatly among the species: The 'Genolevure Program' revealed that *S. bayanus* var. *uvarum*, *S. exiguus* and *S. kluyveri* show significant homology with *S. cerevisiae* TY1 or TY2, and *S. servazzii* show little homology. Fig. 3 presents amplification profiles obtained for these species. Only *S. cerevisiae* and *S. carlsbergensis* strains gave rich profiles, whereas other yeasts gave no signal. Amplification obtained from *S. carlsbergensis* is consistent with the hybrid nature of that species described elsewhere [29].

Table 3
Theoretical fragments and size obtained from Blast analysis of each primer[a]

Chromosome number	5′ Primer	5′ Coordinate	T_m[b]	3′ Primer	3′ Coordinate	T_m[b]	Fragment size
3	D12	83 100	41.6	D12	82 989	41.6	115
3	D21	84 366	28.6	D12	82 989	41.6	1385
3	D21	84 140	28.6	D12	82 989	41.6	1159
3	**D21**	**84 882**	**28.6**	**D12**	**84 584**	**29.3**	**310**
3	D21	169 276	40.5	D21	168 854	26.3	430
4	**D12**	**513 724**	**29.3**	**D21**	**513 507**	**40.5**	**225**
4	D21	520 543	28.6	D21	519 569	28.6	986
4	**D12**	**651 707**	**42.9**	**D21**	**651 340**	**28.6**	**376**
4	**D12**	**878 341**	**41.6**	**D12**	**877 728**	**35.7**	**620**
4	**D12**	**878 341**	**41.6**	**D21**	**877 964**	**28.6**	**385**
4	D12	987 187	41.6	D12	987 077	35.7	117
4	D21	992 674	40.5	D21	992 561	28.6	120
4	D12	1 151 385	35.7	D12	1 151 315	35.7	80
5	D21	449 367	28.6	D21	449 244	28.6	135
7	**D21**	**535 831**	**40.5**	**D12**	**535 490**	**41.6**	**344**
7	**D12**	**568 778**	**35.7**	**D21**	**567 682**	**28.6**	**1107**
7	D12	817 433	41.6	D12	817 358	29.3	84
10	D21	203 839	40.5	D21	203 743	28.6	103
15	**D21**	**664 884**	**28.6**	**D12**	**664 429**	**34.7**	**465**
15	D12	976 290	34.0	D12	976 193	35.7	108
16	**D21**	**62 461**	**28.6**	**D12**	**62 329**	**41.6**	**140**
16	**D21**	**62 747**	**28.6**	**D12**	**62 677**	**35.7**	**81**
16	**D21**	**62 747**	**28.6**	**D12**	**62 329**	**41.6**	**426**
16	D12	804 686	41.6	D12	804 609	40.6	82
16	D12	850 670	41.6	D12	850 289	41.6	385

Lines in bold letters correspond to bands confirmed after purification and sequencing.
[a]Fragments are determined by Blast searches as described in Section 2.
[b]T_m are calculated as described in Section 2.

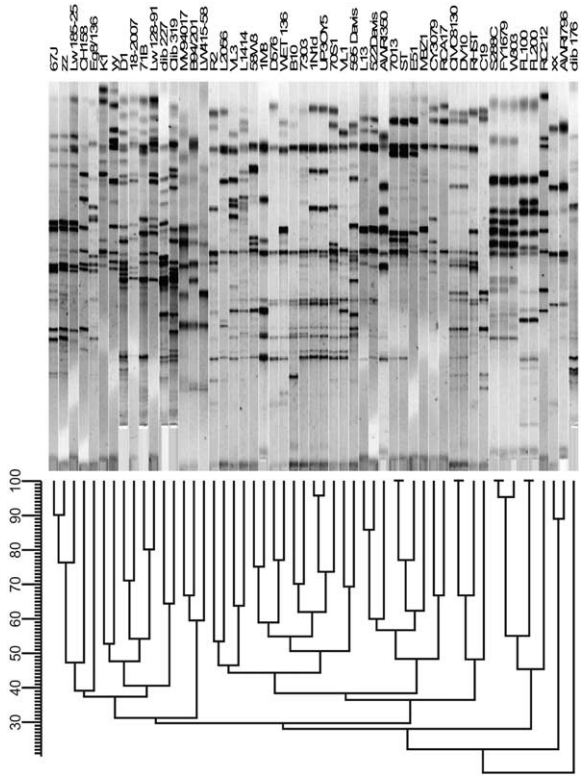


Fig. 4. Electrophoretical patterns and dendrogram showing the clustering of the 53 yeast strains used in the study. Calculated percentages of homology are given on the ruler on the left of the dendrogram.

3.4. *Characterisation of different industrial laboratory and wild-type yeast strains*

As the final aim of the technique is to differentiate the highest number of strains, 53 preparations of industrial yeast, laboratory strains or wild-type isolates were evaluated (Fig. 4).

Numerical data processing of the obtained profiles led to a rapid and easy classification. As expected, *S. carlsbergensis* strain Clib176 gave a pattern differing from the other yeast strains.

Some clustering can also be easily noticed: laboratory strains FL100, S288C, W303, and FY1679 are gathered in a separate branch of the tree, which can easily be understood from the history of laboratory strains. Indeed, these strains arose from a limited set of laboratories and have been extensively crossed and genetically engineered in order to obtain stable mating types or multiple auxotrophic markers. Some strains gave identical patterns DV10 – CIVC8130, 7013-ST, or UP3OY5 isolated from Burgundy in 1987 and 1N1d isolated in 1996 from Alsatian grapes. The identities of those profiles were confirmed by pulsed-field electrophoresis (data not shown). Furthermore, some close profiles, L1414 and B94/201 or CIVC8130 and C19, were also confirmed by pulsed-field electrophoresis, showing the coherence of both techniques. Different samples of DNA from the same yeast (laboratory strain culture, dried yeast) were analysed and showed identical profiles, demonstrating the reliability of this technique.

Comparable work performed on this yeast set with the former version of primer (delta1–delta2) gave poor clustering and led to numerous erroneous associations (data not shown).

4. Conclusion

The *S. cerevisiae* genome sequence was used to design two new primers for interdelta typing. One primer, delta21, shows similar or slightly better results on gels than the delta2 primer proposed formerly [21], but the second primer, delta12, is far more efficient than delta1. The specificity of this method was checked by comparison of the electrophoretic pattern of S288C with the virtual profile calculated from Blast data. With primer pair delta12–delta21, we were able to differentiate unequivocally between 53 industrial, laboratory and wild-type yeast strains. Results are in accordance with data acquired from pulsed-field electrophoresis, thus highlighting the importance of re-analysing yeast strains identified by the sole use of the former delta typing method [21].

Acknowledgements

The authors would like to thank Luc Baudrey for his helpful technical assistance, Dr C. Jackson and Dr Chan Ka Nin for critical discussion of the manuscript.

References

[1] Darriet, P., Lavigne, V. and Dubourdieu, D. (1991) Caractérisation de l'arôme variétal des vins de sauvignon par couplage CPG odométrie. J. Int. Sci. Vigne Vin 25, 167–174.
[2] Chatonnet, P., Dubourdieu, D., Boidron, J.N. and Lavigne, V. (1993) Synthesis of volatile phenols by *Saccharomyces cerevisiae* in wines. J. Sci. Food Agric. 62, 191–202.
[3] Degre, R., Thomas, D.Y., Ash, J., Mailhiot, K., Morin, A. and Dubord, C. (1989) Wine yeast strain identification. Am. J. Enol. Vitic. 40, 309–315.
[4] Pretorius, I.S. and Van Der Westhuizen, T.J. (1991) The impact of yeast genetics recombinant DNA technology on the wine industry. S. Afr. J. Enol. Vitic. 12, 3–31.
[5] Hallet, J.N., Craneguy, B., Zucca, J. and Poulard, A. (1988) Caractérisation de différentes souches industrielles de levures oenologiques par les profils de restriction de leur ADN mitochondrial. Prog. Agric. Vitic. 105, 328–333.
[6] Aizac, T., Delteil, D. and Vezinhet, F. (1991) Etude écologique de souches contaminantes. Application de la technique d'identification par caryotypes. Rev. Fr. Oenol. 130, 16–20.
[7] Frezier, V. and Duburdieu, D. (1992) Ecology of yeast strain *Saccharomyces cerevisiae* during spontaneous fermentation in a Bordeaux winery. Am. J. Enol. Vitic. 43, 375–380.
[8] Vezinhet, F., Hallet, J.N., Valade, M. and Poulard, A.. (1992) Ecological survey of wine yeast strains by molecular methods of identification. Am. J. Enol. Vitic. 43, 83–86.

[9] Versavaud, A., Dulau, L. and Hallet, J.N. (1993) Etude écologique de la microflore levurienne spontanée du vignoble des Charentes et approche moléculaire de la diversité infraspécifique chez *Saccharomyces cerevisiae*. Rev. Fr. Oenol. 142, 20–28.
[10] Querol, A., Barrio, E., Huerta, T. and Ramon, D. (1992) Molecular monitoring of wine fermentations conducted by active dry yeast strains. Appl. Environ. Microbiol. 58, 2948–2953.
[11] Aigle, M., Erbs, D. and Moll, M. (1984) Some molecular structures in the genome of lager brewing yeast. Am. Soc. Brew. Chem. 42, 1–7.
[12] Blondin, B. and Vezinhet, F. (1988) Identification de souches de levures oenologiques par leurs caryotypes en électrophorèse en champ pulsé. Rev. Fr. Oenol. 115, 7–11.
[13] Vezinhet, F., Dulau, L. and Hallet, J.N. (1994) Comparaison de différentes méthodes d'identification moléculaire de levures d'intérêt oenologique. Rev. Fr. Oenol. 149, 13–16.
[14] Grando, M.S., Ubeda, J. and Briones, A.I. (1994) RAPD analysis of wine *Saccharomyces* strains differentiated by pulsed field gel electrophoresis. Biotechnol. Tech. 8, 557–560.
[15] Quesada, M.P. and Cenis, J.L. (1995) Use of random amplified polymorphic DNA (RAPD-PCR) in the characterization of wine yeasts. Am. J. Enol. Vitic. 46, 204–208.
[16] De Barros Lopez, M., Soden, A., Henschke, P. and Langridge, P. (1996) PCR differentiation of commercial yeast strains using intron splice site primers. Appl. Environ. Microbiol. 62, 4514–4520.
[17] Hennequin, C., Thierery, A., Richard, G.F., Lecointre, G., Nguyen, H.-V., Gaillardin, C. and Dujon, B. (2001) Microsatellite typing as a new tool for identification of *Saccharomyces cerevisiae* strains. J. Clin. Microbiol. 39, 551–559.
[18] Gallego, F.J., Perez, G., Martinez, I. and Hidalgo, P. (1998) Microsatellites obtained from database sequences are useful to characterize *Saccharomyces cerevisiae*. Am. J. Enol. Vitic. 49, 350–351.
[19] Perez, M.A., Gallego, F.J. and Hidalgo, P. (2001) Evaluation of molecular techniques for the genetic characterization of *Saccharomyces cerevisiae* strains. FEMS Microbiol. Lett. 205, 375–378.
[20] Azumi, M. and Goto-Yamamoto, N. (2001) AFLP analysis of type strains and laboratory and industrial strains of *Saccharomyces* sensu stricto and its application to phenetic clustering. Yeast 18, 1145–1154.
[21] Ness, F., Lavallee, F., Dubourdieu, D., Aigle, M. and Dulau, L. (1993) Identification of yeast strains using the polymerase chain reaction. J. Sci. Food Agric. 62, 89–94.
[22] Lavallee, F., Salvas, Y., Lamy, S., Thomas, D.Y., Degre, R. and Dulau, L. (1994) PCR and DNA fingerprinting used as quality control in the production of wine yeast strains. Am. J. Enol. Vitic. 45, 86–91.
[23] Sambrook, J. and Russel, D. (2001) Molecular Cloning, 3rd edn., vol. 2, pp. 8–15. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY.
[24] Fernandez-Espinar, M.T., Lopes, V., Ramon, D. and Querol, A. (2001) Study of the authenticity of commercial wine yeast strains by molecular techniques. Int. J. Food Microbiol. 70, 1–10.
[25] Boeke, J.D. (l989) Transposable elements in *Saccharomyces cerevisiae*. In: Mobile DNA (Berg, D.E. and Howe, M.M., Eds.), Chapter l3, pp. 335–374. American Society for Microbiology, Washington, DC.
[26] Siebert, P.D., Chenchik, A., Kellog, D.E., Lukyanov, K.A. and Lukyanov, S.A. (1995) An improved PCR method for walking in uncloned genomic DNA. Nucleic Acids Res. 23, 1087–1088.
[27] Lecerf, F., Foggia, L., Mulsant, P., Bonnet, A. and Hatey, F. (2001) A novel method to isolate the common fraction of two DNA samples: hybrid specific amplification (HSA). Nucleic Acids Res. 29, e87/1–e87/5.
[28] Souciet, J.L., Aigle, M., Artiguenave, F. and Blandin, G. et al. (2000) Genomic Exploration of the Hemiascomycetetous yeasts: 1. A set of yeast species for molecular evolution studies. FEBS Lett. 487, 3–12.
[29] Kielland-Brandt, M.T., Nilsson-Tilgren, T., Gjermansen, C., Holmberg, S. and Pedersen, M.B. (1995) Genetics of brewing yeasts. In: The Yeasts (Rose, A.H., Wheals, A.E. and Harisson, J.S., Eds.), vol. 6, pp. 223–254. Academic Press, London.

Molecular Ecology (2007) **16**, 2091–2102 doi: 10.1111/j.1365-294X.2007.03266.x

Bread, beer and wine: *Saccharomyces cerevisiae* diversity reflects human history

JEAN-LUC LEGRAS,* DIDIER MERDINOGLU,* JEAN-MARIE CORNUET† and FRANCIS KARST*
**INRA/ULP, UMR Santé de la Vigne et Qualité du Vin, 28 rue de Herrlisheim, BP 20507, 68021 Colmar Cedex, France, †INRA/ENSAM/CIRAD/IRD, UMR Centre de biologie et de gestion des populations, Campus International de Baillarguet CS 30016, 34988 Montferrier-sur- Lez Cedex France*

Abstract

Fermented beverages and foods have played a significant role in most societies worldwide for millennia. To better understand how the yeast species *Saccharomyces cerevisiae*, the main fermenting agent, evolved along this historical and expansion process, we analysed the genetic diversity among 651 strains from 56 different geographical origins, worldwide. Their genotyping at 12 microsatellite loci revealed 575 distinct genotypes organized in subgroups of yeast types, i.e. bread, beer, wine, sake. Some of these groups presented unexpected relatedness: Bread strains displayed a combination of alleles intermediate between beer and wine strains, and strains used for rice wine and sake were most closely related to beer and bread strains. However, up to 28% of genetic diversity between these technological groups was associated with geographical differences which suggests local domestications. Focusing on wine yeasts, a group of Lebanese strains were basal in an F_{ST} tree, suggesting a Mesopotamia-based origin of most wine strains. In Europe, migration of wine strains occurred through the Danube Valley, and around the Mediterranean Sea. An approximate Bayesian computation approach suggested a postglacial divergence (most probable period 10 000–12 000 BP). As our results suggest intimate association between man and wine yeast across centuries, we hypothesize that yeast followed man and vine migrations as a commensal member of grapevine flora.

Keywords: domestication, fermentation, microsatellite, population genetics, *Saccharomyces cerevisiae*, wine

Received 13 July 2006; revision received 22 October 2006; accepted 11 December 2006

Introduction

In most societies, fermented beverages and foods have a unique place because of their economical and cultural importance and the development of fermentation technologies is deeply rooted in their history. Archaeologists have found evidence for the production of a fermented beverage in China at 7000 BC (McGovern *et al*. 2004), and of wine in Iran and Egypt at 6000 BC and 3000 BC, respectively (McGovern *et al*. 1997; Cavalieri *et al*. 2003). Since that time, it is believed that these fermentation technologies expanded from Mesopotamia through the world. For example, the cultivation of grapevine and the production of wine has spread all over the Mediterranean Sea towards Greece (2000 BC), Italy (1000 BC), Northern Europe (100 AD) and America (1500 AD) (Pretorius 2000). Beer technology is supposed to be almost as ancient as wine and was acquired from the Middle East by Germanic and Celtic tribes around 1st century AD, whereas lager beer technology appeared more recently in the 16th century. While the transfer of plants by man has favoured pathogen migrations (Galet 1977), the consequences of the spreading out of fermentation technologies on yeast diversity and population structure has never been investigated.

In addition, the question of the natural environment for *Saccharomyces cerevisiae* is still controversial. Because strain isolation from nature or plants is rare (Davenport 1974; Rosini *et al*. 1982; Sniegowski *et al*. 2002), Martini (1993) concluded that wine yeast comes mainly from cellars and described this species as domesticated. Very recently, Fay & Benavides (2005) observed a low diversity among wine yeast as a further argument for domestication and estimated a 2700-year-old divergence within the vineyard yeast

Correspondence: Legras Jean-Luc, Fax: 33 389224989; E-mail: legras@colmar.inra.fr

group. The same question of origin can also be raised for other yeast strains such as those used for ale beer or bread. The rise of the industrialization era for beer, bread or wine making should have led to a standardization of yeast flora. However, the numerous works made on the microflora diversity of wine (Frezier & Dubourdieu 1992; Querol *et al*. 1994; Versavaud *et al*. 1995; …), bread (Pulvirenti *et al*. 2001) and others have revealed a fascinating genetic diversity of *S. cerevisiae* strains. Surprisingly, despite its status as a model species whose genome sequence has been unravelled, no large-scale diversity study of the yeast species *S. cerevisiae* has been performed, and the role of man on this diversity is still unclear.

The biological and genetic characteristics of *S. cerevisiae* have been recently reviewed by Landry *et al*. (2006). Briefly, *S. cerevisiae* is a diplontic yeast with highly clonal reproduction. *S. cerevisiae* is also homothallic, which confers the possibility of regenerating a diploid cell from a haploid, and could be interpreted as a way of genome renewal (Mortimer *et al*. 1994). This mechanism could be responsible for the high rate (28%) of homozygote strains found in vineyards (Mortimer *et al*. 1994). Many studies also pointed out the aneuploidy of wine (Bakalinsky & Snow 1990; Guijo *et al*. 1997; Nadal *et al*. 1999), beer or bread strains (Codon *et al*. 1998). This could be a way for yeast to adapt to the various environments by modifying the dosage of some genes important in adaptation (Bakalinsky & Snow 1990; Salmon 1997). In addition, a high level of karyotype polymorphism has been observed, especially for wine yeast, resulting from various mechanisms such as mitotic or ectopic recombination (Nadal *et al*. 1999; Puig *et al*. 2000) mediated by Ty transposons or other repetitive sequences (Ness & Aigle 1995). As these mechanisms are very likely responsible for a variable sporulation ability and spore viability (Querol *et al*. 2003) the evolutionary importance of mating in yeast is indeed a matter of controversy.

We propose here to investigate the possible effects of human history on yeast diversity from a large-scale evaluation of yeast populations. For that purpose, we characterized 651 strains originating from 56 distinct sources using 12 microsatellite loci, and quantified the genetic differentiation between the most significant origins of yeast strains. We infer possible phylogenetic relationships and further evaluate the effects of major factors acting on this diversity: geographical isolation, and sexual reproduction. Our results give new insights into yeast genetic diversity and the role of man in spreading and selecting this fungus through history.

Material and methods

Strains

Yeast strains were obtained from our own yeast collection, and from several laboratories and yeast public or private collections (Table S1, Supplementary material). Most of them were formerly described as *Saccharomyces cerevisiae*. When no published data was available, species identification was checked by ITS restriction with *Hae*III (White *et al*. 1990). Amplifications at all microsatellite loci were only obtained with *S. cerevisiae*. For *Saccharomyces paradoxus*, we obtained only amplification at loci SCAAT5 and YKL172w. Hybrids were not searched in this work, but previous results showed that only *S. cerevisiae* alleles are detected from *Saccharomyces uvarum* × *S. cerevisiae* or *Saccharomyces kudriavzevii* × *S. cerevisiae* hybrids and should not interfere with the analysis.

The origins of the 651 strains used here are shown in Table 1. They were isolated from different substrates (wine, beer, bread, sake, palm wine, rum …). Strains were maintained in frozen stocks (glycerol, 15% v/v) at –80 °C, or for short-term storage on YPD agar medium (yeast extract, 1% w/v, peptone, 1% w/v and glucose, 2% w/v) at +4 °C.

Microsatellite characterization

Yeast cell cultures and DNA extraction were performed as previously described (Legras *et al*. 2005). Some samples were directly analysed from the DNA kindly provided by the contacted laboratory. The 12 loci used in this study have been described elsewhere (Legras *et al*. 2005). Two multiplex of six primers pairs corresponding to loci C5, C3, C8, C11, C9, SCYOR267c and YKL172w, ScAAT1, C4, SCAAT5, C6, YPL009c, were amplified using the QIAGEN multiplex (polymerase chain reaction) PCR kit according to the manufacturer's instructions (Table S2, Supplementary material). PCRs were run in a final volume of 12.5 µL containing 10–250 ng yeast DNA. Amplification was carried out using an Stratagene (Amsterdam, The Netherlands) thermalcycler with a three-phase temperature program: phase one, 1 cycle: 95 °C for 15 min; phase two, 34 cycles: 94 °C for 30 s, 57 °C for 2 min, 72 °C for 1 min; phase three, 1 cycle: 60 °C for 30 min.

PCR product analysis

PCR products were sized for 12 microsatellite loci on a capillary DNA sequencer (ABI 310) using the polyacrylamide Pop4 and the size standards HD400ROX. For rare fragments larger than 400 nucleotides, three DNA fragments of 420, 450 and 485 bp amplified from phage M13 were added to the sample. Before the analysis, the PCR amplicons were first diluted 50 fold and then 1 µL of the dilution was added to 18.75 µL of formamide (Applied Biosystem) and 0.25 µL of HD400ROX size marker, and the mixture was denaturated at 92 °C for 3 min. Allele distribution into classes was carried out using GENOTYPER 2.5 software (Applied Biosystems).

Table 1 Strain origin and grouping. The number of strains in the population used for F_{ST} tree is lower as the first group may include several time the same genotype, or because only strains from a well defined geographical origin are retained for further analysis (i.e. Tarragona strains among Spanish strains, or Firenze strains among Italians)

Origin	Number of analysed strains	'Population' for F_{ST} tree	No. of strains in the population
Alpechin, Spain	1		
Ale beer miscellaneous (France, Belgium, England The Netherlands …)	8	Ale beer	8
Bread, Italy Sicily	20	Bread, Italy, Sicily	19
Bread, miscellaneous (France, Japan, Spain …)	9	Bread miscellaneous	9
Cassava and banana, Burundi	2		
Cheese, France Camembert	2	Fermented milk	14
Cheese, France Cantal	12	Fermented milk	14
Cider, France Brittany	8		
Distillery, Australia	1		
Distillery, Brazil	8	Distillery Brazil	8
Distillery, China	8	Distillery, China	7
Fermented milk, Morocco	1	Fermented milk	14
Fruit, Indonesia	1		
Grapes (*Vitis amurensis*), Russia	1		
Laboratory strains (USA, France …)	8		
Lager beer miscellaneous (France, China, USA …)	15		
Miscellaneous, Japan	1		
Natural resources, Vietnam	5		
Oak exudates, USA	2		
Palm wine, Ivory Coast	1		
Palm wine, Nigeria	20	Palm wine (Nigeria)	19
Rice wine, China miscellaneous	6	Rice wine	10
Rice wine, Laos	3	Rice wine	10
Rice wine, Thailand	1	Rice wine	10
Rum, France French Indies	15	Rum French Indies	15
Sake, Japan	14	Sake (Japan)	11
Sorghum beer, Ghana	4		
Trout guts, Norway	1		
Type strain CBS1907 (Italy)	1		
Wine and fruits, Turkey	7		
Wine, Australia	4		
Wine, Austria	17	Wine Austria	13
Wine, Croatia	5		
Wine, France Alsace	100	Wine France Alsace,	71
		Wine France Alsace 'Central Europe group'	14
Wine, France Beaujolais	3		
Wine, France Bordeaux	12	Wine France Bordeaux	9
Wine, France Burgundy	17	Wine France Burgundy	16
Wine, France Champagne	2		
Wine, France Cognac	27	France Cognac wine	27
Wine, France Jura	3		
Wine, France Montpellier	20	Wine France, Montpellier	19
Wine, France Nantes	22	Wine France, Nantes	19
Wine, France Rhone valley	23	Wine France, Rhone valley	21
Wine, Germany	13	Wine Germany (Geisenheim)	11
Wine, Hungary	9	Wine Hungary	9
Wine, India	1		
Wine, Italy (Firenze and misc.)	35	Wine Italy, Florence	18
Wine, Japan	3		
Wine, Lebanon	25	Wine Lebanon	24
Wine, miscellaneous industrial strains	23		
Wine, Portugal	1		
Wine, Romania	10	Wine Romania	10
Wine, South Africa	25	Wine South Africa (Cap)	19
Wine, Spain (Tarragona, Penedes, and miscellaneous)	37	Wine Spain (Tarragona)	18
Wine, Uruguay	1		
Wine, USA	27	Wine USA (California)	16
Total	651		502

Statistical analysis

Strain groups. Strain groups were made from strains isolated in the same type of fermentations, and we tried to obtain groups as large as possible (at least nine for wine yeast). Genotypes were included only once in a group (Table 1). For wine, groups were chosen from a well-defined area. We divided among different regions only when we had enough strains. These groups were hereafter considered as 'populations'. Because of their characteristics, a group of strains called 'Alsace Central Europe strains' has been separated from other strains.

Genetic distances and population analysis. The chord distance Dc (Cavalli-Sforza & Edwards 1967) was calculated between each strain with a laboratory-made program. All trees were obtained from distance matrices derived with NEIGHBOUR of the PHYLIP 3.5 package, using MEGA3 (Kumar *et al*. 2004) for tree-drawing. All trees were rooted by the midpoint method. The reliability of the tree topologies was assayed through a jackknife procedure. The validity of nodes was obtained with the CONSENS program (PHYLIP 3.5 package).

Wine yeast population genetic features (F_{IS}, linkage disequilibrium) were evaluated from a subset of diploid wine yeast strains using the FSTAT 2.9.3 software (http://www.unil.ch/izea/softwares/fstat.html). Population genetic tests were also conducted from the 'GENEPOP on the web' (http://wbiomed.curtin.edu.au/genepop/).

The genetic distances F_{ST} (Reynolds *et al*. 1983), and DAS (Bowcock *et al*. 1994) between all groups were calculated using the program MICROSAT 1.5d (Minch *et al*. 1995) after pooling all alleles detected from one group of strains and considering strains as haploids. The reliability of the tree topologies was assayed through bootstrap analysis (1000 replicates resampling loci), and the validity of nodes was obtained with the CONSENS program. Isolation by distance was evaluated with the ISOLDE software from the 'GENEPOP on the web', after calculation of F_{ST} distances with MICROSAT between all groups. The geographical matrix was built with the help of route-finder software for close origins (such as the European wine groups) or from air distances for more distant countries. Cheese, beer and bread strains were not included in the analysis because of their unclear geographical origin, as well as South African and American wine strains because of human-driven migration.

Time divergence estimation. An estimation of wine yeast divergence was attempted through an approximate Bayesian computation (ABC) approach (Beaumont *et al*. 2002). This consists of three steps, namely: (i) simulation of data sets according to a demographic, historical and mutational model, with parameter values drawn from prior distributions; (ii) rejection of simulated data sets based on the Euclidian distance between standardized summary statistics of the observed data set and those of each simulated data set; and (iii) local linear regression of individual parameters on summary statistics of accepted data sets (see Excoffier *et al*. 2005 for a more detailed description).

The two successive divergences occurred at t_1 and t_2 years in the past. Prior distributions on divergence time were set at $t_1 = 2500$ (first traces of the culture of grapes when Greeks established Massalia (Marseille), Dion 1959) and $t_2 \sim U[3500, 53500]$. A generalized mutation model was assumed for microsatellite loci, with mean mutation rate drawn from a uniform $\cup[0.0001, 0.001]$. The analysis was performed twice: once with the populations *Lebanon*, *Montpellier* and *Central Europe*, and a second time replacing *Montpellier* by *Rhone valley*.

For this analysis, nine loci were retained. We removed Ykl172w locus, because of its specific behaviour (abnormally high F_{ST}, see below), YPL009c (linked to C9) and C4 (aneuploïdy of several strains).

Results

Genotypes and strain biodiversity

From 651 strains that were assigned genotypes at all 12 loci, a total of 575 multilocus genotypes were established, with 76 strains showing genotypes identical to others in the survey. Among grapevine and wine isolates, some clones isolated from various vineyards in several continents resulted in the same genotype. In many cases, these strains corresponded to well-known industrial strains such as the 522 Davies group (Fig. 1), or the CIVC8130 and 'Prise de mousse' group (Champagne group Fig. 1). This finding has already been described before (Legras *et al*. 2005) for the champagne strain CIVC8130. Several clones of sake strains obtained from different collections also displayed the same pattern.

The 12 microsatellite loci recorded from 13 to 54 different alleles per locus. SCAAT1, C4 and SCYOR267c displayed the highest number of alleles in the global population, which was expected given the length of these repeated motifs and their selection for high polymorphism (Legras *et al*. 2005). The number of alleles per locus per strain varied from one to four (Table 2). Half of the bread strains and ale beer strains exhibited four alleles at several loci, suggesting tetraploidy. In contrast to the results with bread strains, 88% of wine isolates presented two alleles maximum for all loci, suggesting a diploid state for most wine yeast strains. In total, 28% of the isolates were homozygous for all loci.

The neighbour-joining tree calculated from the Dc chord distance matrix for all pairs of strains (Fig. 1) reveals a clear clustering linked to the technological origin of the strains (Table 1). In particular, yeast strains used for palm wine, sake, fermented milk and cheese, beer (ale and lager), are

Fig. 1 Neighbour-joining tree showing the clustering of 651 yeast strains isolated from different sources. The tree was constructed from the chord distance between strains based on the polymorphism at 12 loci and is rooted according to the midpoint method. Branches are coloured according to the substrate from which strains have been isolated. The percentage of occurrence of nodes obtained through a Jackknife procedure is given at the basic node of main groups. Type strain CBS1171 is given in black. Color code: wine, dark green; cider, light green; bread, yellow; beer, orange; fermented milk, pink; sake from Japan, dark blue; Chinese rice wine and distillery from Vietnam and Thailand, light blue; sorghum beer or palm wine from Africa, brown; oak tree from America, blue-green; distillery from South America and rum from French Indies, purple; laboratory strains, red. Misc., miscellaneous.

Table 2 Classification of strains according to the maximum number of allele encountered per locus for some origins. (One allele maximum means homozygosity; two, heterozygosity, three suggest aneuploidy, and four, tetraploidy)

Origin	No. of different clones	1 allele	2 alleles	3 alleles	4 alleles
Vine and wine					
Austria — Klosterneuburg	17	4	10	2	1
France —Alsace	86	14	54	9	9
France — Cognac	27	14	12	1	0
France —Montpellier	19	9	10	0	0
France — Nantes	19	8	9	2	0
France — Rhône	21	1	16	3	1
Spain (Taragonna and Penedes)	36	9	23	3	1
Germany	11	2	8	0	1
Italy	23	6	15	1	1
Japan	3	1	1	1	
Lebanon	25	10	11	1	3
USA (UC Davis)	26	10	10	2	4
Cider (France, Bretagne)	8	2	6	0	0
Ale beer (miscellaneous)	8	0	1	2	5
Bread (miscellaneous) (group 1)	9	0	3	1	5
Fermented milk and cheese (France + Morocco)	14	0	7	2	5
Palm Wine (Nigeria)	19	4	6	5	4
Rum (Antilles)	15	3	8	4	0
Rice wine and distillery Asia	13	3	1	9	0
Sake Japan	11	2	7	2	0

gathered in clusters. Bread strains analysed here are divided into two clades; one main clade including strains from various countries (Japan, Spain, France, and Sicily) as well as industrial bread strains, and a second clade including only isolates from Sicily. The main group contains mostly tetraploid strains and is located at the border of the wine group. It contains also some wine yeast and amazingly four strains that are widely used for 'Beaujolais nouveau' wine-making.

Almost 95% of wine yeast strains are found in the upper part of the tree, which also includes cider strains. Industrial or grape strains are scattered all over this clade so that it is not possible to differentiate them from other wine strains. Several subgroups are visible in the wine yeast clade, the largest of which contains strains isolated in Germany, Alsace (France), Hungary, and Romania; we called it the 'Central Europe wine group'. This group also contains some strains from Lebanon, as well as Spanish flor yeast strains. The group containing the 'Prise de mousse' strain, further called 'Champagne group', is also related to the 'Central Europe' group. Most American vineyard strains belong to the wine yeast groups, except three isolates. Among the few strains not included in the wine part of the tree, four Austrian and two Californian strains, are found in a separate cluster and five other isolates are found separately. The American strain UCD13, is found close to two American oak litter isolates obtained by Sniegowski *et al*. in 2002. Similarly, one Russian vine strain (CBS 5287) isolated from grapes of wild endemic vine *Vitis amurensis* in the Russian Far East was not related to other wine strains. French Indies rum and Southern America distillery strains are found together either among wine yeast or dispersed among other origins.

The next main specific group is the fermented milk isolates group. These strains have been isolated on cheese mainly in France but also in Morocco. It is noteworthy that despite its French origin, this group is positioned far from French wine or cider isolates but close to the beer strains group. Interestingly, all beer strains (ale and lager) but one appear clearly different from bread strains. Despite their different genetic characteristics (tetraploid for ale strains, and allotetraploid for lager strains) these two types of beer strains are also found in the same clade.

The geographical effect on the structure of the individual tree can hardly be proven as some strains are specific to one type of fermentation such as African Palm wine yeast (Nigeria and Ivory Coast). However, for Asian strains, we have two clades that combine strains from different countries (China, Laos, Thailand, and Japan), isolated from different fermentations (rice wine, or distillery isolates) as well as a Vietnam sugar cane natural isolate suggesting domestication from a local origin.

Wine yeast population analysis

As most of our strains were obtained from vineyards, we tried to investigate the population structure of wine

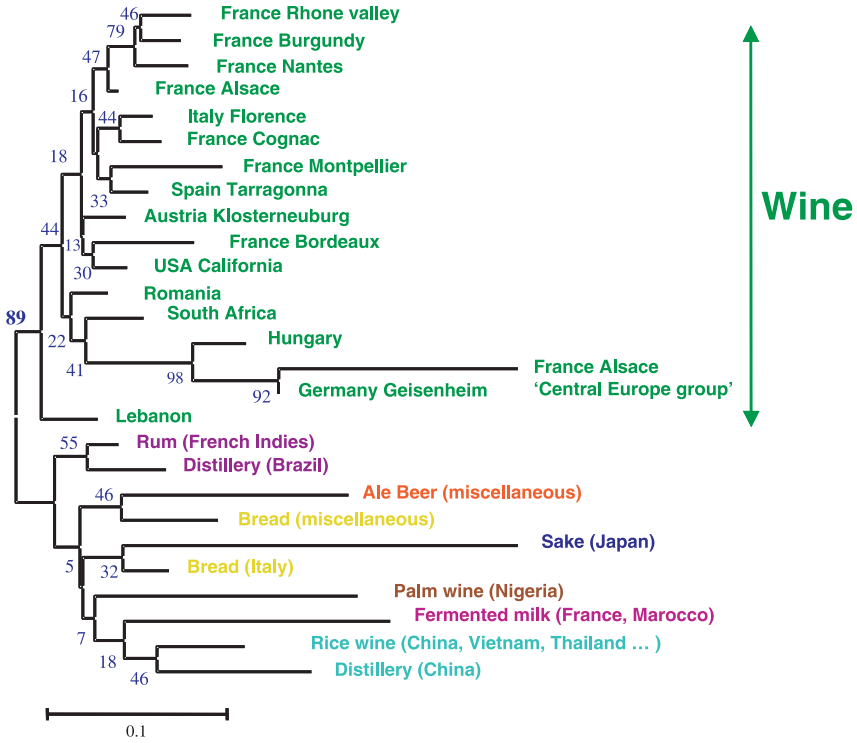


Fig. 2 Consensus tree of populations based on F_{ST} genetic distances obtained after 1000 replicates (resampling loci). The tree was built using the neighbour-joining method, and the root was defined by midpoint rooting.

S. cerevisiae diversity. The 17 most important groups were kept and considered then as pseudopopulations, corresponding to 277 diploid strains. The few aneuploids were not taken into account in the analysis. We first checked associations within loci to determine if genotype frequencies were those expected under Hardy–Weinberg equilibrium. The high number of homozygote strains observed for most groups turned out to be excessive when opposed to Hardy–Weinberg expectations.

In a second step, we tried to estimate how the genetic variation was partitioned within and between *populations* using *F*-statistics (Table S3, Supplementary material). One locus, YKL172w, presented an F_{ST} value three times higher than the other loci, that might be connected to the function of this essential gene so that we have not taken this loci into account for the F_{ST} analysis. The average of F_{IS} values over all loci is very high for most groups and can be explained by inbreeding and may be related here to the effect of homothallism on population genetic structure.

An exact test for association of alleles across loci based on permutation was employed. For wine yeast populations, linkage disequilibrium was not observed for 26% of comparisons among all populations ($P < 0.05$) and 77% inside each population ($P < 0.05$). This suggests that the overall population structure is clonal but with some recombination. Also this can result partly from genetic drift occurring among distantly related populations.

Population relationships

Two trees were built from the F_{ST} and the DAS matrix among groups (Fig. 2, Fig. S1 and Table S4, Supplementary material): both give a structure in agreement with the individual strain tree (Fig. 1).

All wine strain groups gather within the same clade (89% and 90% of F_{ST} and DAS trees obtained by bootstrap analysis) and are separated from other groups. This analysis reveals a clear difference between Mediterranean vineyards and the Central Europe branch (Romania, Hungary, Germany, and some strains of Alsace) that we correlate to the occurrence of strains of the 'Central Europe group'. The Lebanon group is found at the root of the wine group. A further structure can also be observed for Rhone valley–Burgundy and Nantes strains (79% bootstrap) to which is connected the Alsace group (47% bootstrap), and for Cognac and Italian strains or Montpellier and Spanish strains with a lower bootstrap score (44% and 33%, respectively).

Two groups of strains are found close to the wine yeast groups, the French Indies rum and Southern America distillery group suggesting that these groups are related to wine yeasts. In contrast, Asian strains (Chinese and sake) as well as beer strains have a distant position to wine strains.

The analysis of the geographical effect on this yeast diversity was attempted for groups of strains for which we could identify a clear geographical origin: European wine, African palm wine, Southern America distillery strains and French Indies rum, sake, China rice wine and distillery.

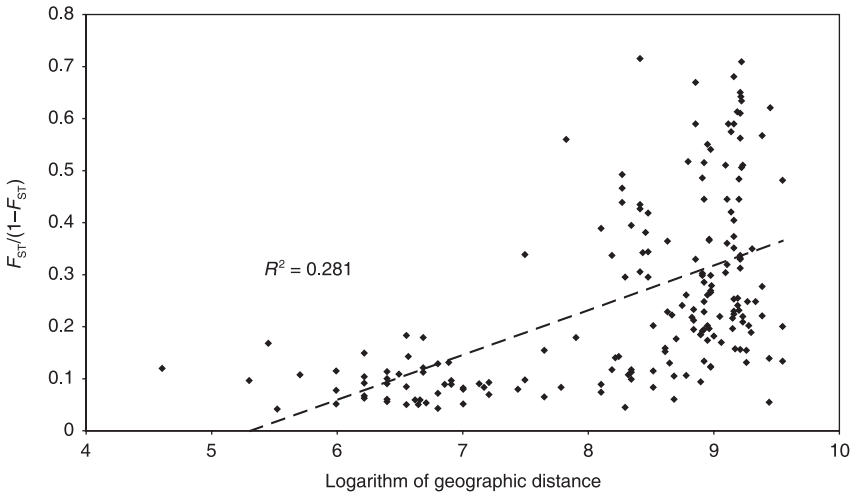


Fig. 3 Evaluation of isolation by distance. Distances between groups were calculated using the MICROSAT program and the isolation by distance evaluated with the software ISOLDE from the GENEPOP website (http://wbiomed.curtin.edu.au/genepop/).

Bread, beer, and cheese strains were discarded because of their nonspecific geographical origin. The F_{ST} distance matrix calculated from MICROSAT software had been tested and plotted against geographical distance. Permutation Mantel test calculated from GENEPOP revealed a global highly significant value ($P < 0.001$), and 28% of the variability could be explained by geographical isolation (Fig. 3).

Time divergence estimations

Assuming Lebanon as a putative origin for all three groups (*Central Europe*, *Montpellier* and *Rhône valley*) and a 2500-year-old divergence with Rhone or Montpellier populations, we obtained two estimations of –10 500 and –11 750 (BP) (Fig. 4), respectively [confidence intervals (–4500, –32 000) and (–4750, –36 000)] between Central Europe and Lebanese yeast strains. As a consequence, yeast migration seems to have occurred after the last ice period.

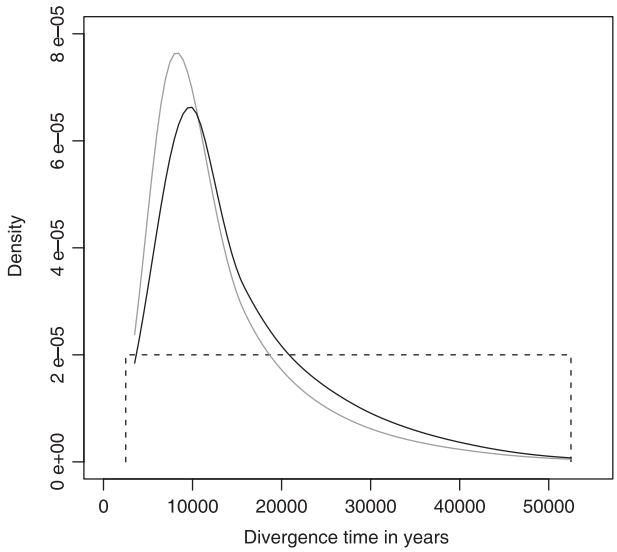


Fig. 4 Distribution of divergence time estimations between the Central Europe and Montpellier (black line) and the Central Europe and Rhone valley (blue line) populations. The dashed line corresponds to the a priori distribution of time divergence.

Discussion

We used both individual and population-based approaches to analyse *Saccharomyces cerevisiae* biodiversity. Multilocus microsatellite typing of strains of different origins revealed a strong structure of yeast strains according to their technological origins. This structuring can be clearly seen from individual as well as population analysis.

Population analysis of wine yeast

The ploidy of wine strains is a remarkable feature of the wine group: almost 84% of strains were deduced to be diploid. The population analysis of these diploid wine yeast groups revealed several original aspects. First of all, as expected, *S. cerevisiae* has a mainly clonal reproduction, as seen from linkage disequilibrium observed between loci. However, a significant proportion of loci are still nonsignificantly linked (26% among all populations and 75% within each group) which suggests that some recombination still exists. A purely clonal population evolving only under mutation would indeed lead to a rapid disappearance of homozygous strains, whereas a clonal population evolving under mitotic recombination would lead to a decrease in heterozygous strains. The high ratio of homozygous strains (30%) and the high F_{IS} positive values suggest that homothallism has a high impact on yeast diversity. A similar pattern has been observed by Fundyga *et al.* (2002) for *Candida albicans* but our results indicate that *S. cerevisiae* has a lower rate of sexual reproduction. Indeed, F_{IS} and linkage disequilibrium are

higher and there is a larger proportion of homozygous strains in our species.

Further parallels can be drawn with *C. albicans*, as we also detected a macrogeographical differentiation of strains between Asian, European and African yeast demes. Isolation by distance accounts for 28% of genetic variation (Fig. 3), which is slightly lower as the 39% estimated for *C. albicans* (Fundyga *et al*. 2002). These genetic differences between yeast groups suggest an ancient divergence leading to local natural populations (in Asia, Mesopotamia, Africa) from which multiplication may have been favoured by humans. This implies also that there must be a natural habitat for yeast which allowed a wide expansion of this species.

Population relationships inferred from individual and population trees

The F_{ST} and DAS consensus trees on populations confirm the global genetic structure of the tree on individuals, with a clear separation between most wine yeast groups and other technological groups (89% bootstrap score for the F_{ST} tree).

This structuring has partially been observed with amplified fragment length polymorphism (AFLP; Azumi & Goto-Yamamoto 2001), microsatellites (Hennequin *et al*. 2001), and very recently by multilocus sequence analysis (MLST) for wine and sake origins by Fay & Benavides (2005) and Ayoub *et al*. (2006) or single nucleotide polymorphism (Ben-Ari *et al*. 2005).

The reference laboratory strain S288C obtained from an American isolate is found very close to Nigerian palm wine strains (Fig. 1), whereas three other American isolates (two oak tree exudates and one Californian wine isolate UCD13) are much closer to CBS 5287 (Asian Russia) and Clib 414 from Japan. This position far from wine yeast is in agreement with data of De Barros Lopes *et al*. (1999) from AFLP, Fay & Benavides (2005) from MLST, and Winzeler *et al*. (2003) from micro-array karyotyping. It must be pointed out, however, that we did not characterize any wine yeast isolate close to S288C as described by Aa *et al*. (2006).

Three main Asian yeast groups of strains were also found: the sake yeast group and two groups including mainly rice wine and some Chinese distillery strains. Sake strains and the two other groups are surprisingly not as close to each other in the individuals' tree as would be expected from sake technology having originated from Korea (Teramoto *et al*. 1993). But the DAS tree suggests that these groups are related and the global position of Asian strains is in agreement with what was described by Azumi & Goto-Yamamoto (2001), Fay & Benavides (2005), and by Ayoub *et al*. (2006). The Nigerian palm wine group represents another well-characterized group including an Ivory Coast strain. However, Ghana sorghum beer strains are closer to beer strains than to palm wine. They are also distinct from Burundi fermented cassava and banana strains, which suggests that genetic differentiation between African yeast populations exists in a similar way to what has been described for wine or bread yeasts.

The wine yeast group is well separated from yeast strains of other technological origins. It contains strain groups from ancient vine areas (Lebanon, Europe) as well as 'new world' recent vineyards, which suggests a migration of wine yeast all over the world, which is revealed by the structure of the F_{ST} tree. In addition to the historical human transport across the Mediterranean Sea, this tree clearly supports the hypothesis of a migration pathway along the Danube valley. The occurrence of some 'special' wine strains outside the main wine yeast group (UCD13, CBS 5287, and Arka) suggests that some autochthonous strains in 'new world' or European vineyards can still be isolated that do not represent the 'standard' wine strains. In a similar manner, half of the strains of the groups of French Indies rum and Brazilian distillery strains found among wine strains are very likely the result of such a human-provoked migration whereas the second half are more distant.

For bread and beer, Azumi & Goto-Yamamoto (2001), and De Barros Lopes *et al*. (1999) found evidence from AFLP data of a close relatedness with wine strains. However, Ayoub *et al*. (2006) found contradictory results for type strain CBS1171 originating from beer. Our results from the individual tree suggest that bread strains are close to wine strains, and far from beer strains. In contrast, the F_{ST} and DAS trees indicate on the contrary that beer and bread (tetraploids) strains are related to each other but are distant from wine yeast. The analysis of the proportion of alleles shared by each group of strains may give a clue: 79% and 67% of alleles of bread strains (main group) are shared by beer and wine strains, respectively, and 96% of bread strains alleles can be found either among beer or wine strains. We propose that the main group of bread strains could originate from a tetraploidization event between an ale beer and a wine yeast strain, which may explain the discrepancy between published data. It must also be pointed out that actual *S. cerevisiae* strains used for making bread come from different geographical origins.

Genetic data vs. historical features

The relative positions of most yeast groups are in good agreement with historical data. The sake technology is supposed to have originated from Korean rice wine technology (Teramoto *et al*. 1993), wine tradition from Mesopotamia (Pretorius 2000), and lager beer from ale brewery knowledge (Corran 1975). The above suggestion that a group of bread strains resulted from a tetraploïdization

event between an ale beer strain and a wine strain also implies that bread technology appeared after beer and wine technology. However, as beer strains are obviously far from wine yeast, our results do not support the classical hypothesis of wine technology as an origin for beer (McGovern 2003). Our results suggest more likely that beer as well as bread have a more oriental origin, which is in agreement with several results of MLST analysis of bread (or bread-related wine strains such as 71B or Levuline Primeur) (Fay & Benavides 2005; Ayoub *et al*. 2006). The position of some beer strains found among bread strains is also logical, considering the historical exchange of strains between beer and bread yeast makers since the end of the 19th century.

The existence of a wine yeast group including 95% of strains, with a Lebanon group close to the root of the F_{ST} tree (Fig. 2) suggest a migration from Mesopotamia with the event of vine domestication and is compatible with known vine migration (Arroyo-Garcia *et al*. 2006). But more strikingly, the substructure inside the wine yeast cluster is also in agreement with historical knowledge. The different wine yeast group locations are consistent with vine migrations routes (Danube valley, Rhone valley Burgundy Alsace and Nantes, or Italy–Cognac; This *et al*. 2006). Indeed, Ugni blanc, main Cognac grape varieties, originated from Italy, wine-making tradition arrived in Burgundy from the Rhône valley, and Muscadet (Nantes) was imported from Burgundy at the 15th century (Viala & Vermorel 1901). With a time divergence estimation of about 11 000 years (BP) between Lebanon and the Central Europe group, we can assess a very early divergence posterior to the last glaciations area but we cannot conclude whether this divergence is connected to a postglaciation colonization route of wild vine (Taberlet *et al*. 1998) or to the culture of vine. The most probable estimation suggest a period corresponding to the advent of wine making as the oldest archaeological site displaying remains of wine technology is 8000 BP. For some more distant yeast strains found in Austria, we can assume that some local strains have been domesticated from the wild local *Vitis sylvestris* which are the progenitors of the actual vine varieties (Levadoux 1956).

Tamed or domesticated yeast?

The question that must be raised is that of yeast domestication as proposed by Martini (1993) or Fay & Benavides (2005). The concept of a domesticated species is often used with different meanings. In a recent review about plant and animal domestication, Diamond (2002) proposed the following definition: 'species bred in captivity and thereby modified from its wild ancestors in ways making it more useful to humans who control its reproduction and (in the case of animals) its food supply'.

Because they have been almost continuously cultivated since very ancient times, rice wine, beer and bread strains are clearly fulfilling these criteria of culture and selection, so that we have at least two different domestication events which occurred in Asia for rice wine and somewhere else for beer.

The way in which wine strains are naturally propagated is, however, poorly understood: flor yeasts which grow almost continuously on the surface of wine during the sherry wine process are very likely an example of domestication. However, for other types of wine strains, we cannot infer such a continuous human control of their culture. Mortimer & Polsinelli (1999) have shown that a population of yeast exists on grapes. However, whether these strains participate in the alcoholic fermentation in the cellar is still controversial: some authors (Rosini *et al*. 1982; Ciani *et al*. 2004) observed that only cellar strains were responsible for the alcoholic fermentation in the vats, whereas others show that 'grapevine strains' can be partially responsible for the alcoholic fermentation (Constanti *et al*. 1997; Gutiérrez *et al*. 1999; Le Jeune *et al*. 2006). We agree with the latter point of view from our own data (Legras, unpublished data). The correlation between grapevine migration and yeast diversity, as well as our time divergence estimation, suggest clearly that wine yeast biology is closely connected to grapevine, which is compatible with the idea of *S. cerevisiae* as a potential pathogen of the vine (Gognies *et al*. 2001) or at least a member of the vine commensal flora. However, the adaptation of yeast observed through the evolution of the SSU1 gene leading to SO_2 resistance (Perez-Ortin *et al*. 2002; Aa *et al*. 2006) demonstrates the adaptation of yeast to the winery environment. Altogether, these results suggest that yeast have adapted to both vine and winery environment.

In conclusion, our data show that yeast genetic diversity has been highly influenced by human technology through history, as well as by natural genetic drift and migration in a similar way to other microorganisms and pathogens, leading to progressively differentiated populations.

For decades, studies on beverages history have relied on vessels comparison, and very recently on chemical analysis. Our results also show that a comprehensive exploration of yeast diversity could provide some new features about the origin of fermentation technology and human history. The ongoing programs using yeast population genomics could give us new clues to these questions.

Acknowledgements

The authors would like to thank A. Alais Perot and G. Butterlin for their helpful technical assistance. We are grateful to C. Schneider for answering vine history questions, and to Dr J. Masson and Pr. R. Gardner for critically reading the manuscript.

We also express gratitude to all the different researchers and Institutes who provided kindly yeast strains: Dr M.J. Ayoub,

University of Beyrouth, Lebanon, Dr S. Berger, Weinbau Institut Klosterneubourg, Austria; Pr. L. Bisson, UC Davis; USA; Dr M. Burcea, University of Bucarest, Roumania; Dr S. Colas, InterRhone, Avignon, France; Dr S. Dequin, UMRSPO INRA Montpellier; Dr J.F. Drilleau, SRC INRA Rennes, France; Dr H. Erten, Cukurova University, Turkey; Pr. O. Ezeronye, DBS MOUA, Umudile, Nigeria; Dr L. Fahrasmane, URTPV INRA Antilles, France; Pr. J. Fay, University of Washington, Saint Louis, USA; Dr N Goto-Yamamoto, NRIB, Japan, Dr L. Granchi, University of Firenze, Italy; Pr. Grossmann, State Research Institute Geisenheim, Germany; Dr I. Masneuf Pommade, Institut d'Oenologie de Bordeaux, France; Dr MC. Montel, URF, INRA Aurillac, France; Dr H.-V. Nguyen, CLIB, INRA INAPG, France; Dr V. Petravic; University of Zagreb, Croatia; Dr A. Poulard, ITV, Nantes, France; Dr C Roulland, BNIC, Cognac, France; Dr N. Rozes, Universitat Rovira i Virgili, Tarragona, Spain; Pr. M. Sipiczki, University of Debrecen, Debrecen, Hungary; Pr. P. Sniegowski, University of Chicago, USA; Pr. Van Rensburg, University of Stellenbosch, South Africa.

References

Aa E, Towsend JP, Adams RI, Nielsen KM, Taylor JW (2006) Population structure and gene evolution in *Saccharomyces cerevisiae*. *FEMS Yeast Research*, **6**, 702–715.

Arroyo-Garcia R, Ruiz-Garcia L, Bolling L *et al.* (2006) Multiple origins of cultivated grapevine (*Vitis vinifera* L. ssp. *Sativa*) based on chloroplast DNA polymorphism. *Molecular Ecology*, **15**, 3707–3714.

Ayoub MJ, Legras JL, Saliba R, Gaillardin C (2006) Application of multilocus sequence typing to the analysis of the biodiversity of indigenous *Saccharomyces cerevisiae* wine yeasts from Lebanon. *Journal of Applied Microbiology*, **100**, 699–711.

Azumi M, Goto-Yamamoto N (2001) AFLP analysis of type strains and laboratory and industrial strains of *Saccharomyces sensu stricto* and its application to phenetic clustering. *Yeast*, **18**, 1145–1154.

Bakalinsky AT, Snow R (1990) The chromosomal constitution of wine strains of *Saccharomyces cerevisiae*. *Yeast*, **6**, 367–382.

Beaumont MA, Zhang W, Balding DJ (2002) Approximate Bayesian computation in population genetics. *Genetics*, **162**, 2025–2035.

Ben-Ari G, Zenvirth D, Sherman A, Simchen G, Lavi U, Hillel J (2005) Application of SNPs for assessing biodiversity and phylogeny among yeast strains. *Heredity*, **95**, 493–450.

Bowcock AM, Ruiz-Linares A, Tomfohrde J, Minch E, Kidd JR, Cavalli-Sforza LL (1994) High resolution of human evolutionary trees with polymorphic microsatellites. *Nature*, **368**, 455–457.

Cavalieri D, McGovern PE, Hartl DL, Mortimer R, Polsinelli M (2003) Evidence for *S. cerevisiae* fermentation in ancient wine. *Journal of Molecular Evolution*, **57**, S226–S232.

Cavalli-Sforza LL, Edwards AWF (1967) Phylogenetic analysis: models and estimation procedures. *Evolution*, **21**, 550–570.

Ciani M, Mannazzu I, Marinangeli P, Clementi F, Martini A (2004) Contribution of winery-restent *Saccharomyces cerevisiae* to spontaneous grape must fermentations. *Antonie Van Leeuwenhoek*, **85**, 159–164.

Codon AC, Benitez T, Korhola M (1998) Chromosomal polyphormism and adaptation to specific industrial environments of *Saccharomyces* strains. *Applied Microbiology and Biotechnology*, **49**, 154–163.

Constanti M, Poblet M, Arola L, Mas A, Guillamon JM (1997) Analysis of yeast populations during alcoholic fermentation in a newly established winches. *American Journal for Enology and Viticulture*, **48**, 339–344.

Corran HS (1975) *A History of Brewing*. Newton Abbot, David & Charles, London.

Davenport RR (1974) Microecology of yeast and yeast-like organisms associated with an English vineyard. *Vitis*, **13**, 123–130.

De Barros Lopes M, Rainieri S, Henschke PA, Landridge P (1999) AFLP fingerprinting for analysis of yeast genetic variation. *International Journal of Systematic Bacteriology*, **49**, 915–924.

Diamond J (2002) Evolution, consequences and future of plant and animal domestication. *Nature*, **418**, 700–707.

Dion R (1959) *Histoire de la Vigne et du Vin en France*. Flammarion, France.

Excoffier L, Estoup A, Cornuet JM (2005) Bayesian analysis of an admixture model with mutations and arbitrarily linked markers. *Genetics*, **169**, 1727–1738.

Fay JC, Benavides JA (2005) Evidence for domesticated and wild populations of *Saccharomyces cerevisiae*. *Plos Genetics*, **1**, 66–71.

Frezier V, Dubourdieu D (1992) Ecology of yeast strain *Saccharomyces cerevisiae* during spontaneous fermentation in a Bordeaux winery. *American Journal of Enology and Viticulture*, **43**, 375–380.

Fundyga RE, Lott TJ, Arnold J (2002) Population structure of *Candida albicans*, a member of the human flora as determined by microsatellite loci. *Infection Genetics and Evolution*, **2**, 57–68.

Galet P (1977) *Les Maladies et les Parasites de la Vigne. Tome 1*. Imprimerie du Paysan du Midi, Montpellier.

Gognies S, Belarbi A, Barka EA (2001) *Saccharomyces cerevisiae* a potential pathogen towards grapevine, *Vitis vinifera*. FEMS. *Microbiology Ecology*, **37**, 143–150.

Guijo S, Mauricio JC, Salmon JM, Ortega JM (1997) Determination of the ploidy in different *Saccharomyces cerevisiae* strains used for fermentation and flor film ageing of dry sherry-type wine. *Yeast*, **13**, 101–117.

Gutiérrez AR, Santamaria P, Epifano S, Garijo P, Lopez R (1999) Ecology of spontaneous fermentation in one winery during consecutive years. *Letters in Applied Microbiology*, **29**, 411–415.

Hennequin C, Thierry A, Richard GF *et al.* (2001) Microsatellite typing as a new tool for identification of *Saccharomyces cerevisiae* strains. *Journal of Clinical Microbiology*, **39**, 551–559.

Kumar S, Tamura K, Nei M (2004) MEGA3: integrated software for molecular evolutionary genetic analysis and sequence alignment. *Briefings in Bioinformatics*, **5**, 150–163.

Landry CR, Towsend JP, Hartl DL, Cavalieri D (2006) Ecological and evolutionary genomics of *Saccharomyces cerevisiae*. *Molecular Ecology*, **15**, 575–591.

Le Jeune C, Erny C, Demuyter C, Lollier M (2006) Evolution of the population of *Saccharomyces cerevisiae* from grape to wine in a spontaneous fermentation. *Food Microbiology*, **23**, 709–716.

Legras JL, Ruh O, Merdinoglu D, Karst F (2005) Selection of hypervariable microsatellite loci for the characterization of *Saccharomyces cerevisiae* strains. *International Journal of Food Microbiology*, **102**, 73–83.

Levadoux L (1956) Les populations sauvages et cultivées de *Vitis vinifera*. *Annales d'Amélioration des Plantes*, **1**, 59–117.

Martini A (1993) Origin and domestication of the wine yeast *Saccharomyces cerevisiae*. *Journal of Wine Research*, **4**, 165–176.

McGovern PE (2003) *Ancient Wine: the Scientific Search for the Origins of Viniculture*. Princeton University Press, Princeton, New Jersey.

McGovern PE, Hartung U, Badler VR, Glusker DL, Exner LJ (1997) The beginnings of winemaking and viniculture in the ancient Near East and Egypt. *Expedition*, **39**, 3–21.

McGovern PE, Zhang JZ, Tang JG *et al*. C (2004) Fermented beverages of pre- and proto-historic China. *Proceedings of the National Academy of Sciences, USA*, **101**, 17593–17598.

Minch E, Ruiz-Linares A, Goldstein D, Feldman M, Cavalli-Sforza M (1995) MICROSAT, Version 1.5d. *A Computer Program for Calculating Various Statistics on Microsatellite Allele Data*. Department of Genetics, Stanford University, Palo Alto, California.

Mortimer R, Polsinelli M (1999) On the origin of wine yeast. *Research in Microbiology*, **150**, 199–204.

Mortimer R, Romano P, Suzzi G, Polsinelli M (1994) Genome renewal: a new phenomenon revealed from a genetic study of 43 strains of *Saccharomyces* cerevisiae from natural fermentation of grape musts. *Yeast*, **10**, 1543–1552.

Nadal D, Carro D, Fernandez-Larrea J, Pina B (1999) Analysis and dynamics of the chromosomal complements of wild sparkling wine yeast strains. *Applied and Environmental Microbiology*, **65**, 1688–1695.

Ness F, Aigle M (1995) RTM1: a member of a new family of telomeric repeated genes in yeast. *Genetics*, **140**, 945–956.

Perez-Ortin JE, Querol A, Puig S, Barrio E (2002) Molecular characterization of a chromosomal rearrangement involved in the adaptative evolution of yeast strains. *Genome Research*, **12**, 1533–1539.

Pretorius I (2000) Tailoring wine yeast for the new millennium: novel approaches to the ancient art of winemaking. *Yeast*, **16**, 675–729.

Puig S, Querol A, Barrio E, Perez-Ortin JE (2000) Mitotic recombination and genetic changes in *Saccharomyces cerevisiae* during wine fermentation. *Applied and Environmental Microbiology*, **66**, 2057–2061.

Pulvirenti A, Caggia C, Restuccia C, Gullo M, Giudici P (2001) DNA fingerprinting methods used for identification of yeasts isolated from Sicilian sourdoughs. *Annals of Microbiology*, **51**, 107–120.

Querol A, Barrio E, Ramon D (1994) Population dynamics of natural *Saccharomyces* strains during wine fermentation. *International Journal of Food Microbiology*, **21**, 315–323.

Querol A, Fernandez-Espinar MT, del Olmo M, Barrio E (2003) Adaptative evolution of wine yeast. *International Journal of Food Microbiology*, **86**, 3–10.

Reynolds J, Weir BS, Cockerham CC (1983) Estimation of the coancestry coefficient: Basis for a short-term genetic distance. *Genetics*, **105**, 767–779.

Rosini G, Federici F, Martini A (1982) Yeast flora of grape berries during ripening. *Microbiology Ecology*, **8**, 83–89.

Salmon JM (1997) Enological fermentation kinetics of an isogenic ploidy series derived from an Industrial *Saccharomyces cerevisiae* strain. *Journal of Fermentation and Bioengineering*, **83**, 253–260.

Sniegowski P, Dombrowski PG, Fingerman E (2002) *Saccharomyces cerevisiae* and *Saccharomyces paradoxus* coexist in a natural woodland site in North America and display different levels of reproductive isolation from European conspecifics. *FEMS Yeast Research*, **1**, 299–306.

Taberlet P, Fumagalli L, Wurst-Saucy AG, Cossons JF (1998) Comparative phylogeography and postglacial colonization routes in Europe. *Molecular Ecology*, **7**, 453–464.

Teramoto Y, Okamoto K, Ueda S, Kayashima S (1993) Rice wine brewing with sprouting rice, sprouting rice infected with *Aspergillus oryzae* and rice koji. *Journal of the Institute of Brewing*, **99**, 467–471.

This P, Lacombe T, Thomas MR (2006) Historical origins and genetic diversity of wine grapes. *Trends in Genetics*, **22**, 511–519.

Versavaud A, Courcoux P, Roulland C, Dulau L, Hallet JN (1995) Genetic diversity and geographical distribution of wild *Saccharomyces cerevisiae* strains from the wine-producing area of Charentes, France. *Applied and Environmental Microbiology*, **61**, 3521–3529.

Viala P, Vermorel V (1901) *Traité d'Ampélographie, Tome II. Protat*. Frères, Macon, France.

White TJ, Lee S, Taylor J (1990) PCR Protocols. A guide to methods and applications. In: *Amplification and Direct Sequencing of Fungal Ribosomal RNA Genes for Phylogenetics* (eds MA Innis, Gelfand DH Sninsky JJ White TJ), pp. 315–322. Academic Press, San Diego, California.

Winzeler EA, Castillo-Davis CI, Oshiro G *et al*. (2003) Genetic diversity in yeast assessed with whole-genome oligonucleotide arrays. *Genetics*, **163**, 79–89.

JL Legras is interested in understanding how yeast diversity is generated and affects the technological properties of wine yeast. F Karst is a yeast geneticist who studies isoprenoid biosynthesis. D Merdinoglu is a plant geneticist and has a special interest for grapevine resistance to its pathogens. JM Cornuet is a population geneticist with interest in developing methods and softwares for statistical inference in population genetics.

Supplementary material

The following supplementary material is available for this article:

Fig. S1 Consensus tree of populations based on DAS genetic distances obtained after 1000 replicates (resampling loci). The tree was built using the neighbour-joining method, and the root was defined by midpoint rooting.

Table S1 Strains used in the study.

Table S2 Microsatellite loci description and primers.

Table S3 *F*-statistics in diploid wine yeast populations (computed with Genepop).

Table S4 F_{IS} per locus and wine yeast population (computed with *F*stat).

Table S4 F_{ST} between pairs of yeast populations (computed with Microsat 1.5d).

This material is available as part of the online article from:
http://www.blackwell-synergy.com/doi/abs/10.1111/j.1365-294X.2007.03266.x
(This link will take you to the article abstract).

Please note: Blackwell Publishing are not responsible for the content or functionality of any supplementary materials supplied by the authors. Any queries (other than missing material) should be directed to the corresponding author for the article.

Applied and Environmental Microbiology, Nov. 2010, p. 7526–7535 Vol. 76, No. 22
0099-2240/10/$12.00 doi:10.1128/AEM.01280-10

Activation of Two Different Resistance Mechanisms in *Saccharomyces cerevisiae* upon Exposure to Octanoic and Decanoic Acids▽†

J. L. Legras,[1,2,4]* C. Erny,[3] C. Le Jeune,[3] M. Lollier,[3] Y. Adolphe,[3] C. Demuyter,[3] P. Delobel,[4] B. Blondin,[4,5,6] and F. Karst[1,2]

INRA, UMR1131 Santé de la Vigne et Qualité du Vin, F-68000 Colmar, France[1]; Université de Strasbourg, UMR1131, F-67000 Strasbourg, France[2]; Université de Haute Alsace, EA3991 Laboratoire Vigne Biotechnologies et Environnement, F-68000 Colmar, France[3]; INRA, UMR1083 Sciences pour l'Oenologie, F-34000 Montpellier, France[4]; SupAgro, UMR1083 Sciences pour l'Oenologie, F-34000 Montpellier, France[5]; and Université Montpellier 1, UMR1083 Sciences pour l'Oenologie, F-34000 Montpellier, France[6]

Received 31 May 2010/Accepted 10 September 2010

Medium-chain fatty acids (octanoic and decanoic acids) are well known as fermentation inhibitors. During must fermentation, the toxicity of these fatty acids is enhanced by ethanol and low pH, which favors their entrance in the cell, resulting in a decrease of internal pH. We present here the characterization of the mechanisms involved in the establishment of the resistance to these fatty acids. The analysis of the transcriptome response to the exposure to octanoic and decanoic acids revealed that two partially overlapping mechanisms are activated; both responses share many genes with an oxidative stress response, but some key genes were activated differentially. The transcriptome response to octanoic acid stress can be described mainly as a weak acid response, and it involves Pdr12p as the main transporter. The phenotypic analysis of knocked-out strains confirmed the role of the Pdr12p transporter under the control of *WAR1* but also revealed the involvement of the Tpo1p major facilitator superfamily proteins (MFS) transporter in octanoic acid expulsion. In contrast, the resistance to decanoic acid is composite. It also involves the transporter Tpo1p and includes the activation of several genes of the beta-oxidation pathway and ethyl ester synthesis. Indeed, the induction of *FAA1* and *EEB1*, coding for a long-chain fatty acyl coenzyme A synthetase and an alcohol acyltransferase, respectively, suggests a detoxification pathway through the production of decanoate ethyl ester. These results are confirmed by the sensitivity of strains bearing deletions for the transcription factors encoded by *PDR1*, *STB5*, *OAF1*, and *PIP2* genes.

The completion of alcoholic fermentation is one of the major objectives of most wine makers. Several factors can lead to stuck fermentation, such as insufficient nitrogen, the low lipidic content of the grape must, and high concentrations of medium-chain fatty acids (MCFA) in the fermenting must (3). The main MCFA, octanoic and decanoic acids, are produced by yeasts during alcoholic fermentation as by-products of lipid synthesis (40). The exposure of yeast cells to these acids in synthetic media leads to a fast decline of cell viability that is enhanced by low pH and the presence of ethanol (49). Medium-chain fatty acids as well as long-chain fatty acids penetrate inside the cell by passive diffusion in a nonionized form (21) and dissociate at the higher internal pH, leading to a decrease of the intracellular pH (50). The toxicity of short- to medium-chain fatty acids also is correlated with their lipophilic properties, indicating a probable effect on the cell membrane (45). Indeed, decanoic acid, the more toxic of the two compounds (49), has been shown to increase membrane permeability (3). However, the rapid exposure of yeast cells to sublethal concentrations of octanoic acid provokes an adaptive response, allowing the cells to resist larger amounts of inhibitor (7).

Two mechanisms have been hypothesized to explain yeast adaptation to MCFA, including their detoxification into ethyl esters (30) and the activation of a membrane transporter, which has not been identified (7). The *WAR1*-regulated Pdr12p transporter, which is responsible for the resistance to weak lipophilic organic acids such as sorbate or benzoate (24, 32), is a natural candidate for this function. Nevertheless, results concerning its ability to expulse octanoic acid are contradictory (18, 19), while it clearly is not involved in the decanoic acid response (19). So the question of the transporter involved in octanoic and/or decanoic acid expulsion has not been solved.

In this study, we present an investigation of the yeast response to the exposure to octanoic and decanoic acids. The transcriptome analysis of these two responses combined with the screening of deleted strains enabled us to identify the key transporters involved in yeast resistance to MCFA, and it reveals that ethyl ester synthesis is a possible detoxification pathway for these two acids. In addition, the comparison of these transcriptional responses to the responses caused by other lipophilic compounds enabled us to point out the role of several transcription factors.

MATERIALS AND METHODS

Strains and culture conditions. The *Saccharomyces cerevisiae* strains used in this study are listed in Table S1 in the supplemental material. They were strains

* Corresponding author. Mailing address: UMR 1083 Sciences pour l'Oenologie INRA, SupAgro, 2 Place Viala, F-34060 Montpellier cedex 1, France. Phone: 33499613170. Fax: 33499612857. E-mail: jean-luc.legras@supagro.inra.fr.

† Supplemental material for this article may be found at http://aem.asm.org/.

▽ Published ahead of print on 17 September 2010.

TABLE 1. Wine strain resistance level codification used for octanoic or decanoic acid resistance

Resistance	Maximal MCFA concn allowing growth (mM)	
	Octanoic acid	Decanoic acid
1	0.4	0.2
2	0.6 (+)[a]	0.25 (+)
3	0.6 (+++)[b]	0.25 (+++)
4	0.8 (+)	0.3 (+)
5	0.8 (+++)	0.3 (+++)

[a] Impaired growth.
[b] Normal growth.

that were indigenous to wine and were collected in Alsatian wineries or were derivates of the By laboratory strain and eventually were deleted for one or two genes. All strains were cultivated aerobically at 30°C in YPD medium (1% [wt/vol] yeast extract [Difco], 2% Bacto peptone [Difco], 2% glucose [Euromedex]). Unless otherwise indicated, all yeast strains were grown routinely at 28°C.

Strains harboring multiple deletions have been obtained after crossing Euroscarf deletant strains of the opposite sexual type, followed by sporulation and spore dissection. The nonparental ditype (npdt) was verified by the amplification of the wild-type allele and control of kanamycin resistance.

MCFA susceptibility assay. Tests for medium-chain fatty acid (MCFA) resistance phenotypes were performed with cells grown to the exponential growth phase (optical density at 660 nm [OD_{660}] of 0.8) and diluted to an OD_{660} of 0.08. Identical volumes of yeast suspensions were spotted onto agar plates containing 0 to 0.25 mM decanoic acid or 0 to 0.8 mM octanoic acid. Growth was evaluated after 72 h of incubation. The resistance levels of strains indigenous to wine were ranked on a scale of 1 to 5 according to their resistance to octanoic or decanoic acids (Table 1).

DNA microarray profiling experiment. (i) Medium-chain fatty acid exposure and cell harvest. For the fatty acid treatment, the U13 wine strain was precultivated overnight in 10 ml agitated YPD media. Cells then were diluted in 300 ml fresh YPD (in 1-liter Erlenmeyer flasks) to obtain an OD_{660} of 0.05 and grown aerobically at 28°C until an OD_{660} of 1 to 1.1 was reached. Cultures were split in three, and 10 g/liter octanoic or decanoic acid (Sigma) dissolved in ethanol was added at a final concentration of 0.05 mM to one-third of the culture. The same volume of ethanol was added to the control culture. After 20 min, untreated and treated cultures were harvested by centrifugation at room temperature (2 min, 4,000 × *g*), and cells were immediately washed in ice-cold water, centrifuged at 4°C at 10,000 rpm/min, and frozen at −80°C. For each treatment, four independent samples were prepared and analyzed.

(ii) mRNA extraction and reverse transcription. Total RNA was extracted using Trizol reagent (Gibco BRL, Life Technologies). For each sample, 10^9 cells were pelleted by centrifugation (5,000 × *g* for 5 min) in two microcentrifuge tubes, resuspended in 400 μl Trizol, and broken by being vortexed for 4 min with 300-μl glass beads. The two extracts were pooled, and the total volume was adjusted to 8 ml with Trizol reagent. After incubation for 5 min at room temperature, 1.6 ml chloroform was added to separate the aqueous and the organic phases with a brief agitation. After incubation for 3 min at room temperature, the mixture was centrifuged at 10,000 × *g* for 15 min, and the aqueous phase was recovered. The RNA was precipitated by the addition of an equal volume of cold (−20°C) isopropyl alcohol and centrifugation at 10,000 × *g* for 10 min. The precipitate was further dissolved in 150 μl of RNase-free water, and 100 μg of RNA was cleaned up with an RNeasy kit cartridge (Qiagen).

Fluorescent cDNAs were prepared using a ChipShot direct labeling and cleanup system (Promega Z4100) by direct labeling using dCTPs labeled with Cy3-Cy5 according to the manufacturer's instructions (Amersham).

The three modalities (octanoic acid, decanoic acid, and control) were compared in a triangular design in which each sample has been analyzed once, representing four biological replicates per modality.

Hybridization and microarray analysis. Microarray slides were obtained from Eurogentec. They were washed with the buffers provided with a Pronto! universal hybridization kit for 25 slides (Corning 40026) according to manufacturer's instructions. Approximately 1,200 ng of labeled cDNA was deposited on the slide and hybridized overnight at 42°C. After being washed, the arrays were read with a Genepix 4000B scanner (Axon Instrument Inc.) and analyzed with Genepix Pro 3.0 (Axon Instrument Inc.). Artifactual or saturated spots were excluded from the analysis.

Statistical analysis. Raw data were further analyzed using the LIMMA GUI R package (51). Data first were normalized for each slide according to a print-tip-group loess and then normalized between slides according to the quantile procedure. A linear analysis was further performed to detect the genes differentially expressed. Only genes for which a *q* value higher than 0.05 after Benjamini and Hochberg's (6) false-discovery rate (FDR) adjustment for multiple tests has been applied were retained for further analysis. The full data set has been deposited at GEO with accession number GSE18480 (http://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?token=bzglpecmoiywudm&acc=GSE18480). For the different responses, the list of transcription factors involved in the regulation of each gene was obtained from the Yeastract database web site (http://www.yeastract.com) (43). The percentages of genes in each response regulated by every transcription factor obtained from Yeastract then were compared in a factorial correspondence analysis made with Statbox (Grimmersoft, Issy Les Moulineaux, France).

RESULTS

Screening of wine-indigenous strains reveals different responses to octanoic and decanoic acid stress. In an attempt to evaluate the variability of the resistance of wine strains to MCFA, we compared the sensitivities of 76 indigenous strains to these inhibitors in a drop test (Table 1). The strains were gathered in groups of similar sensitivities (see Table S1 in the supplemental material). Figure 1 shows that wine strains present a high variability in their ability to resist MCFA. Resistance to decanoic acid generally is associated with a medium to strong resistance to octanoic acid. In contrast, four of the strains showing the highest resistance to octanoic acid were sensitive to decanoic acid. These results suggest strongly that wine yeast activate two partially overlapping mechanisms to resist these MCFA, and that some mechanisms involved in decanoic acid resistance may contribute to octanoic acid resistance.

Transcriptome analysis reveals that octanoic and decanoic acids activate two partially overlapping sets of genes. To get further insights into the genes involved in the *S. cerevisiae* response to octanoic and decanoic acids, we studied the transcriptome of wine yeast strain U13, chosen for its high resistance to both inhibitors in the former experiment. This strain was exposed for 20 min to 50 μM each acid. For both acids these conditions were found to be sufficient for the induction of high resistance in a preliminary experiment.

We carried out competitive hybridizations in a triangular design between cDNA obtained from nonexposed cells (T) and cDNA obtained from octanoic acid (C8)- or decanoic acid (C10)-treated cells, as well as between cDNAs from the two acid-exposed conditions. Considering all the genes whose expression was significantly altered between the two conditions, we found that exposure to octanoic or decanoic acid affected 81 and 620 genes, respectively, compared to control cells, with 76 being common to both responses. The comparative hybridization of C8 and C10 modalities revealed that 71 genes were differently affected by these organic acids. The imbalance between the numbers of genes involved in each response suggested that cells were exposed to different stress intensities. To determine if the two responses were correlated, we selected the genes induced by decanoic acid (i.e., genes with significant C10/T log ratios) but that were not detected in the C8/C10 comparison. The C10/T log ratio of these genes was plotted against the C8/T log ratio of the same genes (Fig. 2). The two ratios were highly correlated ($R^2 = 0.89$), indicating that the two acids similarly affected this set of genes, and the slope of

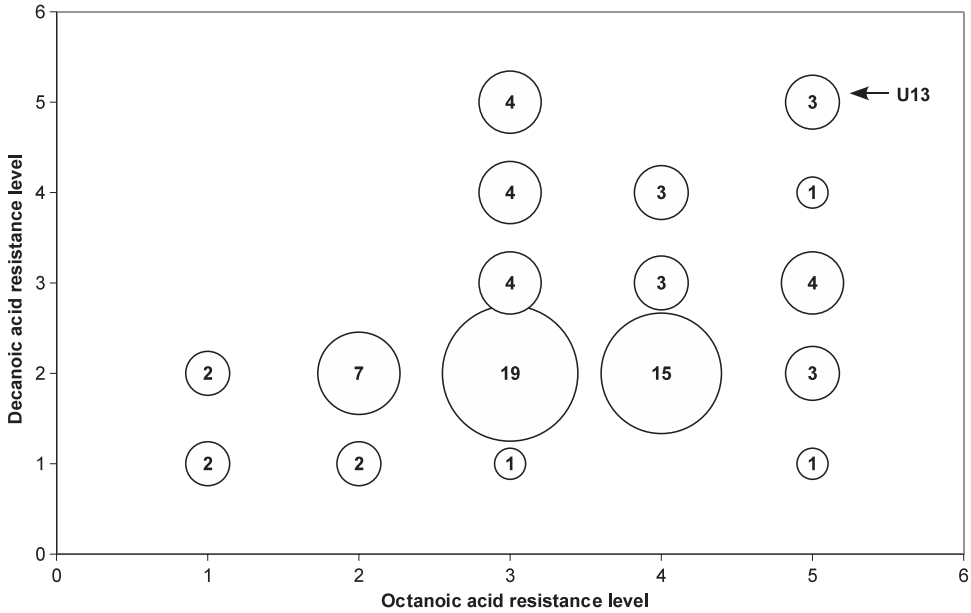


FIG. 1. Variability of *Saccharomyces cerevisiae* wine strain resistance to octanoic acid (*x* axis) and decanoic acid (*y* axis) as revealed by drop test (resistance levels are given in Table 1). The dimension of the spots is related to the number of strains in each category indicated in the spots. The arrow indicates the group containing U13.

0.6 reveals that the C8 response was weaker than the C10 response at the tested concentrations.

To minimize the biological noise, we restricted the analysis to the genes whose expression significantly differed by a minimum log ratio of 0.5 for C10 acid compared to that of the control. A ratio of 0.3 was chosen for C8 acid compared to that of the control to obtain similar cutoffs for the two acids. As a consequence, compared to the control, 75 genes were significantly modulated by octanoic acid and 165 by decanoic acid, with 53 genes being shared by the two responses (Fig. 3). The

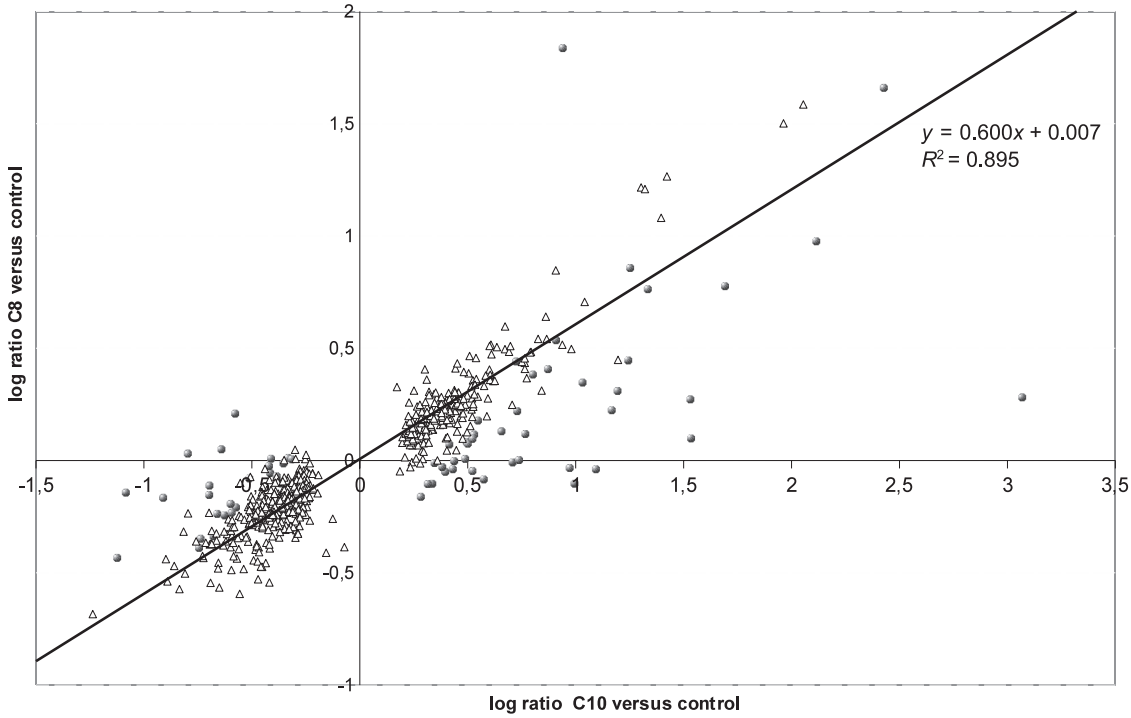


FIG. 2. Correlation between the two responses. Triangles, expression ratio of genes whose expression level varied significantly compared to that of the control for one modality; spheres, expression ratio of genes whose expression level varied significantly for C8 compared to that of C10; these points were excluded for the estimation of the correlation between the two responses.

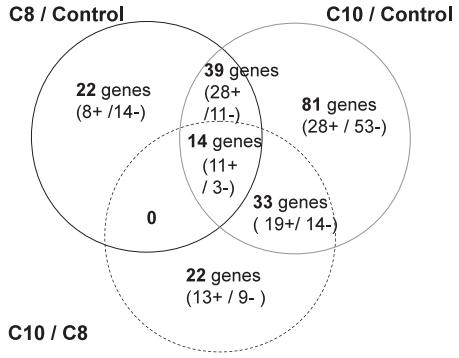


FIG. 3. Venn diagram presenting the genes differentially expressed in these three conditions. C8, 20-min exposure to octanoic acid at 50 mM; C10, 20-min exposure to octanoic acid at 50 mM; C8/C10, comparison of exposure to C10 to exposure to C8. Repressed genes are followed by a minus sign; induced genes are followed by a plus sign.

analysis of the functional categories through gene ontology (MIPS) (Table 2) shows that the yeast responses to these two organic acids share almost the same subset of genes involved in cell energy supply. Decanoic acid response also includes sets of genes involved in ribosome biogenesis and RNA processing. In agreement with this, GO biological processes for decanoic acid response include ribosome and large ribosomal subunit biogenesis (GO:0042254 and GO:0042273) and oxidation reduction (GO:0055114). The major facilitator superfamily proteins (MFS) also are especially well represented in this response. In addition, the analysis of these responses with Eu.Gene 1.2.1 (10) showed the significant activation of the fatty acid beta oxidation pathway.

Direct comparison of octanoic and decanoic acid-treated cells (with a log ratio of 0.3) allowed us to gain further indications on the analogies/differences between the two responses. We found that 68 genes presented significantly different expression profiles (Fig. 3). As the activation of genes was not similar for octanoic acid and decanoic acid, we divided the activated genes in three sets: genes activated by both C8 and C10 but not by the control were qualified as shared responses, the genes activated by C8 but not the control or C10 were called C8-specific responses, and the genes activated by C10 but not the control or C8 combined with genes differentially expressed after exposure to C10 were called C10-specific responses (Fig. 3).

Among the 53 genes whose expression was affected by both acids (C8 and C10 shared response) (see Table S2 in the supplemental material), 39 were upregulated and 14 repressed. Among them, C8 more efficiently induced *PDR12* (3.5 times increase), whereas C10 more specifically induced *ALD4*, *CWP1*, *TMA17*, and *HXK1*.

The octanoic acid-specific response included 22 genes; 8 were upregulated (i.e., *IDH2*, *ATP3*, *ALG12*, *TRX2*, etc.), whereas 14 were repressed (*EFT1*, *EFT2*, *ZRT1*, *FAS1*, etc.).

The exposure to C10 resulted in the specific modulation of the expression of 114 genes (C10-specific response), among which *EEB1*, coding for an ethyl ester synthase, tops the list for its high induction (8.4 times increase). Two other genes (*FAA1* and *ELO1*) involved in fatty acid metabolism also were induced (1.7 and 2.4 times increase), suggesting a potential metabolism of the fatty acid. However, two transporters involved in cell detoxification, *TPO4* and *PDR12* (two times higher expression), and, to a lesser extent, *TPO1* (1.5 times increase), also were activated.

Comparison of octanoic and decanoic acid responses to stress responses already described. The octanoic and decanoic acid responses were compared to those already described for different stresses (Table 3) after a similar incubation period: sorbic acid (a weak acid response has been described [36]), sodium dodecyl sulfate (39), octanol (14), fluphenazine (12), benomyl (27), 2,4-deoxyphenoxyacetic acid (44), polyoxyethylene-9-laurylether (POELE), 2,4-dichlorophenol (DCP) (37), oleic acid oxidative stress (23) and H_2O_2 (15). For each pair of stresses, we have counted the number of genes significantly induced or repressed by both stresses. When the significance information was not available, we selected genes induced (or repressed) by at least a factor of 2. The highest similarities were observed between the responses to octanoic and decanoic acid (71% of C8 responses are shared with C10 responses), but about half of the genes involved in both responses also are shared with H_2O_2 oxidative stress. Significant portions of these responses also are activated by detergent stresses: i.e., 45 and 35% of genes activated by octanoic and decanoic acids also are activated by SDS. Octanoic acid response presents 29% of genes in common with the sorbic acid response. In contrast with these different stresses, the oleic acid oxidative stress involved few of the genes activated by C8 or C10.

As a consequence, our results suggest that the responses to octanoic and decanoic acid are composite responses involving the organic weak acid response and a detergent response, with both of them presenting many similarities with an oxidative stress.

Search for known transcription factors involved in MCFA response. For both acid responses as well as for the other stresses cited in the former paragraph, we obtained from the Yeastract web site (43) the transcription factors involved in the regulation of each gene and scored the number of genes regulated per transcription factor. These scores were compared in a correspondence analysis. The result plotted in Fig. 4 revealed that the octanoic acid response presents, as expected, similarities to the weak acid response involving War1 and Msn4 (Fig. 4). In contrast, the C10 response appeared much closer to the SDS stress response. C8 and C10 responses also appeared as potentially regulated by transcription factors *HAP1*, *HAP4*,

TABLE 2. Classification of genes involved in response to octanoic and decanoic acid according to MIPS functional categories

MIPS category	*P*	
	C8	C10
Sugar, glucoside, polyol and carboxylate catabolism (01.05.02.07)	2.3e-07	4.3e-06
Electron transport (02.11)	2.9e-07	
Electron transport and membrane-associated energy conservation (02.11)	3.2e-06	1.3e-06
Ribosome biogenesis (12.01)		2.5e-11
rRNA processing (11.04.01)		7.7e-10
Purine nucleotide/nucleoside/nucleobase anabolism (01.03.01.03)		1.9e-08
Tricarboxylic-acid pathway (02.10)	2.1e-05	

TABLE 3. Comparison of different transcriptional responses[a]

Stress[b]	% Genes shared between the different responses											
Octanoic acid (0.05 mM)	100											
Decanoic acid (0.05 mM)	71	100										
Sorbic acid (8 mM) (36)	29	19	100									
POELE (1 mM) (37)	43	31	43	100								
DCP (3 mM) (37)	19	22	20	28	100							
Octanol (1%) (14)	0	23	18	14	19	100						
SDS (1%) (39)	45	39	26	21	20	21	100					
Fluphenazine (1 mM) (12)	12	17	12	26	27	17	37	100				
Benomyl (7 mM) (27)	12	14	12	17	25	16	23	13	100			
2,4-Dichlorophenoxyacetic acid (3 mM) (44)	16	14	17	20	13	15	21	11	6	100		
Oleic acid (23)	7	6	8	8	10	6	17	16	12	5	100	
H_2O_2 (3 mM) (15)	52	47	42	27	35	23	40	38	41	30	22	100

[a] The percentage of genes shared between two different responses is given as the number of genes shared by the two responses divided by the number of genes of the smaller response.

[b] The number of genes involved in each response was as follows: octanoic acid, 75; decanoic acid, 165; sorbic acid, 137; POELE, 548; DCP, 342; octanol, 621; SDS, 463; fluphenazine, 132; benomyl, 69; 2,4 dichlorophenoxyacetic acid, 172; oleic acid, 267; H_2O_2, 936.

and *HAP5*. This could be interpreted as a sign of the activation of the fatty acid beta-oxidation pathway.

Phenotyping screening of Euroscarf deletion mutant strains allows us to identify genes involved in resistance. (i) Transporters. Looking for transporters involved in octanoic and decanoic acid expulsion, we screened a collection of haploid strains deleted for their PDR genes as well as for other transporters, including *ADP1*, *AQR1*, *ATM1*, *AUS1*, *AZR1*, *BPT1*, *DIP5*, *FLR1*, *NFT1*, *PDR5*, *PDR10*, *PDR11*, *PDR12*, *PDR15*, *PDR18*, *PXA2*, *SNQ2*, *TPO1*, *TPO2*, *TPO3*, *TPO4*, *YBT1*, and *YOR1*. Figure 5 shows the results obtained for the sensitivity test of some of these strains in the presence of 0 to 0.6 mM C8 and 0 to 0.25 mM C10.

The highest sensitivity to octanoic acid was obtained for the

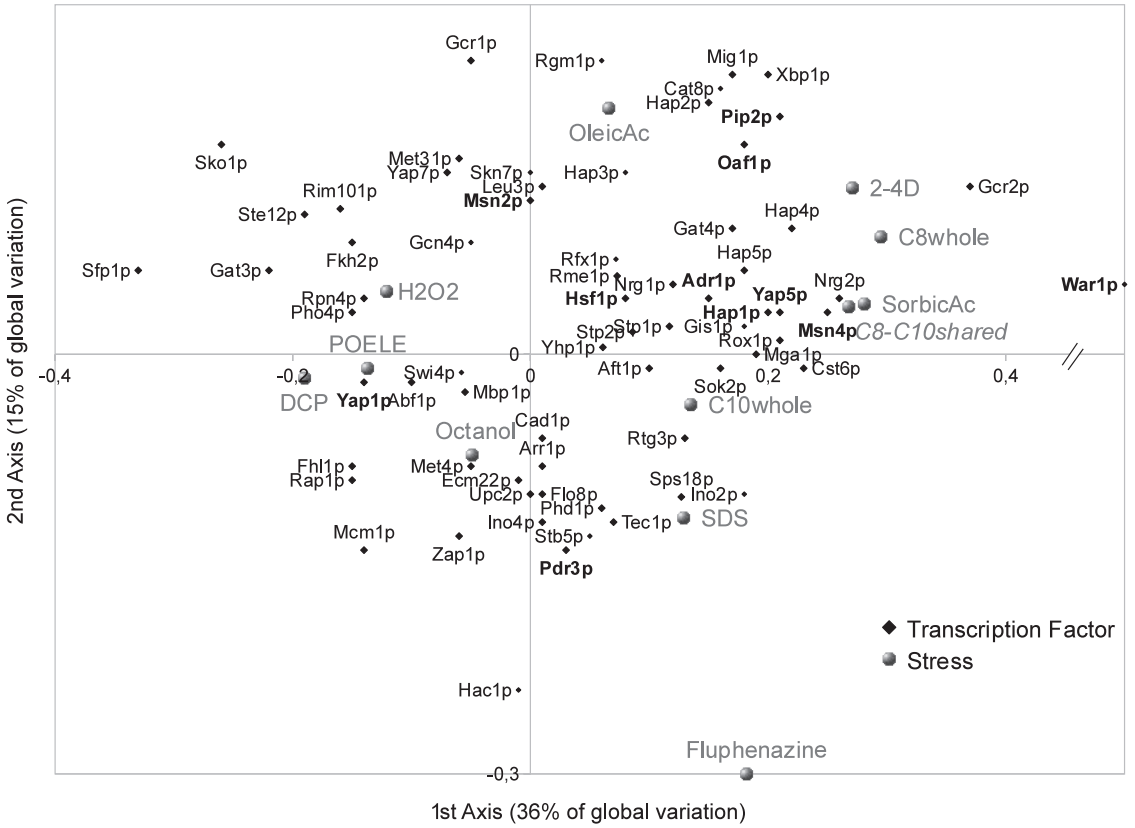


FIG. 4. Factorial component analysis comparing the involvement of each transcription factor in the different stress responses. The proportion of each stress response explained by one transcription factor has been calculated from the Yeastract website. Codes for the different stresses: OleicAc, oleic acid; POELE, polyoxyethylene-9-laurylether; SorbicAc, sorbic acid; 2-4D, 2,4-dichlorophenoxyacetic acid; DCP, 2,4-dichlorophenol; C10whole, decanoic acid; C8whole, octanoic acid. The octanoic and decanoic shared response (C8-C10 shared) is analyzed as a supplementary individual and given in italic. Transcription factors with a $\cos^2$ lower than 0.1 are not drawn. Main transcription factors involved in the stress response are in boldface.

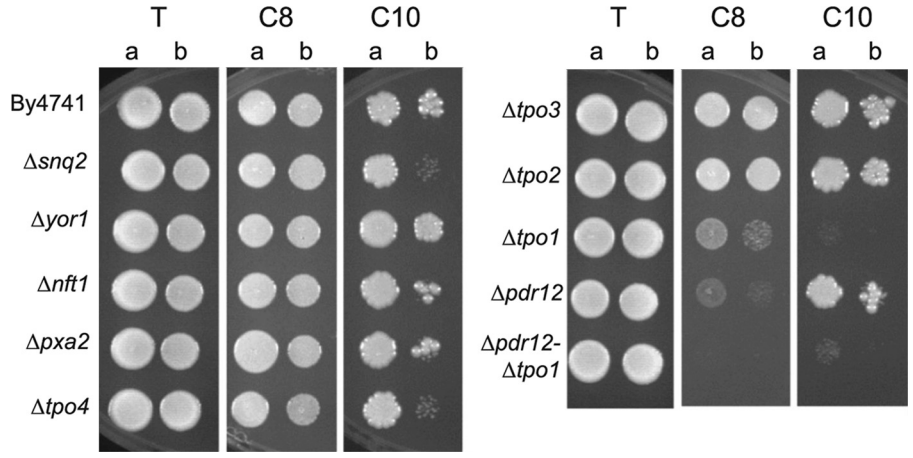


FIG. 5. Drop test presenting the sensitivity provoked by the deletion of different transporters on YPD (pH 4.5) medium containing 0.6 mM octanoic acid (C8) or 0.25 mM decanoic acid (C10) compared to that of the control (T). Cells used to prepare the spots were grown on liquid YPD (pH 4.5) medium until a standardized OD_{660} of 0.8 (a) and diluted to an OD_{660} of 0.08 (b). The growth observed for the nondiluted spot of the wild-type By laboratory strain corresponds to a rank of three in the first figure.

Δ*pdr12* strain; however, the deletion of *TPO1* also resulted in an increased sensitivity to this inhibitor. In addition, the Δ*pdr12*-Δ*tpo1* double deletant strain was found to be more sensitive than the two single-deletion strains, showing a cumulative effect of the two transporters in the expulsion of this acid. In contrast to octanoic acid, the Δ*tpo1* strain exhibited the highest sensitivity to decanoic acid, while the deletion of PDR12 did not affect this phenotype. However, Δ*tpo4* and Δ*snq2* strains also were slightly affected, indicating a possible ability of these transporters to expulse decanoic acid. None of the other strains tested showed modified resistance to octanoic or decanoid acid. Since Pdr12p and Tpo1p were the main transporters of octanoic and/or decanoic acid, we constructed diploid strains harboring one or two deleted copies of the genes (Fig. 6). In a background where two alleles of a given transporter were present, the presence of a single allele of the other transporter was sufficient to regenerate the wild-type phenotype. When the two tested genes were present as a single copy, the growth of yeast strains was injured on both octanoic and decanoic acids. Moreover, when a single copy of the *TPO1* gene was present, decanoic acid resistance was correlated with the number of copies of the *PDR12* gene, indicating that Pdr12p plays a part, though discrete, in resistance to decanoic acid. This observation was confirmed by analyzing the growth of the diploid on liquid medium complemented with inhibitors (Bioscreen analysis) (results not shown).

(ii) Transcription factors. We also tested Euroscarf strains deleted for regulatory genes, including main regulators of stress responses (*PDR1*/*PDR3*, *MSN2*/*MSN4*, *HSP30*, etc.), transcription factors already described in weak organic acid stress or in acid stress (*WAR1*, *HAA1*, etc.), and other transcription factors suggested by the transcriptome analysis. Tested transcription factors were *ADR1*, *AFT1*, *ARR1*, *CIN5*, *CRZ1*, *KFH2*, *FLO8*, *GCN4*, *GCR2*, *HAA1*, *HAC1*, *HAP2* to *HAP5*, *HSF1*, *HSP30*, *MSN2*, *MSN4*, *NRG1*, *OAF1*, *PDR1*, *PDR3*, *PIP2*, *RPN4*, *SEF1*, *SKO1*, *SOK2*, *STB5*, *TOS8*, *WAR1*, *YAP1*, *YAP2*, and *YRR1*. Figure 7 shows the results obtained for the drug sensitivity test of some of these strains in the presence of 0 to 0.6 mM C8 and 0 to 0.25 mM C10.

The octanoic acid response is clearly under the control of the *WAR1* transcription factor. The deletion of *PDR3* also has a great impact on this phenotype, while *PDR1* deletion does not modify strain resistance to octanoic acid. None of the other transcription factors that were tested seemed to be involved in the modulation of the octanoic acid response, including *HAA1* (not shown), which was described to be involved in a weak acid response (13).

The response to decanoic acid is clearly under the control of *PDR1*, while *PDR3* deletion does not result in higher sensitiv-

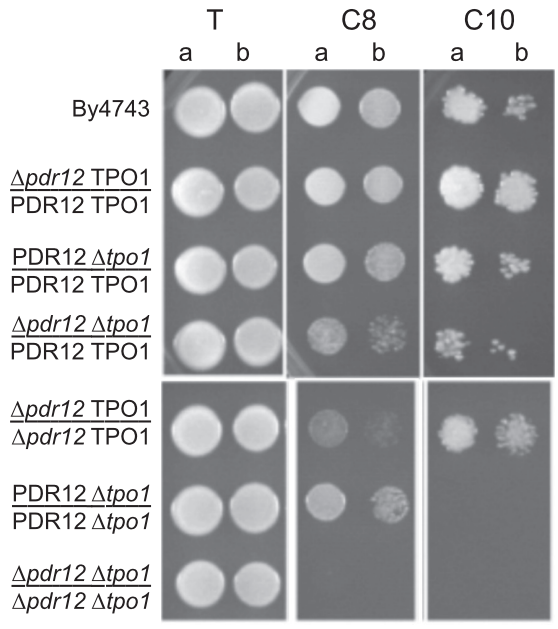


FIG. 6. Drop test presenting the sensitivity of diploid strains deleted from one or two copies of *PDR12* and *TPO1* transporter genes on YPD (pH 4.5) medium containing 0.6 mM octanoic acid (C8) or 0.25 mM decanoic acid (C10) compared to that of the control (T). Cells used to prepare the spots were grown on liquid YPD (pH 4.5) medium until a standardized OD_{660} of 0.8 (a) and diluted to an OD_{660} of 0.08 (b).

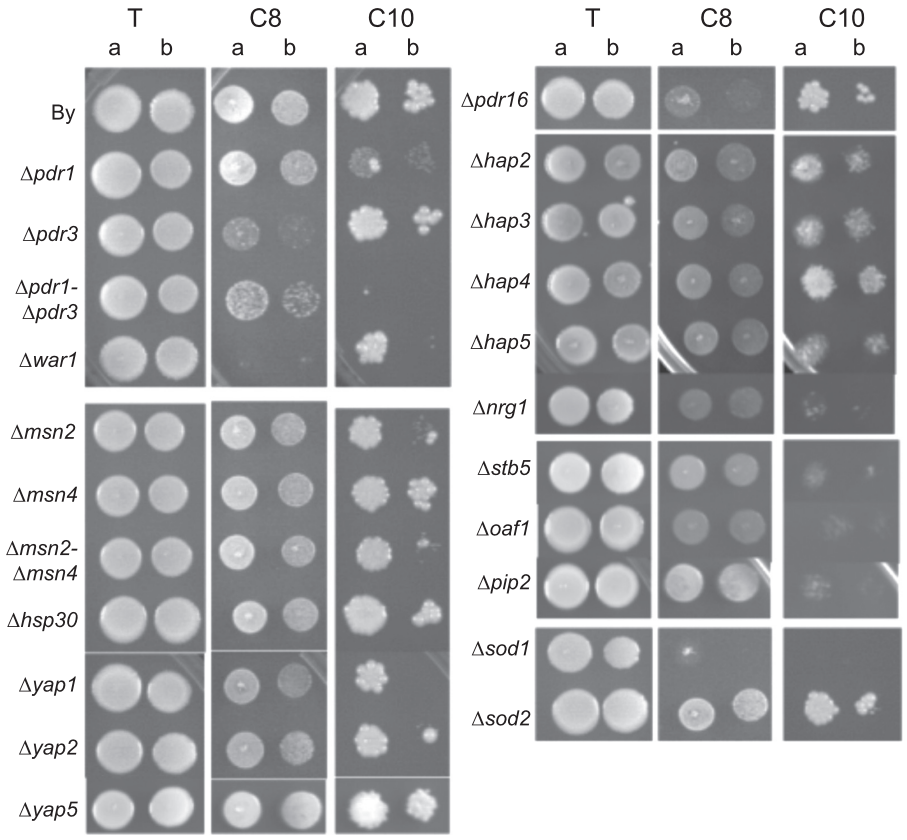


FIG. 7. Drop test presenting the sensitivity of strains deleted for different transcription factors on YPD (pH 4.5) medium containing 0.6 mM octanoic acid (C8) or 0.25 mM decanoic acid (C10) compared to that of the control (T). Cells used to prepare the spots were grown on liquid YPD (pH 4.5) medium until a standardized OD_{660} of 0.8 (a) and diluted to an OD_{660} of 0.08 (b).

ity. However, the Δ*pdr1*-Δ*pdr3* double mutant is more sensitive to decanoic acid than the Δ*pdr1* mutant, indicating a slight role of *PDR3*. Moreover, the deletion of *STB5* also lowers strain resistance to decanoic acid. This transcription factor is known to form a heterodimer with Pdr1p but not with Pdr3p and to regulate the pentose phosphate pathway and NADPH production in response to oxidative stress (25). Three other transcription factors are clearly involved in response to decanoic acid, namely, *NRG1*, *OAF1*, and *PIP2*.

Finally, the deletion of *MSN2* (but not *MSN4*) or *YAP1* seems to have little effect on decanoic acid resistance.

(iii) Other genes. We also tested some other genes already described to be involved in weak acid resistance or in lipophilic compound resistance. We observed that *PDR16* deletion had an impact on octanoic acid resistance, while the deletion of the homologous *PDR17* gene had no impact.

Finally, *SOD1* deletion had a drastic effect on both octanoic and decanoic acid sensitivity, while *SOD2* deletion had no impact on these phenotypes.

DISCUSSION

Octanoic and decanoic acids are two compounds synthesized by yeast during alcoholic fermentation and are potent inhibitors, and they have been suggested to be involved in sluggish fermentation (4). The toxicity of these lipophilic molecules has been attributed to internal medium acidification (8). In the meantime, because toxicity is correlated with lipophilicity, an effect on membrane organization also has been suggested (3). Nevertheless, *S. cerevisiae* is able to adapt to these inhibitors through the induction of transporters (7) and possibly other mechanisms. The negative influence of these inhibitors on the fermentation process is offset by the fact that they are supposed to be the precursors of ethyl esters, which enrich wine flavor with fruit aromas (17), even though this has never been clearly shown (5). Our work was aimed at gaining insights into the mechanisms involved in the adaptation of *S. cerevisiae* to octanoic and decanoic acids. In a first approach we analyzed the resistance of a set of wild strains to both acids. The variability of the responses observed revealed unambiguously that two different and partially overlapping mechanisms were activated in response to the two inhibitors.

Transcriptome analysis reveals composite responses. To get further information on the genes involved in the octanoic or decanoic acid responses, we used a transcriptomic analysis. Both responses appeared highly correlated, which we expected from the closely related structures of the two acids, but they also presented differences. We were able to distinguish a common response involving genes that were modulated by both acids and more restricted octanoic- or decanoic-specific responses. The responses to octanoic or decanoic acid also were compared to some already-described stress responses, including response to weak acids, to several detergents (SDS,

POELE, etc.), to oxidative stress (H_2O_2), and to oleic acid as a unique carbon source. The octanoic and decanoic responses share more genes with each other than with any other stress response. They both present features in common with the oxidative stress response, but with an original component: the response triggered by octanoic acid exposure presents similarities to the weak acid response (36), whereas the decanoic acid response additionally presents some similarities to the oleic acid early stress response (23) and to the SDS stress response (39). This can be related to the chemical structures of the two inhibitors, and one can suggest that octanoic acid is perceived as an acid, while the hydrophobic part of decanoic acid is more prominent than its acidic part and induces a detergent-like response.

Phenotypic analysis of deletant strains highlights the main genes involved in resistance. Transcriptome analysis reveals a number of genes induced or repressed as a response to an environmental change, turning the cell metabolism to the new conditions. However, many of these genes do not play a key role in the resistance to the stress agents (15). Several of the genes induced by MCFA according to transcriptome analysis did not participate significantly in MCFA resistance in the drop test phenotypic screening. As an example, the expression of *TPO4* and *PDR12* transporters was much more increased by exposure to decanoic acid (2.3 and 1.9 times increase, respectively) than the expression of *TPO1* (1.5 times increase); however, the latter was the only transporter that was really efficient in the triggering of decanoic acid resistance. The weak induction of TPO1 is not related to a high basal expression but may be related to the mild intensity of the challenge. The higher activation of *TPO4*, which is not involved in the resistance, has been described likewise for 2,4-dichlorophenoxyacetic acid (44). In contrast, some of the genes involved in inhibitor resistance could be undetected in the transcriptome due to a delayed response or a posttranscriptional regulation. Drop tests of deleted strains submitted to octanoic and decanoic acids allowed us to identify genes involved in the resistance to both acids (common resistance mechanism) or to one of them (specific resistance mechanism).

Common resistance mechanism. Among the numerous genes we tested, only *TPO1*, *YAP1*, and *SOD1* seemed to be involved in both octanoic and decanoic acid resistance.

The MFS transporter Tpo1p is the key player in decanoic acid resistance and contributes significantly to octanoic acid resistance. Tpo1p was described to protect cells against a broad range of structurally unrelated molecules, including spermidine, cycloheximide, nystatin, artesunate, ibuprofen, 2,4-dichlorophenoxyacetic acid, etc. (for a review, see reference 33). To our knowledge, this work is the first evidence of Tpo1p being involved in the resistance to octanoic and decanoic acids. In our analysis of the resistance of wild strains to octanoic and decanoic acid, we observed that a high level of decanoic acid resistance always was associated with a high level of octanoic acid resistance. Strains with more efficient Tpo1p transport would exhibit such a phenotype. Interestingly, the *TPO1* gene (as well as *TPO4* and *SNQ2*) is activated progressively during the course of alcoholic fermentation (28), suggesting a key role during alcoholic fermentation. In analyzing the genomic differences between laboratory yeast strains and wine-making commercial strains by the CGH array, Dunn and coauthors (11) observed the amplification of *TPO1* in all of the commercial wine strains tested, which could be the sign of an adaptive evolution, even if this amplification was not observed in a later study (9). The genome of the U13 wine strain that was used for transcriptomic analysis was compared to the S288c genome by a CGH array and did not reveal any amplification of the *TPO1* gene (unpublished data), suggesting the gain of efficiency by another mechanism.

YAP1 is a transcription factor involved in oxidative stress adaptation (26) and the regulation of the expression of antioxidant genes such as thioredoxin, thioredoxin reductase, and glutathione reductase (reviewed in reference 20). *YAP1* deletion hampers the cell resistance to octanoic and decanoic acid, suggesting that these inhibitors cause oxidative damage to the cell. Consistently with this hypothesis, MCFA resistance also is impaired in the *SOD1*-deleted strain but not in the mitochondrial *SOD2*-deleted strain. This indicates that, at least under conditions where respiration is glucose repressed, MCFA toxic action is due to superoxide anion production (42).

Octanoic acid resistance-specific mechanism. The Pdr12p ABC transporter under the control of *WAR1* is the main factor of octanoic acid resistance, which is enhanced by the activity of Tpo1p in a cumulative manner as revealed by the drop test (Fig. 6). Moreover, *PDR16* deletion resulted in enhanced sensitivity to octanoic acid. The ABC transporter Pdr12p has been described to be the weak organic acid transporter (32) that is able to confer resistance to the C1 to C7 organic acids but not to the longer fatty acids (C8 to C10) (19), although the expression of *PDR12* was induced by C3 to C8 acids (18). Our experiments show unambiguously that Pdr12p is indeed involved in octanoic acid resistance. The discrepancy observed between our results and previous ones could be due to the cumulative effects of the two transporters Pdr12p and Tpo1p or to the use of different genetic backgrounds.

Strikingly, the activity of these two transporters is based on two different energetic supplies. Pdr12p function is dependent on ATP consumption, while the activity of Tpo1p is linked to the proton gradient across the membrane. The second mechanism could be especially well adapted to the acidic pH of fermenting medium.

Among the other genes we tested, only *PDR16* seemed to be involved in the octanoic acid response. This gene affects the lipid composition of the plasma membrane, limiting the passive uptake of the drug across the membrane (48). Curiously, the *PDR16* homologous gene *PDR17* has no effect on octanoic acid resistance, nor do other genes described to be involved in cell response to weak organic acids, such as *SPI1* (38) or *HAA1* (13).

Considering the regulation of the octanoic acid response, our results clearly highlight the essential role of *WAR1*, probably through the activation of Pdr12p. *TPO1* and *PDR16* genes are under the control of *PDR1*, but our phenotypic screening revealed a poor effect of *PDR1* deletion on octanoic acid resistance, while the *PDR3* deletion was more effective. Moreover, none of the other transcription factors tested had an effect on octanoic acid resistance. Octanoic acid seems to act essentially as a short-chain organic acid and to induce a relatively simple response, including expulsion through Pdr12p and Tpo1p and perhaps membrane adaptation through Pdr16p.

Decanoic acid resistance-specific mechanism. The yeast response to decanoic acid is much more complex. Besides the key effect of *TPO1* deletion, a slight effect of *SNQ2* and *TPO4* transporter deletion can be observed. It is noteworthy that the deletion of the *PDR5* transporter, which often is associated with *TPO1* in drug resistance (41, 2), had no impact on decanoic acid resistance. The deletion of the genes coding for the transcription factors *PDR1*, *STB5*, *MSN2*, *NRG1*, *OAF1*, and *PIP2* resulted in impaired decanoic acid resistance.

PDR1, PDR3, and *STB5* are three transcription factors that form homo- or heterodimers (1) and regulate many ABC transporters. The *PDR1*/*PDR3* complex has been shown to be activated by membrane-active compounds (37) and, in this context, to act mainly through the Pdr5 transporter. In our case, neither *PDR3* nor *PDR5* deletions had any impact on decanoic acid resistance, suggesting another regulation network. The transcription factor Stb5p can form heterodimers with Pdr1p (1) but also is able to act without Pdr1p or Pdr3p to regulate the pentose phosphate pathway and NADPH production as an answer to oxidative agents (25). It is noteworthy that one of the Stb5p targets is the *SNQ2* transporter, which has a (slight) effect on decanoic acid resistance. These results reinforce the idea of MCFA impairing the oxidative state of the cell, which is evidenced by the activation of *PRX1* (also obtained during the screening of a multicopy expression library) as well as *GPX2* (a phospholipid hydroperoxide glutathione peroxidase) that protects cells from phospholipid hydroperoxide during oxidative stress.

The *MSN2*/*MSN4* complex is the mediator of the general stress response that is induced by different environmental changes. Moreover, Msn2p nucleus translocation and the activation of STRE genes also can be obtained by a range of membrane-disturbing agents (29). In our experiments, *MSN2* deletion has more effect on decanoic acid resistance than *MSN4* deletion, reinforcing the hypothesis of decanoic acid acting through membrane perturbation as well as internal medium acidification.

Finally, *OAF1* or *PIP2* deletion also resulted in impaired decanoic acid resistance. These transcription factors act as dimers to positively regulate genes encoding peroxisomal proteins in response to oleate induction in glucose-free medium. The induction of genes involved in fatty acid metabolism, such as *EEB1*, *FAA1*, and *ELO1* (8.3, 1.7, and 2.4 times increase, respectively), is in agreement with the beta oxidation of fatty acids taking place in the peroxisome. One of the targets of these transcription factors is the gene *EEB1* (22), which is the most induced by decanoic acid in our study. The high activation of *EEB1* after exposure to decanoic acid as well as the slight (35%) but significant induction of *YMR210w*, coding for a putative acyltransferase with similarity to *EEB1*, suggests that ethyl ester synthesis plays a complementary role in the detoxification of culture media. Indeed, the deletion of *EEB1* only or *EEB1* and *YMR210w* in a laboratory strain causes 45 and 80% decrease in the production of ethyl ester during alcoholic fermentation, respectively (34). However, the same team (35) also found that the addition of octanoic acid during alcoholic fermentation resulted in an increase of *EEB1* expression but not the addition of decanoic acid. The low sensitivity of defective *EEB1* strains after MCFA exposure suggests that this metabolic route participates moderately in global resistance. However, we observed that during alcoholic fermentation strains possessing an inactivated copy of *EEB1* presented a higher lag phase when exposed to decanoic acid (unpublished data), suggesting a more-significant role during alcoholic fermentation.

Activation mechanism of fatty acid resistance. Fatty acids from C12 were shown to be able to trigger the transcriptional signal (47) through the direct activation of Oaf1p in a ligand-dependent manner (31). In a similar manner, various hydrophobic inhibitors are able to bind to a discrete xenobiotic-binding domain of Pdr1p and Pdr3p (46). In analogy with these recent results, we can hypothesize that these MCFA activate Pdr1p and Oaf1p. This activation of Oaf1p/Pip2p and Pdr1p/Stb5p is mediated by the complex Gal11p/MED15, and we also observed that the deletion of *GAL11* results in a hypersensitivity of the Δ*gal11* mutant strain.

The mechanism activating the weak acid response is not understood. Gregori et al. (16) hypothesized the direct activation of War1p by the acid, but this has not been observed until now.

In conclusion, we have shown that the resistance to octanoic and decanoic acids, two potent fermentation inhibitors, results from the activation of three different mechanisms. As for many drugs, the expulsion of these two acids by the two transporters Tpo1p and Pdr12p represents the main part of the resistance. The adaptation to these acids also involves an oxidative stress response similar to what has been observed for other acids (i.e., sorbic acid) or detergents (SDS), but the activation of beta-oxidation can explain a part of this oxidative stress. However, contrary to former observations, we also could observe, unexpectedly for decanoic acid (and not octanoic acid), the activation of the genes involved in ethyl ester synthesis, which are key odorant compounds of wine aroma.

ACKNOWLEDGMENTS

This work has been supported by a grant from the Alsace Region, France.

We are grateful to F. Devaux and C. Rodrigues-Pousada for kindly providing strains. We thank A. Alais Perot for her helpful technical assistance.

REFERENCES

1. **Akache, B., S. MacPherson, M. Sylvain, and B. Turcotte.** 2004. Complex interplay among regulators of drug resistance genes in *Saccharomyces cerevisiae*. J. Biol. Chem. **279:**27855–27860.
2. **Alenquer, M., S. Tenreiro, and I. Sá-Correia.** 2006. Adaptive response to the antimalarial drug artesunate in yeast involves Pdr1p/Pdr3p-mediated transcriptional activation of the resistance determinants TPO1 and PDR5. FEMS Yeast Res. **6:**1130–1139.
3. **Alexandre, H., B. Mathieu, and C. Charpentier.** 1996. Alteration in membrane fluidity and lipid composition, and modulation of H^+-ATPase activity in *Saccharomyces cerevisiae* caused by decanoic acid. Microbiology **142:**469–475.
4. **Alexandre, H., and C. Charpentier.** 1998. Biochemical aspects of stuck and sluggish fermentation in grape must. J. Ind. Microbiol. Biotechnol. **20:**20–27.
5. **Bardi, L., C. Crivelli, and M. Marzona.** 1998. Esterase activity and release of ethyl esters of medium-chain fatty acids by *Saccharomyces cerevisiae* during anaerobic growth. Can. J. Microbiol. **44:**1171–1176.
6. **Benjamini, Y., and Y. Hochberg.** 1995. Controlling the false discovery rate: a practical and powerful approach to multiple testing. J. R. Stat. Soc. B **57:**289–300.
7. **Cabral, M. G., C. A. Viegas, and I. Sá-Correia.** 2001. Mechanisms underlying the acquisition of resistance to octanoic-acid-induced-death following exposure of *Saccharomyces cerevisiae* to mild stress imposed by octanoic acid or ethanol. Arch. Microbiol. **175:**301–307.
8. **Carmelo, V., H. Santos, and I. Sá-Correia.** 1997. Effect of extracellular acidification on the activity of plasma membrane ATPase and on the cytosolic and vacuolar pH of *Saccharomyces cerevisiae*. Biochim. Biophys. Acta **1325:**63–70.

9. **Carreto, L., M. F. Eiriz, A. C. Gomes, P. M. Pereira, D. Schuller, and M. A. Santos.** 2008. Comparative genomics of wild type yeast strains unveils important genome diversity. BMC Genomics **9:**524.
10. **Cavalieri, D., C. Castagnini, S. Toti, K. Maciag, T. Kelder, L. Gambineri, S. Angioli, and P. Dolara.** 2007. Eu. Gene Analyzer a tool for integrating gene expression data with pathway databases. Bioinformatics **23:**2631–2632.
11. **Dunn, B., R. P. Levine, and G. Sherlock.** 2005. Microarray karyotyping of commercial wine yeast strains reveals shared, as well as unique, genomic signatures. BMC Genomics **6:**53.
12. **Fardeau, V., G. Lelandais, A. Oldfield, H. Salin, S. Lemoine, M. Garcia, V. Tanty, S. Le Crom, C. Jacq, and F. Devaux.** 2007. The central role of PDR1 in the foundation of yeast drug resistance. J. Biol. Chem. **282:**5063–5074.
13. **Fernandes, A. R., N. P. Mira, R. C. Vargas, I. Canelhas, and I. Sá-Correia.** 2005. *Saccharomyces cerevisiae* adaptation to weak acids involves the transcription factor Haa1p and Haa1p-regulated genes. Biochem. Biophys. Res. Commun. **337:**95–103.
14. **Fujita, K., A. Matsuyama, Y. Kobayashi, and H. Iwahashi.** 2004. Comprehensive gene expression analysis of the response to straight-chain alcohols in S*accharomyces cerevisiae* using cDNA microarray. J. Appl. Microbiol. **97:**57–67.
15. **Gasch, A. P., P. T. Spellman, C. M. Kao, O. Carmel-Harel, M. B. Eisen, G. Storz, D. Botstein, and P. O. Brown.** 2000. Genomic expression programs in the response of yeast cells to environmental changes. Mol. Biol. Cell **11:** 4241–4257.
16. **Gregori, C., C. Schüller, I. E. Frohner, G. Ammerer, and K. Kuchler.** 2008. Weak organic acids trigger conformational changes of the yeast transcription factor War1 in vivo to elicit stress adaptation. J. Biol. Chem. **283:**25752–25764.
17. **Guth, H.** 1997. Quantitation and sensory studies of character impact odorants of different white wine varieties. J. Agric. Food Chem. **45:**3027–3032.
18. **Hatzixanthis, K., M. Mollapour, I. Seymour, B. E. Bauer, G. Krapf, C. Schüller, K. Kuchler, and P. W. Piper.** 2003. Moderately lipophilic carboxylate compounds are the selective inducers of the *Saccharomyces cerevisiae* Pdr12p ATP-binding cassette transporter. Yeast **20:**575–585.
19. **Holyoak, C. D., D. Bracey, P. W. Piper, K. Kuchler, and P. J. Coote.** 1999. The *Saccharomyces cerevisiae* weak-acid-inducible ABC transporter Pdr12 transports fluorescein and preservative anions from the cytosol by an energy-dependent mechanism. J. Bacteriol. **181:**4644–4652.
20. **Ikner, A., and K. Shiozaki.** 2005. Yeast signaling pathways in the oxidative stress response. Mutat. Res. **569:**13–27.
21. **Kamp, F., W. Guo, R. Souto, P. F. Pilch, B. E. Corkey, and J. A. Hamilton.** 2003. Rapid flip-flop of oleic acid across the plasma membrane of adipocytes. J. Biol. Chem. **278:**7988–7995.
22. **Karpichev, I. V., and G. M. Small.** 1998. Global regulatory functions of Oaf1p and Pip2p (Oaf2p), transcription factors that regulate genes encoding peroxisomal proteins in *Saccharomyces cerevisiae*. Mol. Cell. Biol. **18:**6560–6570.
23. **Koerkamp, M. G., M. Rep., H. J. Bussemaker, G. P. Hardy, A. Mul, K. Piekarska, A. Szigyarto, J. M. De Mattos, and H. F. Tabak.** 2002. Dissection of transient oxidative stress response in *Saccharomyces cerevisiae* by using DNA microarrays. Mol. Biol. Cell **13:**2783–2794.
24. **Kren, A., Y. M. Mamnun, B. E. Bauer, C. Schüller, H. Wolfger, K. Hatzixanthis, M. Mollapour, C. Gregori, P. Piper, and K. Kuchler.** 2003. War1p, a novel transcription factor controlling weak acid stress response in yeast. Mol. Cell. Biol. **23:**1775–1785.
25. **Larochelle, M., S. Drouin, F. Robert, and B. Turcotte.** 2006. Oxidative stress-activated zinc cluster protein Stb5 has dual activator/repressor functions required for pentose phosphate pathway regulation and NADPH production. Mol. Cell. Biol. **26:**6690–6701.
26. **Lee, J., C. Godon, G. Lagniel, D. Spector, J. Garin, J. Labarre, and M. B. Toledano.** 1999. Yap1 and Skn7 control two specialized oxidative stress response regulons in yeast. J. Biol. Chem. **274:**16040–16046.
27. **Lucau-Danila, A., G. Lelandais, Z. Kozovska, V. Tanty, T. Delaveau, F. Devaux, and C. Jacq.** 2005. Early expression of yeast genes affected by chemical stress. Mol. Cell. Biol. **25:**1860–1868.
28. **Marks, V. D., S. J. Ho Sui, D. Erasmus, G. K. van Der Merwe, J. Brumm, W. W. Wasserman, J. Bryan, and H. J. van Vuuren.** 2008. Dynamics of the yeast transcriptome during wine fermentation reveals a novel fermentation stress response. FEMS Yeast Res. **8:**35–52.
29. **Moskvina, E., E. M. Imre, and H. Ruis.** 1999. Stress factors acting at the level of the plasma membrane induce transcription via the stress response element (STRE) of the yeast *Saccharomyces cerevisiae*. Mol. Microbiol. **32:**1263–1272.
30. **Peddie, H. A. B.** 1990. Ester formation in brewery fermentations. J. Inst. Brew. **96:**327–331.
31. **Phelps, C., V. Gburcik, E. Suslova, P. Dudek, F. Forafonov, N. Bot, M. MacLean, R. J. Fagan, and D. Picard.** 2006. Fungi and animals may share a common ancestor to nuclear receptors. Proc. Natl. Acad. Sci. U. S. A. **103:**7077–7081.
32. **Piper, P., Y. Mahé, S. Thompson, R. Pandjaitan, C. Holyoak, R. Egner, M. Mühlbauer, P. Coote, and K. Kuchler.** 1998. The pdr12 ABC transporter is required for the development of weak organic acid resistance in yeast. EMBO J. **17:**4257–4265.
33. **Sá-Correia, I., S. C. dos Santos, M. C. Teixeira, T. R. Cabrito, and N. P. Mira.** 2009. Drug:H+ antiporters in chemical stress response in yeast. Trends Microbiol. **17:**22–31.
34. **Saerens, S. M., K. J. Verstrepen, S. D. Van Laere, A. R. Voet, P. Van Dijck, F. R. Delvaux, and J. M. Thevelein.** 2006. The *Saccharomyces cerevisiae* EHT1 and EEB1 genes encode novel enzymes with medium-chain fatty acid ethyl ester synthesis and hydrolysis capacity. J. Biol. Chem. **281:**4446–4456.
35. **Saerens, S. M., F. Delvaux, K. J. Verstrepen, P. Van Dijck, J. M. Thevelein, and F. R. Delvaux.** 2008. Parameters affecting ethyl ester production by *Saccharomyces cerevisiae* during fermentation. Appl. Environ. Microbiol. **74:**454–461.
36. **Schüller, C., Y. M. Mamnun, M. Mollapour, G. Krapf, M. Schuster, B. E. Bauer, P. W. Piper, and K. Kuchler.** 2004. Global phenotypic analysis and transcriptional profiling defines the weak acid stress response regulon in *Saccharomyces cerevisiae*. Mol. Biol. Cell **15:**706–720.
37. **Schüller, C., Y. M. Mamnun, H. Wolfger, N. Rockwell, J. Thorner, and K. Kuchler.** 2007. Membrane-active compounds activate the transcription factors Pdr1 and Pdr3 connecting pleiotropic drug resistance and membrane lipid homeostasis in *Saccharomyces cerevisiae*. Mol. Biol. Cell **18:**4932–4944.
38. **Simões, T., N. P. Mira, A. R. Fernandes, and I. Sá-Correia.** 2006. The SPI1 gene, encoding a glycosylphosphatidylinositol-anchored cell wall protein, plays a prominent role in the development of yeast resistance to lipophilic weak-acid food preservatives. Appl. Environ. Microbiol. **72:**7168–7175.
39. **Sirisattha, S., Y. Momose, E. Kitagawa, and H. Iwahashi.** 2004. Toxicity of anionic detergents determined by *Saccharomyces cerevisiae* microarray analysis. Water Res. **38:**61–70.
40. **Taylor, G. T., and B. H. Kirsop.** 1977. The origin of medium chain length fatty acids present in beer. J. Inst. Brew. **83:**241–243.
41. **Teixeira, M. C., and I. Sá-Correia.** 2002. *Saccharomyces cerevisiae* resistance to chlorinated phenoxyacetic acid herbicides involves Pdr1p-mediated transcriptional activation of TPO1 and PDR5 genes. Biochem. Biophys. Res. Commun. **292:**530–537.
42. **Teixeira, M. C., J. P. Telo, N. F. Duarte, and I. Sá-Correia.** 2004. The herbicide 2,4-dichlorophenoxyacetic acid induces the generation of free-radicals and associated oxidative stress responses in yeast. Biochem. Biophys. Res. Commun. **324:**1101–1107.
43. **Teixeira, M. C., P. Monteiro, P. Jain, S. Tenreiro, A. R. Fernandes, N. P. Mira, M. Alenquer, A. T. Freitas, A. L. Oliveira, and I. Sá-Correia.** 2006. The YEASTRACT database: a tool for the analysis of transcription regulatory associations in *Saccharomyces cerevisiae*. Database issue. Nucleic Acids Res. **34:**D446–D451.
44. **Teixeira, M. C., A. R. Fernandes, N. P. Mira, J. D. Becker, and I. Sá-Correia.** 2006. Early transcriptional response of *Saccharomyces cerevisiae* to stress imposed by the herbicide 2,4-dichlorophenoxyacetic acid. FEMS Yeast Res. **6:**230–248.
45. **Tenreiro, S., P. A. Nunes, C. A. Viegas, M. S. Neves, M. C. Teixeira, M. G. Cabral, and I. Sá-Correia.** 2002. AQR1 gene (ORF YNL065w) encodes a plasma membrane transporter of the major facilitator superfamily that confers resistance to short-chain monocarboxylic acids and quinidine in *Saccharomyces cerevisiae*. Biochem. Biophys. Res. Commun. **292:**741–748.
46. **Thakur, J. K., S. Pan, W. S. Moye-Rowley, H. Arthanari, F. Yang, A. M. Näär, X. Fan, J. Breger, D. P. Frueh, K. Gulshan, D. K. Li, E. Mylonakis, K. Struhl, B. P. Cormack, and G. Wagner.** 2008. A nuclear receptor-like pathway regulating multidrug resistance in fungi. Nature **452:**604–609.
47. **Thakur, J. K., H. Arthanari, F. Yang, K. H. Chau, G. Wagner, and A. M. Näär.** 2009. Mediator subunit Gal11p/MED15 is required for fatty acid-dependent gene activation by yeast transcription factor Oaf1p. J. Biol. Chem. **284:**4422–4428.
48. **van den Hazel, H. B., H. Pichler, M. A. do Valle Matta, E. Leitner, A. Goffeau, and G. Daum.** 1999. PDR16 and PDR17, two homologous genes of *Saccharomyces cerevisiae*, affect lipid biosynthesis and resistance to multiple drugs. J. Biol. Chem. **274:**1934–1941.
49. **Viegas, C. A., M. F. Rosa, I. Sá-Correia, and J. M. Novais.** 1989. Inhibition of yeast growth by octanoic and decanoic acids produced during ethanolic fermentation. Appl. Environ. Microbiol. **55:**21–28.
50. **Viegas, C. A., and I. Sá-Correia.** 1997. Effects of low temperatures (9–33 degrees C) and pH (3.3–5.7) in the loss of *Saccharomyces cerevisiae* viability by combining lethal concentrations of ethanol with octanoic and decanoic acids. Int. J. Food Microbiol. **34:**267–277.
51. **Wettenhall, J. M., and G. K. Smyth.** 2004. limmaGUI: a graphical user interface for linear modeling of microarray data. Bioinformatics **20:**3705–3706.

OPEN ACCESS Freely available online

PLOS ONE

Population Structure and Comparative Genome Hybridization of European Flor Yeast Reveal a Unique Group of *Saccharomyces cerevisiae* Strains with Few Gene Duplications in Their Genome

Jean-Luc Legras[1,2,3]*, Claude Erny[4], Claudine Charpentier[5]

1 INRA, UMR1083 Science pour l'Œnologie, Montpellier, France, **2** Montpellier SupAgro, UMR1083 Science pour l'Œnologie, Montpellier, France, **3** Université Montpellier 1, UMR1083 Science pour l'Œnologie, Montpellier, France, **4** Université de Haute Alsace, Laboratoire Vigne Biotechnologies et Environnement, Colmar, France, **5** Institut Universitaire de la Vigne et du Vin, Université de Bourgogne, Dijon, France

Abstract

Wine biological aging is a wine making process used to produce specific beverages in several countries in Europe, including Spain, Italy, France, and Hungary. This process involves the formation of a velum at the surface of the wine. Here, we present the first large scale comparison of all European flor strains involved in this process. We inferred the population structure of these European flor strains from their microsatellite genotype diversity and analyzed their ploidy. We show that almost all of these flor strains belong to the same cluster and are diploid, except for a few Spanish strains. Comparison of the array hybridization profile of six flor strains originating from these four countries, with that of three wine strains did not reveal any large segmental amplification. Nonetheless, some genes, including *YKL221W/MCH2* and *YKL222C*, were amplified in the genome of four out of six flor strains. Finally, we correlated *ICR1* ncRNA and *FLO11* polymorphisms with flor yeast population structure, and associate the presence of wild type *ICR1* and a long Flo11p with thin velum formation in a cluster of Jura strains. These results provide new insight into the diversity of flor yeast and show that combinations of different adaptive changes can lead to an increase of hydrophobicity and affect velum formation.

Citation: Legras J-L, Erny C, Charpentier C (2014) Population Structure and Comparative Genome Hybridization of European Flor Yeast Reveal a Unique Group of *Saccharomyces cerevisiae* Strains with Few Gene Duplications in Their Genome. PLoS ONE 9(10): e108089. doi:10.1371/journal.pone.0108089

Editor: Alan M. Moses, University of Toronto, Canada

Received April 14, 2014; **Accepted** August 11, 2014; **Published** October 1, 2014

Data Availability: The authors confirm that all data underlying the findings are fully available without restriction. All .cel files and analyzed data are available from GEOGSE55925 at http://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE55925. Sequences are available at Gene-Bank under accession number HG965200–HG965204.

Funding: The authors have no support or funding to report.

Competing Interests: The authors have declared that no competing interests exists.

* Email: jean-luc.legras@supagro.inra.fr

Introduction

Numerous fermented beverages have been developed all over the world during history. In addition to alcoholic fermentation, some beverages are obtained through a specific aging process called flor wine aging. During this process, which takes place only after the completion of alcoholic fermentation, a biofilm called velum is formed by yeast at the surface of the wine leading to the progressive oxidation of alcohol and remaining carbohydrates. This yeast oxidative metabolism generates many aromatic compounds (ethanal, sotolon, solerone…)[1–3], which give these wines their unique flavor.

Flor aging (or biological aging) is performed traditionally in several vineyards in Europe, including Hungary (Tokaj Hegyalja) to produce Szamorodni, Italy (Sardinia) to produce Vernaccia di Oristano, Spain (Jerez area) to produce Xeres, and France (Jura) to produce Vin Jaune. The apparition of the velum is generally spontaneous [4] but some French wine makers use selected flor starters. Flor yeast belong to the species *Saccharomyces cerevisiae* [5], and the population of flor yeast isolated from the velum of Sherry wines differs from the population of strains that perform alcoholic fermentation [5,6]. These two populations are genetically isolated [7], as shown by the ITS1 region in Spanish and Jura flor strains, which have specific alleles of ITS1 caused by a 24 bp deletion [5] and a G insertion [8], respectively. Furthermore, various molecular techniques used to explore the diversity of flor yeast populations in several countries suggest a large genetic diversity [8–11].

Yeast strains adopt a specific lifestyle during flor aging, and adaptation to this ecological niche has long remained the focus of many investigations. Aneuploidies have been described [12,13] as a major genetic feature of Spanish flor strains and were hypothesized to explain adaptation to flor aging. Indeed, yeast are able to adapt to stressful conditions due to the amplification of specific regions of their genome [14,15]. The main adaptive feature of flor yeast is their ability to develop a velum on wine when sugars are depleted, which is an activity that is carried out only by some yeast strains [16]. The build-up of the biofilm is obtained by the aggregation of single cells, permitted by their high hydrophobicity. The high hydrophobicity of flor cells results from modifications of the lipid content and the activation of *FLO11* [17], which encodes a GPI anchored protein with a serine and threonine rich central region. Flor strains carry specific *FLO11*

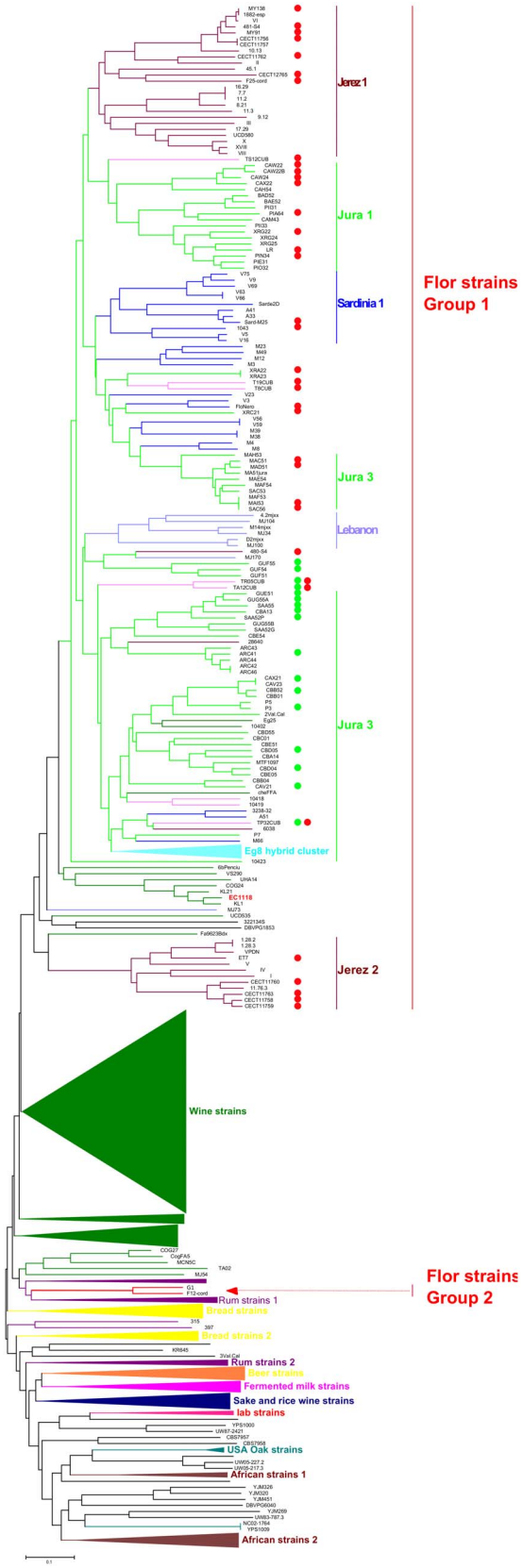


Figure 1. Neighbor joining tree presenting the diversity of flor strains evaluated at 12 microsatellite loci, in comparison with strains of other origins. The tree was built from the Dc chord distance and drawn with MEGA5.22. The wine cluster has been condensed due to its large size. Red dots indicate the presence of a 111 bp deletion in the *FLO11* promoter, and a green dot indicates that this deletion is missing.
doi:10.1371/journal.pone.0108089.g001

alleles that encode a protein with an expanded central hydrophobic core, which facilitates the adaptation of yeast cells to the velum environment [18,19]. In addition, sherry flor strains have a deletion in the long noncoding RNA *ICR1* located upstream from *FLO11*. ICR1 functions as a switch that regulates the expression of *FLO11* and its disruption stimulates the expression of *FLO11* [18,20].

Flor aging is encountered in highly distant vineyards, which raises the question of the relatedness and origin of these strains. The similar conditions faced by various strains in European vineyards implies that these strains share a similar genomic makeup and features of aneuploidy. In this paper, we compared flor yeast populations from Hungary (Tokaj), France (Jura), Italy (Sardinia) and Spain (Jerez). We used various molecular genetic techniques to investigate the genetic composition of these strains. The polymorphism of microsatellite markers allowed us to infer the structure of the flor yeast population. We measured the ploidy of strains and compared the genomes of several flor strains by CGH on array, which enabled us to detect aneuploidies specific to flor strains. Finally, we also examined polymorphisms within the promoter and protein central core region of *FLO11* and link these polymorphisms to the ability to grow on velum media.

Material and Methods

1. Strains and growth conditions

The strains of this study originated from several laboratories in Spain, Hungary, Italy and France. They are described in detail in Table S1. The two first letters of Jura strains indicate the cellar from which each strain was isolated.

Yeast cells were cultivated in 10 ml of YPD medium (36 h, 28°C, 160 rpm). Velum growth was verified on Fornachon medium [21] (Yeast extract 1 $g.l^{-1}$, $(NH_4)_2SO_4$ 0.5 $g.l^{-1}$, $MgSO_4$ 1 $g.l^{-1}$, $CaCl_2$ 0.5 $g.l^{-1}$, pH adjusted to 3.2 with HCl, autoclaved 35 minutes at 110°C, following which 4% (v/v) ethanol was added aseptically after cooling), after 8 days of incubation at 28°C.

2. Microsatellite typing and determination of population structure

S. cerevisiae microsatellite loci were amplified as described previously [22]. Genomic DNA was isolated by phenol/chloroform extraction, after cell grinding with glass beads, and isopropanol precipitation as described previously [23]. Allelic variation at 12 microsatellite loci was examined in 142 strains as described previously [22]. The chord distance Dc [24] matrix was calculated for each couple of strains with a laboratory-made program. The tree was obtained from the distance matrices with the Neighbor program of the Phylip 3.67 package, and drawn with MEGA5.22 [25]. The tree was rooted by the midpoint method. To assess the assignment of flor strains to a particular origin, InStruct [26] was used to evaluate the number of populations in the set of strains and a graphical display was obtained with R software version 2.15.1 [27].

3. Analysis of *FLO11* polymorphisms and cell hydrophobicity

The polymorphism of the length of Flo11p was measured from the amplification of *FLO11* alleles with a pair of primers located −

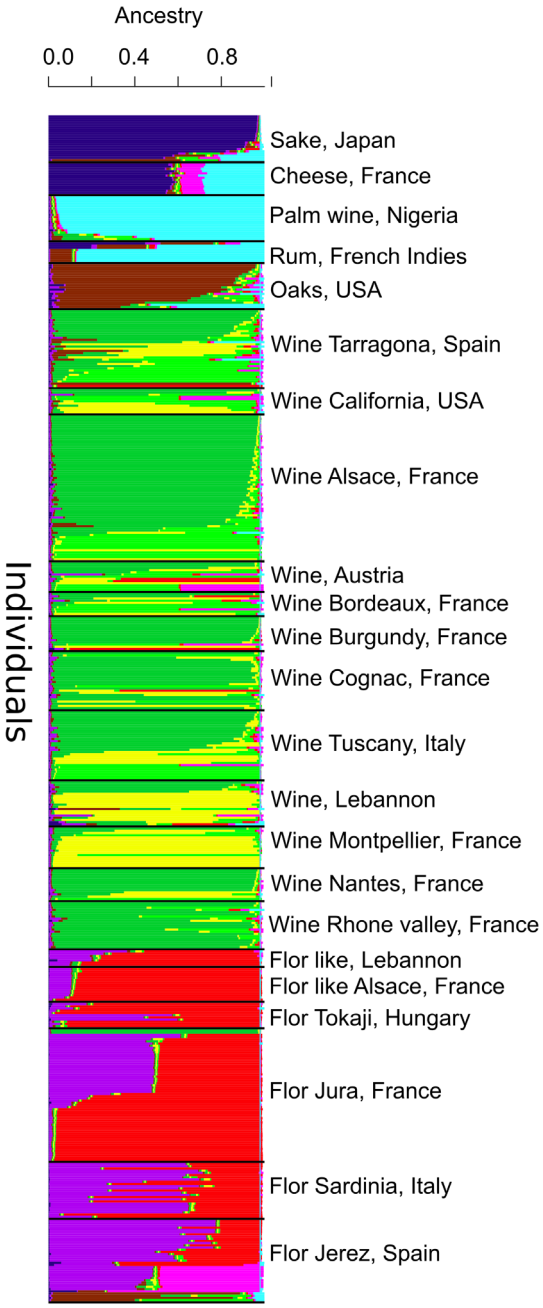


Figure 2. Clustering of flor strains with InStruct population structure inference software for K = 9 populations. Each color corresponds to one inferred ancestral group. The proportion of each colors gives the proportion of the corresponding ancestral genome in the genome of each strain. The name of the isolated population is shown at the top of each cluster.
doi:10.1371/journal.pone.0108089.g002

53 bp in the 5′ of *FLO11* (Flo11IntFw CTCCCTCATCATGTTGTGGTTC), and +3126 bp in the terminal part of *FLO11* (Flo11IntRv AACGACGGTGGTTGAGACAA). ExTaq DNA polymerase (TaKaRa) was used to amplify this long DNA fragment. The PCR temperature program was 95°C for 5 min, followed by 30 cycles with an initial denaturation step of 95°C for 30 sec, annealing at 61°C for 30 sec, and elongation at 72°C for 6 min.

The presence of the 111 bp deletion in *ICR1* ncRNA was examined by the amplification of this region with the primer pair Flo11promFw CAGCCCCAGAGTATGTTCTCACAG and Flo11promRv AATCACCTTCTAAACGCTCGGA. This PCR was performed with regular MBI Fermentas Taq DNA polymerase. The PCR temperature program was 95°C for 5 min, followed by 30 cycles with a first denaturation step 95°C for 30 sec, annealing at 56°C for 45 sec, and elongation at 72°C for 1 min. The presence of the deletion was detected from the band size of the amplified fragment in gel electrophoresis.

For 5 strains (CAV21, LRJura, CECT11758, TR05CUB, T8CUB), the amplified fragment was sequenced with the same primers. These five sequences are available in GenBank under the accession number (HG965200–HG965204).

Cell hydrophobicity was evaluated following the procedure of Ishigami et al. [17], which relies on the measure of the partition of yeast cells between a buffer solution and an organic solvent. Yeasts strains were cultivated for 48 h with shaking in Fornachon's media containing 4% ethanol, and then harvested, washed three times with water and suspended in 4 ml of McIlvaine buffer, pH 3.5. The cell population was adjusted to an optical density of approximately 0.5 at 660 nm (OD660). Four ml of this suspension was transferred to a test tube (15·150 mm) with a stopper. An equivalent volume of hexane was gently layered over the buffer. This test tube was vigorously vortexed for 5 min, with care taken to avoid emulsification. The OD660 of the initial and the residual buffer layers were measured, and the degree of hydrophobicity of the yeast cell surfaces (HD) was calculated from the equation:

$$HD(\%) = 100(I - R)I$$

Where I and R are the OD660 of the initial and the residual layers, respectively.

4. CGH on array

Genomic DNA was labeled and hybridized against GeneChip Yeast Genome 2.0 Array from Affymetrix (Santa Clara, CA), which covers all *S. cerevisiae* S288C genes [28]. Labeled fragments were prepared from 200 to 500 ng of genomic DNA with the BioPrime DNA Labeling System (Invitrogen). The hybridization and detection steps were performed at the IGBMC Microarray and Sequencing Platform (Illkirch, France). Two arrays were used for each strain. Intensity data of perfect match probes were obtained with apt1.12.0 Affymetrix software, after RMA background subtraction and quantile normalization [29]. After filtering for probes with insufficient signal, the final number of probes used for the analysis was 38863. Signal intensities were scaled across arrays and log ratios were calculated using S288C as a reference. The log ratios were averaged by groups of three consecutive probes, to reduce probe to probe variation and facilitate analysis with DNAcopy. The best results were obtained after RMA background subtraction and quantile normalization of array data. Array Data were analyzed with the package DNAcopy [30] and R software version 2.15.1 [27]. A custom script was used to associate the mean log ratio calculated per chromosome segment with each ORF it contained. Gene clustering was performed with Cluster 3.0 [31], using a filter of 0.5 minimum difference in log ratio between all strains, and limiting missing data to six strains. Uncentered correlation and the centroid clustering were chosen as parameters, and dendrograms were drawn with TreeView. Gene ontology enrichment analysis was performed with Gene Codis 3.0 available at http://genecodis.cnb.csic.es/analysis [32].

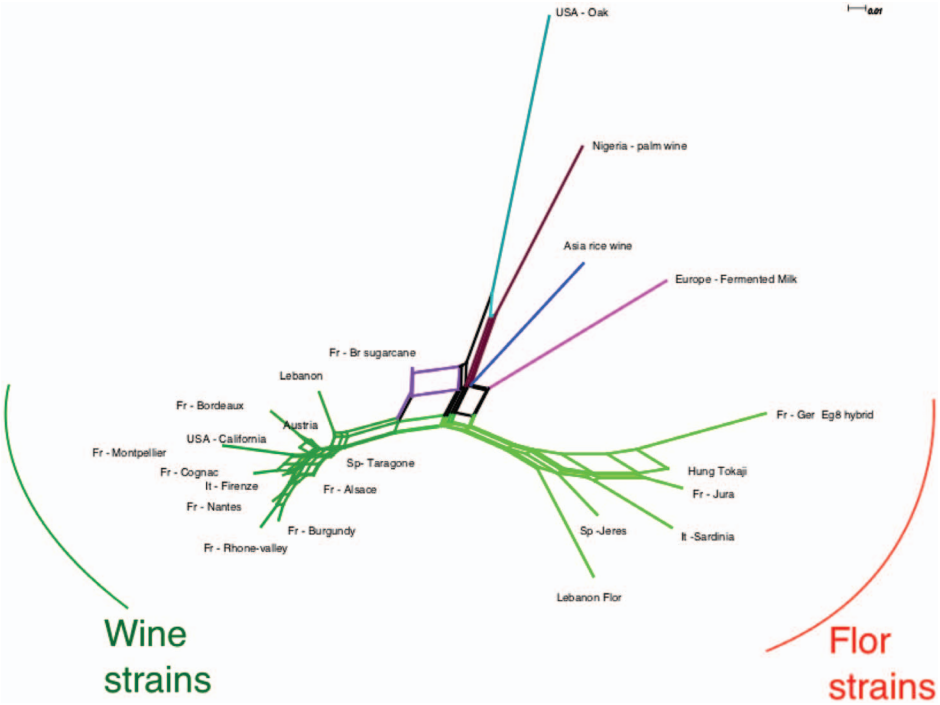


Figure 3. Neighbor net representing the differentiation between populations measured by Fst distance matrices.
doi:10.1371/journal.pone.0108089.g003

The full data set has been deposited at the NCBI Gene Expression Omnibus (GEO) with GEO accession number (GSE55925) http://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE55925.

5. Ploidies

For the analysis of cell DNA content, yeast cells were prepared in 96 well plates as described previously [33]. DNA content per cell was determined with an BD Accuri C6^{TM} flow cytometer. However, Syber Green was used instead of Sytox Green because of the minimum variation observed with this fluorescent dye [34]. Both dyes give sharper peaks than propidium iodide, which has been used in most studies until now, and provide a more accurate evaluation of ploidy [34,35]. By4741 and By4742 were used as haploid references and S288C and By4743 were used as diploid references.

Results

1. Diversity of flor strains from various countries

We collected 142 flor stains from various countries. The 64 French strains from Jura were characterized previously by pulsed field gel electrophoresis and inter delta typing [8]. The other flor strains were provided by research groups from Spain (40 strains from the Jerez region and three strains from the Cordoba region), Italy (29 strains from Sardinia) and Hungary (6 strains from the Tokaj region). We evaluated the diversity of these 142 flor strains from polymorphisms detected at 12 microsatellite loci and were able to differentiate 131 genotypes. We compared these strains with 497 strains isolated from other sources (wine, palm wine, sake, oak bark) genotyped previously [22,36] and 35 strains sequenced recently [37]. Flor strains clustered into one main group in a neighbor joining tree (Figure 1), with the exception of two Spanish flor strains isolated from Cordoba. Interestingly, subclusters formed inside the main group of flor strains according to geographical origin: three clusters of Jura strains, two clusters of sherry wine strains (Jerez 1 and 2), and one main cluster of Sardinian strains. In addition, Jura strains were grouped according to the cellar from which they were isolated. One Jura flor strain, MAA52, did not cluster with the other flor strains, and was thus considered as a wine strain.

To confirm the global structure observed from microsatellite typing, we used the software InStruct to detect population structure and assign the various flor strains to a particular origin. InStruct [26] is an alternative program to Structure [38] that takes into account partial self-fertilization and inbreeding; therefore, it is well suited for such an analysis because a high rate of inbreeding has been inferred from Fis values for yeast populations [22,39,40]. We selected groups of strains with sufficient members, reducing our strain set to 520, with the aim of limiting spurious clustering caused by an unbalanced effectives of the different origins. When evaluating the optimal number of ancestral lineages, DIC decreased sharply up to 9 and then continued to decrease up to 14, whereas a high variability appeared between 9 and 14 ancestral populations (FigureS1); therefore, $K = 9$ is the most probable partition inferred by InStruct. At $K = 9$, flor strains were assigned to two specific clusters (different from wine) (Figure 2). It is noteworthy that the separation of flor and wine clusters from strains of other origins already occurred at $K = 3$ (Figure S2).

The possible relationship between the different groups of flor strains can also be evaluated from the Fst genetic distance between each population. The neighbor-net network obtained with Splittree [41] from this distance matrix (Figure 3) separates clearly wine, flor and other strains into different groups, as suggested by InStruct. Interestingly, French and Hungarian flor populations are present at the end of the branches, whereas Lebanese and Spanish groups are the most basal.

Table 1. Ploidy of flor strains from various countries (Spain, Italy, Hungary, and France) estimated from the DNA content measured in Flow cytometry.

Strain	Ploidy	CV %	Strain	Ploidy	CV %
Spain			**France**		
FINO 7.7	1.9	8.0	ARC42	2.0	4.3
FINO 11.3	2.1	7.8	ARC44	2.0	5.0
FINO 1.282	2.9	4.6	ARC46	2.0	6.3
Manzanilla-II	1.9	6.1	BAE52	2.1	8.0
Manzanilla-III	2.0	7.9	CAW24	2.1	5.0
Manzanilla-VI	2.0	9.0	CBA13	2.1	12.6
Manzanilla-VIII	2.1	9.2	CBB01	2.0	4.4
Manzanilla-X	2.6	4.2	CBB52	2.0	4.3
My138	1.9	6.6	CBD05	2.1	6.9
My91	1.9	6.9	CBD55	2.1	4.5
F25	2.9	5.0	GUF54	2.0	9.5
1682-S4	2.0	4.2	LRJura	2.0	5.4
CECT11761	2.0	4.7	MAC51	1.9	8.7
CECT11764	2.0	4.8	MAD51	2.1	5.7
G1	2.0	5.2	MAE53	2.1	5.2
Italy			MAE54	2.1	5.0
2D	2.0	8.3	MAF53	2.0	5.7
FloraNero	2.1	6.1	MAF54	2.0	5.0
A33	1.9	7.7	P3	2.0	7.1
A41	2.0	6.3	PIA64	2.1	5.2
A51	2.1	7.7	PII31	2.0	4.8
A9	2.0	8.4	PII33	2.0	6.5
M23	2.1	4.8	PIN34	2.0	6.5
M3	1.9	13.0	PIO32	2.0	5.5
M38	2.1	7.9	SAA52 g	2.0	7.2
M39	2.1	7.7	SAA55	2.0	6.1
M4	2.1	4.6	SAC56	2.1	6.7
M49	2.1	4.6	XRG25	2.1	5.5
M66	2.1	5.4			
M8	2.1	6.4	**Hungary**		
V23	2.2	7.7	T19CUB	2.0	6.3
V5sard	2.1	8.5	T8CUB	2.0	5.5
V63	2.0	5.6	TA12CUB	2.1	8.9
V75	2.0	4.9	TR5CUB	2.0	5.1
V80	2.0	4.9	TS12CUB	2.2	7.4
V9	2.0	5.0			

doi:10.1371/journal.pone.0108089.t001

2. Ploidies of flor strains

Flor strains have been described as aneuploid [13,42] and variations in ploidy may explain differences in the properties of flor strains. We measured the DNA content per cell of 70 strains, which indicated that almost all strains were diploid, except three Spanish flor strains: F25 and Fino 1.28 that were triploid and Manzanilla X that was 2.6n (Table 1).

3. Comparative Genome Hybridization on array

Aneuploidy and gene amplification have been hypothesized as major sources of variation explaining adaptation to flor media [12]. We searched for a shared pattern of deletion or amplification specific to flor strains. We hybridized the genomic DNA of 11 strains of yeast to 2.0 Affymetrix chips using S288C as a reference. We tested six flor strains representing the four countries (LRJura from cluster "Jura 1", P3 from cluster "Jura 3", CECT11758 and My138 from cluster "Jerez 1", TA12CUB from Hungary, and FloraNero from Sardinia) and four French wine strains (Eg25 and UHA13 isolated in Alsace, the haploid spore V5 from the champagne strain CIVC8130, and Eg8). The wine strain Eg8, a *Saccharomyces *S. kudriavzevii* hybrid, displays substantial aneuploidy [33] and was therefore chosen to verify our ability to detect large chromosomal imbalance. In addition, this strain has a

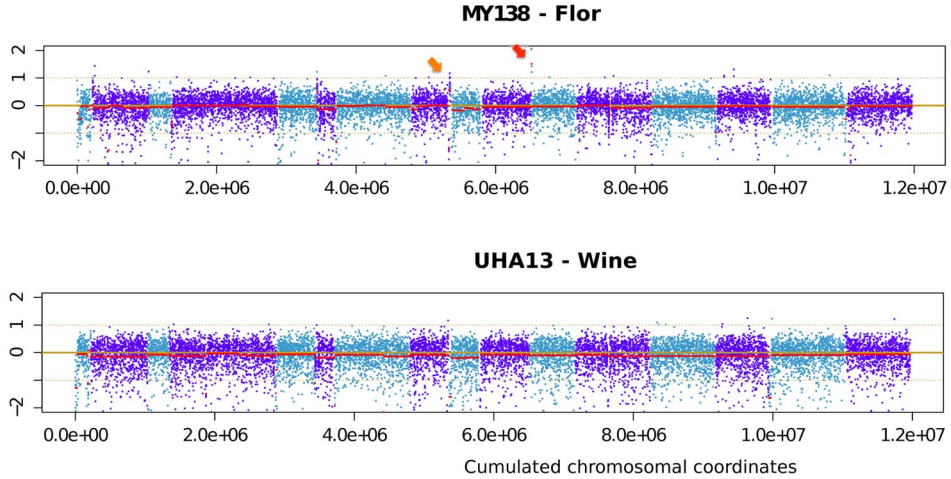


Figure 4. Karyoscope obtained with DNAcopy, showing variations in hybridization signal along the chromosome for flor strain My181 and wine strain UHA13. Chromosomes are colored in blue (uneven numbers) or dark blue (even numbers). Mean segment level estimated by DNAcopy is shown as a red line. The red arrow indicates the *YKL221W/MCH2* and *YKL222C* region, and the orange arrow indicates the *PHO12* and *IMD2* region.
doi:10.1371/journal.pone.0108089.g004

microsatellite profile indicating that the *S. cerevisiae* moiety of its genome belongs to the flor yeast group.

A first analysis carried out with different normalization methods dedicated to Affymetrix arrays (RMA, GCRMA, MAS5) indicated that 1606, 834, and 218 probe sets, respectively varied significantly between strains after correction for multiple tests (adj. p. value $<$ 0.01). Although we were able to detect the main aneuploidies of Eg8, the high gene to gene variation in hybridization necessitated the use of a sliding window smoothing over three genes to reduce noise [33]. This explains why we used directly the signal of PM probes. We then chose to evaluate variation in copy number by detecting discontinuities of log ratios along the chromosomes with the Rpackage DNAcopy [30]. The hybridization patterns of each strain and discontinuities detected with DNA copy outputs for flor strain My138 and wine strain UHA13 are presented in Figure 4; other karyotypes are shown in Figure S3. As expected, aneuploidies were detected for *S. cerevisiae x S. kudriavzevii* hybrid Eg8 [33], and for the wine strain Eg25, isolated in Alsace. The microsatellite profile of Eg25 suggests that it is also present in the flor cluster. This strain has three main aneuploidies: two at chromosome III (there is only one copy of *YCL073C* to *YCL036W*, encompassing HMLALPHA1, but three copies of *YCR028W* to *YCR102W*, encompassing HMRA1 and 2) and one at chromosome XVI (from *YPL278C* to *YPL094C*). The anomaly of chromosome XVI involves a trisomy of the left arm of the chromosome starting at *YPL094C*, close to the promoter of *SSU1 (YPL092W)*[43].

In contrast with these aneuploid strains, we did not find substantial aneuploidy in the six flor strains tested. A low hybridization signal for chromosome I suggested the presence of only one copy in the CECT11758 strain, making it the only flor strain with a typical aneuploidy. Interestingly, a low hybridization signal at each subtelomeric region leading to an inverted U hybridization profile occurred in three of the six flor strains tested (TA12CUB, P3, and FloraNero), suggesting divergent alleles or missing genes in these regions.

For all strains, the hybridization signal of several genes was lower than that of the reference strain S288C. This suggests either the existence of divergent genes or genes with a low number of copies. We defined three thresholds to differentiate regions with zero, one, two or three copies: -1, -0.38 and $+0.3$, taking into account the average values observed for aneuploidies of CECT1158, Eg25 and Eg8 strains (Chromosome I of CECT11758, chromosome III and XVI of Eg25 and chromosomes IV, V, VIII and XVI of Eg8) and the dispersions around this average ratio. Accordingly, we divided regions with a low hybridization signal into two categories according to their log ratio: log ratio between -0.38 and -1, indicating one copy, and regions with hybridization signal lower than -1, indicating no copies. The gene lists corresponding to these thresholds are shown in Table S2, and the results of the comparison of these lists is shown in Table S3. We performed a clustering of the log ratio profiles, which revealed three main clusters (Figure 5). Interestingly, the global clustering separates flor and wine strains, suggesting that flor strains share copy number variation (CNV) profiles.

Cluster A (Figure 5A) contains 109 genes with a low hybridization signal. Twenty-four genes were apparently missing in all strains (flor and wine), and another 24 were missing in eight out of the nine strains. Among these genes, the cluster containing *ASP3–1/YLR155C* and *YLR157W–E* that was missing in all strains, and the neighboring genes *ASP3–3/YLR158C, ASP3–4/YLR160C, YLR161W,* and *YLR162W* that were missing in the genome of four flor strains, were detected previously in wine isolates [44,45]. A second block of 15 genes from *HPF1*/YOL155C to *AAD15/YOL166C* on the left subtelomeric zone of chromosome XV, including the ferric enterobactin transporter *ENB1*, and the hexose transporter *HXT11/YOL156W*, is also missing among wine strains [44,45]. We also observed the loss of a block of seven genes on chromosome VII, including *MAL13/YGR288W* and *MAL11/YGR289C* which are involved in maltose metabolism, and another cluster located on the left end of chromosome X containing an isomaltose α-glucosidase *IMA5/YJL216C*, three other genes *REE1/YJL217W*, *YJL218W* and the hexose transporters *HXT9/YJL219W*. Two other subtelomeric regions detected in wine strains analyzed previously by other groups [44,45] were missing: a region containing eight genes from *HXT13/YEL069C* to *YEL075W–A* and another containing five genes from *IMA3/YIL172W* to *YIL169C*. *CUP1–2/YHR54C* was missing in all flor strains (except FloraNero), and in the wine

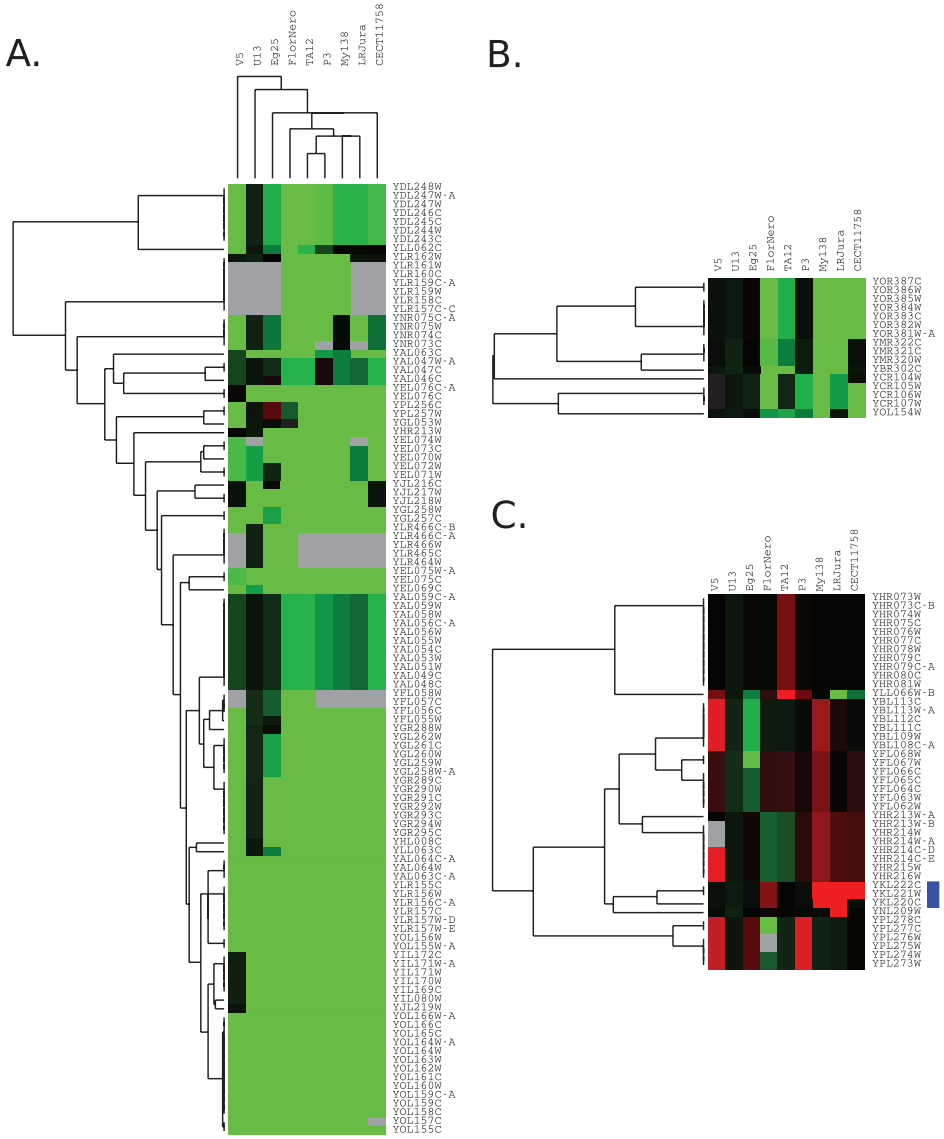


Figure 5. Hierarchical clustering of array CGH profile. Main clusters of gene with inter strain variability. A. Genes with a low hybridization signal for most strains. B. Cluster of genes with a low hybridization signal specifically for flor strains. C. Clusters of genes potentially amplified (Log ratio>0.3) in comparison with S288C.
doi:10.1371/journal.pone.0108089.g005

strain Eg25, and *CUP1–1/YHR053C* was present in only two of the six flor strains.

In addition to the set of genes showing low hybridization, some genes showed moderately low hybridization, as exemplified by two subtelomeric clusters. The first cluster includes *AAD4/YDL243C, HXT15/YDL245C, MPH2/YDL247W, SOR2/YDL246C, COS7/YDL248W, YDL247W–A,* which was only present in wine strain UHA13. Interestingly this region was noted as giving a high amplification signal for wine strains, thus differentiating wine strains from strains of another origin [46] A second cluster, *PEX22/YAL055W, GPB2/YAL056W, YAL056C–A, CNE1/YAL058W, ECM1/YAL059W, YAL059C–A, BDH1/YAL060W, BDH2/YAL061W* and, *GDH3/YAL062W,* showed moderately low hybridization for five out of six strains, whereas other wine strains presented a hybridization log ratio close to 0 for this region.

Cluster B (Figure 5B) contains genes that are either missing or present with a low copy number in the genome of flor strains. One cluster of seven genes is located close to the right end of chromosome XV and contains several genes involved in iron import into the cell. These include the siderophore retaining proteins *FIT2/YOR382C* and *FIT3/YOR383C*, the siderophore Ferric reductase *FRE5/YOR384*W, and genes with other functions: *YOR381W–A*, *YOR385*W, *PHR1/YOR386*W, and *YOR387C*. A second subtelomeric cluster contains *PAU3/YCR104W*, ADH7/*YCR105W*, and RDS1/*YCR106W*, AAD3/*YCR107W* and a third cluster located at the right end of chromosome XIII contains *YMR320W*, *YM321C* and *SNO4/YMR322C*. The low hybridization of genes from the first and second clusters was detected previously by Caretto et al. [45] in the genome of two clinical isolates.

The presence of several clusters with low hybridization signals in subtelomeric regions is puzzling. These clusters explain the typical "inverted U" observed in Figure 4 (and in Figure S3) for several chromosomes of three flor strains: TA12CUB, P3 and in particular, FloraNero.

Table 2. *FLO11* promoter and ORF polymorphisms, and hydrophobicity of the various flor strains.

Cluster	Strain	FLO11 diversity				Hydrophobicity	replicates
		Promoter	Flo11p length	mean per strain	mean per Cluster		
Jura 1	BAE52	del	3.2	3.2	3.7±0.38		
	CAH54	del	3.2	3.2			
	CAW22	del	4.2	4.2			
	CAW24	del	4.2	4.2			
	CAX22	del	3.5	3.5			
	LRJura	del	3.8	3.8		91.3±2.8	3
	PIA64	del	3.2	3.2		93.5±4.4	2
	PIN34	del	3.6	3.6		88.4±8.0	3
	SAC56	del	3.5; 4.2	3.9			
	XRG22	del	3.8	3.8		90.0±2.7	3
Jura 2	MAC51	del	3.6; 4	3.8	3.6±0.18	94.7±0.8	2
	MAD51	del	3.5	3.5			
	MAI53	del	3.5	3.5			
	XRA22	del	3.7	3.7		92.8±3.3	3
	XRC21	del	3; 3.7	3.35			
	SAC53	del	*ND				
Jura 3	GUF55	WT	4.7	4.7	4.7±0.41	90.5±6.6	3
	GUF51	WT	5.0	5		88.9±5.3	3
	CBB52	WT	4.5	4.5			
	CAV23	WT	4.7	4.7		90.7±4.8	2
	P5	WT	4.7	4.7		89.9±7.9	2
	CBD04	WT	4.8	4.8			
	GUE51	WT	4.8	4.8			
	GUG55A	WT	4.9	4.9			
	CBA13	WT	5.0	5.0			
	SAA52G	WT	6.0	6.0		94.7±1.8	2
	CBD05	WT	4.5; 5	4.8			
	CBE05	WT	4.5; 5	4.8			
	CAV21	WT	3.7; 4.5	4.1			
	ARC41	WT	4.5; 5	4.8		76.8±3.0	3
	SAA55	WT	*ND				
Jerez 2	CECT11758	del	4.5; 3.8	4.15	3.7±0.68	95.3±0.6	2
	CECT11759	del	2.7	2.7		92.4±2.8	3
	CECT11760	del	4.5; 3.8	4.15			
	CECT11763	del	*ND			94.1±0.8	3
	ET7	del	3.7	3.7		88.8±3.6	3
Jerez 1	480-SL	del	3.5	3.5	3.8±0.50	92.1±3.4	3
	481-SL	del	4.8; 2.7	3.75		94.0±3.7	3
	CECT11756	del	*ND				
	CECT11757	del	*ND				
	CECT11762	del	*ND				
	CECT12765	del	4.2	4.2			
	CECT1882	del	5; 4	4.5			
	My138	del	3.3	3.3		95.3±0.7	2
	My91	del	3.3	3.3			
Sardinia	1043	del	2.5		2.8±0.52	88.4±10.8	3
	FloraNero	del	3.4			91.8±5.4	3
	M25	del	2.5			35.4±5.5	3

Table 2. Cont.

Cluster	Strain	FLO11 diversity				Hydrophobicity	replicates
		Promoter	Flo11p length	mean per strain	mean per Cluster		
Hungary	T19CUB	del	2.0		3.0±0.62		
	T8CUB	del	3.3				
	TS12CUB	del	2.4			93.9±1.2	3
	TA12CUB	WT del	3.5			94.8±0.9	3
	TP32CUB	WT del	3.0			95.0±2.0	3
	TR05CUB	WT del	3.5				
Spanish Flor 2	G1	WT	2.2	2.2	2.2	10.0±7.8	3
Wine Cluster	MAA52	WT	2.4	2.4	2.9±0.81		
	MTF2-K1	WT	2.4	2.4		8.7±7.0	3
	RM11	WT	3.8	3.8			
Lab	S288C	WT	3.2			16.7±5.8	3

del: presence of the deletion in ICR1, WT: Wild type allele. The size of the core region of Flo11p alleles is given, as well as the mean size per strain. The mean size of Flo11p per cluster is given with standard variation. Hydrophobicity was measured according to Ishigami et al. [17], and is expressed as mean of replicates +/− standard deviation. The number of replicates is given in the last column.
*ND: could not be amplified.
doi:10.1371/journal.pone.0108089.t002

Few functional categories were associated with these genes. Genes involved in maltose metabolism were significantly affected (GO:0000023: maltose metabolic process, p. value = 4.6 e^{-6}), as well as other hexose transporters. Nine of these genes encode proteins that are located in plasma membrane (GO:0016021: integral to membrane, p. value = 0.0062), including several involved in iron uptake.

In addition to gene loss, gene amplification may also drive adaptation in response to a selective constraint [47]. We analyzed genes showing a higher hybridization signal for tested strains than for the reference control (Figure 5C); however, we found that only three genes were amplified in some, but not all, flor strains. These included *MCH2/YKL221W* and *YKL222W* that were amplified in LRJura, My138, CECT11758 and FloraNero strains (red arrows in Figure 4 and blue square in Figure 5), and *FRE2/YKL220W* in LRJura, My138, and CECT11758. The hybridization signal indicated that these genes were present in four copies in My138, LRJura and CECT11758, and three copies in FloraNero. Five other genes, *YAR064W, YAR068W, YHR214W, PHO12, IMD2* showed a high hybridization signal in three flor strains (LRJura, My138, and CECT11758). A second cluster of genes including *YHR213W–A*, *YHR213W–B*, *YHR214W–A*, *YHR214W, YHR214C–D*, *YHR214C–E*, *PHO12/YHR215W, IMD2/YHR216W* (red arrow) showed a high hybridization signal in the V5 strain, which was described previously for the wine strains EC1118 and ICV D254 [45]. The average log ratio in this region suggests three copies for LRJura, My138, and four copies for V5. Another cluster of six genes, located at the extremity of the left arm of chromosome XVI, containing the genes *SAM4/YPL273W, SAM3/YPL274C, FDH2/(YPL275W, YPL276W), YPL277C, YPL278C* was amplified in the genome of three strains (wine and flor): P3, Eg25 and V5. Dunn et al. observed previously the amplification of this region in several wine strains [44]. Another subtelomeric cluster encompassing *YFL062C* to *YFL068W* presented a high hybridization signal in strain My138 (log ratio 0.48). This was also the case for strains CECT11758, TA12CUB, FloraNero and V5; however, the log ratio for these strains was below 0.3 (between 0.23 to 0.24), thus the genes were not considered as amplified. Interestingly, except for the cluster containing *YHR073W* to *YHR081W* that was amplified only in TA12CUB, all the clusters containing amplified genes were subtelomeric.

4. Variability in velum production and *FLO11* polymorphism

The ability to develop a velum is an essential trait of flor yeast and requires high hydrophobicity at the surface of yeast cells. This trait has been related previously to polymorphisms of the *FLO11* gene [18]. Two modifications have been reported to enhance *FLO11* expression. These comprise a 111 bp deletion inside the *ICR1* non coding RNA located in the *FLO11* promoter and an increase in the size of the central part of *FLO11*. We investigated both these phenomena. First, the amplification of a fraction of *ICR1* ncRNA enabled us to detect the presence of the 111 bp deletion in the genome of 36 flor strains from the four countries, including 18 strains from the Jura 3 cluster. The cluster 2 of Spanish flor carried the wild type allele (wt) (Table 2, Figure 1). Three strains from Hungary carried both mutated and wt alleles. We sequenced the PCR amplification products of three strains originating from France (LRJura), Hungary (T8CUB), or Spain (CECT11757). Comparison of the resulting sequences with those described previously [18,48] showed that these strains had the same deletion (Figure 6) as Spanish and Sardinian flor strains. The sequencing of this locus in two strains carrying the wild type allele, one from Jura (CAV21) and one from Hungary (TR05CUB), revealed a sequence devoid of deletion and similar to S288C.

We amplified the core region of Flo11p for 59 strains and obtained DNA fragments for 53 strains, with sizes varying from 2.5 to 6 kb. We did not obtain amplification for four Spanish strains and two Jura strains. The mean size for wine strains was 2.9 kb, similar to Hungarian flor strains at 3.0 kb. The core region of Flo11p was longer in other flor groups, including Jura 1 and 2 at 3.6 kb and Jerez 1 and 2 at 3.7 kb,Jura 3 cluster strains had the longest Flo11p core region (4.8 kb). We obtained a mean value of 2.8 kb for three Sardinian strains. The size of the variable central core of *FLO11* was evaluated previously [48] with a different primer pair for Sardinian strains. These primers were closer to the

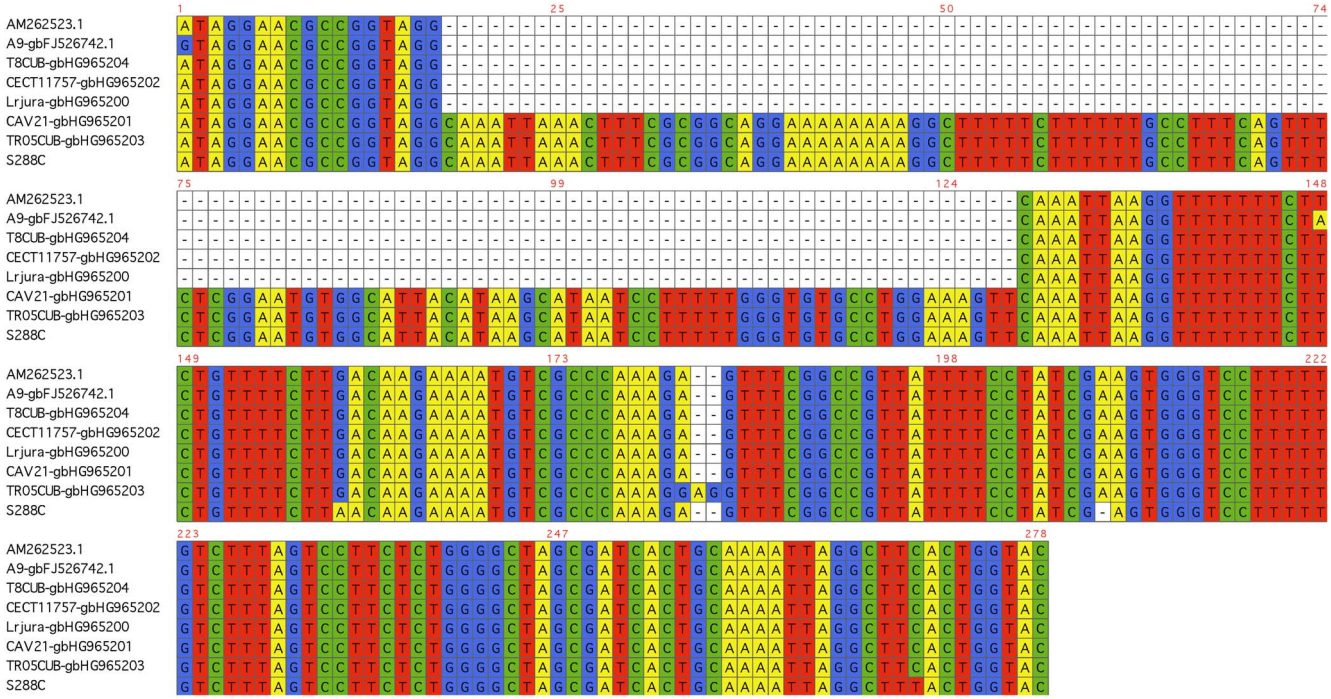


Figure 6. Alignment of 278 bp of ICR1 containing the deletion of 111 bp described by Fidalgo et al. [18]**.** The first sequences were obtained from Genbank and correspond to Spanish and an Italian flor strains [18,48] that carry this deletion. The Spanish strain CECT11758, the Hungarian strain T8CUB and the Jura strain LRJura share the same deletion. The alleles of the Jura strain CAV21 and the Hungarian strain TR05CUB are similar to that of S288C.
doi:10.1371/journal.pone.0108089.g006

central variable core of Flo11p than those used in our study; therefore, we recalculated the mean size obtained with our primers as 3.4 kb for the 22 genotyped strains, which is similar to most flor yeast groups, and lower than the size measured for Jura 3 cluster strains. This indicates that most flor strains contain a deletion in *ICR1* and have a core region of Flo11p that is longer than that of wine strains. Strains of the Jura 3 cluster have a particular combination of *FLO11* alleles with a full length *ICR1* and a very long core region.

We measured cell hydrophobicity and velum formation to examine the effect of *FLO11* polymorphism on phenotype (Table 2). We measured the hydrophobicity of 28 strains, and found highly significant differences both between strains (*p*. value of a one factor ANOVA $<2\ e^{-16}$) and groups (*p*. value of a Kruskal Walis Test = $4.63\ e^{-5}$). As expected, the flor and the "non-flor" group, including K1 wine strains, the Spanish flor 2 G1 strain and the reference strain S288C, showed the largest differences in hydrophobicity. We also found significant differences between the three Sardinian strains and the "non-flor" group (*p*. value of Kruskal Walis Test = 0.00034) but not with other flor yeast. The hydrophobicity of Jura 3 cluster strains was similar to that of other flor groups.

We assessed the ability of 29 strains to produce a velum by cultivating them on Fornachon's media. All strains of clusters Jura 1, Jura 2 and Jerez 2 produced a velum (Table 3). The growth of strains of clusters Jerez 1, Jura 3, Sardinia, and Hungary was variable. Five out of six strains from the Jura 3 cluster, and several strains from Hungary (TR05CUB, TP32CUB, TA12CUB, TS12CUB) either produced a thin velum or no velum at all. Wine strains and the two atypical Spanish flor strains were unable to develop a velum in this media. We previously correlated velum thickness and color in Jura flor yeast with genetic group assessed by interdelta typing [8]. Almost all of the strains analyzed in this prior study were genotyped; therefore, we were able to evaluate the correlation between genetic structure revealed by microsatellite typing and the ability to produce a velum for these strains. The correlation between microsatellite structure and the production of thin velum in Jura 3 cluster strains (*p*. value of χ^2 test$<6.7\ e^{-10}$ and $1.0\ e^{-07}$ for color and thickness respectively, for 55 strains) was substantially higher than that we obtained previously between velum production and delta clusters (*p*. value of χ^2 test<0.0007 and 0.0076 for color and thickness respectively) [8].

Discussion

Flor strains are found in several countries in Europe; however, until now no global approaches had been undertaken to compare strains from various vineyards. We showed previously that Jura flor strains carry a specific allele of *ITS1*, which differs from that characterized in Spanish strains [5,8], suggesting the existence of separate populations. In addition, a previous study on Spanish flor yeast revealed that flor yeast are genetically isolated from wine fermentation yeast during the aging process [7], suggesting that flor strains represent a separate family of *Saccharomyces cerevisiae*.

In this study, we used microsatellite typing, InStruct clustering and population analysis to reveal for the first time that most flor strains share the same unique origin. Lebanese and Spanish strains showed the most basal position within the population structure; therefore, it is difficult to infer the origin of flor yeast. Interestingly, a flor yeast population was recently characterized in Georgian aged wines produced by the "Kakhetian" method [49]. Nonetheless, it is still possible that all flor strains have a Mesopotamian origin because wine making is an ancient process in Georgia and this country is close to origin of vine domestication. However, the comparison of a larger number of strains is necessary. The position

Table 3. Growth of the various strains on Fornachon's media. Intensity of velum formation is scored from 0 (no velum) to 4 (thick velum).

Microsatellite	Duration	Incubation (days)				
cluster	Strain	2	4	6	8	10
Jura 1	BAE52	2	4	4	4	
	LR	1	4	4	4	
	PIN34		2	3	3	fell
	MAC51		4	4	4	4
Jura 2	GUF55		1	1	1	0
	MAD 51		4	4	4	fell
	MAI53	4	4	4	4	
Jura 3	ARC41	0	0	0	0	0
	CAV21	0	0	0	0	0
	CBD04		0	0	0	0
	GUG55	0	4	4	4	4
	GUE 51		1	0	0	1
	P5	0	0	1	1	0
Jerez 1	480 SL		0	0	0	0
	481 SL		4	4	4	0
	MY138	0	3	3	2	
Jerez 2	CECT11758	0	0	1	1	
	CECT11763	0	4	4	4	
	ET7	0	0	1	1	
Sardinia	1043	0	0	0	0	0
	Flora Nero	0	1	3	3	3
	M25	0	0	0	0	0
Hungary	T19CUB	0	3	3	3	4
	T8CUB	2	4	4	4	4
	TR05CUB	0	0	0	0	0
	TP32CUB	0	1	0	1	0
	TA12CUB	0	0	0	1	1
	TS12CUB	0	0	0	0	0
Spanish Flor 2	G1	0	0	0	0	0
Lab	S288C	0	0	0	0	0

doi:10.1371/journal.pone.0108089.t003

of the Jura strains at the end of the branch of the population Fst network suggests that Jura flor strains have a lower diversity than Spanish or Italian populations, indicating that this vineyard received strains from other vineyards. This can be seen also from InStruct clustering: Spanish flor strains are mostly mosaics of two origins (at $K = 9$), with a third origin for some strains, whereas half of Jura strain are associated with only one cluster. Hungarian strains are closely related to the Jura population as shown by the network and InStruct output. The second Spanish flor cluster is associated with some rum strains as seen from the individual tree and these strains share ancestry according the InStruct output (three individuals at the right of the Spanish flor strain cluster).

Wine is a much harsher environment than must for yeast cells during flor aging. During alcoholic fermentation, yeast cells metabolize almost all fermentable sugars and assimilate most nitrogen sources (except proline) and vitamins. As a result, wine contains a high concentration of alcohol (starting from 13% v/v in Jura, and 14–15% in Sardinia and Spain) and a low nitrogen and vitamin content. In addition, yeast cells have an aerobic biofilm lifestyle, and use glycerol and ethanol as carbon sources. Many experiments have shown how yeast are able to adapt to particular environmental conditions [50–53] through various adaptive genetic changes [54,55]. The intense stressful conditions of flor aging to which flor yeast cell are subjected for years of growth may drive such adaptation.

Aneuploidy is a mechanism that fuels adaptation to environmental changes [15,50]. Comparative Genome Hybridization on array (aCGH) has enabled the exploration of gene copy number variations. This technique revealed that wine yeast share a genomic signature [44,45]. Aneuploidies have also been detected in the genome of flor yeast [12,13] and proposed as a motor for adaptation. In addition, recent studies show that gene duplication or loss is specific of certain lineages, suggesting that it can offer a shortcut to evolutionary adaptation [47]. A recent aCGH study examined the genetic constitution of strains of different origins including one flor yeast [56]. The array technology and data processing method used in this study differs from that used here; nonetheless, findings for the triploid flor strain GB-FlorC are

similar between the two studies: Ibanez et al. found that the genes *YKL221W/MCH2* and *YKL222C* were among 81 genes showing a log ratio greater than 0.5 with S288C used as a reference strain. The *YHR215W/PHO12* gene, which is amplified in LRJura and My138, was also included in this list. In addition, for the flor strain GB-FlorC, half of all genes with a log ratio lower than −0.7 were also included in the list of genes with a low hybridization signal of flor strains analyzed here. Our investigation has two limits: (1) we cannot exclude the possibility that some genes were missed by our data analysis; and (2) our findings are limited to comparison with the S288C genome; therefore, we did not take into account genes detected specifically in wine yeast such as A, B, and C regions identified in EC1118 [57]. Our aCGH analysis and that of Ibanez et al. [56] do not support the view that many gene amplification events must occur to enable the adaptation of yeast to the flor aging environment. We hypothesize that the substantial differences observed previously [12] originate partially from differences in ploidy between the two Spanish strains and that these differences are a specific feature of this pair of strains as opposed to a general adaptive pattern. However, recent observations show that aneuploidies appear in the first steps of adaptation [58], but are subsequently replaced by other mutations, probably because of the cost of aneuploidy. Pulsed field gel electrophoresis to examine the genetic variability of wine and flor yeast has also revealed the importance of aneuploidy in yeast adaptation. It is possible that the numerous variations observed with this technique result from translocations, which can also generate new phenotypes as shown previously for *SSU1* [59], or from specific gene clusters such as those detected in EC1118 [57]. Such clusters may be inserted at different loci with a variable number of copies [60,61]. However, we successfully identified amplified genes shared by flor yeasts, including two genes: *YKL221W/MCH2* and *YKL222C*. *MCH2* is a putative monocarboxylic acid transporter with homology with mammalian transporters, although its involvement in monocarboxylic acid transport has not been shown experimentally [62]. Nonetheless, a recent study showed that this gene is important for yeast survival during the second phase of alcoholic fermentation (during alcohol accumulation) [63]. In addition, Zara et al. found that succinic, lactic and acetic acids could not provide consistent growth as a sole carbon source under aging conditions [64]. The role of *YKL222C* is also unknown; however, a recent overexpression screen to identify genes involved in endocytic trafficking, suggested a role for Ykl222cp in the early endosome or during endocytosis [65].

Several genomic regions showed a low hybridization signal indicating that these regions are missing or contain variations hampering hybridization. One of the most puzzling aspects was the location of most of these events in subtelomeric regions, which was observed previously by other groups [44,45,56]. The low number of copies of several genes in contrast with the amplification of other genes suggests translocation between subtelomeric regions. Indeed, several translocations have been shown to play a key role in the adaptation of yeast to selective pressure [15], especially in the response of wine yeast to sulfite exposure [59,66]. Unfortunately, we were unable to detect directly translocation events from our data. In addition, linkage analysis has revealed that these regions play a key role in defining individual quantitative variation and thus in the adaptation of natural populations [67].

Polymorphism of *FLO11* is also a key feature of flor strains. The global hydrophobicity of flor cells is determined by the level of *FLO11* expression and Flo11p length [18,48]. Our results are in line with these findings we correlated flor yeast population structure data with *FLO11* polymorphisms. We detected the 111 bp deletion, first observed by Fidalgo [18], in Spanish, Italian, Hungarian, and French strains, suggesting that it is extremely old. Only two Hungarian strains were heterozygote at this locus indicating that this deletion has probably been selected for by most flor strains. As a result, the wild type allele has nearly disappeared from flor strains, except in particular groups such as the Jura 3 cluster. Thus, various adaptive strategies enabling yeast cells to overcome the stressful conditions of flor aging co-exist, similar to what has been observed in experiments of adaptive evolution [54].

In conclusion, our results reveal that flor yeast are a unique family. Flor strains are mainly diploids, with some polyploid Spanish strains. We detected a shared pattern of amplification for two genes in four out of six flor strains (*MCH2* and *YKL222w*) and identified genomic regions with low hybridization to probes based on the S288C genome. These regions were mainly located in subtelomeric regions, which may be associated with a high level of divergence and thus explain adaptation to flor aging. In addition, *FLO11* polymorphisms suggest that several alternative strategies can lead to adaptation to flor aging. Further investigation is required to unravel the mechanisms of flor yeast adaptation, in particular studies involving genome sequencing.

Supporting Information

Figure S1 Evolution and variability of Deviance Information Criteria for different values of K.
(PDF)

Figure S2 Clustering of flor strains with InStruct population structure inference software for K = 3 populations. Each color corresponds to one inferred ancestral group. The proportion of each color gives the proportion of the corresponding ancestral genome in the genome of each strain. The name of the isolated population is shown at the top of each cluster.
(TIFF)

Figure S3 Karyoscope obtained with DNAcopy, showing variations in hybridization signal along the chromosome for 8 other strains: 5 flor strains P3, FloraNero, LRJura, CECT11758, TA12CUB, 2 wine strains Eg25, V5, and aneuploidy hybrid *Saccharomyces cerevisiae*S.kudriavzevii* Eg8. Chromosomes are colored in blue (uneven numbers) or dark blue (even numbers). Mean segment level estimated by DNAcopy is shown as a red line. The red arrow indicates the *YKL221W/MCH2* and *YKL222C* region, and the orange arrow indicates the *PHO12* and *IMD2* region.
(EPS)

Table S1 Origin of the different strains analyzed in this study.
(XLSX)

Table S2 List of genes showing variation in hybridization between tested strains and the reference strain S288C.
(XLSX)

Table S3 Comparison of the various gene lists obtained according to hybridization signal.
(XLSX)

Acknowledgments

The authors are extremely grateful to Prof M. Budroni, University of Sassari, Italy, Prof A. Maráz, Corvinus University of Budapest, Hungary, Dr. Maria Esther Rodríguez Jiménez, Universidad de Cádiz, Spain, Prof J. Jimenez Universidad Pablo de Olavide, Sevilla, Prof JC Garcia Mauricio,

Universidad de Cordoba Spain, Prof JA Suárez-Lepe Universidad Politécnica de Madrid, Madrid, Spain, for sending us flor strains from the different vineyards. We also acknowledge two reviewers for their constructive critics that substancially improved this manuscript.

The authors would like also to thank Anne Alais for her helpful technical assistance with microsatellite typing, Pierre Delobel for the analysis of flow cytometry data, and Dr. Frédéric Bigey for mapping affymetrix array probes on S288C reference genome.

Author Contributions

Conceived and designed the experiments: JLL CE CC. Performed the experiments: JLL CC. Analyzed the data: JLL CC. Contributed reagents/materials/analysis tools: JLL CC. Contributed to the writing of the manuscript: JLL CE CC.

References

1. Cortes MB, Moreno JJ, Zea L, Moyano L, Medina M (1998) Changes in Aroma compounds of Sherry wines during their Biological Aging carried out by Saccharomyces cerevisiae races bayanus and capensis. J Agric Food Chem 46: 2389–2394.
2. Thuy PT, Elisabeth G, Pascal S, Claudine C (1995) Optimal Conditions for the Formation of Sotolon from.alpha.-Ketobutyric Acid in the French "Vin Jaune." J Agric Food Chem 43: 2616–2619.
3. Collin S, Nizet S, Claeys Bouuaert T, Despatures P-M (2012) Main odorants in Jura flor-sherry wines. Relative contributions of sotolon, abhexon, and theaspirane-derived compounds. J Agric Food Chem 60: 380–387.
4. Pozo-Bayon MA, Moreno-Arribas MV (2011) Sherry wines. Adv Food Nutr Res 63: 17–40.
5. Esteve-Zarzoso B, Fernández-Espinar MT, Querol A (2004) Authentication and identification of Saccharomyces cerevisiae "flor" yeast races involved in sherry ageing. Antonie Van Leeuwenhoek 85: 151–158.
6. Naumova ES, Ivannikova YV, Naumov GI (2005) Genetic differentiation of the sherry yeasts Saccharomyces cerevisiae. Appl Biochem Microbiol 41: 578–582.
7. Sancho ED, Hernandez E, Rodriguez-Navarro A (1986) Presumed Sexual Isolation in Yeast Populations during Production of Sherrylike Wine. Appl Environ Microbiol 51: 395–397.
8. Charpentier C, Colin A, Alais A, Legras J-L (2009) French Jura flor yeasts: genotype and technological diversity. Antonie Van Leeuwenhoek 95: 263–273.
9. Mesa JJ, Infante JJ, Rebordinos L, Cantoral JM (1999) Characterization of Yeasts Involved in the Biological Ageing of Sherry Wines. Leb Wiss und Technol 32: 114–120.
10. Budroni M, Giordano G, Pinna G, Farris GA (2000) A genetic study of natural flor strains of Saccharomyces cerevisiae isolated during biological ageing from Sardinian wines. J Appl Microbiol 89: 657–662.
11. Martínez P, Codón AC, Pérez L, Benítez T (1995) Physiological and molecular characterization of flor yeasts: polymorphism of flor yeast populations. Yeast 11: 1399–1411.
12. Infante JJ, Dombek KM, Rebordinos L, Cantoral JM, Young ET (2003) Genome-wide amplifications caused by chromosomal rearrangements play a major role in the adaptive evolution of natural yeast. Genetics 165: 1745–1759.
13. Guijo S, Mauricio JC, Salmon JM, Ortega JM (1997) Determination of the relative ploidy in different Saccharomyces cerevisiae strains used for fermentation and "flor" film ageing of dry sherry-type wines. Yeast 13: 101–117.
14. Adams J, Puskas-Rozsa S, Simlar J, Wilke CM (1992) Adaptation and major chromosomal changes in populations of Saccharomyces cerevisiae. Curr Genet 22: 13–19.
15. Dunham MJ, Badrane H, Ferea TL, Adams J, Brown PO, et al. (2002) Characteristic genome rearrangements in experimental evolution of Saccharomyces cerevisiae. Proc Natl Acad Sci U S A 99: 16144–16149.
16. Zara S, Antonio Farris G, Budroni M, Bakalinsky AT (2002) HSP12 is essential for biofilm formation by a Sardinian wine strain of S. cerevisiae. Yeast 19: 269–276.
17. Ishigami M, Nakagawa Y, Hayakawa M, Iimura Y (2004) FLO11 is essential for flor formation caused by the C-terminal deletion of NRG1 in Saccharomyces cerevisiae. FEMS Microbiol Lett 237: 425–430.
18. Fidalgo M, Barrales RR, Ibeas JI, Jimenez J (2006) Adaptive evolution by mutations in the FLO11 gene. Proc Natl Acad Sci U S A 103: 11228–11233.
19. Zara S, Bakalinsky AT, Zara G, Pirino G, Demontis MA, et al. (2005) FLO11-based model for air-liquid interfacial biofilm formation by Saccharomyces cerevisiae. Appl Environ Microbiol 71: 2934–2939.
20. Bumgarner SL, Dowell RD, Grisafi P, Gifford DK, Fink GR (2009) Toggle involving cis-interfering noncoding RNAs controls variegated gene expression in yeast. Proc Natl Acad Sci U S A 106: 18321–18326.
21. Fornachon JC (1953) Studies on the sherry flor. Australian Wine Board A, editor Adelaide.
22. Legras JL, Merdinoglu D, Cornuet JM, Karst F (2007) Bread, beer and wine: Saccharomyces cerevisiae diversity reflects human history. Mol Ecol 16: 2091–2102.
23. Legras JL, Karst F (2003) Optimisation of interdelta analysis for Saccharomyces cerevisiae strain characterisation. FEMS Microbiol Lett 221: 249–255.
24. Takezaki N, Nei M (1996) Genetic distances and reconstruction of phylogenetic trees from microsatellite DNA. Genetics 144: 389–399.
25. Tamura K, Peterson D, Peterson N, Stecher G, Nei M, et al. (2011) MEGA5: Molecular Evolutionary Genetics Analysis using Maximum Likelihood, Evolutionary Distance, and Maximum Parsimony Methods. Mol Biol Evol 28: 2731–2739.
26. Gao H, Williamson S, Bustamante CD (2007) A Markov chain Monte Carlo approach for joint inference of population structure and inbreeding rates from multilocus genotype data. Genetics 176: 1635–1651.
27. R Development Core Team (2011) R: A language and environment for statistical computing. R Found Stat Comput Vienna, Austria ISBN 3–900051–07–0.
28. Winzeler EA, Richards DR, Conway AR, Goldstein AL, Kalman S, et al. (1998) Direct allelic variation scanning of the yeast genome. Science 281: 1194–1197.
29. Liu W, Mei R, Di X, Ryder TB, Hubbell E, et al. (2002) Analysis of high density expression microarrays with signed-rank call algorithms. Bioinformatics 18: 1593–1599.
30. Venkatraman ES, Olshen AB (2007) A faster circular binary segmentation algorithm for the analysis of array CGH data. Bioinformatics 23: 657–663.
31. Eisen MB, Spellman PT, Brown PO, Botstein D (1998) Cluster analysis and display of genome-wide expression patterns. Proc Natl Acad Sci U S A 95: 14863–14868.
32. Nogales-Cadenas R, Carmona-Saez P, Vazquez M, Vicente C, Yang X, et al. (2009) GeneCodis: interpreting gene lists through enrichment analysis and integration of diverse biological information. Nucleic Acids Res 37: W317–22.
33. Erny C, Raoult P, Alais A, Butterlin G, Delobel P, et al. (2012) Ecological Success of a Group of Saccharomyces cerevisiae/Saccharomyces kudriavzevii Hybrids in the Northern European Wine-Making Environment. Appl Environ Microbiol 78: 3256–3265.
34. Delobel P, Tesnière C (2014) A Simple FCM Method to Avoid Misinterpretation in Saccharomyces cerevisiae Cell Cycle Assessment between G0 and Sub-G1. PLoS One 9: e84645.
35. Haase SB, Reed SI (2002) Improved flow cytometric analysis of the budding yeast cell cycle. Cell Cycle 1: 132–136.
36. Stefanini I, Dapporto L, Legras J-L, Calabretta A, Di Paola M, et al. (2012) Role of social wasps in Saccharomyces cerevisiae ecology and evolution. Proc Natl Acad Sci U S A 109: 13398–13403.
37. Liti G, Carter DM, Moses AM, Warringer J, Parts L, et al. (2009) Population genomics of domestic and wild yeasts. Nature 458: 337–341.
38. Pritchard JK, Stephens M, Donnelly P (2000) Inference of population structure using multilocus genotype data. Genetics 155: 945–959.
39. Cubillos FA, Vásquez C, Faugeron S, Ganga A, Martínez C (2009) Self-fertilization is the main sexual reproduction mechanism in native wine yeast populations. FEMS Microbiol Ecol 67: 162–170.
40. Goddard MR, Anfang N, Tang R, Gardner RC, Jun C (2010) A distinct population of Saccharomyces cerevisiae in New Zealand: evidence for local dispersal by insects and human-aided global dispersal in oak barrels. Environ Microbiol 12: 63–73.
41. Huson DH, Bryant D (2006) Application of phylogenetic networks in evolutionary studies. Mol Biol Evol 23: 254–267.
42. Alexandre H (2013) Flor yeasts of Saccharomyces cerevisiae-their ecology, genetics and metabolism. Int J Food Microbiol. 167: 269–75.
43. Brion C, Ambroset C, Sanchez I, Legras J-L, Blondin B (2013) Differential adaptation to multi-stressed conditions of wine fermentation revealed by variations in yeast regulatory networks. BMC Genomics 14: 681.
44. Dunn B, Levine RP, Sherlock G (2005) Microarray karyotyping of commercial wine yeast strains reveals shared, as well as unique, genomic signatures. BMC Genomics 6: 53.
45. Carreto L, Eiriz MF, Gomes AC, Pereira PM, Schuller D, et al. (2008) Comparative genomics of wild type yeast strains unveils important genome diversity. BMC Genomics 9: 524.
46. Dunn B, Richter C, Kvitek DJ, Pugh T, Sherlock G (2012) Analysis of the Saccharomyces cerevisiae pan-genome reveals a pool of copy number variants distributed in diverse yeast strains from differing industrial environments. Genome Res 22: 908–924.
47. Ames RM, Rash BM, Hentges KE, Robertson DL, Delneri D, et al. (2010) Gene duplication and environmental adaptation within yeast populations. Genome Biol Evol 2: 591–601.
48. Zara G, Zara S, Pinna C, Marceddu S, Budroni M (2009) FLO11 gene length and transcriptional level affect biofilm-forming ability of wild flor strains of Saccharomyces cerevisiae. Microbiology 155: 3838–3846.
49. Capece A, Siesto G, Poeta C, Pietrafesa R, Romano P (2013) Indigenous yeast population from Georgian aged wines produced by traditional "Kakhetian" method. Food Microbiol 36: 447–455.
50. Gresham D, Desai MM, Tucker CM, Jenq HT, Pai DA, et al. (2008) The repertoire and dynamics of evolutionary adaptations to controlled nutrient-limited environments in yeast. PLoS Genet 4: e1000303.

51. Gresham D, Usaite R, Germann SM, Lisby M, Botstein D, et al. (2010) Adaptation to diverse nitrogen-limited environments by deletion or extrachromosomal element formation of the GAP1 locus. Proc Natl Acad Sci U S A 107: 18551–18556.
52. Brauer MJ, Saldanha AJ, Dolinski K, Botstein D (2005) Homeostatic adjustment and metabolic remodeling in glucose-limited yeast cultures. Mol Biol Cell 16: 2503–2517.
53. Ferea TL, Botstein D, Brown PO, Rosenzweig RF (1999) Systematic changes in gene expression patterns following adaptive evolution in yeast. Proc Natl Acad Sci U S A 96: 9721–9726.
54. Kao KC, Sherlock G (2008) Molecular characterization of clonal interference during adaptive evolution in asexual populations of Saccharomyces cerevisiae. Nat Genet 40: 1499–1504.
55. Lang GI, Rice DP, Hickman MJ, Sodergren E, Weinstock GM, et al. (2013) Pervasive genetic hitchhiking and clonal interference in forty evolving yeast populations. Nature 500: 571–574.
56. Ibáñez C, Pérez-Torrado R, Chiva R, Guillamón JM, Barrio E, et al. (2014) Comparative genomic analysis of Saccharomyces cerevisiae yeasts isolated from fermentations of traditional beverages unveils different adaptive strategies. Int J Food Microbiol 171: 129–135.
57. Novo M, Bigey F, Beyne E, Galeote V, Gavory F, et al. (2009) Eukaryote-to-eukaryote gene transfer events revealed by the genome sequence of the wine yeast Saccharomyces cerevisiae EC1118. Proc Natl Acad Sci 106: 16333–16338.
58. Chang S-L, Lai H-Y, Tung S-Y, Leu J-Y (2013) Dynamic Large-Scale Chromosomal Rearrangements Fuel Rapid Adaptation in Yeast Populations. PLoS Genet 9: e1003232.
59. Pérez-Ortín JE, Querol A, Puig S, Barrio E (2002) Molecular characterization of a chromosomal rearrangement involved in the adaptive evolution of yeast strains. Genome Res 12: 1533–1539.
60. Borneman AR, Desany BA, Riches D, Affourtit JP, Forgan AH, et al. (2011) Whole-Genome Comparison Reveals Novel Genetic Elements That Characterize the Genome of Industrial Strains of Saccharomyces cerevisiae. PLoS Genet 7: e1001287.
61. Galeote V, Bigey F, Beyne E, Novo M, Legras J-L, et al. (2011) Amplification of a Zygosaccharomyces bailii DNA Segment in Wine Yeast Genomes by Extrachromosomal Circular DNA Formation. PLoS One 6: e17872.
62. Makuc J, Paiva S, Schauen M, Krämer R, André B, et al. (2001) The putative monocarboxylate permeases of the yeast Saccharomyces cerevisiae do not transport monocarboxylic acids across the plasma membrane. Yeast 18: 1131–1143.
63. Novo M, Mangado A, Quirós M, Morales P, Salvadó Z, et al. (2013) Genome-Wide Study of the Adaptation of Saccharomyces cerevisiae to the Early Stages of Wine Fermentation. PLoS One 8: e74086.
64. Zara S, Gross MK, Zara G, Budroni M, Bakalinsky AT (2010) Ethanol-independent biofilm formation by a flor wine yeast strain of Saccharomyces cerevisiae. Appl Environ Microbiol 76: 4089–4091.
65. Arlt H, Perz A, Ungermann C (2011) An overexpression screen in Saccharomyces cerevisiae identifies novel genes that affect endocytic protein trafficking. Traffic 12: 1592–1603.
66. Zimmer A, Durand C, Loira N, Durrens P, Sherman DJ, et al. (2014) QTL Dissection of Lag Phase in Wine Fermentation Reveals a New Translocation Responsible for Saccharomyces cerevisiae Adaptation to Sulfite. PLoS One 9: e86298.
67. Cubillos FA, Billi E, Zörgö E, Parts L, Fargier P, et al. (2011) Assessing the complex architecture of polygenic traits in diverged yeast populations. Mol Ecol 20: 1401–1413.

Steyer et al. BMC Genomics 2012, 13:573
http://www.biomedcentral.com/1471-2164/13/573

BMC Genomics

RESEARCH ARTICLE Open Access

QTL mapping of the production of wine aroma compounds by yeast

Damien Steyer[1,2,3], Chloe Ambroset[4,5,6], Christian Brion[4,5,6], Patricia Claudel[1,2], Pierre Delobel[4,5,6], Isabelle Sanchez[4,5,6], Claude Erny[7], Bruno Blondin[4,5,6], Francis Karst[1,2] and Jean-Luc Legras[4,5,6*]

Abstract

Background: Wine aroma results from the combination of numerous volatile compounds, some produced by yeast and others produced in the grapes and further metabolized by yeast. However, little is known about the consequences of the genetic variation of yeast on the production of these volatile metabolites, or on the metabolic pathways involved in the metabolism of grape compounds. As a tool to decipher how wine aroma develops, we analyzed, under two experimental conditions, the production of 44 compounds by a population of 30 segregants from a cross between a laboratory strain and an industrial strain genotyped at high density.

Results: We detected eight genomic regions explaining the diversity concerning 15 compounds, some produced *de novo* by yeast, such as nerolidol, ethyl esters and phenyl ethanol, and others derived from grape compounds such as citronellol, and cis-rose oxide. In three of these eight regions, we identified genes involved in the phenotype. Hemizygote comparison allowed the attribution of differences in the production of nerolidol and 2-phenyl ethanol to the *PDR8* and *ABZ1* genes, respectively. Deletion of a *PLB2* gene confirmed its involvement in the production of ethyl esters. A comparison of allelic variants of *PDR8* and *ABZ1* in a set of available sequences revealed that both genes present a higher than expected number of non-synonymous mutations indicating possible balancing selection.

Conclusions: This study illustrates the value of QTL analysis for the analysis of metabolic traits, and in particular the production of wine aromas. It also identifies the particular role of the *PDR8* gene in the production of farnesyldiphosphate derivatives, of *ABZ1* in the production of numerous compounds and of *PLB2* in ethyl ester synthesis. This work also provides a basis for elucidating the metabolism of various grape compounds, such as citronellol and cis-rose oxide.

Keywords: Saccharomyces cerevisiae, QTL mapping, Wine aroma, Citronellol, Rose oxide, Nerolidol, Farnesene, Ethyl esters, 2-phenyl ethanol, *PDR8*, *ABZ1*, *PLB2*, *QDR2*

Background

The wide diversity that can be observed among individuals of the same species is one of the most remarkable aspects of life. Deciphering the mechanisms explaining this phenotypic variety is among the major aims of evolutionists and geneticists. Quantitative genetics has been applied to untangle these issues, and over the last 30 years numerous studies have illustrated the power of these genetic approaches, and in particular quantitative trait locus (QTL) mapping, with the characterization of many genomic regions linked to or containing genes responsible for quantitative variations in a phenotype. These approaches have been extensively used in plant and cattle breeding programs; they have contributed to the understanding of resistance to several diseases [1] and also led to a significant improvement in crop yields and cattle breeding. Surprisingly quantitative genetic approaches have been applied only recently to budding yeast, initially to elucidate various complex mechanisms, including sporulation efficiency [2], thermotolerance [3,4], and drug resistance [5]. Even more recently, this quantitative approach has been used to decipher complex traits [6,7] at high resolution [8,9]. It has also been applied successfully to the analysis of variations in gene

* Correspondence: jean-luc.legras@supagro.inra.fr
[4]INRA, UMR1083, Montpellier F-34060, France
[5]Montpellier SupAgro, UMR1083, Montpellier F-34060, France
Full list of author information is available at the end of the article

expression [10,11]. The QTL approach is now being used to study features important for the beverages industry, for example wine fermentation [12,13], sake technological traits [14] and ethanol tolerance for ethanol production [15]. All these studies have implicated defective alleles, of for example *AMN1* [10], *ASP1* [16] or *ABZ1* [11], in the diversity of the phenotypes of segregants. Strain By4741 possesses a defective allele of *AMN1* which leads to faster daughter cell separation; wine strain SB possesses a defective allele of the *ASP1* gene involved in asparagine catabolism; and S288C possesses a defective allele of *ABZ1* that codes for an enzyme which catalyzes the synthesis of 4-amino-4- deoxychorismate from chorismate, a step in the synthesis of paraminobenzoic acid. This defective allele of *ABZ1* modulates the fermentation rate by controlling nitrogen utilization [11].

Wine aroma is complex and results from the blending of numerous compounds synthesized by vines, some of which are transformed by yeast, together with compounds directly produced by yeast as a result of its primary metabolism [17,18]. The metabolic pathways leading to the synthesis of these yeast volatile compounds are numerous and incompletely described. The roles of some of the key genes, such as *ATF1* for acetates and *EEB1* for ethyl esters, has been demonstrated [19,20]. Nevertheless, little is known about the factors explaining large strain-to-strain differences in the production of volatile compounds [21-23]. Holistic approaches [24] have given new insights into the roles of various key genes in the diversity of production of some volatile compounds. Further work from the same group highlighted how a few key players, such as transcription factors, may explain some of the differences between strains [25].

To analyze the differences in the production of wine aroma compounds linked to yeast strain diversity, we used QTL analysis with a population of 30 segregants arising from a cross between the laboratory strain S288C and 59A, a spore isolated from the industrial wine strain EC1118. This population of segregants has been genotyped with Affymetrix YGS98 microarrays to obtain a high density genetic map and was used for the first quantitative analysis of transcriptome variations during enological fermentation [11]. We tested this population of segregants for the production of aromatic compounds in two different experimental conditions: synthetic musts mimicking white and red wine fermentations. These analyses enabled us to detect the involvement of eight genomic regions in the production of various volatile compounds explaining 39% to 72% of the diversity. As examples, we characterized the role of two genes by hemizygote analysis and identified another candidate gene by analysis of the phenotype of a deleted mutant. Our findings provide new insights into the genetic architecture underlying the production of wine aroma by yeast.

Results

Each of the 30 segregants was tested in two experimental designs. In the first design (experiment A), white wine fermentation was simulated by fermentation at 20°C in medium with a low lipid content, whereas in the second design, mimicking red wine fermentation [11], the fermentations were run at 28°C in medium with a high lipid content (experiment B).

The fermentation kinetics of the 30 segregants presented significant diversity, from typical wine fermentations to clearly sluggish as observed for S288C (which presented the longest fermentation). Unlike the parental strains, several segregants presented a clear ability to flocculate. This resulted in large and significant diversity in the concentrations of volatile compounds at the end of the alcoholic fermentations. We measured a set of 27 compounds in experiment A, and 33 compounds in experiment B. We performed a principal component analysis to reduce the multidimensional data set of experiment A into three more informative dimensions (Figure 1). The first three axes explained 51% of the global variance (37.6% for axes 1 and 2 in Figure 1A and 33.7% for axes 1 and 3 in Figure 1B) and in this analysis the various compounds are grouped according to chemical family. Ethyl esters and medium chain fatty acids were correlated to the first axis, the various acetates correlated together with the second axis and 2-phenyl ethanol and isoamylalcohol were correlated to axis three. The representation of individual progeny strains in the factorial plan indicated a substantial diversity in the ability to produce volatile compounds. This was especially clear for acetates, as many strains were able to produce more acetates in the media than either of the two parents (S288C and 59A). In addition, some strains more than others metabolized geraniol into citronellol or into the high olfactive impact compound cis-rose oxide. This indicates that the characteristics of the yeast strain have a significant and variable impact on the grape aroma fraction. A similar picture was obtained from the analysis of experiment B.

Genetic analysis of the volatile compounds production

From experiment A, heritability was estimated to be greater than 70% for 21 of 27 compounds, which included the grape aroma compounds geraniol, linalool, citronellol and geraniol acetate.

To identify QTL for these technological features, we performed a linkage analysis with the previously reported genotypes for these progeny [11]. The concentrations of most compounds obtained for the population of segregants did not follow normal distributions (Additional file 1), so we performed linkage analysis with both parametric and non parametric models. We identified four and six regions involved in variations in the production of different compounds in experiments A and B, respectively

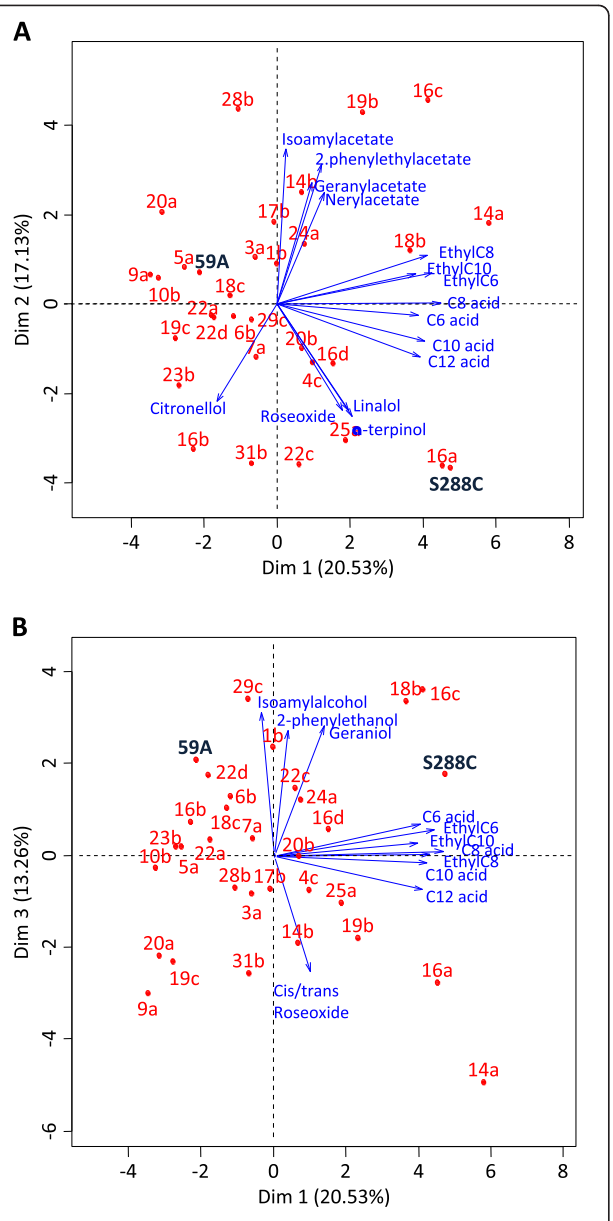


Figure 1 Principal component analysis presenting the variability in the concentration of volatile compounds produced by the various segregants and parent strains (experiment A). A: Components 1 and 2 representing 37.7% of the global variation; **B:** Components 1 and 3 representing 33.8% of the global variation. Segregants are indicated in red, aroma compound vectors are given in blue.

(Tables 1 and 2, respectively). However, additional factors may contribute to wine aroma production: flocculation is one [26]; and the presence of the *ABZ1*-S288C allele, which is responsible for large variations in fermentation kinetics [11], may also have an indirect effect. To overcome the potential effects of these factors, we performed a second linkage analysis taking these two factors into account as covariables in the model. This enabled us (i) to improve the significance for some QTL detected after a simple scan (such as for ethyl octanoate), (ii) to detect a genetic effect for additional compounds (ethyl hexanoate) of one region already found and (iii) to detect three and one additional genomic regions in experiments A and B, respectively, for other compounds. The effects of flocculation and of *ABZ1* allele on aroma production for each QTL are given as Additional file 2.

For several compounds we were unable to identify any QTL despite a high heritability. This were the case for instance for isoamyl-alcohol and its acetate ester. In the case of isoamyl-alcohol, this might be due to two isomeric compounds (3-methyl-1-butanol and 2-methyl-1-butanol) both being involved. Nevertheless, we detected several regions involved in the diversity of the production of various compounds in the acid, alcohols ethyl ester and isoprenoid chemical families. As a whole, these metabolic QTL (mQTL) explained between 43 and 73% of the metabolite variation.

One region on chromosome XII was identified in both experiments with high Lod score values and explained as much as 46 and 72% of the variations in the production of nerolidol. The same region was identified for other isoprenoids characterized only in design B. Another region, on chromosome XIV, was also detected in both experiments and was associated with various phenotypes: ethyl hexanoate in experiment A and several compounds (including 2-phenyl ethanol and ethyl octanoate) in experiment B. The *ABZ1* gene which maps in this region has been reported to be responsible for variations in the rate of fermentation [11]. When *ABZ1* polymorphism was used as a covariable, we detected other QTL for more compounds indicating that *ABZ1* allelic variations may affect the production of several metabolites. These mQTL are scattered through the genome of strain S288C such that we did not detect a major region associated with all the compounds analyzed (Figures 2 and 3).

The other regions detected for several compounds were each found in only one experiment. These regions also explained a smaller part of the variations in the production of the various volatile metabolites despite a high heritability. Possibly, the production of most of these compounds is under multigenic control and the corresponding regions cannot be detected with such a small population of segregants.

In addition to compounds produced by yeast during alcoholic fermentation, we also studied the fate of geraniol, which is present in grape must at up to 3 mg/L [27]. We did not detect any region explaining variations in the metabolism of geraniol. However, we detected one QTL explaining 43% of the variations in the concentration of citronellol a compound produced from geraniol during alcoholic fermentation. The synthetic pathway for citronellol has not been clearly described and our

Table 1 QTL analysis of volatile compounds produced during alcoholic fermentation (experiment A, geraniol 5mg/L)

Compounds	Localization	Single QTL scan		Flocculation as a covariable		Fraction of variation explained by the QTL	Heritability
		LOD	p-value	LOD	p-value		
Isoamyl acetate							33
Isoamyl alcool							83
Ethyl hexanoate	Chr XIV 634-687			4.52	0.034	50.1	28
Ethyl octanoate							-
Ethyl decanoate							71
Ethyl myristate							70
2-phenylethyl acetate							80
2 phenyl ethanol	Chr VIII 422-469	3.23	0.04			39.1	99
Hexanoic acid							-
Octanoic acid							90
Decanoic acid							99
Myristic acid							97
Ethyl 9-decenoate							99
Nerolidol	Chr XII 675-704	8.28	<0.004			71.9	93
Farnesol							74
Ethyl 3-hydoxydecanoic acid							-
α-terpineol							22
Linalol							78
Citronellol	Chr XIII 290-342	3.69	0.033			43.3	78
Geraniol							98.
Nerol							-
Citronellyl acetate							79
Geranyl acetate							87
nerylacetate							63
Isobutanol							-
Cis-rose oxide	Chr I 21-55			4.59	0.02	51.1	62
Cis-rose oxide	Chr VII 47-85			4.27	0.04	48.6	
Trans-rose oxide							90
Cis/trans rose oxide ratio	Chr XIV 537-589	3.94	0.01			45.4	99

results may indicate new targets to investigate. Two other QTL explained variations in the content of cis-rose oxide and in the ratio between cis and trans isomers of rose oxide. These QTLs may be technologically interesting as the two isomers of rose oxide present different olfactive thresholds.

Evaluation of the role of various candidate genes in the QTL

***PDR8* is responsible for variations in nerolidol production during alcoholic fermentation** The major QTL responsible for variations in the concentration of nerolidol in experiment A and of nerolidol, farnesene and bisabolene in experiment B maps to a short region of 20 kb containing 26 ORFs. Nerolidol, farnesene and bisabolene are all derived from farnesyl diphosphate, an intermediate in isoprenoid and ergosterol biosynthesis (Figure 2): at acidic pH, the instability of the diphosphate group leads to the release of farnesol and its isomer nerolidol.

It seemed likely that the gene involved in the modulation of nerolidol, farnesene and bisabolene production is involved in ergosterol biosynthesis or in farnesol/nerolidol transport because of the size of these molecules. One of the genes mapping in this region is *PDR8*, a transcription factor that modulates the expression of 16 genes [28] including transporters (*AZR1*, *PDR15*, *QDR2*, *YOR1*),

Table 2 QTL analysis of volatile compounds produced during alcoholic fermentation (experiment B, Ambroset et al. 2011)

Compound	Localization	Single QTL scan		Flocculation (1) or *ABZ1* (2) as a covariable		Fraction of variation explained by the QTL
	(coordinates in kb)	LOD	p-value	LOD	p-value	
Ethyl octanoate	Chr XIII 255-305	3.80	0.044	4.86	0.007 (**2**)	46.4
Ethyl decanoate	Chr XIII 245-304			4.62	0.024 (**2**)	
Ethyl myristate	Chr XIII 230-290	3.92	0.010			47.6
2-phenyl ethanol	Chr XIV 657-702	4.01	0.022			48.3
Dodecanoic acid	Chr VII 332-370			5.45	0.039 (**1**)	54.2
Nerolidol	Chr XII 674-705	3.94	0.004			46.0
Isoamyl octanoate	Chr VIII 423-481	3.46	0.033			43.4
Methyl oleate	Chr XIII 234-285	3.79	0.029			46.4
Farnesol						
(E,E)-Farnesol						
(E,Z)- or (Z,E)-Farnesol	Chr II 593-646	4.23	0.018	4.7	0.022 (**2**)	50.2
trans-β-farnesene	Chr XII 711-750	3.45	0.050			
(Z,E)-α-farnesene	Chr XII 693	3.31	0.050			
α-bisabolene	*Chr XII 735*	*3.46*	*0.059*			
β-bisabolene	Chr XII 706-757	3.71	0.011	3.89		45.6
(E,E)-α-farnesene	Chr XII 675-704	3.47	0.044	3.62		43.5

Regions above the 0.05 threshold are indicated in italics. Only compounds with differences that are significant or close to the significance threshold are given.

a gene of the ergosterol biosynthesis pathway (*ERG8*), and enzymes involved in oxido-reduction processes (*CTT1*, *GTT2*, *YMR315w*). This transcription factor was clearly a good candidate. The nucleotide sequences of the *PDR8* genes in strains S288C and 59A show numerous single nucleotide differences. These SNPs generate five non-synonymous substitutions between the Pdr8p proteins in 59A and S288C.

To confirm the involvement of the *PDR8* gene in the observed phenotype, we compared two reciprocal

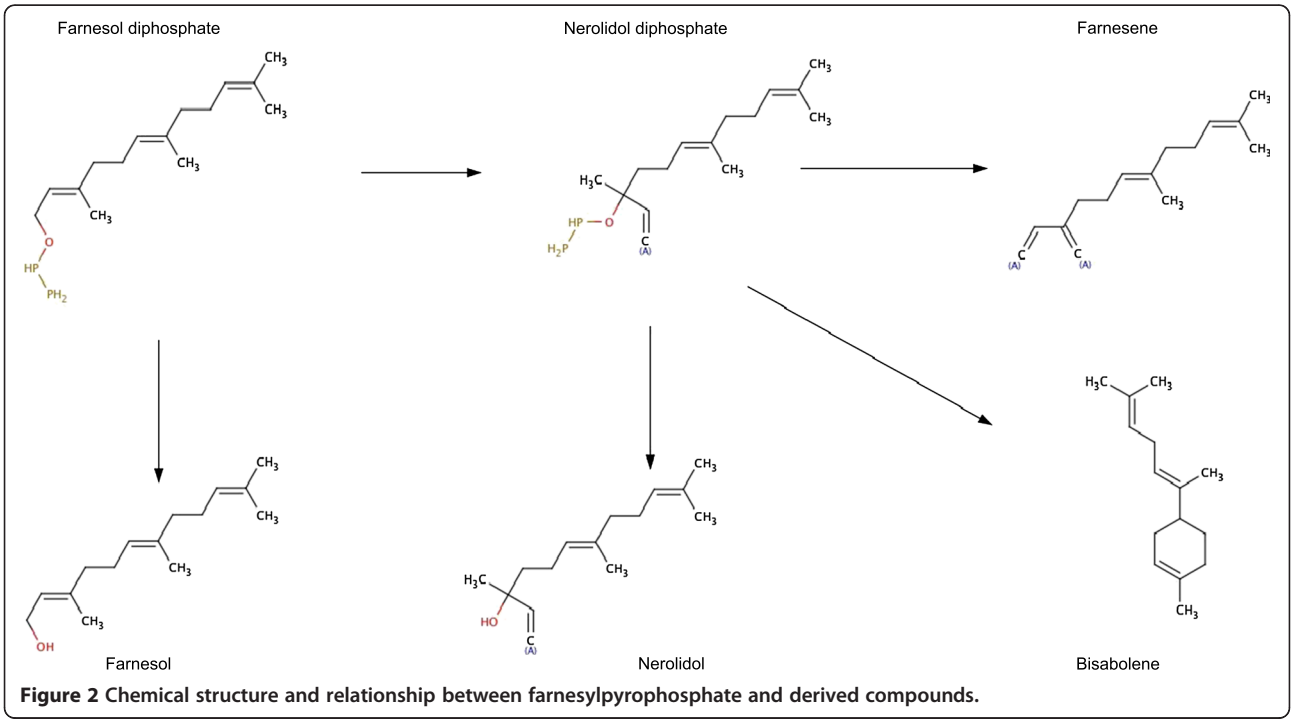


Figure 2 Chemical structure and relationship between farnesylpyrophosphate and derived compounds.

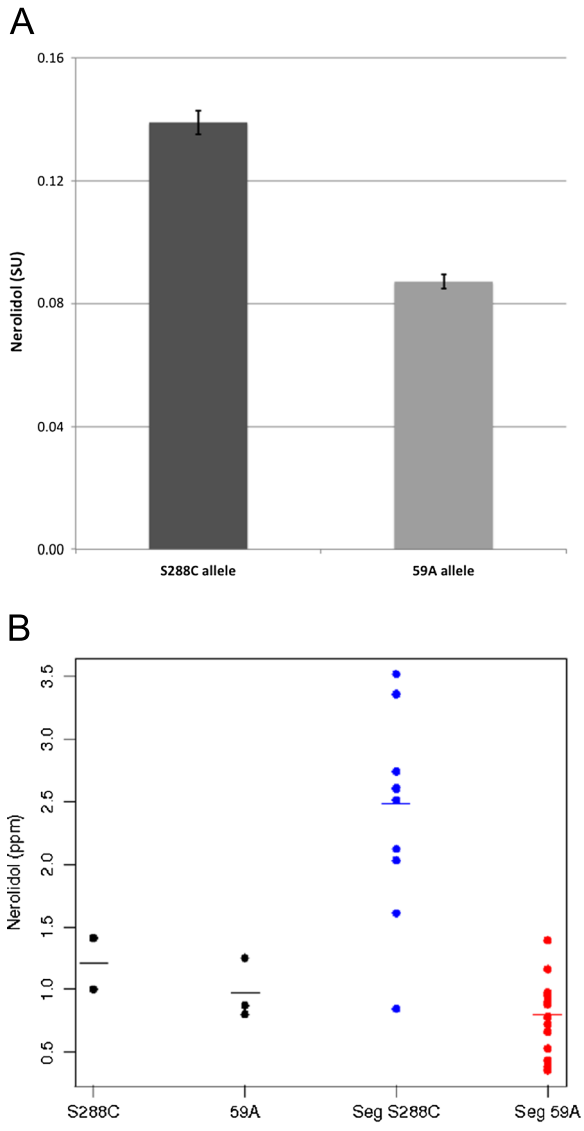


Figure 3 A: Production of nerolidol by two hemizygote strains carrying only the 59A or the S288C - *PDR8* allele. Dark gray: S288C, Light Gray: 59A. Differences are highly significant (p value<0.001). For each experiment, three fermentations were performed as described for experiment B. **B**: Production of nerolidol by the different segregants, relative to that measured for the parental strains indicated as 59A and S288C. (Data from experiment A). Segregants carrying 59A allele of *PDR8* are indicated as Seg 59A, and segregants carrying S288C allele of *PDR8* are indicated as Seg S288C.

hemizygotes between S288C and 59A containing only one of the parental alleles. These hemizygotes presented the different phenotypes observed in the population of segregants, with the enological *PDR8* allele of 59A leading to a lower production of nerolidol (Figure 3A). These results are in agreement with those obtained for the whole population (Figure 3B). However, the parental strain S288C produced less nerolidol than most of the segregants, indicating further interactions with the genetic background.

Characterization of PDR8 targets involved in the phenotype

To identify which of the targets of *PDR8* explain the observed variations in nerolidol production, we measured the production of nerolidol by the corresponding 16 deletant strains in the By4741 background (Figure 4). The deletion of *YOR1* led to an increase of nerolidol production whereas the deletion of four other *PDR8* target genes (*QDR2*, *PDR15*, *GPH1* and *YMR135W*) led to decreases of nerolidol production similar to that observed after the deletion of *PDR8*. The genes *QDR2* and *PDR15* encode transporters that may be involved in the export of nerolidol or derived compounds from the cell. The deletion of two other genes, *GPH1* and *YMR315W*, resulted in a similar reductions in nerolidol production indicating other possible mechanisms. *GPH1* is a glycogen phosphorylase required for the mobilization of glycogen, and *YMR315W* is an oxidoreductase enzyme thay may be involved in the reduction of farnesol (data not shown). *ERG8*, encoding mevalonate phosphate kinase, is an essential gene for isoprenoid and ergosterol biosynthesis, so it was not possible to conduct the appropriate tests with the deleted haploid strain. Deletion of only one copy of *ERG8*, in the diploid strain By4743, did not lead to any relevant change so we evaluated the effect of the overexpression of *ERG8*: no significant increase of the production of nerolidol was detected (data not shown). We did not detect any variation in the expression of *QDR2*,

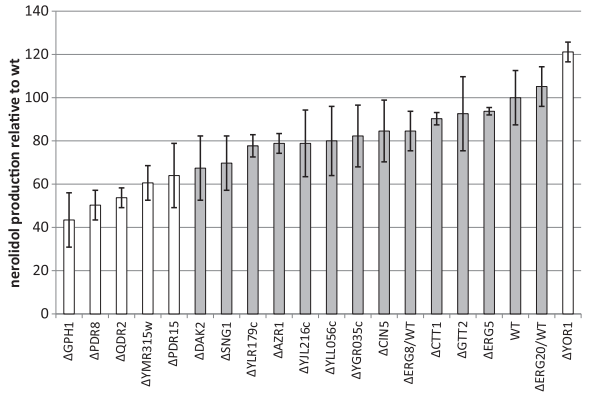


Figure 4 Production of nerolidol by mutant strains deleted for *PDR8* targets, reported relative to the production by the wild-type strain By4741. The Dunnett test was used to compare the production by each deletant strain to that by the corresponding wild type strain (p-value <0.05). Each experiment was repeated at least twice. For By4742 ΔERG20/wt and BY4742 ΔERG8/wt the wild-type control strain is By4743. Bars in gray indicate that the production was not significantly different to that by the wt. White bars indicate that the production of nerolidol was significantly different from that by wt.

PDR15, *GPH1* and *YMR135W* associated with the *PDR8* allelic form reported in the experiment by Ambroset et al. [11], probably because of the high FDR rate. Therefore, we replaced the *PDR8* allele in strain 59A and we compared the expression of these four genes between the strains containing each of the two alleles of *PDR8*. Quantative PCR (Figure 5) indicated that only *QDR2* was more strongly expressed in the strain carrying the S288C-*PDR8* allele.

ABZ1 allelic variations affect production of 2-phenylethanol and ethyl esters by yeast during fermentation

The variations in the concentrations of 2-phenyl ethanol and of ethyl hexanoate esters were linked to another mQTL corresponding to a 33 kb region of chromosome XIV. This region overlaps a region involved in differences in fermentation kinetics due to allelic variations of the *ABZ1* gene [11]. There are five non synonymous mutations between the S288C and 59A alleles of *ABZ1*. We compared two reciprocal hemizygotes between strains S288C and 59A containing only one allele of each origin to confirm the role of this gene in 2-phenylethanol production. The hemizygote which carried the enological allele of 59A, produced more 2-phenylethanol than the hemizygote which carried the S288c allele (Figure 6) . The addition of 1 mg/L of p-aminobenzoic acid to the fermentation media suppressed the differences in the rates of fermentation of the two strains; it caused a reduction of only 15% of the difference in the production of 2-phenylethanol (Additional file 3: Table S3), but completely abolished the differences in 2-phenyl acetate production. Abz1p uses chorismate as a substrate, which is also one of the precursors of 2-phenylethanol synthesis. We tested for the effects of the two alleles on the concentration of the various compounds analyzed during mQTL analysis (Table 3). We observed significant effects on the concentrations of many volatile compounds, including ethyl esters, confirming the involvement of *ABZ1* in their variations. These results also validate the use of *ABZ1* as an additive covariable in the model used to search for mQTLs.

PLB2 allelic variations may affect ethyl ester production

A 60kb region of chromosome XIII was linked with variations in the production of ethyl esters, and we identified two candidate genes with two allelic forms in this region: *PLB1* and *PLB2*. These genes code for phospholipase B which displays transacylase activity *in vitro* [29]. Plb1p in 59A presents some minor differences to that in S288C, whereas Plb2p of S288C carried a P378A substitution with respect to that in 59A. This proline residue is conserved in other *Saccharomyces* species and the mutation was not found in other available *S. cerevisiae* genome sequences. The Δ*PLB2*-By4741 strain produced much less octanoic ethyl ester than the control (Figure 7), and the difference was greater than that associated with deletion of *EEB1*, one of the key genes involved in the synthesis of decanoic ethyl ester [20]. These findings are consistent with the involvement of *PLB2* in this phenotype. Deletion of *PLB2* also led to a decrease in decanoic ethyl ester production and an increase of decanoic acid production, which was not observed after the deletion of *EEB1* [20].

Polymorphism of the various genes and adaptation

The two major QTL detected in this study, *PDR8* and *ABZ1*, show substantial polymorphism with numerous differences between the allelic forms in S288C and 59A. We investigated whether the differences between the alleles originated from the introgression from a specific lineage by comparing the corresponding alleles from other yeast genome sequences. The phylogeny (Figure 8)

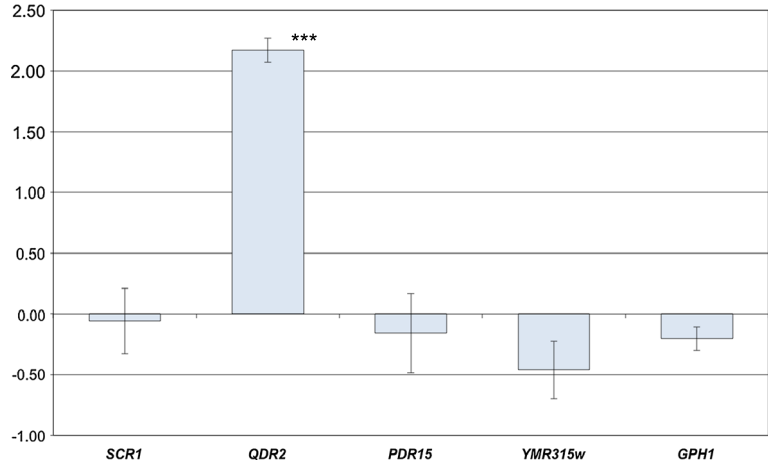


Figure 5 Q-PCR analysis of the expression of *PDR8* targets in 59A strains bearing 59A or S288C alleles during alcoholic fermentation. Differences in expression are given as fold ratio in comparison to 59A. Only the expression of QDR2 was highly significantly different (pvalue <0.001) for both strains. Other differences were not significant.

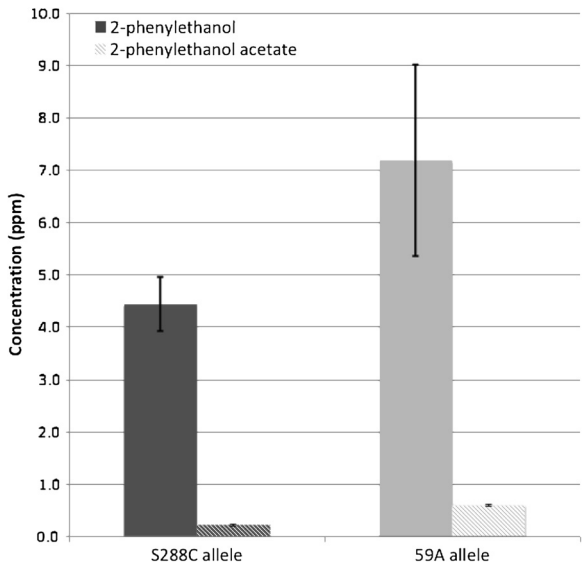


Figure 6 Production of 2-phenylethanol and 2-phenyethylacetate by two hemizygote strains carrying only the 59A or the S288C – *ABZ1* allele. Fermentation was performed as described for experiment B. Dark gray: S288C. Light gray; 59A. Differences are highly significant (p value<0.001). Errors bars correspond to one SD.

reveals that the *PDR8* allele of S288C is related to Malaysian or Asian alleles, and apparently one of the closest to its *S. paradoxus* ortholog; by contrast, the allele in 59A is a typical wine allele. As polymorphism may also result from specific adaptation, we performed a McDonald Kreitman test [34]. This test compares the ratio of nonsynonymous to synonymous polymorphism (intra species) to the ratio of nonsynonymous to synonymous divergence with the nearest species. This ratio is called the neutrality index (NI). An NI lower than one reflects a paucity of nonsynonymous polymorphism relative to nonsynonymous divergence, and is indicative of positive selection; an NI greater than one indicates negative selection of deletorious alleles driving divergence between species or balancing selection. This test was applied to a set of 15 *PDR8* alleles from strains isolated from various substrates and NI was 2.30, indicating a significant excess of non-neutral mutations (p value=0.009). This suggests that *PDR8* is subject to the accumulation of slightly deleterious mutations that are eliminated by negative selection during speciation, or alternatively that *PDR8* presents substantial diversity that might be associated with balanced selection resulting from specific adaptation to different niches.

In contrast with *PDR8*, the overall phylogeny (Figure 9) revealed that the S288C *ABZ1* sequence is related to copies from clinical isolate 322134S and bread strains YS2 and YS4. However, the S288C allelic form of *ABZ1* is located at the end of a long branch such that it appears to be the result of the accumulation of numerous mutations. Similarly, the McDonald Kreitman test with a set of 15 *ABZ1* sequences from strains isolated from various substrates indicated an excess of non-neutral mutations (NI = 3.00, p value<10^{-3}).

Discussion

We report 13 regions linked to variations in the production of wine volatile compounds. This study is the first demonstration of the potential usefulness of QTL analysis for understanding the origin of the variations in the concentrations of wine aroma compounds and deciphering this "intricate lattice of chemical and biological interactions" [24]. It was not possible to detect QTL for all relevant compounds, despite high heritability. Presumably, the synthesis of many of these compounds is under multigenic control, such that the small size of our segregant population prevented exploration of their complexity. Until now, few key technological traits for alcoholic fermentation have been characterized [11,13,14,16].

Several of the QTL found here are related to terpenoids, which constitute a large family of compounds. They include monoterpenes, which with their corresponding alcohols present useful properties, such as fragrances (in essential oils) or variety aroma (in wines), and even antimicrobial and cancer chemopreventive properties [35]. In yeast, these compounds are synthesized through the mevalonic acid pathway from acetyl-coA, which is converted to isopentenylpyrophosphate (IPP) and its isomer dimethylallyl pyrophosphate (DMAPP), the building blocks of isoprenoids. The main product of this pathway is ergosterol, and geranyldiphosphate and farnesyl diphosphate are intermediate metabolites.

We did not detect any QTL explaining variations in residual geraniol. However, one QTL explained some of the variation in the concentrations of citronellol; this QTL maps to a region of chromosome XIII containing several candidate genes. We did not find a candidate explaining the variations in the concentration of cis-rose oxide in the media or in the ratio between the cis and trans isomers. This compound is significant to wine-making because of its high odor activity [36] and it has been shown recently that yeast can produce cis-rose oxide in wine [37].

We demonstrate that the alleles of *PDR8* found in S288C and 59A differently regulate the *QDR2* gene responsible for the release of nerolidol into the media. Farnesol and its isomer nerolidol arise from farnesyl diphosphate instability at low pH, like that in the yeast vacuole or in the exocellular medium [38]. Therefore, it is possible that the transporter Qdr2p is responsible of the export of either farnesyl diphosphate or of nerolidol.

Enological strains have a *PDR8* allele more divergent from *S. paradoxus* than that of S288C. The neutrality index [34] we calculated for this gene is incompatible with its neutral evolution: it presented a higher number of replacement polymorphisms than expected under

Table 3 Effects of the different alleles of the ABZ1 gene on the concentrations of several fermentation compounds (relative units)

Compounds	S288C ABZ1			59A ABZ1			p-value
Isoamylacetate	**0.521**	±	**0.015**	**1.011**	±	**0.045**	**0.006**
farnesyl acetate	**0.439**	±	**0.013**	**0.780**	±	**0.046**	**0.006**
Isoamyl octanoate	0.057	±	0.005	0.060	±	0.006	0.729
Isoamyl decanoate	**0.076**	±	**0.001**	**0.092**	±	**0.004**	**0.048**
Ethyl hexanoate	0.510	±	0.005	0.459	±	0.011	0.060
Ethyl octanoate	**3.349**	±	**0.107**	**2.613**	±	**0.065**	**0.006**
Ethyl decanoate	**3.846**	±	**0.065**	**4.317**	±	**0.055**	**0.028**
Ethyl dodecanoate	**0.709**	±	**0.005**	**1.334**	±	**0.057**	**0.006**
Ethyl myristate	**0.035**	±	**0.001**	**0.060**	±	**0.002**	**0.002**
Ethyl hexadecanoate	**0.205**	±	**0.020**	**0.328**	±	**0.015**	**0.040**
Ethyl octadecanoate	0.068	±	0.013	0.085	±	0.012	0.399
Ethyl 9-decenoate	0.011	±	0.000	0.010	±	0.001	0.738
Ethyl 4-hydroxybutanoate	0.017	±	0.001	0.038	±	0.007	0.090
Ethyl 3-hydroxyoctanoate	0.052	±	0.004	0.039	±	0.002	0.086
Ethyl 3-hydroxydecanoate	**0.133**	±	**0.010**	**0.086**	±	**0.011**	**0.008**
Ethyl 9-hexadecenoate	**0.142**	±	**0.015**	**0.456**	±	**0.010**	**0.002**
2-phenylethyl acetate	**0.221**	±	**0.011**	**0.599**	±	**0.022**	**0.015**
2-phenylethyl hexanoate	**0.035**	±	**0.002**	**0.031**	±	**0.002**	**0.001**
2-phenylethyl octanoate	**0.011**	±	**0.001**	**0.022**	±	**0.005**	**0.004**
Acetic acid	**0.069**	±	**0.004**	**0.059**	±	**0.023**	**0.016**
Octanoic acid	**0.450**	±	**0.018**	**0.558**	±	**0.049**	**0.034**
Decanoic acid	**1.671**	±	**0.068**	**2.092**	±	**0.095**	**0.067**
Dodecanoic acid	**0.474**	±	**0.030**	**0.823**	±	**0.018**	**0.013**
methyl Oleate	**0.172**	±	**0.012**	**0.445**	±	**0.020**	**0.014**
isobutanol	1.508	±	0.327	1.230	±	0.067	0.408
isoamyl alcohol	17.440	±	1.703	20.582	±	0.238	0.150
1-octanol	**0.031**	±	**0.001**	**0.042**	±	**0.003**	**0.036**
2-phenyl ethanol	**4.435**	±	**0.369**	**7.201**	±	**0.238**	**0.003**
Nerolidol	1.318	±	0.105	1.428	±	0.053	0.334
2,3-dihydrofarnesol	3.245	±	0.396	2.895	±	0.220	0.411
farnesol	**3.948**		**0.220**	**2.361**	±	**0.087**	**0.006**
Trans-β-farnesene	0.057	±	0.005	0.060	±	0.006	0.623
Trans-α-farnesene	0.034	±	0.003	0.037	±	0.006	0.609
Cis-β-farnesene	0.036	±	0.005	0.043	±	0.007	0.387
Cis-bisabolene	0.006	±	0.001	0.007	±	0.001	0.372

Compounds whose concentration varies significantly are given in bold. Mean of 3 triplicates +/- standard deviation.

neutral selection. This may be the result of adaptation to different niches or the results of isolation and multiple migrations as suggested by Aa et al. for *SSU1* [39]. The role of farnesol and nerolidol production by *S. cerevisiae* is not clear. Under the anaerobic conditions of wine fermentation ergosterol synthesis is blocked, however *ERG20* expression correlates with fermentation speed [11] even in a fermentation medium containing ergosterol (experiment B). The synthesis of farnesol diphosphate is essential for (i) the synthesis of other compounds including dolichol which is necessary for cell wall assembly [40,41], (ii) protein prenylation such as that of Skt5p [42] involved in chitin synthase activity, and (iii) ubiquinone synthesis which may be less important during fermentation. Furthermore, farnesol is a biologically active compound that at concentrations higher than 50 μM

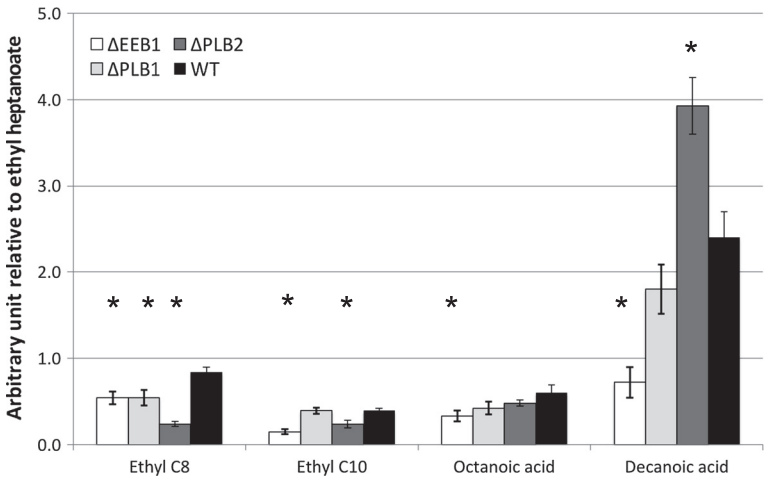


Figure 7 Effects of the deletion of *PLB1*, *PLB2*, and *EEB1* on the production of octanoic and decanoic acids and their corresponding ethyl esters, reported relative to the wild-type strain. Fermentations were as for experiment B. White bars: By**Δ***EEB1*, Light gray bars: By**Δ***PLB1*, Dark gray bars: By**Δ***PLB2*, Black bars: By4741wild-type strain. The Dunnett test was used to compare the production by each deletant strain to that by the corresponding wild-type strain (p-value<0.05). *= results significantly different to wt [30-33].

inhibits *S. cerevisiae* and *C. albicans* growth [43,44] and at lower concentrations is involved in quorum sensing by *C. albicans* [45,46].

In addition to the regions affecting terpenoid production, we also linked several other regions to variations in the concentrations of various volatiles. The *ABZ1* gene in one of these regions seems to have the widest effect as its allelic variations affected 2-phenylethanol and ethyl ester synthesis in experiment B, and explained as much as 50% of the variations in the concentration of ethylhexanoate in experiment A. The lower production of 2-phenylethanol was only partially restored by the addition of p-aminobenzoic acid to the fermentation media, and this suggests that this phenotype is not solely the consequence of the substantial effect of the *ABZ1*-S288C allele on fermentation speed via its impact on nitrogen metabolism [11]. The phylogeny of *ABZ1* was clearly different from that of *PDR8*: the *ABZ1*-S288C allele is located at the end of a long branch whose branch point is close to that of wine strains (relative to other origins). Again, the neutrality index [34] calculated for this gene was significantly higher than 1. Possibly, the allelic form of *ABZ1* has accumulated several deleterious mutations leading to a loss of activity. This feature may explain the very particular phenotypic behavior observed for S288C [47].

The two experimental sets we report generated complementary results. We observed effects of *PDR8* and *ABZ1* allelic variations in both experiments. Nevertheless, the impact of *ABZ1* polymorphism was much less pronounced in experiment A than B, as it explained only variation for ethylhexanoate production after correction for the flocculation effect. We also detected one additional region in experiment A only, and four in experiment B only; presumably the different sets of QTL identified in the two experiments reflect the effects of the different environmental conditions. By considering genes mapping in these regions, we identified *PLB2* as possibly involved in the variations of ethyl ester content in experiment B. This gene may have a role complementary to *EEB1* in the synthesis of ethyl esters during alcoholic fermentation.

Conclusion

This study shows that linkage analysis can give valuable information about the metabolic pathways involved in production of volatile compounds in yeast, even with only a small population of segregants. We identified the involvement of the *PLB2* gene in the metabolism of ethyl esters, and evidenced the role of *PDR8* in the release of nerolidol into the media via the regulation of *QDR2* expression. We also showed that the weak activity of SC288C Abz1p allele leads to a lower production of many metabolites, including 2-phenyl ethanol, and that this effect was only partially relieved by supplementation with paminobenzoic acid. Other candidate genes are currently being evaluated (i.e. for citronellol synthesis). However, we could not find candidate genes in all regions detected, and despite a high heritability, we did not find any regions associated with the production diversity of many of the compounds considered. Possibly, a larger number of segregants is necessary for a more exhaustive analysis. Our results identify potential new targets for a marker-aided breeding strategy in yeast for the optimization of the production of volatile compounds during fermentation.

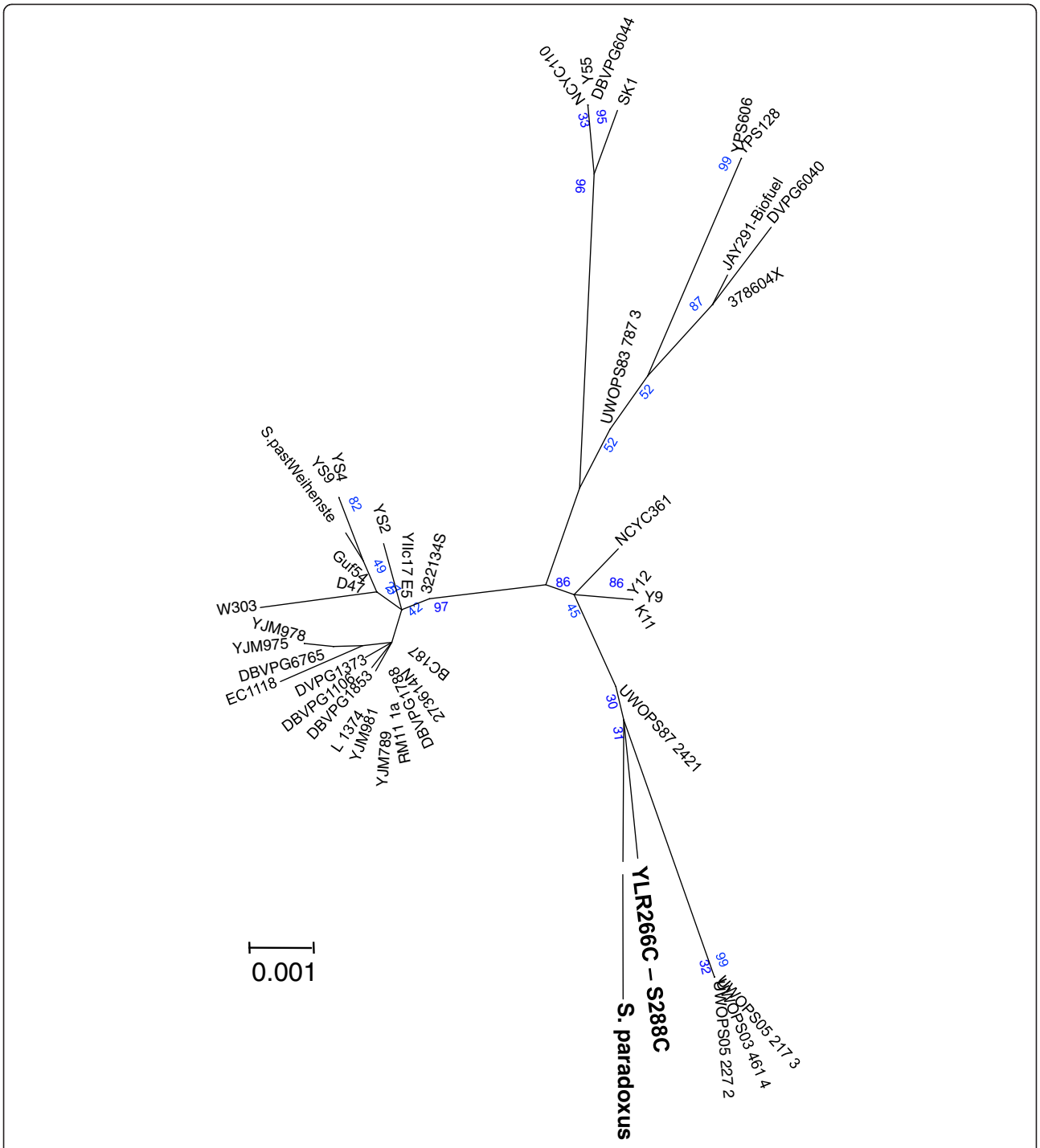


Figure 8 *PDR8* molecular phylogenetic tree. The evolutionary history was inferred by the maximum likelihood method based on the Kimura 2-parameter model and using 43 nucleotide sequences from the genome sequences available [30-33].

Interestingly, our genetic analysis revealed the particular evolution of the *PDR8* gene. This may reflect a specific adaptation to wine fermentation conditions, but raises questions about the role of farnesol and nerolidol for *S. cerevisiae* during fermentation.

Methods

Strains, growth conditions, and fermentation conditions

The two parental *Saccharomyces cerevisiae* strains studied were the standard strain S288c (MAT**a**; SUC2; gal2) and a haploid derivative of the industrial wine strain EC1118

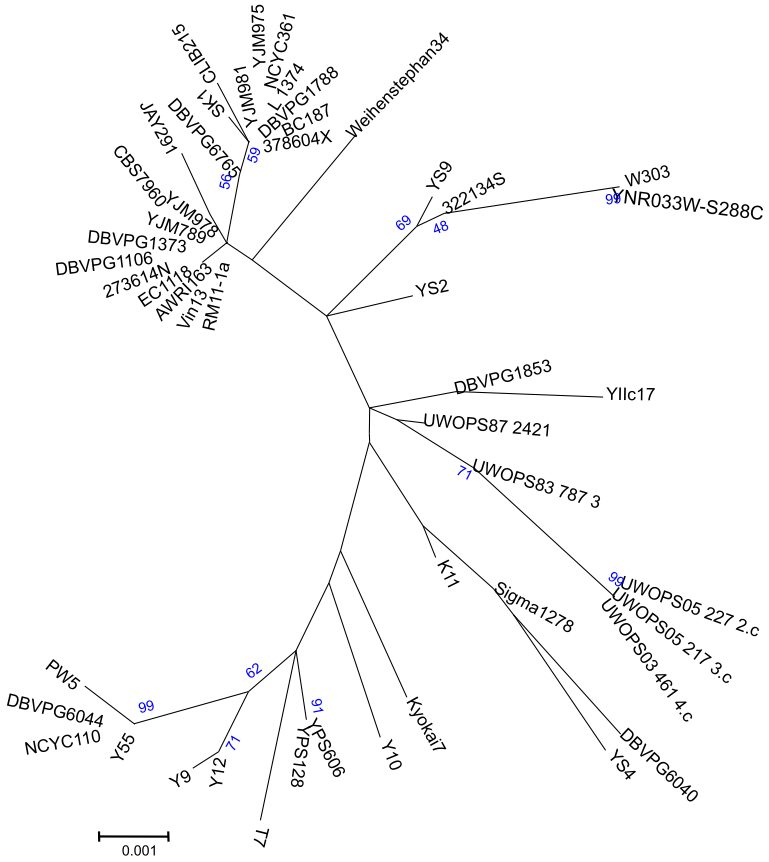


Figure 9 *ABZ1* molecular phylogenetic tree. The evolutionary history was inferred by the maximum likelihood method based on the Kimura 2-parameter model using 43 nucleotide sequences from the genome sequences available [30-33].

(HO/ho), herein referred to as 59A (MATa; ho). This strain is prototrophic and has fermentation properties similar to the diploid strain EC1118. The population of 30 segregants obtained from these two parental strains used for QTL analysis have been genotyped after hybridization on high density oligonucleotide microarrays Affymetrix YGS98 oligoarrays.

The strains BY4742 (MAT**a**; his3Δ1; leu2Δ 0; lys2Δ 0; ura3Δ 0) and BY4742**Δ**ABZ1 (Mat**a**; his3Δ 1; leu2Δ 0; lys2Δ 0; ura3Δ 0; YNR033w::kanMX4), and BY4742 (MAT**a**; his3Δ 1; leu2Δ 0; lys2Δ 0; ura3Δ 0) and BY4742**Δ**PDR8 (Mat**a**; his3Δ 1; leu2Δ 0; lys2Δ 0; ura3Δ 0; YNR033w::kanMX4) were used for hemizygous constructions.

Allelic replacement at *PDR8* in 59A was obtained in three steps: 1) deletion of *PDR8* from 59A using the hphMX4 cassette for hygromycin resistance (pAG32). Primers for cassette amplification and verification were obtained form Euroscarf. 2) preparation of a replacement cassette containing *PDR8*-loxP-kanMX4-loxP by the insertion of loxP-KanMX4-loxP into the terminator of *PDR8* in strain S288c (primers are given in Additional file 4: Table S4). 3) replacement of the hphMX4 cassette from 59A *PDR8Δ::hph* with the *PDR8*-loxP-kanMX4-loxP replacement cassette from S288c and selection on YPD containing G418 (200 $\mu g.l^{-1}$). The loss of hphMX4 cassette was verified by PCR and the absence of growth on hygromycin.

YPD medium was used for precultures at 28°C for 24h in 125 mL flasks with shaking.

Synthetic MS300 medium, which mimics a natural must [48] and [11], was used for fermentation experiments. The first experimental design mimicked white wine fermentation (20°C, low lipid content and containing sitosterol; experiment A). Geraniol, one of the key aroma compounds found in Gewürztraminer wine, was added to study its metabolism during alcoholic fermentation. We also analyzed the production of volatile compounds during fermentation as described in Ambroset et al 2011, which differed by the higher lipid content of the synthetic must and fermentation temperature 28°C (Table 4; experiment B). In some experiments, p-aminobenzoic acid was added to the fermentation media at 1 mg/L to study the effect of *ABZ1* alleles in the hemizygotes.

Fermentations were performed in 250 mL flasks equipped with airlocks to maintain anaerobiosis without stirring (design A) and in 1 L fermenters with constant stirring (design B). Small flask fermentations were

Table 4 Differences in the two experimental designs used in this work (adapted from [48])

Experimental design	A	B [11]
Temperature	20°C	28°C
Stirring	no	yes
Fermentation volume	150 mL	1 L
Anaerobic factors for 1L	Tween 5 μL	Tween 0.5 mL
	Oleic acid 0.05 μg	Oleic acid 5 μg
	Sistosterol: 15 μg/L	Ergosterol 1500 μg/L
Geraniol content (mg/L)	5	0
Number of fermentations	2	1

weighed twice daily and stopped as soon as the daily loss was less than 1% of the expected total loss.

Volatile compounds analysis

Wine aroma compounds were analyzed by the Stir Bar Sorptive Extraction method [49] adapted to our laboratory conditions, with a 1 μL injection volume. The analyses were performed with an Agilent 6890N gas chomatograph equipped with an Agilent 7683 automatic liquid sampler coupled to an Agilent 5975B inert MSD (Agilent Technologies). The gas chomatograph was fitted with a DB-Wax capillary column (60 m × 0.32 mm i.d. × 0.50 μm film thickness, J&W Scientific) and helium was used as carrier gas (1 mL min^{-1} constant flow). The GC oven temperature was programmed without initial hold time at a rate of 2.7°C min^{-1} from 70°C to 235°C (hold 10 min). The injector was set to 250°C and used in pulsed splitless mode (25 psi for 0.50 min). The temperatures of the interface, MS ion source and quadrupole were 270°C, 230°C and 150°C, respectively. The mass spectrometer was operated in electron impact ionization mode (EI, 70 eV) and the masses were scanned over a m/z range of 29 – 300 amu. Agilent MSD chemStation software (G1701DA, Rev D.03.00) was used for instrument control and data processing. The mass spectra were compared with the Wiley's library reference spectral bank

The following compounds were analyzed: isoamyl alcohol, isoamyl acetate, isobutanol, 2-3 butanediol, 2-phenylethyl acetate, ethyl hexanoate, ethyl octanoate, ethyl decanoate, ethyl dodecanoate, ethyl myristate, ethyl palmytate, ethyl laurate, 2-phenyl ethanol, hexanoic acid, octanoic acid, decanoic acid, dodecanoic acid, ethyl 9-decenoate, isoamyl octanoate, 2-phenyl ethyl hexanoate, 2-phenylethyl octanoate, 2-phenylethyl decanoate, ethyl 4 hydroxy butanoate, ethyl 3-hydroxydecanoate, ethyl 3-hydroxyoctanoate, nerolidol, farnesol, 2-3 dihydro farnesol, (E, Z)- or (Z, E)- farnesol (A), (E, E)- farnesol (A), farnesyl acetate, isoamyl octanoate, isoamyl decanoate, isoamyl dodecanoate, methyl oleate, trans-β-farnesene, (Z, E)-α-farnesene, α-bisabolene, ß-bisabolene, (E, E)-α-farnesene, α-terpineol, linalol, citronellol, geraniol, nerol, citronellyl acetate, geranyl acetate, neryl acetate, cis-rose oxide, trans-rose oxide.

Statistical and QTL analysis

Heritability was calculated according to the method of Brem et al. [50]. Statistical analyses were performed using R software version 2.13.1 [51]. QTL analysis was done for each phenotype of the two datasets (experiments A and B) using the genetic map of 1834 markers genotyped previously [11]. The distribution of each phenotype was verified using a Shapiro-Wilk normality test: the normality of the distribution was rejected for 17 of the 29 compounds analyzed in experiment A and 30 of 40 in experiment B, for a threshold of 0.05

As the distribution of most phenotypes was not normal and due to the small sample size, linkage analysis was performed using both parametric and non-parametric models to evaluate the robustness of the parametric model. The parametric model consists of a linkage analysis performed using a normal model with the Haley-Knott regression method implemented in the R/qtl package [52,53]. As the results of the two analyses were concordant, only the normal analysis is presented.

To overcome the potential effects of flocculation [26] and of the presence of the *ABZ1*-S288C allele which provokes large variations in fermentation kinetics [11], we performed a second linkage analysis using a normal model with the Haley-Knott regression method, first with flocculation as an interactive covariate, and then with the *ABZ1* marker (Chr 14, position 689.4 kb) as an additive covariate. For these regions, a significant effect was indeed observed for both flocculation and the *ABZ1*-specific markers.

For the three models and the two datasets analyzed, logarithm of odds (LOD) scores were computed for each marker every 2.5cM. An interval estimate of the location of each QTL was obtained as the 1-LOD support interval. The LOD significance threshold was estimated after permutation tests that were replicated 1000 times. The percentage of variance explained by each QTL was estimated from a drop-one-term analysis of results in the global model.

Q-PCR analysis of the expression of PDR8 targets after allelic replacement

Fermentations (900 mL of MS300 medium) were performed in triplicate with strains 59A and S288C-*PDR8* 59A, and cells were sampled when 70% of the glucose had been fermented. RNA was extracted with trizol as described previously [54]. cDNA was produced by reverse transcription and a 1 in 25 dilution of the resulting cDNA was used for the realtime PCR assays with gene-specific primers and Strategene's Brilliant II SYBR Green QPCR

Master Mix (Santa Clara, CA) and an ABI7300 QPCR machine. Expression levels were measured relative to those of *UBC6* and *SCR1*, both giving similar results.

Sequence analysis and phylogeny

The comparison of the sequences of the 59A and S288C genomes and the differences between them can be found at http://genome.jouy.inra.fr/genyeastrait/ [11].

To infer the evolutionary history of *ABZ1* and *PDR8*, we collected their sequences from genomes available at SGD (http://www.yeastgenome.org/). All uncompleted or frameshift-containing sequences where discarded from this set. The phylogenies were inferred with MEGA [55] by the Maximum Likelihood method based on the Kimura 2-parameter model [56]. The trees with the highest log likelihood are shown. The trees are drawn to scale, with branch lengths proportional to the number of substitutions per site. The significance of the Neutrality Index [34] test was calculated using the http://bioinf3.uab.cat/mkt/MKT.asp website.

The list of the sequences used for the two analyses is given in supplementary data (Additional file 5).

Additional files

Additional file 1: Distribution of the different phenotypes for the population of segregants and parental strains.

Additional file 2: Impact of flocculation and ABZ1 on aroma production at the different QTL and variablity of fermentation length.

Additional file 3: Table S3. Effect of the addition of p-aminobenzoic acid on the production of 2phenylethanol and 2-phenylethanolacetate.

Additional file 4: Table S4. Primers used in this study.

Additional file 5: Sequenced used for McDonald Kreitman test.

Competing interests

The authors declare that they have no competing interests.

Author's contribution

DS: performed fermentation (experiment A), aroma and QTL analysis , candidate gene search and validation (*PDR8*, *PLB2*, . . .). DS wrote a first draft of the manuscript. CA produced the strain set and performed fermentations (experiment B). CB performed allelic replacement for S288c- *PDR8* 59A strain fermentations and QPCR expression analysis. PC, analyzed aroma compounds. PD built *ABZ1* hemizygote strains and performed fermentations. IS, performed statistical analysis (QTL). JLL performed phylogenic analysis and tests. CE, BB, FK, JLL conceived the study ,designed and coordinated the research JLL wrote the manuscript. All authors analyzed the data. All authors read and approuved the final manuscript.

Acknowledgments

This work has been supported by a grant from "Comité Interprofessionnel des Vins d'Alsace", Colmar, France and from Region Alsace, France.

Author details

[1]INRA, UMR1131, Colmar F-68021, France. [2]Université de Strasbourg, UMR1131, Strasbourg F-68021, France. [3]Twistaroma, Colmar F-68021, France. [4]INRA, UMR1083, Montpellier F-34060, France. [5]Montpellier SupAgro, UMR1083, Montpellier F-34060, France. [6]Université Montpellier 1, UMR1083, Montpellier F-34060, France. [7]Université de Haute Alsace, EA3991 Laboratoire Vigne Biotechnologies et Environnement, Colmar F-68021, France.

Received: 29 May 2012 Accepted: 4 October 2012
Published: 30 October 2012

References

1. Young ND: **QTL mapping and quantitative disease resistance in plants.** *Annu Rev Phytopathol* 1996, **34:**479–501.
2. Deutschbauer AM, Davis RW: **Quantitative trait loci mapped to single-nucleotide resolution in yeast.** *Nat Genet* 2005, **37:**1333–1340.
3. Sinha H, Nicholson BP, Steinmetz LM, McCusker JH: **Complex genetic interactions in a quantitative trait locus.** *PLoS genetics* 2006, **2:**e13.
4. Steinmetz LM, Sinha H, Richards DR, Spiegelman JI, Oefner PJ, McCusker JH, Davis RW: **Dissecting the architecture of a quantitative trait locus in yeast.** *Nature* 2002, **416:**326–330.
5. Perlstein EO, Ruderfer DM, Roberts DC, Schreiber SL, Kruglyak L: **Genetic basis of individual differences in the response to small-molecule drugs in yeast.** *Nat Genet* 2007, **39:**496–502.
6. Demogines A, Smith E, Kruglyak L, Alani E: **Identification and dissection of a complex DNA repair sensitivity phenotype in Baker's yeast.** *PLoS genetics* 2008, **4:**e1000123.
7. Ansel J, Bottin H, Rodriguez-Beltran C, Damon C, Nagarajan M, Fehrmann S, François J, Yvert G: **Cell-to-cell stochastic variation in gene expression is a complex genetic trait.** *PLoS genetics* 2008, **4:**e1000049.
8. Ehrenreich IM, Torabi N, Jia Y, Kent J, Martis S, Shapiro JA, Gresham D, Caudy AA, Kruglyak L: **Dissection of genetically complex traits with extremely large pools of yeast segregants.** *Nature* 2010, **464:**1039–1042.
9. Cubillos FA, Billi E, Zörgö E, Parts L, Fargier P, Omholt S, Blomberg A, Warringer J, Louis EJ, Liti G: **Assessing the complex architecture of polygenic traits in diverged yeast populations.** *Mol Ecol* 2011, **20:**1401–1413.
10. Yvert G, Brem RB, Whittle J, Akey JM, Foss E, Smith EN, Mackelprang R, Kruglyak L: **Trans-acting regulatory variation in *Saccharomyces cerevisiae* and the role of transcription factors.** *Nat Genet* 2003, **35:**57–64.
11. Ambroset C, Petit M, Brion C, Sanchez I, Delobel P, Guérin C, Chiapello H, Nicolas P, Bigey F, Dequin S, Blondin B: **Deciphering the Molecular Basis of Wine Yeast Fermentation Traits Using a Combined Genetic and Genomic Approach.** *G3* 2011, **1:**263–281.
12. Marullo P, Yvert G, Bely M, Aigle M, Dubourdieu D: **Efficient use of DNA molecular markers to construct industrial yeast strains.** *FEMS yeast research* 2007, **7:**1295–1306.
13. Marullo P, Mansour C, Dufour M, Albertin W, Sicard D, Bely M, Dubourdieu D: **Genetic improvement of thermo-tolerance in wine *Saccharomyces cerevisiae* strains by a backcross approach.** *FEMS yeast research* 2009, **9:**1148–1160.
14. Katou T, Namise M, Kitagaki H, Akao T, Shimoi H: **QTL mapping of sake brewing characteristics of yeast.** *J Biosci Bioeng* 2009, **107:**383–393.
15. Hu XH, Wang MH, Tan T, Li JR, Yang H, Leach L, Zhang RM, Luo ZW: **Genetic dissection of ethanol tolerance in the budding yeast *Saccharomyces cerevisiae*.** *Genetics* 2007, **175:**1479–1487.
16. Marullo P, Aigle M, Bely M, Masneuf-Pomarède I, Durrens P, Dubourdieu D, Yvert G: **Single QTL mapping and nucleotide-level resolution of a physiologic trait in wine *Saccharomyces cerevisiae* strains.** *FEMS yeast research* 2007, **7:**941–952.
17. Loscos N, Hernandez-Orte P, Cacho J, Ferreira V: **Release and formation of varietal aroma compounds during alcoholic fermentation from nonfloral grape odorless flavor precursors fractions.** *J Agric Food Chem* 2007, **55:**6674–6684.
18. Ferreira V, Lopez R, Cacho JF: **Quantitative determination of the odorants of young red wines from different grape varieties.** *J Sci Food Agric* 2000, **80:**1659–1667.
19. Lilly M, Lambrechts MG, Pretorius IS: **Effect of increased yeast alcohol acetyltransferase activity on flavor profiles of wine and distillates.** *Appl Environ Microbiol* 2000, **66:**744–753.
20. Saerens SMG, Verstrepen KJ, Van Laere SDM, Voet ARD, Van Dijck P, Delvaux FR, Thevelein JM: **The *Saccharomyces cerevisiae* EHT1 and EEB1 genes encode novel enzymes with medium-chain fatty acid ethyl ester synthesis and hydrolysis capacity.** *J Biol Chem* 2006, **281:**4446–4456.
21. Wondra M, Berovic M: **Analyses of Aroma Components of Chardonnay Wine Fermented by Different Yeast Strains.** *Food Technology and Biotechnology* 2001, **39:**141–148.
22. Romano P: **Function of yeast species and strains in wine flavour.** *International Journal of Food Microbiology* 2003, **86:**169–180.

23. Camarasa C, Sanchez I, Brial P, Bigey F, Dequin S: **Phenotypic Landscape of Saccharomyces cerevisiae during Wine Fermentation: Evidence for Origin-Dependent Metabolic Traits.** *PLoS One* 2011, **6:**e25147.
24. Rossouw D, Naes T, Bauer FF: **Linking gene regulation and the exo-metabolome: a comparative transcriptomics approach to identify genes that impact on the production of volatile aroma compounds in yeast.** *BMC Genomics* 2008, **9:**530.
25. Rossouw D, Jacobson D, Bauer FF: **Transcriptional regulation and the diversification of metabolism in wine yeast strains.** *Genetics* 2012, **190:**251–261.
26. Pretorius IS: **Gene Technology in Winemaking: New Approaches to an Ancient Art.** *Agric Conspec Sci* 2001, **66:**27–47.
27. Duchene E, Legras J-L, Karst F, Merdinoglu D, Claudel P, Jaegli N, Pelsy F: **Variation of linalool and geraniol content within two pairs of aromatic and non-aromatic grapevine clones.** *Australian Journal of Grape and Wine Research* 2009, **15:**120–130.
28. Hikkel I, Lucau-Danila A, Delaveau T, Marc P, Devaux F, Jacq C: **A general strategy to uncover transcription factor properties identifies a new regulator of drug resistance in yeast.** *J Biol Chem* 2003, **278:**11427–11432.
29. Merkel O, Fido M, Mayr JA, Prüger H, Raab F, Zandonella G, Kohlwein SD, Paltauf F: **Characterization and function in vivo of two novel phospholipases B/lysophospholipases from *Saccharomyces cerevisiae*** *J Biol Chem* 1999, **274:**28121–28127.
30. Akao T, Yashiro I, Hosoyama A, Kitagaki H, Horikawa H, Watanabe D, Akada R, Ando Y, Harashima S, Inoue T, Inoue Y, Kajiwara S, Kitamoto K, Kitamoto N, Kobayashi O, Kuhara S, Masubuchi T, Mizoguchi H, Nakao Y, Nakazato A, Namise M, Oba T, Ogata T, Ohta A, Sato M, Shibasaki S, Takatsume Y, Tanimoto S, Tsuboi H, Nishimura A, Yoda K, Ishikawa T, Iwashita K, Fujita N, Shimoi H: **Whole-genome sequencing of sake yeast Saccharomyces cerevisiae Kyokai no. 7.** *DNA research: an international journal for rapid publication of reports on genes and genomes* 2011, **18:**423–434.
31. Borneman AR, Desany BA, Riches D, Affourtit JP, Forgan AH, Pretorius IS, Egholm M, Chambers PJ: **Whole-Genome Comparison Reveals Novel Genetic Elements That Characterize the Genome of Industrial Strains of *Saccharomyces cerevisiae*.** *PLoS Genetics* 2011, **7:**e1001287.
32. Novo M, Bigey F, Beyne E, Galeote V, Gavory F, Mallet S, Cambon B, Legras J-L, Wincker P, Casaregola S, Dequin S: **Eukaryote-to-eukaryote gene transfer events revealed by the genome sequence of the wine yeast *Saccharomyces cerevisiae* EC1118.** *Proc Natl Acad Sci* 2009, **106:**16333–16338.
33. Argueso JL, Carazzolle MF, Mieczkowski PA, Duarte FM, Netto OVC, Missawa SK, Galzerani F, Costa GGL, Vidal RO, Noronha MF, Dominska M, Andrietta MGS, Andrietta SR, Cunha AF, Gomes LH, Tavares FCA, Alcarde AR, Dietrich FS, McCusker JH, Petes TD, Pereira GAG: **Genome structure of a *Saccharomyces cerevisiae* strain widely used in bioethanol production.** *Genome Res* 2009, **19:**2258–2270.
34. McDonald JH, Kreitman M: **Adaptive protein evolution at the Adh locus in Drosophila.** *Nature* 1991, **351:**652–654.
35. Oswald M, Fischer M, Dirninger N, Karst F: **Monoterpenoid biosynthesis in *Saccharomyces cerevisiae*.** *FEMS yeast research* 2007, **7:**413–421.
36. Guth H: **Quantitation and Sensory Studies of Character Impact Odorants of Different White Wine Varieties.** *J Agric Food Chem* 1997, **45:**3027–3032.
37. Koslitz S, Renaud L, Kohler M, Wüst M, Chem F, The C, However L: **Stereoselective formation of the varietal aroma compound rose oxide during alcoholic fermentation.** *J Agric Food Chem* 2008, **56:**1371–1375.
38. Muramatsu M, Ohto C, Obata S, Sakuradani E, Shimizu S: **Alkaline pH enhances farnesol production by *Saccharomyces cerevisiae*** *J Biosci Bioeng* 2009, **108:**52–55.
39. Aa E, Townsend JP, Adams RI, Nielsen KM, Taylor JW: **Population structure and gene evolution in *Saccharomyces cerevisiae*.** *FEMS yeast research* 2006, **6:**702–715.
40. Juchimiuk M, Pasikowska M, Zatorska E, Laudy AE, Smoleńska-Sym G, Palamarczyk G: **Defect in dolichol-dependent glycosylation increases sensitivity of *Saccharomyces cerevisiae* towards anti-fungal drugs.** *Yeast (Chichester, England)* 2010, **27:**637–645.
41. Orłowski J, Machula K, Janik A, Zdebska E, Palamarczyk G: **Dissecting the role of dolichol in cell wall assembly in the yeast mutants impaired in early glycosylation reactions.** *Yeast (Chichester, England)* 2007, **24:**239–252.
42. Grabińska KA, Magnelli P, Robbins PW: **Prenylation of *Saccharomyces cerevisiae* Chs4p Affects Chitin Synthase III activity and chitin chain length.** *Eukaryotic cell* 2007, **6:**328–336.
43. Fairn GD, Macdonald K, McMaster CR: **A chemogenomic screen in Saccharomyces cerevisiae uncovers a primary role for the mitochondria in farnesol toxicity and its regulation by the Pkc1 pathway.** *J Biol Chem* 2007, **282:**4868–4874.
44. Langford ML, Hasim S, Nickerson KW, Atkin AL: **Activity and toxicity of farnesol towards *Candida albicans* are dependent on growth conditions.** *Antimicrob Agents Chemother* 2010, **54:**940–942.
45. Hornby JM, Jensen EC, Lisec AD, Tasto JJ, Jahnke B, Shoemaker R, Dussault P, Nickerson KW: **Quorum sensing in the dimorphic fungus *Candida albicans* is mediated by farnesol.** *Appl Environ Microbiol* 2001, **67:**2982–2992.
46. Palková Z, Váchová L: **Life within a community: benefit to yeast long-term survival.** *FEMS Microbiol Rev* 2006, **30:**806–824.
47. Cubillos FA, Zia A, Gjuvsland A, TJ, Warringer J, Forsmark A, Durbin R, Omholt SW, Louis EJ, Liti G, Moses A, Blomberg A: **Trait Variation in Yeast Is Defined by Population History.** *PLoS Genetics* 2011, **7:**e1002111.
48. Bely M, Sablayrolles J-M, Barre P: **Automatic detection of assimilable nitrogen deficiencies during alcoholic fermentation in oenological conditions.** *J Ferment Bioeng* 1990, **70:**246–252.
49. Coelho E, Rocha S, Delgadillo I, Coimbra M: **Headspace-SPME applied to varietal volatile components evolution during Vitis vinifera L. cv. "Baga" ripening.** *Anal Chim Acta* 2006, **563:**204–214.
50. Brem RB, Yvert G, Clinton R, Kruglyak L: **Genetic dissection of transcriptional regulation in budding yeast.** *Science (New York, N.Y.)* 2002, **296:**752–755.
51. R Core Team: **R: A language and environment for statistical computing, Version 2.15.1.** *R Foundation for Statistical Computing* 2012, .
52. Broman KW, Wu H, Sen S, Churchill G: **a.: R/qtl: QTL mapping in experimental crosses.** *Bioinformatics* 2003, **19:**889–890.
53. Broman KW, Sen Ś: *b.: A Guide to QTL Mapping with R/qtl:*. Germany/New York: Springer-Verlag, Berlin/Heidelberg; 2009.
54. Legras J-L, Erny C, LeJeune C, Lollier M, Adolphe Y, Demuyter C, Delobel P, Blondin B, Karst F: ***Saccharomyces cerevisiae* activates two different resistance mechanisms when exposed to octanoic and decanoic acids.** *Appl Environ Microbiol* 2010, **76:**7526–7535.
55. Tamura K, Peterson D, Peterson N, Stecher G, Nei M, Kumar S: **MEGA5: Molecular Evolutionary Genetics Analysis using Maximum Likelihood, Evolutionary Distance, and Maximum Parsimony Methods.** *Mol Biol Evol* 2011, **28:**2731–2739.
56. Kimura M: **A simple method for estimating evolutionary rate of base substitutions through comparative studies of nucleotide sequences.** *J Mol Evol* 1980, **16:**111–120.

doi:10.1186/1471-2164-13-573
Cite this article as: Steyer *et al.*: QTL mapping of the production of wine aroma compounds by yeast. *BMC Genomics* 2012 **13**:573.

MOLECULAR ECOLOGY

Molecular Ecology (2017) doi: 10.1111/mec.14053

SPECIAL ISSUE: MICROBIAL LOCAL ADAPTATION

Genomic signatures of adaptation to wine biological ageing conditions in biofilm-forming flor yeasts

A. L. COI,* F. BIGEY,† S. MALLET,† S. MARSIT,† G. ZARA,* P. GLADIEUX,‡ V. GALEOTE,† M. BUDRONI,* S. DEQUIN† and J. L. LEGRAS†

Dipartimento di Agraria, Università di Sassari, 07100 Sassari, Italy, †SPO, INRA, SupAgro, Université de Montpellier, 34060 Montpellier, France, ‡INRA, UMR BGPI, 34398 Montpellier, France

Abstract

The molecular and evolutionary processes underlying fungal domestication remain largely unknown despite the importance of fungi to bioindustry and for comparative adaptation genomics in eukaryotes. Wine fermentation and biological ageing are performed by strains of *S. cerevisiae* with, respectively, pelagic fermentative growth on glucose and biofilm aerobic growth utilizing ethanol. Here, we use environmental samples of wine and flor yeasts to investigate the genomic basis of yeast adaptation to contrasted anthropogenic environments. Phylogenetic inference and population structure analysis based on single nucleotide polymorphisms revealed a group of flor yeasts separated from wine yeasts. A combination of methods revealed several highly differentiated regions between wine and flor yeasts, and analyses using codon-substitution models for detecting molecular adaptation identified sites under positive selection in the high-affinity transporter gene *ZRT1*. The cross-population composite likelihood ratio revealed selective sweeps at three regions, including in the hexose transporter gene *HXT7*, the yapsin gene *YPS6* and the membrane protein coding gene *MTS27*. Our analyses also revealed that the biological ageing environment has led to the accumulation of numerous mutations in proteins from several networks, including *Flo11* regulation and divalent metal transport. Together, our findings suggest that the tuning of *FLO11* expression and zinc transport networks are a distinctive feature of the genetic changes underlying the domestication of flor yeasts. Our study highlights the multiplicity of genomic changes underlying yeast adaptation to man-made habitats and reveals that flor/wine yeast lineage can serve as a useful model for studying the genomics of adaptive divergence.

Keywords: adaptation, biofilm, biological ageing, domestication, FLO11, flor yeast, genome, *Saccharomyces cerevisiae*, *ZRT1*

Received 31 March 2016; revision accepted 31 January 2017

Introduction

Domestication is a specific case of adaptive divergence in response to human-mediated artificial selection, and an opportunity to study the genetic basis of adaptation (Ross-Ibarra *et al.* 2007; Gladieux *et al.* 2014). Despite the importance of domesticated fungi to bioindustry, and as model eukaryote species, relatively few studies – compared to plants and animals – have investigated the genomic and evolutionary changes underlying fungal domestication (Galagan *et al.* 2003; Gibbons & Rinker 2015; Ropars *et al.* 2015). Domesticated *Saccharomyces cerevisiae* yeasts have emerged as a preeminent fungal group for comparative evolutionary genomics of domestication and subsequent strain improvement, due to (i) their close connection with the history of agriculture and human migrations, (ii) the existence of multiple domestication events (Fay & Benavides 2005; Baker *et al.* 2015; Gallone *et al.* 2016) and (iii) unrivalled set of tools to explore genotype–phenotype–fitness relationships (Liti *et al.* 2009; Scannell *et al.* 2011; Ho & Zhang

Correspondence: J. L. Legras, Fax: +33 4 99 61 28 57; E-mail: jean-luc.legras@inra.fr

2014). For example, the comparison of sake and wine *Saccharomyces cerevisiae* yeasts indicated that the recurrent loss of functional aquaporins was involved in their adaptation to contrasted anthropogenic environments (Will *et al.* 2010). The acquisition by horizontal gene transfer of peptide transporters with high affinity for grape peptides (Marsit *et al.* 2015) or strong ability to expel sulphites from the cell (Pérez-Ortín *et al.* 2002) are other examples of yeast adaptation to the winemaking environment. More generally, winemaking exposes yeast populations to contrasted environmental conditions inside wineries and between wineries and vineyards, providing unique opportunities for investigating the genetic bases of yeast adaptation to man-made habitats.

Winemaking was invented more than 7000 years before present and technological developments have continuously transformed the winemaking process to improve wine quality and conservation. The use of sulphites by Egyptians and Romans to clean wine containers (Romano & Suzzi 1993), or the spread of bottles with corks in the 17th century (Estreicher 2006) are examples of technical developments that have drastically improved oxygen management and paved the way to the modern practice of protecting wine aromas from oxygen after alcoholic fermentation. Before the advent of the efficient methods for limiting the impact of oxygen on aromas, winemakers were making use of the spontaneous oxidative changes occurring during wine ageing and this practice has continued in some European countries (Hungary, France, Italy and Spain). The process of wine ageing, referred to as biological ageing, takes place after the completion of alcoholic fermentation and it leads to the production of specific flavours and aromas along with the progressive oxidation of alcohol and remaining carbohydrates. Biological ageing involves a biofilm called velum that forms on the surface of wine and consists in specific *Saccharomyces cerevisiae* yeast strains called 'flor yeasts' (Ibeas & Jimenez 1997; Legras *et al.* 2014).

Wine and flor yeasts are intimately associated with wine fermentation and biological ageing processes. Wine yeasts, which perform alcoholic fermentation, represent a genetically homogeneous group that experienced a recent demographic expansion (Fay & Benavides 2005; Legras *et al.* 2007; Liti *et al.* 2009; Schacherer *et al.* 2009), preceded by a domestication bottleneck that greatly reduced genetic variability (Fay & Benavides 2005). A recent population genomic study suggests that *S. cerevisiae* associated with Mediterranean oaks is the main progenitor of the wine yeast lineage (Almeida *et al.* 2015). Flor yeasts have the specific ability of weaving a biofilm, in addition to their ability to perform alcoholic fermentation (Fidalgo *et al.* 2006; Legras *et al.* 2014). The close association of flor yeast with the ageing process, which has been used for centuries of almost continuous growth, also suggests that flor yeasts represent a domesticated group of microbes and that they may display specific genomic features allowing them to thrive in the harsh conditions of biological ageing (Legras *et al.* 2007). Fidalgo *et al.* (2006) suggested that the process of biological ageing impacted biofilm formation and led to the fixation of adaptive mutations in the promoter and coding region of the gene *FLO11* (coding for a flocculin), resulting in a higher hydrophobicity of yeast cells. The Flo11 protein is also central to the mechanism enabling yeasts to invade agar or to produce a pseudohypha under nutritional deprivation (Ryan *et al.* 2012). The regulation of *FLO11* expression is highly complex, involving several pathways, such as the Ras/cAMP/PKA and the MAP-kinase pathways, and also two noncoding RNAs (ICR1 and PWR1; Brückner & Mösch 2011). To date, no other allelic variation has been associated with the specific habitat of flor yeasts. Here, we used whole-genome resequencing to investigate the evolutionary origins of flor yeasts and to identify candidate genomic regions involved in the adaptation of flor yeasts to their specific habitat. For some of the candidate genomic regions identified by our analyses, we provide evidence for the phenotypic impact of allelic differences between wine and flor yeast through allelic replacement or reciprocal hemizygosity tests.

Materials and methods

Yeast strains, media and growth conditions

Yeast strains used in this work are listed in Table S1a and b (Supporting information). For some strains, we chose to sequence diploid derivatives from the original wine or flor strain to obtain assemblies from homozygous strains and to improve the phasing of other diploid genomes.

For the phenotypic validation experiments, yeasts were precultivated in YEPD (2% glucose, 1% yeast extract and 2% peptone), and after 24 h, they were inoculated at OD_{600} = 0.1 in YNB (Difco) + 4% ethanol (v/v) in 24-well plates (Coi *et al.* 2016). Photographs of plates were taken after 5–7 days of growth, and the surfaces of the biofilms were measured using images taken with IMAGEJ Version 2.0.0 under the FIJI environment (http://imagej.net).

Construction of mutants

The primers used to construct the deletion cassette and to verify the deletion are listed in the Supplementary

Material (Table S2, Supporting information). The allelic replacement of *RGA2* and *ZRT1* in P3-D5 MATα was performed in two steps, as follows. *RGA2* or *ZRT1* was deleted using a kanamycin cassette and then replaced with the wine functional allele amplified in tandem with hygromycin for *RGA2* or with the wine functional allele for *ZRT1*. Selection was then performed in SD medium supplemented with 1 mM EDTA to chelate zinc (Gitan *et al.* 1998). The resulting P3-D5*K1-RGA2*:Hyg MATα and P3-D5*K1-ZRT1* MATα strains were checked by sequencing. For constructing *SFL1* reciprocal hemizygotes, we deleted *SFL1* using a Kanamycin cassette in P3-D5 MATα and in K1-280-2B MATα and crossed each deleted strain with the wild type of the other background. *SLN1* hemizygotes were obtained by the deletion of one copy of *SLN1* in the P3-D5 MATα x K1-280-2B MATα zygote and identification of the remaining allele by sequencing. All cassettes were amplified by PCR using KAPA HiFi HotStart Taq polymerase (KAPA BIOSYSTEMS) or conventional Taq DNA Polymerase (Fermentas).

Genome sequencing, population analysis and SNP detection

We used the genomic sequences of nine flor strains from Spain, Italy, France and Hungary and nine wine strains from Italy and France (Table S1a, Supporting information) recently published by our group (Marsit *et al.* 2015). We added new genome sequences obtained with the same technology for EC1118 and 59A, the haploid derivative of EC1118. Genomic DNA was prepared following a standard phenol–chloroform extraction protocol. Autosomal DNA was separated from mitochondrial DNA via isopycnic CsCl ultracentrifugation at 289 000 *g* for 20 h (Barth & Gaillardin 1996) and processed to generate libraries of 300-bp inserts. The 20 libraries were multiplexed in different lanes of Illumina HiSeq 2000. Paired-end (2 × 100 nucleotides) sequencing resulted in an average theoretical coverage of 77–730×. Image analysis and data extraction were performed using ILLUMINA RTA version 1.13.48.0 and CASAVA version 1.8.2. Read sequences were processed to improve their global quality; they were trimmed for the first nine nucleotides using FASTX Toolkit version 0.0.13 and for low-quality regions using SICKLE version 1.000 (quality below 25 in a 20-bp sliding window). Reads shorter than 50 bp were removed. Genomic sequences were assembled with SOAPDENOVO version 1.05 (Li *et al.* 2010) using different Kmers and by keeping the assembly with the highest N50.

We used the Genome Analysis Toolkit (GATK; McKenna *et al.* 2010), version 2.3-9, for calling single nucleotide polymorphisms (SNPs) and indels, using Best Practice Variant Detection with GATK version 4, release 2.0, available online. In short, the workflow is divided into the following four sequential steps: initial mapping, refinement of the initial reads, multisample indel and SNP calling and, finally, variant quality score recalibration (VQSR). First, reads were aligned to the S288c reference genome (release number R64-1-1, downloaded from the SACCHAROMYCES GENOME DATABASE (SGD; http://www.yeastgenome.org/)) using BWA version 0.6.2 (Li & Durbin 2009). Second, optical and PCR duplicates were removed using MARKDUPLICATE from the PICARD TOOLS version 1.84 (http://picard.sourceforge.net). Reads encompassing indels were realigned to increase the indel call accuracy using INDELREALIGNER (GATK). The base quality was recalibrated using BASERECALIBRATOR/PRINTREADS (GATK) to bring this probability closer to the probability of mismatching the reference genome. At the end of this step, we obtained the analysis-ready reads. Third, we performed SNP and indel discovery across all the sequences simultaneously using UNIFIEDGENOTYPER (GATK). Fourth, we used VQSR to build an adaptive error model using known variant sites and then applied this model to estimate the probability that each variant in the callset is a true genetic variant or a machine/alignment artefact. This step was performed with the VariantRecalibrator/ApplyRecalibration tools of GATK, using a data set of known SNPs and indels obtained from 86 available genomes to train the model (including EC1118 and reference genome S288c). This genotyping pipeline provided us a final multisample VCF file that contained all the variant sites discovered across samples with which a genotyping quality for each strain was associated. The final set was obtained when keeping biallelic variants corresponding to the truth sensitivity threshold of 100% to limit false positives and filtering genotypes with a minimum genotype quality of 30 and a minimum sequencing depth of 10 (HighQuality-SNP set). The genotyping of the 20 strains revealed 59 973 variant positions (indels and biallelic positions). The potential impact of each of these variants corresponding to nonsynonymous positions was then evaluated with the Sorting Intolerant from TOLERANT program (SIFT; Ng & Henikoff 2001), which predicts the probability that a mutation impacts protein structure from the physical properties and conservation of amino acids.

To compare these sequences with 84 other genomes available, genome sequences were downloaded from SGD (Wei *et al.* 2007; Argueso *et al.* 2009; Liti *et al.* 2009; Akao *et al.* 2011; Borneman *et al.* 2011; Engel & Cherry 2013) and aligned pairwise to the S288C reference genome sequence using MUMMER 3.0 (Kurtz *et al.* 2004). SNPs were extracted with a custom script and combined with the set of SNPs obtained from genotyping,

resulting in a set of 427 587 variant positions after filtering out a maximum of 20% missing sites per locus (Allyeast-set). This combined set of variants has been used for the analysis of global phylogeny, population structure analysis, variant differentiating flor strains from other populations, and the subset of wine and flor strains for sweep detection with cross-population composite likelihood ratio (XP-CLR). These sets of variants were then further phased with BEAGLE 3.3 (Browning & Browning 2007).

Genetic analyses

To infer relationships among the strains, the table of SNPs (vcf format) was converted to a fasta file using SNPHYLO version 20140701 (Lee *et al.* 2014) and a maximum-likelihood genealogy was estimated using RAXML version 8.1.3 (Stamatakis 2014) with the GTRGAMMA model of sequence evolution. The population structure was analysed using principal components analysis (PCA) and discriminant analysis of principal components (DAPC) using the R package ADEGENET version 2.0 (Jombart *et al.* 2010; Jombart & Ahmed 2011) and the ADMIXTURE version 1.22 software (Alexander *et al.* 2009).

Comparative analyses of wine and flor strains were carried out only on the set of variants resulting from genotyping with GATK. Metrics of absolute divergence (D_{XY}) and relative divergence [$D_a = D_{XY} - (\pi_{wine} + \pi_{flor})/2$ and F_{st} (Excoffier *et al.* 1992)], as well as a measure of the skewness of the allele frequency spectrum (Tajima's D), were computed using POPGENOME version 2.1.6 R package (Pfeifer *et al.* 2014). F_{IS} were calculated using an R script written by Eva Chan (http://www.evachan.org). Linkage disequilibrium (r^2) was analysed on phased data for wine and flor populations using PLINK version 1.9, and r^2 was averaged in classes of 10 bp distances surrounding pairs of SNPs. Recombination rates were estimated using the PAIRWISE and INTERVAL programs in LDHAT version 2.2a (Auton & McVean 2007). The selection of the most divergent regions (D_{XY}, D_a, PCA differentiating flor strains from all other yeast) was performed using a threshold of 2% to obtain lists of segments corresponding to at most 5% of the total number of ORFs (330). A threshold of 20% was chosen from the distribution of the contribution of nonsynonymous potentially impacting SNPs to the first principal component (Fig. S9, Supporting information) for the selection of SNPs differentiating best wine and flor strains.

Tests for detecting positive selection

We used a combination of approaches for detecting positive selection to identify adaptive events that occurred at different time frames (Oleksyk *et al.* 2010). Positive selection over the long term was investigated based on the rate of accumulation of amino acid changes (approach based on the ratio of nonsynonymous to synonymous substitutions, $\omega = d_N/d_S$) or based on the contrast between intraspecies polymorphism and interspecies divergence (McDonald–Kreitman test). Recent positive selection was investigated based on distinctive footprints on the site frequency spectrum and patterns of linkage disequilibrium.

Tests for positive selection based on the $\omega = d_N/d_S$ ratio were carried out using the codeml program in PAML version 4 (Yang 2007). We used the branch-site test, which is aimed at detecting positive selection that impacted nucleotide substitutions in a specific lineage (the foreground branch; i.e. the flor yeasts) in comparison to the rest of a phylogeny (the background branches). The branch-site test compares the likelihood of the fit of the alignment to the following two models: model A, which assumes two classes of sites with ω_0 between 0 and 1 and $\omega_1 = 1$ in the background branches and one supplementary class with $\omega_2 > 1$ in the foreground branch; in the second (null) model, the ω_2 parameter is set to 1. The test was performed on a subset of genes identified by at least two methods. For each gene, an alignment was obtained with MUSCLE version 3.6, including the alleles of P3-D5 and K1-280-2B (obtained previously for the McDonald–Kreitman test, with the exception of *ZRT1*, for which we used the assembly sequence given its too high divergence from the reference genome, leading to missing variant positions in the variant data set) and the orthologs available at http://sss.genetics.wisc.edu/cgi-bin/s3.cgi for *S. paradoxus* CBS432, *S. uvarum* CBS7001, *S. mikate* IFO1815 and *S. kudriavzevii* IFO1802 or obtained from *S. arboricola* CBS 10644 and *S. eubayanus* CBS 12357 genome sequences (http://www.ncbi.nlm.nih.gov/genome/) when available. For each gene, maximum-likelihood phylogenies were estimated using RAXML v8.1.3 (Stamatakis 2014) with the GTRGAMMA model of sequence evolution.

To carry out the McDonald–Kreitman test, pseudo-DNA sequences were generated for all genes using the phased variant file and the reference sequence, and divergence was measured against *S. paradoxus* (www.saccharomycessensustricto.org). The coding sequence was then translated into amino acids before alignment using MAFFT version 7 (Katoh & Standley 2013). Significance was computed using the MKTEST program in the analysis package version 0.8.4 (Thornton 2003) using a Fischer exact test with correction for multiple testing (Benjamini & Hochberg 1995).

Genome scans for selective sweeps were performed using the XP-CLR (Chen *et al.* 2010). This method relies

on the modelling of the evolutionary trajectory of an allele under linked selection and under neutrality and enables the detection of regions in the genome for which the changes in allelic frequencies occurred too quickly to be due to random drift. The significance threshold was obtained from 1000 neutral simulations performed with ms (Hudson 2002) using the best demographic scenarios inferred for flor yeasts (see Demographic inferences section) and using the global recombination rate inferred for the flor population with the pairwise module of LDhat. For each simulation, the highest value of the XP-CLR statistic was kept and the threshold was chosen as the 95% quantile of simulations.

Demographic inferences

We performed demographic inferences separately for the split between wine and jura/Hungarian flor strains on the one hand and for the split between wine Spanish/Italian flor strains on the other hand. Preliminary analyses indicated that the joint demographic history of the two flor yeast lineages could not be inferred, likely owing to the limited sample size of each group. The analysis was run on a subset of SNPs for which ancestral states could be determined based on alignments with *S. paradoxus* CBS432 and *S. mikate* IFO 1815^T genome sequences. The folded joint allele frequency spectrum was calculated and fitted to different scenarios using a diffusion approximation based method as implemented in δaδi (Gutenkunst *et al.* 2009). Demographic models are presented in Fig. S5A and B (Supporting information).

Translocation and chimeric allele detection

We searched for translocation events from the analysis of discordantly mapped paired-ends with DELLY version 0.2.2 (Rausch *et al.* 2012) using a cut-off of 50 minimum events per translocation site. Translocations were checked by sequencing candidate translocated alleles and by examining variations in sequencing depth.

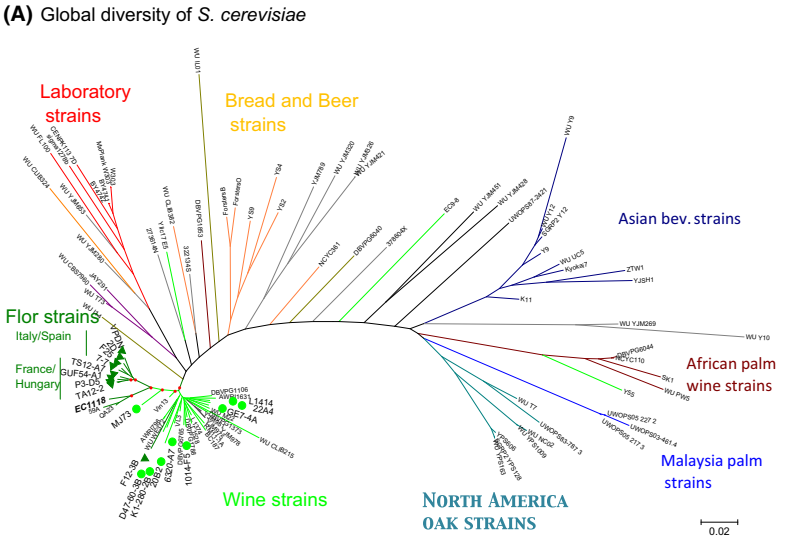
Results

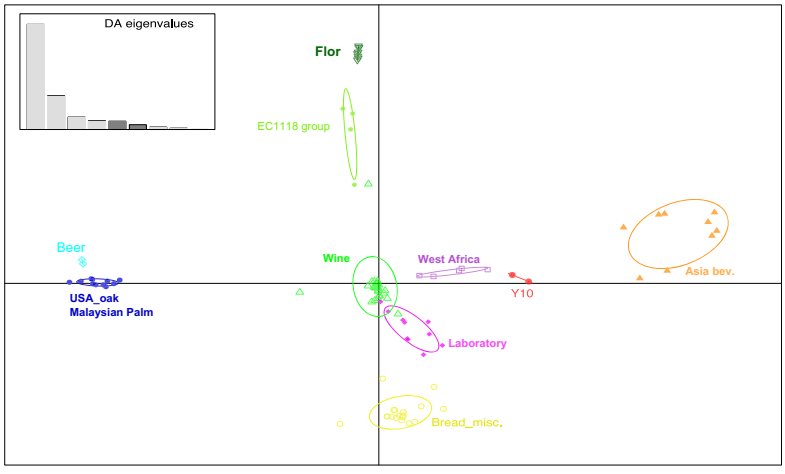
Origins of flor yeasts

To examine the evolutionary history of flor strains in comparison to strains isolated from wine and other origins, we first inferred a maximum-likelihood genome genealogy (Fig. 1A). The genome genealogy revealed the main yeast lineages previously described, such as the lineages associated with wine, Asian beverages, North American oak (Liti *et al.* 2009; Cromie *et al.* 2013) or beer (Gallone *et al.* 2016). Eight of the nine flor strains formed a new cluster that was close but distinct from the cluster including wine yeasts. The flor strain F12-3B did not belong to the main flor lineage and was instead included in the wine group; F12-3B was thus considered as a wine strain for subsequent analyses. The industrial strains EC1118, QA23, Vin13 and wine strain MJ73 formed a group that was basal to the group of flor strains and included recently described Champagne and fructophilic isolates (Borneman *et al.* 2016). Absolute and relative divergences between flor and wine were D_{XY} = 1.29 per kb and D_a = 0.7, respectively, which is lower than those estimated between wine strains and their putative progenitors isolated from Mediterranean oaks (D_{XY} = 2.0, D_a = 1.5 per kb; Almeida *et al.* 2015). Nucleotide diversity was 0.94 per kb in wine strains and 0.53 or 0.28 per kb in flor strains with or without EC1118, respectively (Table 1).

Analyses of population subdivision and admixture

We used a combination of model-based and nonparametric analyses of population structure to analyse more finely the subdivision within the lineage as well as the differentiation between, flor and wine lineages. We first used discriminant analysis of principal components (DAPC; Jombart & Ahmed 2011), which does not rely on any population genetic model and is therefore not affected by deviations from Hardy–Weinberg assumptions, such as those caused by selfing, which is relatively frequent in *S. cerevisiae* (Magwene *et al.* 2011; Fig. 1B). The model with $K = 10$ clusters, delineating isolates from Africa, US Oaks, Asia, and wine/Europe (Liti *et al.* 2009) as well as flor and Champagne strains (closely related to EC1118), was the best model according to the Bayesian information criterion (BIC). For all K values, all individuals had high membership probabilities in a single cluster, suggesting limited admixture. However, because admixture is not specifically addressed in DAPC analyses, a second analysis was performed using the model-based clustering algorithm implemented in the ADMIXTURE software (Alexander *et al.* 2009), which jointly infers K clusters at Hardy–Weinberg and linkage equilibrium and estimates membership proportions of all individuals in the K clusters. The BIC suggested optimal clustering at $K = 9$ clusters (Fig. 1C), separating flor strains from those isolated from wine and other origins. Wine and flor genotypes were not grouped in separate clusters until models with $K > 5$ clusters were used, and in contrast to DAPC, strains related to industrial strain EC1118 were inferred to result from admixtures between the wine and flor yeast gene pools. The mixed ancestry of EC1118 and related strains was confirmed by analyses of shared





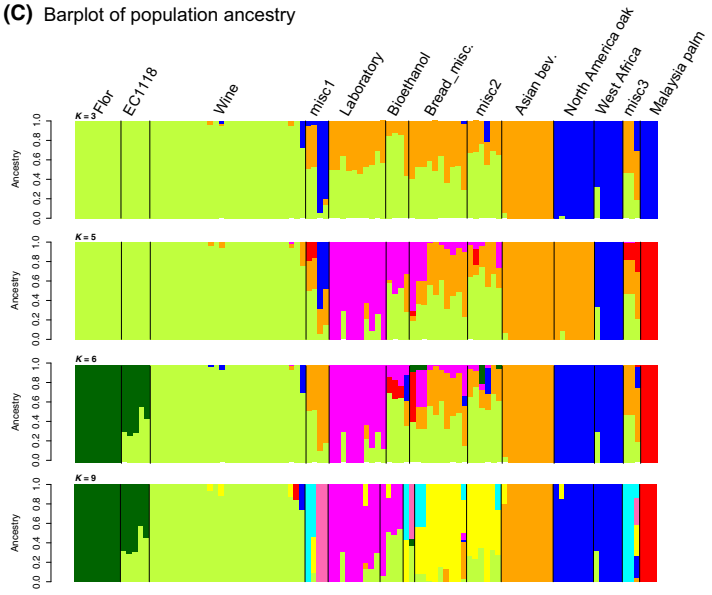


Fig. 1 Whole-genome genealogical relationships among flor and wine strains. (A) Maximum-likelihood genealogy showing the relationships among the nine flor strains (triangles) and nine wine strains (circles) sequenced, in comparison to other groups of yeasts. The genealogy was inferred based on 427 587 SNPs with less than 20% missing. Flor strains are coloured in green, wine strains are in bright green, sake and Asian strains are in dark blue, Oak strains are in blue-green, bread and beer strains are in orange, laboratory strains are in red and African strains are in brown. Bootstrap values greater than 95 are indicated as red dots for the wine and the flor clusters. (B) Analyses of population subdivision based on discriminant analysis of principal components (Jombart *et al.* 2010). (C) Ancestry estimations from ADMIXTURE (Alexander *et al.* 2009) assuming $K = 3$ to $K = 9$ of ancestral clusters.

Table 1 Summary statistics of genomic variation within and between populations of flor and wine *S. cerevisiae* yeasts

	Wine population	Flor population
Number of analysed strains	8	9
Segregating sites*	34 553	8712
Pi (×1000)*	0.942	0.278
Watterson theta (×1000)*	1.126	0.284
Tajima's D†	−0.884	−0.115
F_{IS}†,‡	0.69	0.61
Outcrossing rate	0.18	0.24
Divergence between Wine and Flor populations (×1000)		
D_{XY}†		1.31
Relative divergence between wine and flor populations (×1000)		
D_a†		0.70
Φst†		0.43

*Calculated for the whole genome.
†Values averaged across chromosomes.
‡F_{IS} estimates were obtained from diploid individuals.

haplotype segments implemented in the fineStructure program (Fig. S1, Supporting information) and by gene genealogies of two loci that are well differentiated between wine and flor strains (see section 'Genomic regions of elevated divergence between wine, flor and other yeasts'). Gene genealogies indicated that strains EC1118 and MJ73 harbour two ancestral phases of the high-affinity zinc transporter gene *ZRT1* (Fig. S2a, Supporting information) and of the osmosensor gene *SLN1* (Fig. S2b, Supporting information).

Life cycle of flor yeasts

As a low spore viability has been reported for several flor yeasts (Ibeas & Jimenez 1996; Budroni *et al.* 2000) in comparison to *S. cerevisiae* from other origins, suggesting a mainly asexual life cycle, we compared patterns of linkage disequilibrium and estimates of the recombination rate to gain insight into the contribution of recombination, and thus likely sexual reproduction, to the genomic variability of wine and flor populations. Linkage disequilibrium declined rapidly with distance, reaching half of its maximum value at 1.7 and 5.4 kb in flor and wine populations, respectively (Fig. S3, Supporting information). Estimates of the recombination rate also indicated genomic heterogeneity in rates of recombination within the genomes of each population (Fig. S4, Supporting information). Using the inbreeding coefficient F_{IS} (0.69 for flor yeasts), we estimated an outcrossing rate in the flor population that was similar to previous estimates for the wine population (Magwene *et al.* 2011), with 18% of individuals resulting from an outcrossing event. The decay of linkage disequilibrium with distance in the flor population, similar to wine population, and positive value of F_{IS} do not support that the actual flor population is a clonal lineage.

Demographic inference

To gain insight into the demographic history of wine and flor yeasts and because deviations from the standard neutral model assuming constant population size can bias inferences of positive selection, we used the diffusion-based approach implemented in the δaδi program to infer the historical demography of wine and flor yeasts based on their joint allele frequency spectra. We compared different demographic models, including postdivergence gene flow and population size change (Gutenkunst *et al.* 2009). The most likely scenarios for the divergence between the wine group and the Spanish/Italian and French/Hungarian flor populations, respectively, included demographic expansion of the wine group constant size followed by a recent bottleneck for the two flor populations, and with asymmetrical gene flow (Fig. S5a, b Supporting information).

Differences in genomic content between wine and flor yeasts

We previously identified three genomic regions in the genome of EC1118 and other wine strains (called A, B and C) that have been acquired by horizontal gene transfer from distant yeasts (Novo *et al.* 2009). The latter region, which results from an initial transfer of at least 158 kb from *Torulaspora microellipsoides* into *S. cerevisiae*, provides an ecological advantage in grape must by enhancing the assimilation of oligopeptides from grape musts (Marsit *et al.* 2015). We thus searched for these three regions in flor yeasts. Region A was not found in any flor strain, whereas region B was found in two strains and region C in five strains. Compared to EC1118, which contains a region C of 65 kb, truncated forms were found in the wine strains analysed here, all carrying *FOT1* and *FOT2* genes in addition to one to five other ORFs. In contrast, the four flor strains that contain this fragment (P3-D5, GUF54, TA12-2, 7.7, F25) have the full 65-kb region C described for EC1118 that contains 18 ORFs, including the high-affinity fructose transporter *FSY1* (Galeote *et al.* 2010).

Specific translocations and recombined alleles in flor yeast

A reciprocal translocation between the sulphite export pump *SSU1* and *ECM34* has been shown to be involved

into the adaptation of wine yeast to the adjunct of sulphites in grape must (Pérez-Ortín *et al.* 2002). We thus searched for such translocation events in the genome of flor yeasts. The *ECM34-SSU1* translocation was detected in four flor strains. In addition, recombinations between orthologs were detected in flor strains (Table S3, Supporting information), for example between *FRE2* and *FRE3*, encoding siderophore ferric reductases, and between *HXT3* and *HXT1*, encoding for low-affinity glucose transporters (Fig. S6, Supporting information). The flor *HXT3* allele is identical to a fructophile allele of *HXT3* (Guillaume *et al.* 2007), which enables faster fermentation.

Genomic regions of elevated divergence between wine, flor and other yeasts

Divergent genomic regions between populations exchanging migrants adapted to distinct environments are expected to be enriched in genes involved in adaptive divergence (Akey *et al.* 2002, 2010; Amato *et al.* 2009; Moradi *et al.* 2012) because genomic divergence between populations should be higher in regions resistant to introgression due to divergent selection than in regions homogenized by gene flow. We thus searched for such regions using different methods. As principal components analysis (PCA) has been widely used for the detection of population structure (Patterson *et al.* 2006; McVean 2009) and provides the contribution of each SNP to the axis that differentiate two populations, we applied PCA to the set of genomes representing strains from wine, flor and various other origins (Allyeast-set). Flor strains were differentiated from all other yeast along the sixth axis of the PCA (Fig. 2A). The summed contributions of SNPs to these axes per 500-bp region (Fig. 2B) were used to detect highly differentiated regions between flor and wine strains or between flor and all other strains. Based on the PCA, we identified 319 genes or tRNAs differentiating the flor group from other groups (Table S3, Supporting information). We also computed metrics of absolute and relative genomic divergence (D_{XY} and D_a) between populations (Figs. 2C, S7, Supporting information; Cruickshank & Hahn 2014). Genome scans using D_{XY} and D_a revealed 320 and 311 highly divergent genes or tRNAs between wine and flor, and 159 of 507 genes were common to the sets of genes identified using PCA, D_{XY} and D_a (Fig. 3), including key genes involved in biofilm formation such as the major regulator of cAMP level, the GTPase *IRA1*; *SFG1*, a transcription factor required for growth of superficial pseudohyphae; *HMS2*, a protein with similarity to the heat-shock transcription factor; the osmosensor *SLN1*; and three genes of the MAP-kinase pathway (the MAP-kinase *STE7*; *KDX1*, the protein kinase involved in cell wall integrity; and the CDC42 GTPase-activating MAP-kinase *RGA2*).

Functional impact of genetic variants

The unique ability of flor strains to produce a biofilm suggests that several key molecular functions are modified in this group compared to the wine group. Genetic drift occurring relatively independently in the wine and flor groups should result in the random accumulation of mutations and not in mutations enriched in functional categories. Because divergence-based analyses do not account for the synonymy of SNPs, we focused our analysis on nonsynonymous SNPs, differentiating wine and flor strains for which we could predict an impact on protein function. Almost one-third (2086) of the 6607 ORFs of wine and flor *S. cerevisiae* strains (HighQuality-SNP set) contained nonsynonymous mutations that may impact their function. However, focusing on SNPs in the top 20% fraction differentiating wine and flor strains, we could detect an enrichment for 19 'biological process' and 15 'molecular functions' GO categories, including the 'fungal cell wall organization', 'signal transduction' and 'zinc ion transport' categories (adj. *P* value = 0.008, 0.017 and 0.006, respectively) relevant in biological ageing. In addition, these categories encompassed 44.4% of the initial set containing 583 genes.

Signatures of positive selection in flor yeasts

To identify genes potentially involved in adaptation to the biological ageing environment, we used different methods aimed at detecting positive selection. We first carried out the McDonald–Kreitman test, which compares the ratio of nonsynonymous to synonymous polymorphisms of the flor groups to the ratio of nonsynonymous to synonymous divergence with *S. paradoxus*, the closest species to *S. cerevisiae*, and thus searches for positive selection on a wide time frame. When carrying out this test for the 561 genes with enough polymorphism in flor strains, none deviated from neutral expectations after correction for multiple tests.

We then searched for positive selection in the flor branch based on the ratio of nonsynonymous to synonymous substitution rates ($\omega = d_N/d_S$). This approach is inaccurate for intrapopulation studies (Kryazhimskiy & Plotkin 2008), but it may be applied here given the isolation of wine and flor populations, which probably occurred after more than *Ne* generations according to our demographic inferences (Kryazhimskiy & Plotkin 2008). The branch-site test of positive selection was shown to be more robust and powerful than branch-based tests that average substitution rates over all codons (Yang & Dos Reis 2011). To reasonably limit the

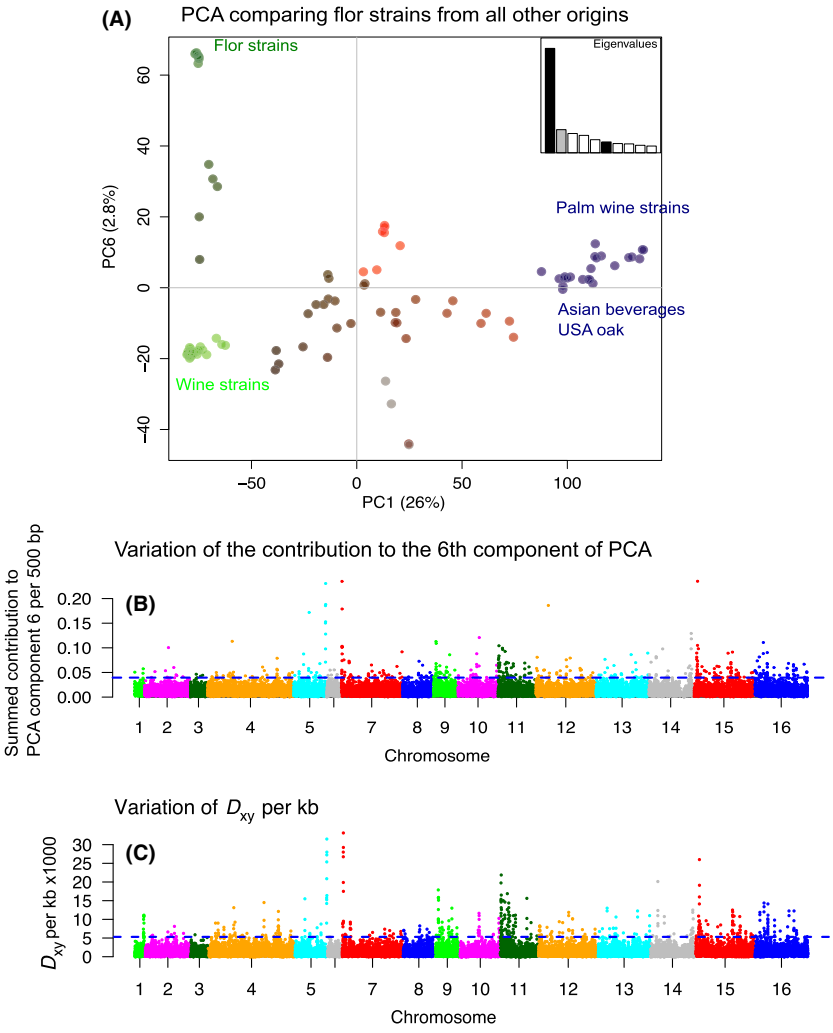


Fig. 2 (A) Principal components analysis (PCA) of genomic variation in flor yeasts, wine yeasts and yeasts from multiple origins. The first and sixth axes (PC1, PC2) represent 26% and 2.8% of the global variance, respectively. Wine strains are in dark green, flor strains are in light green and sake, palm wine and USA oak strains are in blue. (B) Genome scan showing the regions that differentiate wine and flor yeasts. (A) Contribution of individual SNPs to the sixth axis of PCA per 500-bp segment. The blue dashed line indicates the 2% segment that has the highest contribution. (C) Genome scan of absolute divergence (D_{XY}) between the wine and flor strains (×1000 per kb), per 1-kb sliding window (500 bp step).

computational burden, the analysis was performed using a subset of 232 genes that were identified as highly divergent in analyses based on PCA or divergence metrics. After correction for multiple tests, positive selection was inferred for the flor allele of *ZRT1* (adj. *P* value = 0.013) at the following three residues: S15, E16 and Q306 (Table 2). In addition, three genes with interesting functions displayed *P* values close to significance levels after correction for multiple tests: *HMS2* can restore pseudohyphal growth in a strain defective for ammonium transport (*P* value 0.005); *CCS1* (*P* value = 0.007), a chaperon for superoxide dismutase *SOD1*, is involved in oxidative stress protection; and *ADH6* (*P* value = 0.003) provides yeast resistance to hydroxymethyl furfural through its reduction into 5-hydroxymethylfurfuryl alcohol (Petersson *et al.* 2006) and might offer detoxification against the various aldehydes produced during biological ageing.

For detecting genes under recent selective pressures, we used the cross-population composite likelihood ratio, XP-CLR (Chen *et al.* 2010), which relies on the extent of multilocus genetic differentiation around a selected variant and was suggested recently to be one of the most sensitive methods for detecting recent selective sweeps (Vatsiou *et al.* 2016). We performed a genomic scan for selective sweeps in the Mediterranean flor strains group and the Jura/Hungary flor strains group, using the results of demographic inference to set significance thresholds (see above demographic inferences). Three genomic regions shared by the two groups of flor strains displayed a selective sweep signature (Table S3, Supporting information). These regions encompassed the high-affinity glucose transporter gene *HXT7*; the gene *MST27*, which is involved in vesicle formation; and the gene *YPS6*, which codes for a putative GPI-anchored aspartic protease involved in cell wall growth and maintenance.

Phenotypic impact of genetic variation in candidate genes from flor strains

To assess the importance of genes detected in analyses of divergence or positive selection for biological ageing

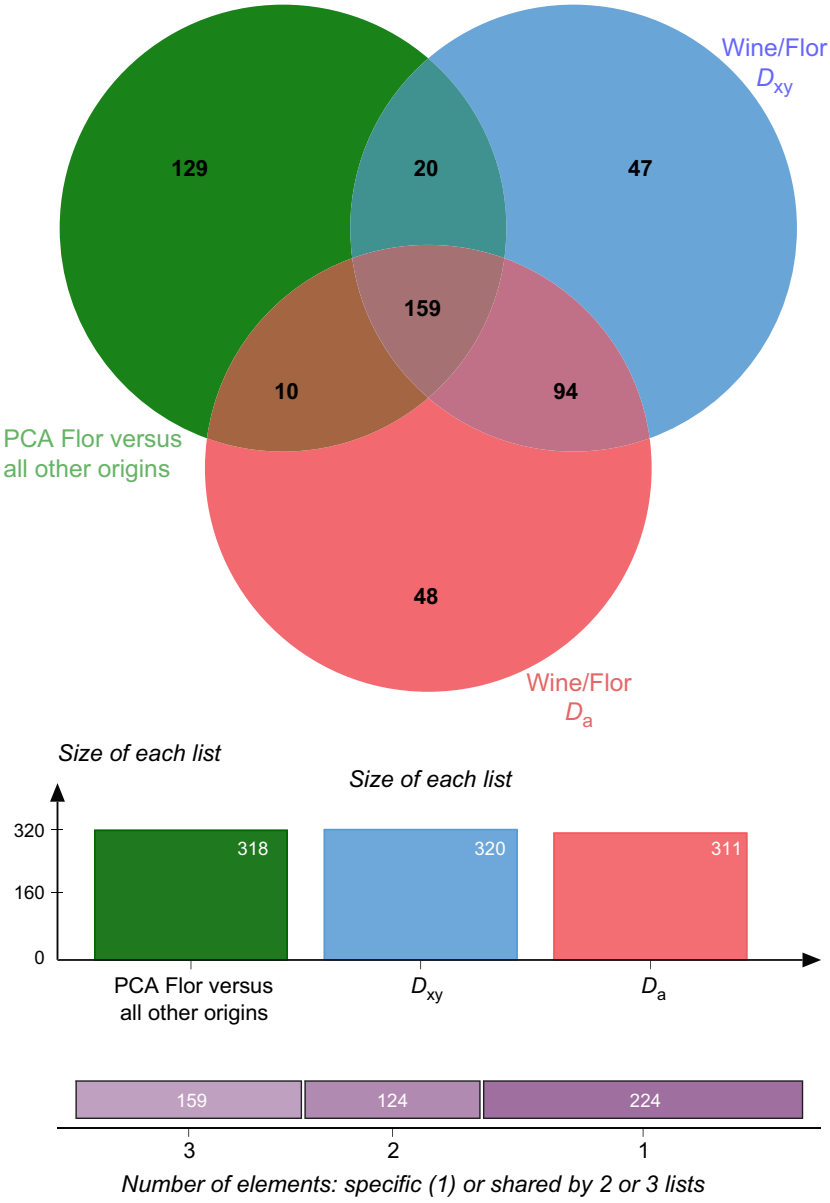


Fig. 3 Venn diagram comparing the three sets of highly divergent genes between flor and wine yeasts, identified using principal components analysis, absolute divergence (D_{XY}) and relative divergence (D_a).

and to examine how they contribute to the adaptation of flor yeasts to their habitat, we performed various physiological assays. We focused on genes known to be associated with biofilm formation, osmotic pressure and zinc transport (*RGA2*, *SFL1*, *SLN1* and *ZRT1*). We found an impact of allelic variation on the two genes (*SFL1* and *RGA2*) participating in the regulation of *FLO11* (Fig. 4). The presence of the flor allele of *SFL1* enhanced velum formation in comparison to a strain carrying a wine allele. Similarly, flor strains present a unique allele of *RGA2* that enhances biofilm formation on liquid media (Fig. 4) in an equivalent manner to a strain carrying an inactivated allele of *RGA2*. Because Rga2 acts as an inactivator of Cdc42, a GTPase member of the MAPK cascade that regulates *FLO11* expression (Ueno 1999; Sopko *et al.* 2007), the flor allele of *RGA2* could present an attenuated activity in comparison to the wine allele. In addition, a frameshift in the *RGA2* sequence that leads to a premature stop codon was observed in one Spanish flor strain.

The comparison of hemizygous strains carrying either the flor or wine allele of *SLN1* (Fig. 4) revealed differences in the growth of the velum produced by the two strains, but surprisingly, the growth was better for the strain carrying the wine allele under our test conditions. The two alleles of wine and flor strains differed by 25 amino acid residues (Fig. S8, Supporting information). Finally, introduction of the flor allele of *ZRT1* in a flor

Table 2 Results of tests of positive selection based on nonsynonymous to synonymous substitution rates. Significance after correction of multiple testing is indicated in bold. Nsite: number of sites

Gene	Gene name	N. species in alignment	Nsite in the alignment	Likelihood Model A	Likelihood null Model	*P* value	Adj. *P* value
YJR147W	*HMS2*	8	999	−4695.61	−4699.48	0.00542	0.261
YLR044C	*PDC1*	8	513	−3578.57	−3586.37	0.00008	0.013*
YMR038C	*CCS1*	8	669	−2480.15	−2483.82	0.00679	0.273
YMR318C	*ADH6*	8	813	−2993.65	−2998.14	0.00272	0.164
YGL255W	***ZRT1***	**7**	**1113**	**−3961.97**	**−3969.46**	**0.00011**	**0.013**

*This *P* value is caused by the homologous recombination between orthologous genes *PDC1* and *PDC5* and cannot be considered as indicative of positive selection.

strain did not induce clear differences in velum growth (data not shown).

Discussion

Flor strains represent a specific lineage

We provide the first phylogenetic picture of flor strains obtained from genome-scale data. Our data show that the group of flor strains is divergent from wine strains and is much less diverse (mean nucleotide diversity represents 60% of that calculated for wine strains). Demographic inferences revealed different demographic histories for wine and flor populations. While the wine yeast population presented a clear pattern of demographic expansion (Almeida *et al.* 2015), flor yeast went through a recent bottleneck. The fact that changes in winemaking technology have led to the stricter protection of wines from oxygen, except for wines exposed to biological ageing, is consistent with the inferred demographic histories. We also found that the group of strains related to the Champagne strain EC1118 and related strains (QA23, VIN13, MJ73; Borneman *et al.* 2016) results from a cross between flor and wine strains. Some of the mechanisms underlying the ability of flor strains to grow on nutritionally depleted media and cope with the ethanol stress of wine production may be helpful during the second fermentation of the 'prise de mousse' step that imposes a second anaerobic growth on wine.

Genomic changes underlying adaptation to wine biological ageing

We identified 155 loci that are highly divergent between wine and flor yeasts based on different methods. Thirty-seven of the divergent genes also present nonsynonymous potentially impacting substitutions and therefore may play a key role in adaptation to the biological ageing conditions. Five of the 37 candidate genes (*IRA1*, *SFG1*, *HMS2*, *IME4* coding for a mRNA N6-adenosine methyltransferase and *PIK1* coding for a phosphatidylinositol 4-kinase) were shown to impact pseudohyphal growth (Lorenz & Heitman 1998; Fujita *et al.* 2005; Cullen & Sprague 2012; Adhikari & Cullen 2015).

In addition to highly divergent genes, d_N/d_S analyses revealed signatures of repeated positive selection in *ZRT1* and possibly three other genes (*HMS2*, *ALD6* and *CCS1*). The cross-population composite likelihood ratio (XP-CLR) test, which is aimed at detecting more recent targets of selection, revealed signatures of selective sweeps in genomic regions surrounding *HXT7* and *YPS6*, two genes with functions of interest for biological ageing (high-affinity sugar transporter and protease).

Functional changes associated with adaptation to wine biological ageing

Adaptation to biofilm formation. When exposed to limited nutrients, *S. cerevisiae* can induce pseudohyphal growth biofilm or mat formation (Reynolds & Fink 2001; Karunanithi *et al.* 2012). These responses lead to the induction of *FLO11*, the main actor of cell–cell adhesion, and are mediated by the same main pathways, the MAPK and Ras/cAMP/PKA pathways and nutrient signalization pathways (Fig. 5). In agreement with this finding, the adhesin *FLO11* and the transcription factor *SFL1* (Reynolds & Fink 2001; Zara *et al.* 2005; Fidalgo *et al.* 2006) were differentiated between flor and wine lineages. The differentiation of noncoding RNA *ICR1* (Bumgarner *et al.* 2009) was only observed with D_{XY}, but one wine yeast (6320-A7) contains a flor allele, which may explain a slightly lower rank with other metrics. We demonstrated the role of allelic variation of the genes *RGA2* and *SFL1*, which both present deleterious mutations relieving *FLO11* from negative regulation. The role of *HMS2* and *SFG1* remains to be investigated. In addition, 26 other differentiated genes were found that are important for pseudohyphal

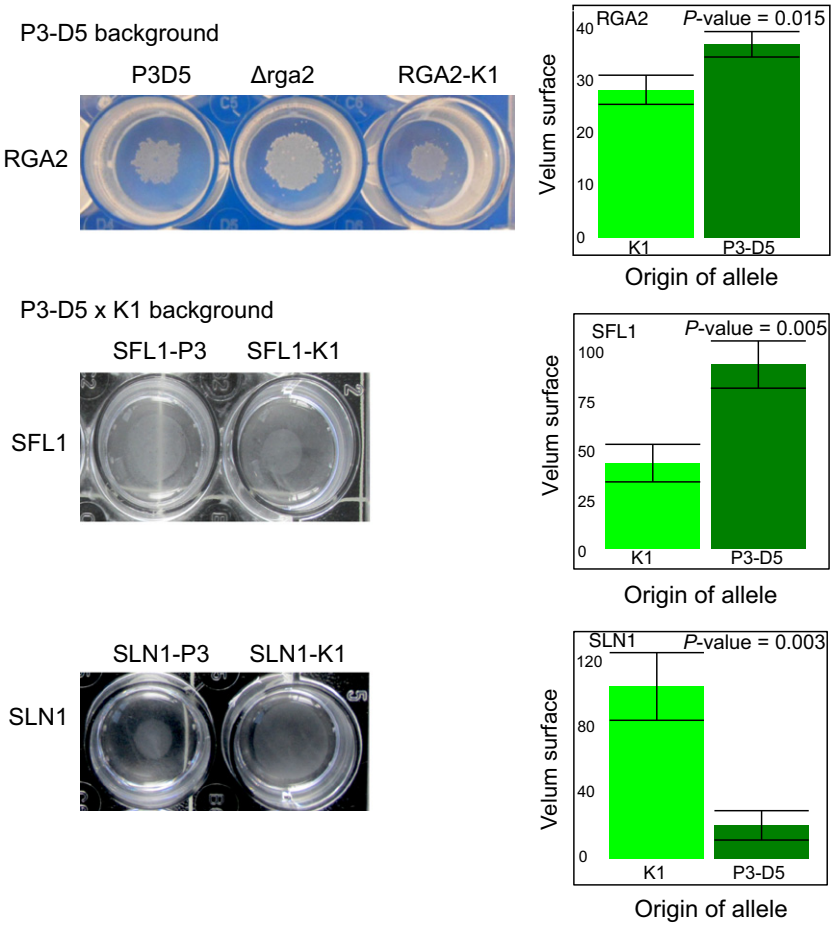


Fig. 4 Comparison of the impact of wine and flor alleles of different genes on biofilm formation. Strains were grown on YNB with 4% ethanol and evaluated after incubation at 23 °C for 3, 5 and 7 days for SLN1, RGA2 and SFL1, respectively, in triplicate. The comparison was performed using the haploid strain P3-D5 mat alpha carrying the wine or flor allele of *RGA2* after allelic exchange, or in hemizygotes of P3-D5 MATα x K1-280-2B MATα zygote in which one copy of the gene has been inactivated for *SFL1* and *SLN1*. Errors bars correspond to one SD.

growth or biofilm formation (Ryan *et al.* 2012), and other genes participating in that mechanism were found to carry nonsynonymous mutations. Allelic divergence was observed in the regulatory pathways associated with *FLO11* regulation, such as the Ras/cAMP/PKA signalling pathway and the MAP-kinase signalling pathways (*IRA1* or *GPB1*; Fig. 5).

Adaptation to carbon sources. At the end of alcoholic fermentation, ethanol is the main carbon source, followed by glycerol and the traces of fructose that might remain. We thus expected signatures of adaptive divergence in genes involved in the use of nonglucose carbon sources. Our study reveals the fructophily of flor yeasts, attested by the presence of a fructophile allele of *HXT3* (Guillaume *et al.* 2007) and the presence of a full region C including the high-affinity transporter of fructose *FSY1* (Galeote *et al.* 2010), which may help cells to utilize the traces of fructose that are present in wine. The finding of a selective sweep signature around the *HXT6-7* region adjacent to *HXT3* is also consistent with adaptation to the limited availability of glucose. The exhaustion of glucose also implies active gluconeogenesis. A candidate for adaptive divergence is the gene *MDH2*, which displays a specific allele in the flor lineage and encodes a cytoplasmic malate dehydrogenase that catalyses the interconversion of malate and oxaloacetate and plays a major role in gluconeogenesis during growth on two-carbon compounds.

Flor yeast and osmotic pressure. Grapes contain a high content of sugars, but given the conversion of each molecule of glucose into two molecules of ethanol, the osmotic pressure in wine is higher than in grape must. Unlike wine yeasts, which have lost functional aquaporins and gained a higher fitness in the grape must environment (Will *et al.* 2010), flor strains contain a functional *AQY2* gene that has also been shown to be involved in the control of cell surface properties and impact the pseudohyphal growth (Furukawa *et al.* 2009). In addition to aquaporins, the homoeostasis of osmotic pressure in yeast is ensured by the high-osmolarity glycerol response (HOG pathway). *SLN1* functions as an osmosensor that permits the adjustment of intracellular glycerol concentration either after an hyperosmotic stress via Ssk1 or in response to an hypoosmotic shock via Skn7 (Saito & Posas 2012). Skn7 participates in the regulation of cell wall integrity that is

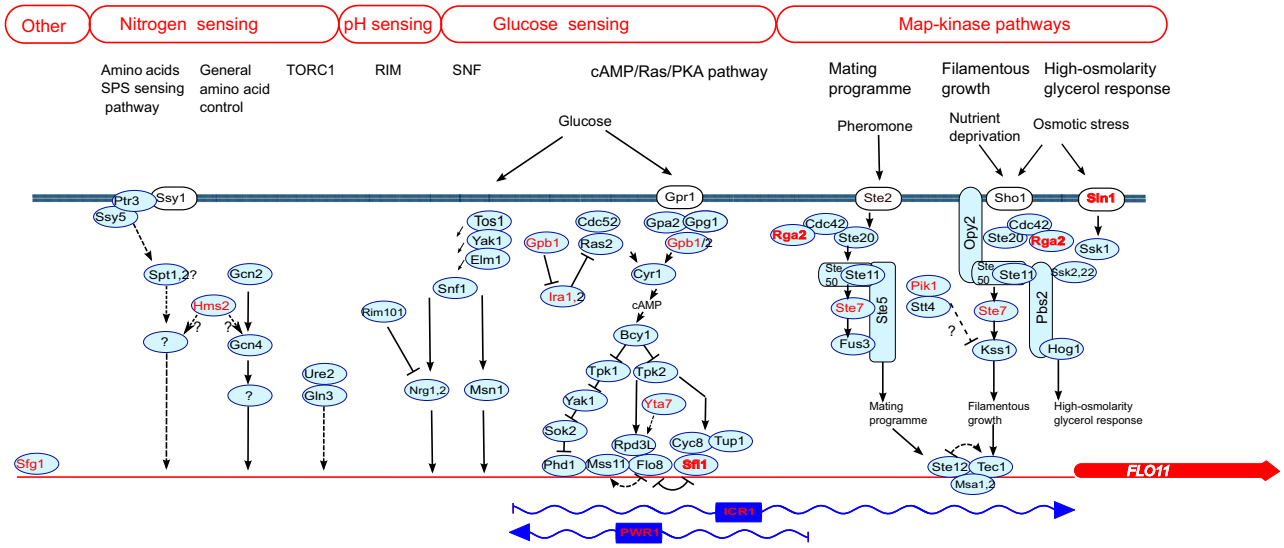


Fig. 5 Schematic representation of the multiple regulatory circuits involved in *FLO11* regulation. Genes differentiated between wine and flor strains are indicated in red. Differentiated genes with unequal velum growth permitted by the two alleles are indicated in bold. Adapted from Brückner & Mösch (2011).

critical for growth under hypo-osmotic conditions. The slower velum growth measured for the hemizygote containing a flor allele in comparison to the wine allele indicates that the two alleles are not equivalent in terms of sensing and osmotic response. We cannot rule out the existence of negative interactions between the flor *SLN1* allele and other genes of the wine background of the hemizygote leading to an apparent lower fitness under biological ageing conditions. Interestingly, Kvitek and Sherlock observed that adaptation to a low-glucose culture leads to a high frequency of mutations in the Ras/cAMP/PKA pathway and the high-osmolarity glycerol (HOG) response pathway (Kvitek & Sherlock 2013).

Flor yeast and divalent metal transport. One of our most puzzling findings is the enrichment of divergent genomic regions in genes involved in zinc ion transmembrane transport and signatures of positive selection at *ZRT1* and possibly *CCS1*. Zinc is an essential nutrient for all organisms, playing a structural role in many proteins and being a catalytic component for over 300 enzymes such as superoxide dismutase *SOD1* (Vallee & Auld 1990). In *C. albicans*, zinc has an essential role in biofilm formation (Nobile *et al.* 2009), and *ZRT1* is involved in pathogenicity via its role in zinc absorption from the host (Citiulo *et al.* 2012). *ZRT1* has already been reported to present evidence of balancing selection in *S. cerevisiae* (Engle & Fay 2013) and the alleles of different origins present large variations. We found that the flor allele of *ZRT1* is much closer to the allele found in African palm wine yeasts than to the allele found in wine yeasts, while the topologies of the tree of neighbouring genes *ADH4* and *FZF1* are in agreement with the global phylogeny (Fig. S2a, Supporting information). This phylogenetic structure suggests that the flor allele may have been introgressed from a distant group (African or US Oak lineages).

Conclusions

Comparative population genomics revealed the phylogenetic origin of flor yeasts and identified genomic regions putatively involved in the adaptive divergence between pelagic wine yeasts fermenting glucose and biofilm-forming flor yeasts utilizing ethanol. Candidate genomic regions involved in adaptation to biological ageing conditions included multiple genes belonging to specific regulatory networks such as genes involved in cell–cell adhesion, zinc transport, hexose transport or signalling pathways. Our results suggest that the tuning of these regulatory networks is a major genomic signature of flor yeast domestication, highlighting the plurality of evolutionary changes underlying adaptation to biological ageing conditions. More generally, the high differentiation and limited gene flow observed between flor and wine lineages make this system an excellent model for studying the early stages of ecological speciation.

Acknowledgements

This work was supported by the grant AIP BioRessources (GENOWINE project) from the French National Institute for Agricultural Research (INRA). Anna Lisa Coi gratefully acknowledges the Sardinia Regional Government for the

financial support of her PhD and postdoc scholarships (P.O.R. Sardegna F.S.E. Operational Programme of the Autonomous Region of Sardinia, European Social Fund 2007–2013 – Axis IV Human Resources, Objective l.3, Line of Activity l.3.1). The authors are also thankful to three anonymous reviewers and to the editor for their insightful comments which helped to improve the manuscript.

References

Adhikari H, Cullen PJ (2015) Role of phosphatidylinositol phosphate signaling in the regulation of the filamentous-growth mitogen-activated protein kinase pathway. *Eukaryotic Cell*, **14**, 427–440.

Akao T, Yashiro I, Hosoyama A *et al.* (2011) Whole-genome sequencing of sake yeast Saccharomyces cerevisiae Kyokai no. 7. *DNA Research: An International Journal for Rapid Publication of Reports on Genes and Genomes*, **18**, 423–434.

Akey JM, Zhang G, Zhang K, Jin L, Shriver MD (2002) Interrogating a high-density SNP map for signatures of natural selection. *Genome Research*, **12**, 1805–1814.

Akey JM, Ruhe AL, Akey DT *et al.* (2010) Tracking footprints of artificial selection in the dog genome. *Proceedings of the National Academy of Sciences of the United States of America*, **107**, 1160–1165.

Alexander DH, Novembre J, Lange K (2009) Fast model-based estimation of ancestry in unrelated individuals Fast model-based estimation of ancestry in unrelated individuals. *Genome Research*, **19**, 1655–1664.

Almeida P, Barbosa R, Zalar P *et al.* (2015) A population genomics insight into the Mediterranean origins of wine yeast domestication. *Molecular Ecology*, **24**, 5412–5427.

Amato R, Pinelli M, Monticelli A *et al.* (2009) Genome-wide scan for signatures of human population differentiation and their relationship with natural selection, functional pathways and diseases. *PLoS ONE*, **4**, e7927.

Argueso JL, Carazzolle MF, Mieczkowski PA *et al.* (2009) Genome structure of a *Saccharomyces cerevisiae* strain widely used in bioethanol production. *Genome Research*, **19**, 2258–2270.

Auton A, McVean G (2007) Recombination rate estimation in the presence of hotspots. *Genome Research*, **17**, 1219–1227.

Baker AE, Wang B, Bellora N *et al.* (2015) The genome sequence of *Saccharomyces eubayanus* and the domestication of lager-brewing yeasts. *Molecular Biology and Evolution*, **11**, 2018–2031.

Barth G, Gaillardin C (1996) Yarrowia lipolytica. In: *Non-Conventional Yeasts in Biotechnology: A Handbook* (ed. Wolf K), pp. 313–388. Springer, Berlin, Heidelberg, New York.

Benjamini Y, Hochberg Y (1995) Controlling the false discovery rate: a practical and powerful approach to multiple testing. *Journal of the Royal Statistical Society*, **57**, 289–300.

Borneman AR, Desany BA, Riches D *et al.* (2011) Whole-genome comparison reveals novel genetic elements that characterize the genome of industrial strains of *Saccharomyces cerevisiae* (G Sherlock, Ed.). *PLoS Genetics*, **7**, e1001287.

Borneman AR, Forgan AH, Kolouchova R, Fraser JA, Schmidt SA (2016) Whole genome comparison reveals high levels of inbreeding and strain redundancy across the spectrum of commercial wine strains of *Saccharomyces cerevisiae*. *G3: Genes|Genomes|Genetics*, **6**, 957–971.

Browning SR, Browning BL (2007) Rapid and accurate haplotype phasing and missing-data inference for whole-genome association studies by use of localized haplotype clustering. *American Journal of Human Genetics*, **81**, 1084–1097.

Brückner S, Mösch H-U (2011) Choosing the right lifestyle: adhesion and development in *Saccharomyces cerevisiae*. *FEMS Microbiology Reviews*, **36**, 25–58.

Budroni M, Giordano G, Pinna G, Farris GA (2000) A genetic study of natural flor strains of *Saccharomyces cerevisiae* isolated during biological ageing from Sardinian wines. *Journal of Applied Microbiology*, **89**, 657–662.

Bumgarner SL, Dowell RD, Grisafi P, Gifford DK, Fink GR (2009) Toggle involving cis-interfering noncoding RNAs controls variegated gene expression in yeast. *Proceedings of the National Academy of Sciences of the United States of America*, **106**, 18321–18326.

Chen H, Patterson N, Reich D (2010) Population differentiation as a test for selective sweeps. *Genome Research*, **20**, 393–402.

Citiulo F, Jacobsen ID, Miramón P *et al.* (2012) Candida albicans scavenges host zinc via Pra1 during endothelial invasion. *PLoS Pathogens*, **8**, e1002777.

Coi AL, Legras J-L, Zara G, Dequin S, Budroni M (2016) A set of haploid strains available for genetic studies of *Saccharomyces cerevisiae* flor yeasts. *FEMS Yeast Research*, **6**, 1–26.

Cromie GA, Hyma KE, Ludlow CL *et al.* (2013) Genomic sequence diversity and population structure of *Saccharomyces cerevisiae* assessed by RAD-seq. *G3 (Bethesda, Md.)*, **3**, 2163–2171.

Cruickshank TE, Hahn MW (2014) Reanalysis suggests that genomic islands of speciation are due to reduced diversity, not reduced gene flow. *Molecular Ecology*, **23**, 3133–3157.

Cullen PJ, Sprague GF (2012) The regulation of filamentous growth in yeast. *Genetics*, **190**, 23–49.

Engel SR, Cherry JM (2013) The new modern era of yeast genomics: community sequencing and the resulting annotation of multiple *Saccharomyces cerevisiae* strains at the Saccharomyces Genome Database. *Database: The Journal of Biological Databases and Curation*, **2013**, bat012.

Engle EK, Fay JC (2013) ZRT1 harbors an excess of nonsynonymous polymorphism and shows evidence of balancing selection in *Saccharomyces cerevisiae*. *G3 (Bethesda, Md.)*, **3**, 665–673.

Estreicher SK (2006) *Wine from Neolithic Times to the 21st Century*. Algora Publishing, New York.

Excoffier L, Smouse PE, Quattro JM (1992) Analysis of molecular variance inferred from metric distances among DNA haplotypes: application to human mitochondrial DNA restriction data. *Genetics*, **131**, 479–491.

Fay JC, Benavides JA (2005) Evidence for domesticated and wild populations of *Saccharomyces cerevisiae*. *PLoS Genetics*, **1**, 66–71.

Fidalgo M, Barrales RR, Ibeas JI, Jimenez J (2006) Adaptive evolution by mutations in the FLO11 gene. *Proceedings of the National Academy of Sciences of the United States of America*, **103**, 11228–11233.

Fujita A, Hiroko T, Hiroko F, Oka C (2005) Enhancement of superficial pseudohyphal growth by overexpression of the SFG1 gene in yeast *Saccharomyces cerevisiae*. *Gene*, **363**, 97–104.

Furukawa K, Sidoux-Walter F, Hohmann S (2009) Expression of the yeast aquaporin Aqy2 affects cell surface properties

under the control of osmoregulatory and morphogenic signalling pathways. *Molecular Microbiology*, **74**, 1272–1286.

Galagan JE, Calvo SE, Borkovich KA *et al.* (2003) The genome sequence of the filamentous fungus Neurospora crassa. *Nature*, **422**, 859–868.

Galeote V, Novo M, Salema-Oom M *et al.* (2010) FSY1, a horizontally transferred gene in the *Saccharomyces cerevisiae* EC1118 wine yeast strain, encodes a high-affinity fructose/H+ symporter. *Microbiology (Reading, England)*, **156**, 3754–3761.

Gallone B, Steensels J, Prahl T *et al.* (2016) Domestication and divergence of *Saccharomyces cerevisiae* Beer Yeasts. *Cell*, **166**, 1397–1410.e16.

Gibbons JG, Rinker DC (2015) The genomics of microbial domestication in the fermented food environment. *Current Opinion in Genetics & Development*, **35**, 1–8.

Gitan RS, Luo H, Rodgers J, Broderius M, Eide D (1998) Zinc-induced inactivation of the yeast ZRT1 zinc transporter occurs through endocytosis and vacuolar degradation. *Journal of Biological Chemistry*, **273**, 28617–28624.

Gladieux P, Ropars J, Badouin H *et al.* (2014) Fungal evolutionary genomics provides insight into the mechanisms of adaptive divergence in eukaryotes. *Molecular Ecology*, **23**, 753–773.

Guillaume C, Delobel P, Sablayrolles J-M, Blondin B (2007) Molecular basis of fructose utilization by the wine yeast *Saccharomyces cerevisiae*: a mutated HXT3 allele enhances fructose fermentation. *Applied and Environmental Microbiology*, **73**, 2432–2439.

Gutenkunst RN, Hernandez RD, Williamson SH, Bustamante CD (2009) Inferring the joint demographic history of multiple populations from multidimensional SNP frequency data. *PLoS Genetics*, **5**, e1000695.

Ho WC, Zhang J (2014) The genotype-phenotype map of yeast complex traits: basic parameters and the role of natural selection. *Molecular Biology and Evolution*, **31**, 1568–1580.

Hudson RR (2002) Generating samples under a Wright-Fisher neutral model of genetic variation. *Bioinformatics*, **18**, 337–338.

Ibeas JI, Jimenez J (1996) Genomic complexity and chromosomal rearrangements in wine-laboratory yeast hybrids. *Current Genetics*, **30**, 410–416.

Ibeas JI, Jimenez J (1997) Mitochondrial DNA loss caused by ethanol in Saccharomyces Flor yeasts. *Applied and Environmental Microbiology*, **63**, 7–12.

Jombart T, Ahmed I (2011) adegenet 1.3-1: new tools for the analysis of genome-wide SNP data. *Bioinformatics (Oxford, England)*, **27**, 3070–3071.

Jombart T, Devillard S, Balloux F (2010) Discriminant analysis of principal components: a new method for the analysis of genetically structured populations. *BMC Genetics*, **11**, 94.

Karunanithi S, Joshi J, Chavel C *et al.* (2012) Regulation of mat responses by a differentiation MAPK pathway in *Saccharomyces cerevisiae*. *PLoS ONE*, **7**, e32294.

Katoh K, Standley DM (2013) MAFFT multiple sequence alignment software version 7: improvements in performance and usability. *Molecular Biology and Evolution*, **30**, 772–780.

Kryazhimskiy S, Plotkin JB (2008) The population genetics of d_N/d_S. *PLoS Genetics*, **4**, e1000304.

Kurtz S, Phillippy A, Delcher AL *et al.* (2004) Versatile and open software for comparing large genomes. *Genome Biology*, **5**, R12.

Kvitek DJ, Sherlock G (2013) Whole genome, whole population sequencing reveals that loss of signaling networks is the major adaptive strategy in a constant environment (J Zhang, Ed.). *PLoS Genetics*, **9**, e1003972.

Lee T-H, Guo H, Wang X, Kim C, Paterson AH (2014) SNPhylo: a pipeline to construct a phylogenetic tree from huge SNP data. *BMC Genomics*, **15**, 162.

Legras J-L, Merdinoglu D, Cornuet J-M, Karst F (2007) Bread, beer and wine: *Saccharomyces cerevisiae* diversity reflects human history. *Molecular Ecology*, **16**, 2091–2102.

Legras J-L, Erny C, Charpentier C (2014) Population structure and comparative genome hybridization of european flor yeast reveal a unique group of *Saccharomyces cerevisiae* strains with few gene duplications in their genome. *PLoS ONE*, **9**, e108089.

Li H, Durbin R (2009) Fast and accurate short read alignment with Burrows-Wheeler transform. *Bioinformatics (Oxford, England)*, **25**, 1754–1760.

Li R, Zhu H, Ruan J *et al.* (2010) De novo assembly of human genomes with massively parallel short read sequencing. *Genome Research*, **20**, 265–272.

Liti G, Carter DM, Moses AM *et al.* (2009) Population genomics of domestic and wild yeasts. *Nature*, **458**, 337–341.

Lorenz MC, Heitman J (1998) The MEP2 ammonium permease regulates pseudohyphal differentiation in *Saccharomyces cerevisiae*. *EMBO Journal*, **17**, 1236–1247.

Magwene PM, Kayıkçı Ö, Granek JA *et al.* (2011) Outcrossing, mitotic recombination, and life-history trade-offs shape genome evolution in *Saccharomyces cerevisiae*. *Proceedings of the National Academy of Sciences of the United States of America*, **108**, 1987–1992.

Marsit S, Mena A, Bigey F *et al.* (2015) Evolutionary advantage conferred by an eukaryote-to-eukaryote gene transfer event in wine yeasts. *Molecular Biology and Evolution*, **32**, 1695–1707.

McKenna A, Hanna M, Banks E *et al.* (2010) The genome analysis toolkit: a MapReduce framework for analyzing next-generation DNA sequencing data. *Genome Research*, **20**, 1297–1303.

McVean G (2009) A genealogical interpretation of principal components analysis. *PLoS Genetics*, **5**, e1000686.

Moradi MH, Nejati-Javaremi A, Moradi-Shahrbabak M, Dodds KG, McEwan JC (2012) Genomic scan of selective sweeps in thin and fat tail sheep breeds for identifying of candidate regions associated with fat deposition. *BMC Genetics*, **13**, 10.

Ng PC, Henikoff S (2001) Predicting deleterious amino acid substitutions. *Genome Research*, **11**, 863–874.

Nobile CJ, Nett JE, Hernday AD *et al.* (2009) Biofilm matrix regulation by Candida albicans Zap1. *PLoS Biology*, **7**, e1000133.

Novo M, Bigey F, Beyne E *et al.* (2009) Eukaryote-to-eukaryote gene transfer events revealed by the genome sequence of the wine yeast *Saccharomyces cerevisiae* EC1118. *Proceedings of the National Academy of Sciencesv*, **106**, 16333–16338.

Oleksyk TK, Smith MW, O'Brien SJ (2010) Genome-wide scans for footprints of natural selection. *Philosophical Transactions of the Royal Society of London. Series B, Biological Sciences*, **365**, 185–205.

Patterson N, Price AL, Reich D (2006) Population structure and eigenanalysis. *PLoS Genetics*, **2**, e190.

Pérez-Ortín JE, Querol A, Puig S, Barrio E (2002) Molecular characterization of a chromosomal rearrangement involved in the adaptive evolution of yeast strains. *Genome Research*, **12**, 1533–1539.

Petersson A, Almeida JRM, Modig T *et al.* (2006) A 5-hydroxymethyl furfural reducing enzyme encoded by the *Saccharomyces cerevisiae* ADH6 gene conveys HMF tolerance. *Yeast (Chichester, England)*, **23**, 455–464.

Pfeifer B, Wittelsbürger U, Ramos-Onsins SE, Lercher MJ (2014) PopGenome: an efficient swiss army knife for population genomic analyses in R. *Molecular Biology and Evolution*, **31**, 1929–1936.

Rausch T, Zichner T, Schlattl A *et al.* (2012) DELLY: structural variant discovery by integrated paired-end and split-read analysis. *Bioinformatics (Oxford, England)*, **28**, i333–i339.

Reynolds TB, Fink GR (2001) Bakers' yeast, a model for fungal biofilm formation. *Science (New York, N.Y.)*, **291**, 878–881.

Romano P, Suzzi G (1993) Sulfur dioxide and wine microorganisms. In: *Wine Microbiology and Biotechnology* (ed. Fleet GH), pp. 373–393. Harwood Academic Publishers, London.

Ropars J, Rodríguez De La Vega RC, López-Villavicencio M *et al.* (2015) Adaptive horizontal gene transfers between multiple cheese-associated fungi. *Current Biology*, **25**, 2562–2569.

Ross-Ibarra J, Morrell PL, Gaut BS (2007) Plant domestication, a unique opportunity to identify the genetic basis of adaptation. *Proceedings of the National Academy of Sciences of the United States of America*, **104**(Suppl. 1), 8641–8648.

Ryan O, Shapiro RS, Kurat CF *et al.* (2012) Global gene deletion analysis exploring yeast filamentous growth. *Science (New York, N.Y.)*, **337**, 1353–1356.

Saito H, Posas F (2012) Response to hyperosmotic stress. *Genetics*, **192**, 289–318.

Scannell DR, Zill OA, Rokas A *et al.* (2011) The awesome power of yeast evolutionary genetics: new genome sequences and strain resources for the *Saccharomyces sensu stricto* Genus. *G3 (Bethesda, Md.)*, **1**, 11–25.

Schacherer J, Shapiro JA, Ruderfer DM, Kruglyak L (2009) Comprehensive polymorphism survey elucidates population structure of *Saccharomyces cerevisiae*. *Nature*, **458**, 342–345.

Sopko R, Huang D, Smith JC, Figeys D, Andrews BJ (2007) Activation of the Cdc42p GTPase by cyclin-dependent protein kinases in budding yeast. *The EMBO Journal*, **26**, 4487–4500.

Stamatakis A (2014) RAxML version 8: a tool for phylogenetic analysis and post-analysis of large phylogenies. *Bioinformatics*, **30**, 1312–1313.

Thornton K (2003) libsequence: a C++ class library for evolutionary genetic analysis. *Bioinformatics*, **19**, 2325–2327.

Ueno Y (1999) RGA2, a putative Rho-type GTPase-activating protein, is regulated by the transcription factor Tec1 during filamentous growth of *Saccharomyces cerevisiae*. Massachusetts Institute of Technology.

Vallee BL, Auld DS (1990) Zinc coordination, function, and structure of zinc enzymes and other proteins. *Biochemistry*, **29**, 5647–5659.

Vatsiou AI, Bazin E, Gaggiotti OE (2016) Detection of selective sweeps in structured populations: a comparison of recent methods. *Molecular Ecology*, **25**, 89–103.

Wei W, McCusker JH, Hyman RW *et al.* (2007) Genome sequencing and comparative analysis of Saccharomyces cerevisiae strain YJM789. *Proceedings of the National Academy of Sciences of the United States of America*, **104**, 12825–12830.

Will JL, Kim HS, Clarke J *et al.* (2010) Incipient balancing selection through adaptive loss of aquaporins in natural Saccharomyces cerevisiae populations. *PLoS Genetics*, **6**, e1000893.

Yang Z (2007) PAML 4: phylogenetic analysis by maximum likelihood. *Molecular Biology and Evolution*, **24**, 1586–1591.

Yang Z, Dos Reis M (2011) Statistical properties of the branch-site test of positive selection. *Molecular Biology and Evolution*, **28**, 1217–1228.

Zara S, Bakalinsky AT, Zara G *et al.* (2005) FLO11-based model for air-liquid interfacial biofilm formation by *Saccharomyces cerevisiae*. *Applied and Environmental Microbiology*, **71**, 2934–2939.

M.B., S.D. and J.L.L. conceived and designed the research; A.L.C., F.B. and J.L.L. obtained genomic data, and performed genetic analysis; A.L.C., V.G., J.L.L. So.M, Sa.M. and G.Z. performed experiments; M.B., A.L.C, J.L.L., So.M., P.G., G.Z. and S.D. analysed the data; S.D., J.L.L., P.G. and M.B. drafted the manuscript with advice and consent from all authors.

Data accessibility

All new sequence reads, assemblies and variant files are available in the European Nucleotide Archive (ENA), and accessions are provided in Table S1 (Supporting information). These data sets are deposited under the study accessions PRJEB6529, PRJEB7675 and PRJEB6586. All other data files including variant files in vcf format, SIFT analysis tables, table of all genes retained by each method, data for recombination rates, data for dadi demographic inferences, scripts for ms simulation and XP-CLR calculation are made available from a Dryad Digital Repository (http://datadryad.org/review?doi=doi:10.5061/dryad.j315n).

Supporting information

Additional supporting information may be found in the online version of this article.

Table S1 (A) List of strains used for sequencing. (B) List of derived strains used for the evaluation of allelic variations

Table S2 List of primers used in this study

Table S3 Main putative translocations found in the genome of at least two strains among wine and flor strain from the discordance of read mapping with DELLY

Table S4 Genes retained by at least 2 methods based on

differentiation, and genes under positive selection according to XPCLR

Fig. S1 FINESTRUCTURE co-ancestry matrix and population structure of wine and flor strains using a set of 54 351 SNPs phased with Beagle.

Fig. S2 Genealogical relationships among haplotypes of Champagne strain EC1118 and flor and wine strains.

Fig. S3 Linkage disequilibrium as a function of physical distance.

Fig. S4 Genomic distribution of recombination rates in wine and flor yeasts.

Fig. S5 Joint allele frequency spectrum of the wine population and two geographic populations of flor yeasts.

Fig. S6 Alignment of the nucleotide sequence of the *HXT3* alleles of wine and flor strains in comparison to the *HXT1* allele of wine strain 20B2.

Fig. S7 Genome scan of relative divergence (D_a) between the wine and flor strains ($\times$1000 per kb) window.

Fig. S8 Alignment of the protein sequence of wine and flor *SLN1* showing the multiple amino acid changes.

Fig. S9 Distribution of the contributions of each non synonymous SNPs with potential impact to the 1st axis of PCA differentiating wine and flor strains.