UNIVERSITÉ LAVAL

Étude de la formation et de l'évolution d'espèces hybrides au sein d'un système de levures sauvages

Thèse

Guillaume Charron

Doctorat en biologie
Philosophiæ doctor (Ph. D.)

Québec, Canada

Étude de la formation et de l'évolution d’espèces hybrides au sein d'un système de levures sauvages

Thèse

Guillaume Charron

Sous la direction de :

Christian Landry, directeur de recherche

RÉSUMÉ

L'hybridation a souvent été considérée comme impossible ou encore comme un faux pas de la part des espèces, donnant naissance à de soi-disant culs-de-sac évolutifs. Les observations de lignées hybrides et l'accumulation de données génomique nous ont permis de comprendre que plusieurs organismes participent à des événements d'hybridation. On reconnaît aujourd'hui l'hybridation comme un mécanisme puissant de génération de nouvelles lignées. Cependant, la contribution de l'hybridation au processus de spéciation est une des questions qui reste en suspens. Quelques exemples de spéciation par hybridation ont été décrits chez les plantes et les animaux, mais peu de données à ce sujet ont été récoltées chez les microorganismes sexués. Les exemples chez les microorganismes se limitent à des organismes ayant un lien intime avec les activités humaines (pathogènes ou ferments). Le manque de données sur les populations naturelles de microorganismes pourrait laisser croire que leurs hybrides sont peu compétitifs ou encore infertiles, menant à leur extinction dans l'environnement. Au cours des travaux effectué dans le cadre cette thèse, nous avons utilisé une approche de génomique des populations sur une collection de souches naturelles de la levure *Saccharomyces paradoxus*. La biogéographie de cette espèce suggère que les deux lignées indigènes de l'Amérique du Nord sont en cours de spéciation. Nos analyses ont révélé une lignée auparavant cryptique qui est le résultat d'un évènement de spéciation par hybridation entre ces deux espèces naissantes. À l'aide de ce système d'étude, nous avons exploré en laboratoire deux aspects de l'hybridation. Premièrement, nous avons comparé la croissance d'hybrides à celle de leurs lignées parentales dans plusieurs environnements à la recherche d'une performance diminuée des hybrides qui pourrait expliquer leur rareté dans leur environnement naturel. Cette approche nous a permis de montrer que les hybrides de souches naturelles ont souvent des phénotypes supérieurs à ceux des

parents. En second lieu, nous avons utilisé une méthode d'évolution expérimentale pour suivre la dynamique de la fertilité après l'hybridation. Les résultats obtenus suggèrent qu'après l'hybridation, les hybrides infertiles peuvent redevenir fertiles rapidement à la suite d'évènements spontanés de duplication du génome. Les résultats présentés dans cette thèse contribuent à l'amélioration des connaissances à propos de la contribution de l'hybridation à la formation de nouvelles espèces, particulièrement chez les organismes unicellulaires. De plus, les souches génétiquement modifiées et évoluées disponibles pourront être utilisées dans le cadre de futures recherches à propos d'autres aspects de l'écologie et de l'évolution des hybrides.

SUMMARY

Hybridization was often considered as impossible or as a blunder for species, as it gave birth to so-called evolutionary dead ends. The observations of hybrid lineages and the accumulation of genomic data lead to the realization that hybridization is rather common in multiple organisms. Hybridization is now recognized as a powerful mechanism for the generation of new lineages. One of the questions still pending is about the contribution of hybridization to the speciation process. The few examples of hybrid speciation remain limited to plants and animals. Little data is available for sexual microorganisms which could lead to the belief that their hybrids are poor competitors or suffer from infertility, leading to their extinction in the environment. In the course of this thesis, we used a population genomics approach on a collection of natural isolates of the yeast *Saccharomyces paradoxus*. The biogeography of this species suggests that the two indigenous lineages found in North America are nascent species. Our analyses revealed a precedently cryptic lineage which rose from the hybridization of the two incipient species. Using this study system in the laboratory, we explored two aspects of hybridization. We first compared the growth of hybrids to their parents' in multiple environments in search of decreased hybrid performance which could explain their rarity in the natural environment. This approach allowed us to show that hybrids between natural strains often show superior phenotypes when compared to their parents. We then used experimental evolution to follow the dynamics of fertility following hybridization. Our results suggest that initially infertile hybrids can rapidly become fertile again following spontaneaous genome duplication events. The results presented in this thesis contribute to a better understanding of how hybridization can shape the formation of new species, particularly in microorganisms. Also, the genetically modified and evolved strains available can be used in future studies about the ecology and evolution of hybrids.

TABLE DES MATIÈRES

LISTE DES FIGURES

LISTE DES TABLEAUX

« Hoc certe non minus prodigium est, quam si vitulum capite lupino praeditum, vacca pareret. »

« Ce n'est certainement pas moins remarquable que si une vache avait donné naissance à un veau à tête de loup. »

« This is certainly no less remarkable than if a cow were to give birth to a calf with a wolf's head. »

-Carolus Linnæus à propos de Peloria (1744)

REMERCIEMENTS

Cette thèse signifie l'achèvement d'une des périodes les plus amusantes de ma vie. J'ai appris beaucoup, notamment comment planifier et faire de la science avec rigueur. Au-delà des activités académiques, travailler avec une vingtaine de personnes, venant de divers endroits, m'a beaucoup appris sur moi-même et sur mes aptitudes au travail d'équipe. Ce fut un exercice d'introspection intéressant qui va me permettre de me fixer de nouveaux objectifs de développement personnel et professionnel. Bien sûr, certains ont été plus impliqués dans mon parcours et je voudrais les remercier personnellement.

Christian, je voulais te remercier de m'avoir accueilli dans ton laboratoire en 2011. Je me considère privilégié d'avoir pu être à la fois témoin et acteur de la progression de ton équipe. J'ai vu la taille du laboratoire tripler en 7 ans, mais tu avais toujours autant de temps à accorder à tes étudiants et tes chercheurs postdoctoraux. Je voudrais exprimer ma gratitude envers ta patience (surtout lors des phases de rédaction, un processus très long pour moi) et ton honnêteté à propos de la critique de mon travail. Merci aussi pour toutes les opportunités que tu m'as offertes que ce soit des collaborations avec d'autres groupes ou encore des présentations en congrès. Toutes ces choses témoignent de la confiance que tu avais en moi et je t'en suis reconnaissant.

Je veux aussi remercier Nadia d'avoir fait partie de mon comité d'examen doctoral et d'avoir assisté à presque toutes mes présentations. Merci aussi à Juile Turgeon et Louis Bernier d'avoir gracieusement accepté de faire partie de mon comité d'examen doctoral, de mon comité d'encadrement et du jury d'évaluation de cette thèse. Vos commentaires sur mon travail, toujours très constructifs, m'ont beaucoup aidé dans la prise de décision concernant le déroulement de ma thèse.

Je veux prendre le temps de remercier les gens avec lesquels j'ai entretenu une relation de collaboration au laboratoire. Merci à Jean-

Baptiste, qui a été (et est toujours) un excellent mentor. Je n'aurais pas pu souhaiter trouver un meilleur guide dans le monde de la recherche. Tes conseils autant scientifiques qu'artistiques m'ont toujours éclairé dans mon travail (même si parfois ton conseil était : « j'ai tout changé ton texte, ça ne faisait pas de sens »).

Merci à Mathieu, le seul de mes stagiaires qui a poursuivi ses études au laboratoire (j'ai sûrement fait peur aux autres!). Ton assiduité au travail, ton sérieux (un peu trop présent à mon goût!) et ton calme font de toi un scientifique exceptionnel. Je vais essayer de rester au courant de tes projets et progrès au cours des prochaines années. Merci aussi à Souhir avec qui on s'est entendu sur le fait qu'on ne s'entendrait pas la majorité du temps, mais dont le travail a été essentiel au bon déroulement de la fin de mon doctorat. Je te respecte énormément d'avoir passé outre le fait que je pouvais parfois te rendre la vie dure et d'avoir continué ton travail avec la même attitude professionnelle. Avec toi et Mathieu, nous avons effectué l'expérience avec la plus grande envergure (sans l'assistance de robot) que j'ai vue en 7 ans au laboratoire et je crois qu'on peut être fier de notre travail.

Alex, merci pour toute l'aide que tu as pu m'apporter pendant les 7 dernières années. Tes opinions sur les méthodes et tes questions à propos des résultats ou du contenu de mes manuscrits ont eu beaucoup de valeur à mes yeux. Le temps que tu m'as accordé (des fois sur ton temps personnel) a été un appui important à la réalisation des différentes étapes de mon doctorat et de l'écriture de cette thèse. Plus personnellement, je veux te remercier toi, Kassy, Antoine et Sophie pour avoir joué le rôle de ma « famille à Québec ». Vous m'avez toujours accueilli les bras ouverts (toujours avec de la bière) dans mes bons comme dans mes mauvais moments, toujours prêt à écouter mes doléances. J'espère pouvoir continuer à vous visiter de temps en temps.

Je tiens aussi à remercier les gars avec lesquels j'ai partagé mon bureau au cours de 4 ou 5 dernières années. Chris, thank you for those

moments we shared in the office (most often not talking about science). We started the PhD at the same time, and I was impressed by your progression throughout your studies. You're a great guy and I wish you the best in both your professional and personal life. Philippe, merci d'avoir occupé le poste d'emmerdeur officiel. Je n'ai jamais éprouvé autant d'amitié en envoyant promener quelqu'un et je crois que c'était réciproque. Continue d'approfondir les connaissances à propos de CRISPR et comment il peut foutre le bordel dans un génome (les jumelles vont en avoir besoin). WikiLou, merci pour ton aide pour mes questions (niaiseuses) de bio-informatique. Merci aussi de ton sens de l'humour intelligent qui a égayé certaines de mes journées. Merci d'avoir été aussi assidu dans ta participation aux « Vendredis pub universitaire ». Sans toi, ces soirées-là auraient souvent pris des tournures beaucoup moins intellectuelles. Hélène, je sais que tu n'as pas partagé mon bureau, mais j'aurais aimé que tu puisses le faire. Merci de ton attitude positive (parce que l'on sait que je peux être plutôt négatif). Ton intégration rapide au groupe du laboratoire m'indique que tu feras une excellente Québécoise. J'espère que tu trouveras un endroit où tu pourras exercer tes multiples talents (origamiste, ou contrôle qualité de « ptit sacs » de mayonnaise).

Emilie B. merci de m'avoir encouragé (et enduré) pendant une bonne partie de mon doctorat. Merci à Émilie B. de m'accompagner en ce moment et de faire la même chose, mais mieux. Je sais que tu commences une grosse étape de ta vie professionnelle et que plusieurs choix difficiles se présenteront. J'espère que je pourrai te fournir le même appui pendant ces périodes stressantes.

Finalement, je veux remercier ma famille. Merci à Olivier et Amélie qui continuent de m'apprendre à être la meilleure personne possible. Je suis reconnaissant de votre écoute, même si l'on ne se parle pas si souvent. Vous aussi avez contribué au bon déroulement de ma thèse par vos questions à propos de ce que je faisais, restez curieux! Désolé Oli, je n'ai pas réussi à inventer une levure qui transforme les gens en zombies. À

vous deux, je ne souhaite que du bonheur et j'espère qu'à partir de maintenant on pourra se voir plus fréquemment. Mes derniers remerciements sont pour Daniel et France, merci de ne m'avoir jamais imposé de choix de carrière ce qui a rendu mon parcours un peu chaotique, mais tellement riche en apprentissages. Votre soutien inconditionnel est un élément important à ma réussite. J'espère que je pourrai un jour vous donner autant que ce que j'ai reçu de votre part. Merci d'être fier de moi peu importe les circonstances.

AVANT-PROPOS

Cette thèse est composée de trois chapitres qui correspondent à trois articles auxquels j'ai participé à titre de premier ou co-premier auteur au cours de mes études doctorales.

Le premier chapitre est composé d'un article publié dans la revue *Nature Microbiology* le 11 janvier 2016 où je suis co-premier auteur avec Jean-Baptiste Leducq (post-doctorant) et Lou Nielly-Thibault (étudiant au doctorat). Ce chapitre rapporte la découverte d'une lignée introgressée de levure sauvage *Saccharomyces paradoxus*. Il présente l'étude de génomique des populations qui a permis de découvrir cette lignée et de montrer que cette dernière était issue d'un évènement de spéciation par hybridation homoploïde. Cet article représente un des cas les mieux documentés pour ce genre de spéciation et le premier exemple chez un eucaryote unicellulaire. Ma contribution à ce chapitre a été de participer à la planification des expériences avec mon directeur ainsi que Jean-Baptiste Leducq. Pour la partie expérimentale, les expériences de génomique des populations ont été effectuées par Jean-Baptiste Leducq. J'ai effectué toutes les expériences de croisements pour tester l'isolement reproducteur entre la lignée introgressée et les autres lignées américaines. J'ai fait les expériences de caryotypage des souches ainsi que la confirmation de la présence de l'inversion dans les différentes souches. J'ai aussi participé à la supervision de Chris Eberlein pendant l'expérience de criblage phénotypique. J'ai fait une partie de l'analyse pour les expériences où j'étais directement impliqué dans les manipulations avec l'aide de mon directeur et de Jean-Baptiste Leducq. Les analyses bio-informatiques ont été effectuées par Jean-Baptiste Leducq et Lou Nielly-Thibault. J'ai participé à la rédaction initiale du manuscrit pour les parties du texte impliquant mes contributions. La rédaction finale a été faite par Jean-Baptiste Leducq et mon directeur Christian Landry. Tous les auteurs ont

contribué aux révisions et à l'édition finale du manuscrit. Les autres auteurs ont contribué en apportant de l'aide technique, des souches ou encore des commentaires à divers stades d'écriture du manuscrit.

Le deuxième chapitre est composé d'un article publié dans la revue *Biology Letters* le 1er juin 2017 où je suis premier auteur. Ce chapitre porte sur l'étude de la comparaison phénotypique entre des hybrides entre les groupes *SpB* et *SpC* de *Saccharomyces paradoxus*. Il rapporte les résultats d'un criblage phénotypique à grande échelle sur une petite collection d'hybrides et leurs souches parentales afin de vérifier s'il était possible qu'un isolement reproducteur extrinsèque isole les levures en nature. Les résultats indiquent la présence d'hétérosis chez les hybrides et excluent l'hypothèse de l'isolement extrinsèque. J'ai participé, sous la supervision de mon directeur, à toutes les étapes de planification et de réalisation des expériences. J'ai fait l'analyse des données et la préparation des figures avec les conseils de mon directeur. J'ai écrit les versions initiale et finale avec l'aide et les conseils de mon directeur. Les membres de mon laboratoire ont contribué à la révision de mon manuscrit.

Finalement, le troisième chapitre est un manuscrit qui a été soumis à la revue *Nature Communication* le 20 juin 2019 où je suis co-premier auteur avec Souhir Marsit (post-doctorante). Ce chapitre rapporte l'étude de l'évolution de la fertilité de plus de 600 lignées expérimentales de levures hybrides. Cette expérience a été réalisée en soumettant les lignées à de forts goulots d'étranglement pour limiter l'efficacité de la sélection pour plus de 700 générations. Les résultats obtenus suggèrent la présence d'une instabilité génomique qui pourrait contribuer à la fois à l'isolement reproducteur en diminuant la fertilité des hybrides et à la duplication complète du génome qui, elle, restaure complètement la fertilité. J'ai participé avec mes collègues Souhir Marsit, Mathieu Hénault et mon directeur à toutes les étapes de planification et de réalisation de

l'expérience d'évolution. J'ai récolté et analysé les données de fertilité des hybrides. J'ai aussi planifié et réalisé les expériences de croisements intratétrades et d'autodiploïdisation. Les expériences de ploïdie et de génotypage par séquençage (GBS) ont été planifiées et réalisées par Souhir. Les expériences de PCR des ADN mitochondriaux ont été planifiées et réalisées par Mathieu. J'ai rédigé le manuscrit initial avec Souhir sous la supervision de mon directeur Christian Landry et les révisions ultérieures ont été faites par tous les auteurs.

Au cours de mon parcours au doctorat, j'ai également eu la chance d'effectuer des expériences ou encore de fournir des souches de notre collection pour différents chercheurs de mon laboratoire ou de l'étranger. Ces collaborations m'ont permis d'être inscrit à titre d'auteur pour les articles suivants:

Freel, K.C., Charron, G., Leducq, J.-B., Landry, C.R. & Schacherer, J., 2015, Lachancea quebecensis sp. nov., a yeast species consistently isolated from tree bark in the Canadian province of Québec. *International journal of systematic and evolutionary microbiology* **65**, 3392-3399.
Contribution: Échantillonnage et envoi de souches de *Lachancea* dont la nouvelle espèce décrite dans cet article.

Peris, D., Langdon, Q.K., Moriarty, R.V., Sylvester, K., Bontrager, M., Charron, G., Leducq, J.-B., Landry, C.R., Libkind, D. & Hittinger, C.T., 2016, Complex ancestries of lager-brewing hybrids were shaped by standing variation in the wild yeast Saccharomyces eubayanus. *PLoS genetics* **12**, e1006155.
Contribution: Isolement, identification et envoi de souches de *Saccharomyces eubayanus* utilisées dans les analyses.

Eberlein, C., Nielly-Thibault, L., Maaroufi, H., Dube, A.K., Leducq, J.B., Charron, G. & Landry, C.R., 2017, The Rapid Evolution of an Ohnolog Contributes to the Ecological Specialization of Incipient Yeast Species. *Mol Biol Evol* **34**, 2173-2186. (doi:10.1093/molbev/msx153).
Contribution: Construction de souches et planification des experiences et révivion du manuscrit initial.

Filteau, M., Charron, G. & Landry, C.R., 2017, Identification of the fitness determinants of budding yeast on a natural substrate. *ISME J* **11**, 959-971. (doi:10.1038/ismej.2016.170).
Contribution: Construction des souches de délétion des gènes de la voie métabolique de l'allantoine et rédaction de la partie méthodologie concernant ces souches.

Leducq, J.B., Hénault, M., Charron, G., Nielly-Thibault, L., Terrat, Y., Fiumera, H.L., Shapiro, B.J. & Landry, C.R., 2017, Mitochondrial Recombination and Introgression during Speciation by Hybridization. *Mol Biol Evol* **34**, 1947-1959. (doi:10.1093/molbev/msx139).
Contribution: Supervision de Mathieu Hénault pendant son stage dont les travaux sont présentés dans cet article.

Peris, D., Moriarty, R.V., Alexander, W.G., Baker, E., Sylvester, K., Sardi, M., Langdon, Q.K., Libkind, D., Wang, Q.M., Bai, F.Y., et al., 2017, Hybridization and adaptive evolution of diverse Saccharomyces species for cellulosic biofuel production. *Biotechnol Biofuels* **10**, 78. (doi:10.1186/s13068-017-0763-7).
Contribution: Envoi de souches de *Saccharomyces*

Xia, W., Nielly-Thibault, L., Charron, G., Landry, C.R., Kasimer, D., Anderson, J.B. & Kohn, L.M., 2017, Population genomics reveals structure at the individual, host-tree scale and persistence of genotypic variants of the undomesticated yeast Saccharomyces paradoxus in a natural woodland. *Molecular Ecology* **26**, 995-1007. (doi:10.1111/mec.13954).
Contribution: Cribblage phénotypique des souches *SpD* qui n'a pas été utilisé dans les analyses

Wolters, J.F., Charron, G., Gaspary, A., Landry, C.R., Fiumera, A.C. & Fiumera, H.L., 2018, Mitochondrial Recombination Reveals Mito-Mito Epistasis in Yeast. *Genetics*, genetics. 300660.302017.
Contribution: Planification, execution et aide dans l'analyse d'un cribblage phénotypique sur la collection de souches avec différentes combinaisons de mitotypes.

Hénault, M., Eberlein, C., Charron, G., Durand, É., Nielly-Thibault, L., Martin, H. & Landry, C.R., 2019, Yeast Population Genomics Goes Wild: The Case of Saccharomyces paradoxus. In *Population Genomics: Microorganisms* (eds. M.F. Polz & O.P. Rajora), pp. 207-230. Cham, Springer International Publishing.
Contribution: Premier jet de la partie sur l'isolement reproducteur chez *S. paradoxus* et révision du manuscrit une fois terminé.

Eberlein, C., Hénault, M., Fijarczyk, A., Charron, G., Bouvier, M., Kohn, L.M., Anderson, J.B. & Landry, C.R., 2019, Hybridization is a recurrent evolutionary stimulus in wild yeast speciation. *Nature Communications*. **10** (923)
Contribution: Planification, exécution et analyse des expériences de croisements de souches pour mesurer l'isolement reproducteur entre *SpD* et les autres groupes de *S. paradoxus*

INTRODUCTION GÉNÉRALE

1-Mise en contexte

Le concept d'espèce biologique décrit les espèces comme des groupes de populations ou d'individus interfertiles, isolés reproductivement d'autres groupes (Mayr 1982). Cette définition implique donc le développement de barrières reproductives entre les espèces au cours du processus de spéciation. Cette dernière est un processus clé dans l'étude de la biologie évolutive. Elle décrit la génération de nouvelles lignées suivant l'action de la sélection naturelle ou de la dérive génétique sur la variation présente à l'intérieur des populations. Le processus commence souvent par la séparation physique ou écologique d'une espèce ancestrale en deux populations distinctes. Après cette séparation, les deux nouvelles populations sont en mesure d'accumuler indépendamment de la variation génétique, entraînant une différenciation. Cette variation est nécessaire, mais n'est cependant pas suffisante à la spéciation. Si les barrières séparant les populations venaient à disparaître, le flux de gènes en résultant pourrait refondre les populations différenciées en une seule population non différenciée. L'isolement reproducteur contribue donc au maintien de l'isolement génétique des lignées évolutives lors d'un contact secondaire. Cette introduction fera d'abord un survol des différents types d'isolements reproducteurs ainsi que leurs bases génétiques connues.

L'accumulation de barrière à la reproduction peut prendre du temps et l'intensité de ces barrières peut être variable. Il n'y a pas de délimitation précise entre espèces indépendantes et populations interfertiles, la spéciation se fait donc le long d'un continuum de divergence génétique. Ceci a pour conséquences que la reproduction entre des membres de populations différenciées, l'hybridation, est possible le long du continuum que représente la spéciation. On rapporte qu'environ 25 % des espèces végétales et 10 % des espèces animales participent à des évènements de reproduction avec des espèces apparentées (Mallet 2007). Les évènements

d'hybridation peuvent avoir des impacts autant positifs que négatifs sur le processus de spéciation. C'est pourquoi étudier la biologie et l'évolution des organismes hybrides peut nous aider à mieux comprendre l'importance de l'hybridation pour la spéciation. Nous aborderons aussi les divers impacts que peut avoir l'hybridation sur la spéciation en portant une attention particulière sur la génération rapide de nouvelles lignées.
L'étude de la biologie chez les espèces hybrides s'est initialement concentrée sur les macroorganismes puisqu'il était aisé de les observer et d'étudier leurs phénotypes. Avec le développement de la microbiologie, des techniques de biologie moléculaire et de génie génétique, l'étude de la spéciation et de l'hybridation a pu être appliquée aux microorganismes. Ces organismes forment un groupe avec des particularités qui en font d'excellents modèles pour la biologie évolutive. Ces particularités ainsi que les connaissances actuelles sur l'hybridation chez les microorganismes seront discutées.

Mes travaux de doctorat se sont basés sur l'étude de la biologie des hybrides avec un modèle eucaryote unicellulaire, la levure sauvage *Saccharomyces paradoxus*. Cet organisme est un modèle émergent dans l'étude de l'écologie et de l'évolution chez les microorganismes. La description du système d'étude utilisé dans les chapitres de cette thèse, les populations nord-américaines de *S. paradoxus*, constituera la dernière partie de l'introduction.

2-Spéciation allopatrique, sympatrique et parapatrique

La spéciation peut se faire principalement selon les modèles allopatrique, sympatrique et parapatrique (Coyne et Orr 2004). Dans le modèle allopatrique, une population ancestrale est séparée par des barrières géographiques au flux de gènes. Un exemple classique de spéciation allopatrique est celui des évènements de glaciation. Pendant les glaciations, les migrations forcées par les changements climatiques ont pu séparer des populations entre des refuges glaciaires différents (Willis *et*

coll. 2004). Autant l'environnement des refuges que les changements démographiques des populations ont pu influencer la divergence de ces populations par la sélection naturelle et la dérive génétique (Hewitt 2008). Avec l'accumulation de divergence, des barrières intrinsèques au flux de gènes vont émerger, processus qui peut être accéléré par une sélection divergente (Schluter 2001). Le modèle allopatrique est le modèle considéré le plus commun (Coyne et Orr 2004). Après le retrait des glaces, les nouvelles lignées ont pu recoloniser les environnements de nouveau accessibles. Pendant cette recolonisation, des contacts secondaires ont eu lieu, agissant comme des tests pour la spéciation.

L'isolement géographique n'est pas une condition *sine qua non* à la spéciation. D'autres modèles comme la spéciation parapatrique, péripatrique et sympatrique impliquent la coexistence spatiale et temporelle entre les populations et une présence de flux de gènes entre les populations. Le modèle de la spéciation parapatrique se base sur l'observation que les aires de répartition des espèces sont souvent plus grandes que la capacité de dispersion des individus (Gavrilets *et coll.* 2000). L'ordre dans lequel les mutations apparaissent est aléatoire (Mani et Clarke 1990) et les régimes de sélection peuvent varier à l'intérieur d'une même région géographique, deux populations contigües peuvent fixer des allèles différents et accumuler de la divergence. Les espèces en anneaux comme les salamandres du genre *Ensatina* sur la côte ouest de l'Amérique du Nord sont un des exemples de patron que pourraient laisser les évènements de spéciation parapatrique (Wake 1997).

Le modèle de la spéciation sympatrique est différent des deux autres puisqu'il n'implique aucune contrainte géographique au flux de gène. Ce modèle a été longtemps controversé puisque les incompatibilités génétiques doivent émerger malgré le pouvoir homogénéisant de la recombinaison (Futuyma et Mayer 1980, Felsenstein 1981). Cependant, Rice et Salt (1990) ont démontré expérimentalement qu'il était possible de développer de l'isolement reproducteur en sympatrie si une sélection

divergente pour la préférence d'habitat fait que les individus se reproduisent dans leur habitat de préférence. Feder *et coll.* (1988) ont montré que les populations naturelles de la mouche *Rhagoletis pomonella* avec des préférences pour des hôtes différents étaient différenciées génétiquement malgré leur sympatrie, renforçant ainsi l'idée que l'initiation de la spéciation en sympatrie était possible en nature.

3-Mécanismes d'isolement reproducteur

Plusieurs barrières sont connues pour interrompre le flux génique entre les populations (Dobzhansky et Tan 1937, Dobzhansky 1951). Ernst Mayr les a catégorisées en deux principaux groupes : les mécanismes agissant avant la fécondation et la formation d'un zygote (prézygotique) et ceux agissant après (postzygotique) (Mayr 1963). Ces mécanismes ne sont pas exclusifs et les espèces peuvent être isolées par plus d'une forme d'isolement reproducteur. Les conséquences de l'isolement reproducteur complet sont que les hybrides ne se forment pas (absence de reproduction interspécifique), ne se reproduisent pas ou le font à des fréquences beaucoup trop faibles pour assurer une descendance (hybrides stériles ou infertiles) (Coyne et Orr 2004). Les populations ainsi isolées poursuivent donc leur trajectoire évolutive indépendamment les unes des autres, menant à l'émergence de nouvelles espèces.

3.1-Isolement reproducteur prézygotique

L'isolement prézygotique peut être divisé en catégories dépendant de la source de l'isolement reproducteur. Les populations peuvent être isolées écologiquement, temporellement, sexuellement ou encore mécaniquement. Ces mécanismes d'isolement reproducteurs sont cruciaux pour amorcer la spéciation en sympatrie puisqu'elles empêchent les espèces de s'hybrider même en absence de barrières postzygotiques, ce qui permet aux populations d'accumuler de la divergence et de développer cet isolement reproducteur (Coyne et Orr 2004). Ils peuvent aussi émerger après

l'isolement postzygotique si ce dernier est partiel et qu'il y a une forte contre-sélection des hybrides. Dans ces cas, on parlera de renforcement de l'isolement reproducteur en sympatrie (Butlin 1987).

3.1.1-Isolement écologique

Les espèces adaptées à différents environnements peuvent développer des préférences quant à la reproduction dans l'habitat auquel elles sont adaptées ou encore de la sélection contre les migrants. Dans ces cas, l'isolement peut être la conséquence indirecte de l'adaptation à l'environnement ou encore être favorisé par la sélection si les traits sélectionnés chez les populations parentales présentent des compromis de fitness entre les deux environnements (Rundle et Nosil 2005). Par exemple, deux populations sympatriques de pucerons *Acyrthosiphon pisum pisum* soupçonnées d'être en cours de spéciation sont spécialisées l'une pour la luzerne et l'autre pour le trèfle rouge. Elles présentent des performances inférieures sur l'hôte pour lequel elles ne sont pas spécialisées (Via 1991). L'isolement reproducteur vient du fait que ces populations se reproduisent exclusivement sur leurs hôtes respectifs (Via 1999) en raison d'une association génétique entre les traits de performances et ceux du choix de la plante hôte (Hawthorne et Via 2001).

3.1.2-Isolement temporel

Une information importante contenue dans le génome est celle qui coordonne temporellement le développement et la maturation d'un organisme (Alberts *et coll.* 2014). Des changements dans la régulation des gènes responsables de la coordination du cycle de vie peuvent donc mener à une variation dans les moments de reproduction (Marshall et Cooley 2000, Villanueva 2016). La synchronisation de comportements comme la migration avec des signaux environnementaux peut aussi avoir des impacts sur la reproduction (Quinn *et coll.* 2007). L'isolement reproducteur causé par ces différences est nommé isolement temporel. Ces différences

d'horaire sont partiellement héritables (Van der Jeugd et McCleery 2002) et peuvent donc être sous sélection entre les populations. Par exemple, deux populations de corail *Mycedium elephantotus* sont isolées temporellement puisque la période de gamétogénèse de l'une est normale pour cette espèce (Dai et Fan 1992) alors que celle de l'autre est décalée d'environ 4 mois (Fan et Dai 1998). Comme les gamètes restent viables seulement quelques heures, cette différence fait que même si elles sont sympatriques les deux populations sont isolées et ont accumulé de la divergence génétique (Dai *et coll.* 2000).

3.1.3-Isolement sexuel

L'isolement sexuel, aussi nommé isolement comportemental, est caractérisé par l'attirance diminuée envers les individus d'une population différente ou encore la non-reconnaissance de membres d'une population différente comme un potentiel partenaire de reproduction (Nosil 2012). Ce choix est souvent basé sur la préférence de phénotypes aux bases génétiques complexes (Chenoweth et Blows 2006). L'émergence des traits préférés par les femelles ou de la préférence des femelles pour certains traits peut être favorisée par la sélection sexuelle, mais pourrait aussi se faire par sélection naturelle et dérive génétique (West-Eberhard 1983, Kirkpatrick et Ryan 1991, Coyne et Orr 2004). Chez les coccinelles *Adalia bipunctata* par exemple, les mâles mélanisés sont préférés aux mâles non mélanisés par les femelles (Majerus *et coll.* 1982). Cette préférence peut être manipulée par des expériences de sélection pour l'augmenter encore la renverser, ce qui signifie une base génétique sous-jacente à cette préférence.

3.1.4-Isolement mécanique

L'évolution de la structure des organes sexuels peut mener à de l'isolement prézygotique. Par exemple chez les plantes, les fleurs différentes peuvent être spécialisées pour des pollinisateurs différents

(Grant 1994). Chez les animaux, l'isolement mécanique est mieux connu chez les insectes, pour qui les organes génitaux sont considérés comme des systèmes « cadenas et clé » (Dufour 1844, Masly 2012). Les différences peuvent directement empêcher la copulation comme chez les demoiselles où la reproduction se fait tout en maintenant deux points de contact entre le mâle et la femelle. Au cours des évènements de reproductions interspécifiques, les individus ont souvent des difficultés à établir ou maintenir le contact (Paulson 1974). D'autres mécanismes comme le manque de stimulation ou encore la reconnaissance de partenaire interspécifique via les organes génitaux pourraient aussi contribuer à cette forme d'isolement (Masly 2012).

3.2-Isolement reproducteur postzygotique

Si les barrières prézygotiques sont absentes ou partielles entre les populations, les barrières postzygotiques peuvent tout de même empêcher les espèces d'échanger des gènes. Ces barrières peuvent agir de manière extrinsèque où l'hybride présente des phénotypes nuisant à sa survie ou ses chances de reproduction dans les environnements parentaux. On parlera de barrières intrinsèques si des incompatibilités génétiques font que les hybrides présentent une inviabilité ou une stérilité indépendante de l'environnement. L'isolement reproducteur postzygotique intrinsèque est une composante importante de la spéciation puisque c'est lorsqu'il est assez fort que l'on considère le processus irréversible (Seehausen *et coll.* 2014).

3.2.1-Isolement postzygotique extrinsèque

L'isolement reproducteur postzygotique extrinsèque est dépendant de l'environnement. Les hybrides sont contre sélectionnés dans les environnements parentaux puisqu'ils possèdent des phénotypes intermédiaires aux phénotypes parentaux ou encore inexistants chez les espèces parentales (Rice et Hostert 1993, Rundle et Nosil 2005). Cet

isolement serait une conséquence importante de la sélection naturelle divergente (Rundle et Nosil 2005, Nosil 2012). Les formes benthiques et limnétiques des épinoches à trois épines (*Gasterosteus aculeatus*) sont isolées par ce genre de mécanisme. Ces deux formes présentent des morphologies qui suggèrent des adaptations à leurs environnements respectifs soit les zones littorales des lacs pour les benthiques et les zones d'eau de surface pour les limnétiques (McPhail 1984). Bien que les deux paires soient séparées par un isolement prézygotique fort (Nagel et Schluter 1998), on retrouve quand même de rares hybrides qui possèdent une morphologie intermédiaire (McPhail 1992). Des expérimentations en laboratoire ont montré que les hybrides avaient un taux de croissance beaucoup plus faible que les parents dans leurs environnements respectifs (Hatfield et Schluter 1999). Les rétrocroisements entre ces F1 et les espèces parentales ont un taux de croissance beaucoup plus élevé dans l'environnement du parent avec lequel ils partagent la majorité de leur génome, suggérant que cette dépendance environnementale du fitness a une base génétique (Rundle 2002).

3.2.2-Isolement reproducteur postzygotique intrinsèque

Les combinaisons d'allèles au sein de populations évoluant indépendamment sont soumises à une coévolution, ce qui limite l'émergence d'incompatibilités. Cependant, les deux jeux d'allèles d'espèces indépendantes n'ont jamais été testés par la sélection dans le même organisme, ce qui peut mener à des incompatibilités génétiques. D'un côté, ces dernières peuvent perturber le développement des hybrides et faire en sorte qu'ils soient non viables. D'un autre côté, les hybrides peuvent être viables. Les incompatibilités interfèrent souvent avec le bon fonctionnement de la gamétogénèse, ce qui rend les hybrides stériles ou infertiles. À la source de l'isolement intrinsèque se trouvent quatre mécanismes génétiques: Les incompatibilités de Bateson-Dobzhansky-

Muller (BDMi), les interactions cytonucléaires, les changements de ploïdie et les réarrangements de chromosomes.

3.2.2.1-Incompatibilités de Bateson-Dobzhansky-Muller

Lorsque deux populations accumulent des mutations indépendamment, l'action de la sélection naturelle ou de la dérive génétique peut mener à la fixation d'allèles différents dans chacun des génomes. Par exemple, une population ancestrale au génotype aabb peut se séparer en deux sous populations. L'une de celles-ci peut alors fixer des mutations et devenir à AAbb alors que l'autre peut fixer le génotype aaBB. La combinaison AB survenant dans un génome hybride (AaBb) n'a donc jamais été exposée à la sélection naturelle (Bateson 1909, Dobzhansky 1936). Ces deux allèles ayant évolué dans des contextes génomiques différents, leur interaction pourrait mener au dysfonctionnement des processus cellulaires les impliquant (Muller 1942). Ces interactions épistatiques entre deux loci et plus sont nommées incompatibilités de Bateson-Dobzhansky-Muller (BDMi). Dans ce modèle, comme les incompatibilités se font entre (au moins) deux gènes, la théorie prédit que leur nombre s'accroîtra de manière plus rapide que linéaire avec la distance génétique (Orr et Turelli 2001). Chaque allèle dérivé peut potentiellement provoquer des incompatibilités à la fois avec d'autres allèles dérivés ou encore des allèles ancestraux (dérivés antérieurement dans leur contexte génétique) (Orr 1995). Bien que des résultats expérimentaux appuient cette théorie (Matute *et coll.* 2010, Moyle et Nakazato 2010), certains suggèrent d'être prudent face à des conclusions hâtives à propos de cette accumulation non linéaire, puisque les modèles disponibles ne considèrent pas le polymorphisme ancestral (Cutter 2012).

3.2.2.2-Incompatibilités cytonucléaires

Les chloroplastes et les mitochondries participent à des activités métaboliques essentielles qui demandent une coordination des génomes

nucléaires et cytoplasmiques lors de la réplication, de la transcription et de la traduction (Ryan et Hoogenraad 2007). Comme les organelles sont vulnérables à l'accumulation de mutations délétère en raison du faible taux de recombinaison, le maintien de ces fonctions est assuré par une étroite coévolution entre les deux génomes (Rand *et coll.* 2004). Pendant cette coévolution, des gènes peuvent être mutés, perdus ou transférés de l'organelle au génome nucléaire (Adams *et coll.* 2002, Khachane *et coll.* 2007). Dans ces conditions, deux populations évoluant indépendamment peuvent accumuler des incompatibilités entre les génomes nucléaires et cytoplasmiques, ce qui constitue un cas spécial des BDMi. Contrairement au modèle original, les incompatibilités cytonucléaires ou impliquant la transmission uniparentale d'éléments génétiques ont le potentiel de produire des patrons asymétriques de fertilité ou de viabilité chez les hybrides réciproques (Turelli et Moyle 2007, Martin *et coll.* 2017). En plus des organelles, les endosymbiotes comme les bactéries du genre *Wolbachia* sont aussi reconnus pour être à la source d'incompatibilités cytonucléaires chez les insectes (Bordenstein *et coll.* 2001).

3.2.2.3-Réarrangements de chromosomes

Les réarrangements de chromosomes, comme les translocations, les fusions chromosomiques et les inversions, contribuent aux barrières génétiques entre les espèces de deux manières différentes. D'une part, les translocations et les fusions de chromosomes peuvent empêcher le bon fonctionnement de la méiose chez les hybrides. Les fusions et les fissions de chromosomes ont le potentiel de changer drastiquement les caryotypes des espèces. Le cheval (32) et l'âne (31) ont presque le même nombre haploïde de chromosomes, mais leurs caryotypes ont subi de multiples réarrangements structurels (Yang *et coll.* 2004). Ces réarrangements font en sorte que le bon appariement des chromosomes est impossible lors de la méiose, ce qui entraîne des interruptions prématurées de la méiose ou la formation de gamètes aneuploïdes (Chandley *et coll.* 1974). D'autre part,

les évènements d'inversion peuvent supprimer la recombinaison dans les régions réarrangées. Cette suppression contribue à l'accumulation de divergence génétique dans ces régions et contribue à l'accumulation d'autres types d'incompatibilités entre les populations même en présence de flux de gènes (Rieseberg 2001). Les régions réarrangées pourraient aussi protéger les incompatibilités qu'elles contiennent et prévenir leur élimination par la sélection naturelle. Ces incompatibilités persistantes pourraient alors laisser assez de temps pour que l'isolement prézygotique soit favorisé par la sélection (Noor *et coll.* 2001).

3.2.2.4-Changements de ploïdie

La ploïdie peut être modifiée chez des individus à la suite de la non-disjonction des chromosomes. Pendant la méiose I, il y a une séparation des chromosomes homologues suivie à la méiose II par la séparation des chromatides sœurs. La non-disjonction peut se faire au cours de ces deux évènements, générant des combinaisons de gamètes diploïdes, haploïdes (normaux) et sans ADN. Plusieurs mécanismes cellulaires sont impliqués dans ces erreurs de ségrégation des chromosomes (De Storme et Mason 2014). Les produits de méiose diploïdes constituent une barrière instantanée au flux de gène puisque la fusion entre des gamètes au contenu dupliqué et des gamètes ancestraux haploïdes mène à la production de zygotes triploïdes généralement infertiles en raison de la production de gamètes aneuploïdes. Cependant, certains gamètes des triploïdes peuvent contenir un nombre normal, doublé ou triplé de chromosomes (Otto et Whitton 2000). Les gamètes avec un nombre triplé de chromosomes procurent une opportunité de générer des tétraploïdes en fusionnant avec des gamètes haploïdes. Les triploïdes peuvent donc constituer un intermédiaire pour la génération de lignées tétraploïdes (Harlan et deWet 1975, Ramsey et Schemske 1998). La polyploïdie est considérée comme un mécanisme majeur de la spéciation chez les plantes où jusqu'à 70 % des angiospermes seraient polyploïdes (Masterson 1994).

4-Hybridation

Un hybride est défini comme étant un organisme résultant de la fertilisation croisée entre deux espèces différentes. Cette définition est un peu contraignante puisqu'elle exige la complétion du processus de spéciation. Au cours de cette thèse, on préférera utiliser la définition plus large d'Harrison (1990) qui décrit les hybrides comme la progéniture d'individus de populations distinguables par au moins un caractère héritable. Comme nous l'avons vu précédemment, l'hybridation constitue un test de l'isolement reproducteur pendant le processus de spéciation. Elle peut avoir trois conséquences selon la force de l'isolement reproducteur entre les espèces impliquées (Abbott *et coll.* 2013). Premièrement, si l'isolement reproducteur est faible, le pouvoir homogénéisant de la recombinaison pourrait mener à l'effondrement des barrières reproductrices. Ce pourrait être le cas particulièrement pour la spéciation écologique, où les barrières à la reproduction émergent par adaptation à des niches écologiques différentes (Seehausen 2006). Dans les cas où l'environnement est perturbé et en absence d'isolement reproducteur intrinsèque, l'hybridation peut se produire. On peut alors assister à la fusion des deux espèces en une seule, ce qu'on appelle « spéciation inverse », causant une perte nette de biodiversité (Seehausen 2006, Taylor *et coll.* 2006). Cet effondrement des espèces a déjà été observé chez les paires d'espèces de cichlidés du lac Victoria (Seehausen *et coll.* 1997).

Deuxièmement, si les hybrides sont viables et font face à une contre sélection en raison d'un isolement reproducteur fort (Butlin 1987), il pourrait y avoir renforcement de l'isolement reproducteur total par le développement d'isolement prézygotique (Dobzhansky 1940). Cette augmentation de l'isolement reproducteur contribue à augmenter le fitness des espèces en prévenant de futurs évènements d'hybridation. En comparant des paires sympatriques et allopatriques d'espèces, il est

possible de mettre en évidence la présence de renforcement (Servedio et Noor 2003, Hopkins 2013).
Finalement, si les hybrides sont viables et fertiles tout en étant isolés des espèces parentales, on parlera alors de spéciation par hybridation.

4.1-Hétérosis

Pour qu'une population d'hybride puisse s'établir en tant qu'espèce, elle doit d'abord survivre à la compétition avec ses espèces parentales. La combinaison de deux génomes divergents peut faire émerger des incompatibilités (intrinsèques et extrinsèques) chez les hybrides. Cependant, les évènements d'hybridation peuvent être de puissants générateurs de variation génétique (Buerkle *et coll.* 2000) parce qu'ils permettent la recombinaison entre des allèles qui n'auraient eu que très peu de chance d'émerger au sein d'un même génome par une succession de mutations *de novo*. Certaines combinaisons de gènes confèrent alors des avantages de fitness par rapport aux espèces parentales. La vigueur hybride, aussi appelée hétérosis (Shull 1914), décrit une amélioration des traits chez l'hybride par rapport aux parents (Shull 1908). Un exemple classique de vigueur hybride est celui de la mule, plus forte physiquement qu'un cheval de taille équivalente et plus grande qu'un âne. Cet exemple n'est toutefois pas un cas de spéciation hybride puisque les caryotypes différents entre le cheval et l'âne sont une source d'isolement reproducteur (Trujillo *et coll.* 1962).
Deux principales hypothèses ont été proposées pour expliquer les bases génétiques de l'hétérosis (Crow 1948, Charlesworth et Willis 2009). L'hypothèse de la dominance repose sur le fait que les lignées consanguines fixeraient des allèles récessifs inférieurs ou légèrement délétères (Bruce 1910, Keeble et Pellew 1910). Lors de l'hybridation entre deux lignées différentes, l'hétérozygotie viendrait alors masquer l'effet de ces allèles, se soldant en un phénotype plus vigoureux (Davenport 1908). Selon l'hypothèse de la surdominance, l'hétérosis serait due à la

combinaison des allèles à un locus précis conférant un avantage de fitness à l'hétérozygote (East 1908, Shull 1908). Les observations faites sur l'alcool déshydogénase par Schwartz et Laughner (1969) suggèrent un mécanisme moléculaire pour la surdominance. Cette enzyme est un homodimère et plusieurs allèles de cette enzyme ségrègent dans les populations de maïs. Puisque c'est une enzyme dimérique, cette enzyme peut être formée avec une sous-unité provenant de chaque allèle dans les hétérozygotes. En couplant un allèle stable à l'activité faible avec un allèle moins stable avec une bonne activité, on génère des enzymes à la fois stables et actives. Une troisième hypothèse récente, celle de l'épistasie, attribue l'hétérosis à l'interaction des allèles divergents entre les loci. L'épistasie positive conférerait alors des phénotypes plus vigoureux que ce à quoi on s'attendrait de l'addition des effets des deux allèles. Des études récentes suggèrent que le rôle de l'épistasie pourrait être important dans l'augmentation des performances des hybrides (Kaeppler 2012). Ces trois hypothèses ne sont pas mutuellement exclusives et pourraient avoir chacune leur rôle dans l'hétérosis.

4.2-Spéciation par hybridation

Bien qu'il soit contre-intuitif par rapport à la définition d'espèce biologique, le processus de spéciation par hybridation participe aussi à l'émergence de nouvelles espèces. Deux modes de spéciation par hybridation sont possibles, un impliquant l'allopolyploïdie et l'autre l'homoploïdie.

4.2.1 Spéciation allopolyploïde

La spéciation allopolyploïde est caractérisée par un changement de ploïdie suivant l'évènement initial d'hybridation. Comme nous l'avons vu précédemment, ce changement de ploïdie constitue un isolement reproducteur entre les populations hybrides et parentales. Il contribue aussi à restaurer la fertilité des hybrides entre les espèces aux caryotypes

très divergents. Comme nous l'avons vu précédemment, chez les hybrides la mauvaise ségrégation des chromosomes est due aux difficultés d'associer des chromosomes ayant plus d'un homologue (dans les cas de fission et de fusion chromosomiques). La duplication du génome assure que chaque chromosome aura un homologue parfaitement identique avec lequel la recombinaison pourra s'effectuer. Ceci permet alors une ségrégation normale des bivalents. La polyploïdie est très présente chez les espèces végétales, où l'hybridation semble être un mécanisme majeur de diversification (Rieseberg 1997) alors que chez les animaux elle peut sembler anecdotique. Les raisons pouvant expliquer la rareté de la polyploïdie chez les animaux sont encore le sujet de débats, mais un point important semble être la séparation des sexes en individus mâles et femelles (Muller 1925, Otto et Whitton 2000, Wertheim *et coll.* 2013). L'asexualité, l'autofertilisation, et la pérennité des espèces végétales pourraient contribuer à une fréquence plus élevée de lignées polyploïdes chez ces organismes (Stebbins 1971, Bell 1982).

4.2.2-Spéciation homoploïde

La spéciation par hybridation homoploïde se fait sans qu'il y ait de changement de ploïdie chez l'hybride par rapport aux espèces parentales. Ce mode de spéciation a longtemps été considéré comme peu probable puisque les hybrides souffrent directement des effets des incompatibilités génétiques et ne sont pas isolés reproductivement des espèces parentales (Barton 2001, Coyne et Orr 2004). Un modèle général de la spéciation homoploïde a été proposé par Templeton (1981) et contient quatre étapes critiques pour la formation des espèces hybrides : 1) L'hybridation est suivie par une phase de consanguinité des hybrides où les incompatibilités diminuent beaucoup leur fitness. 2) Les ségrégants hybrides avec le meilleur fitness sont retenus par la sélection naturelle. 3) Pour que son génotype se stabilise, un hybride doit être isolé reproductivement des lignées parentales. 4) Finalement, l'hybride doit être écologiquement

divergent des parents pour pouvoir coexister avec ses espèces parentales ou coloniser une nouvelle niche écologique. Trois critères ont été proposés pour définir les évènements de spéciation par hybridation homoploïde (Schumer *et coll.* 2014). Les hybrides doivent 1) présenter des traces génomiques d'hybridation, 2) être reproductivement isolés des espèces parentales et 3) l'isolement reproducteur doit être dérivé de l'évènement d'hybridation.

4.3-Processus d'introgression par rétrocroisement

Après l'hybridation initiale, les hybrides fertiles pourront ont alors deux possibilités de reproduction. D'un côté, ils peuvent se reproduire avec d'autres congénères hybrides. De l'autre, ils peuvent faire des rétrocroisements avec l'une ou l'autre des espèces parentales (ou les deux) si l'isolement reproducteur est faible avec ces dernières. La recombinaison entre les génomes hybrides et parentaux mène à des pertes d'hétérozygotie dans la prochaine génération de rétrocroisements. En accumulant assez d'évènements de ce genre, des individus phénotypiquement identiques à une des espèces parentales peuvent émerger, mais leur génome contiendra des fragments de l'autre espèce parentale impliquée dans l'évènement d'hybridation. Ce transfert d'information génétique entre deux espèces par hybridation et rétrocroisement est appelé introgression (Anderson et Hubricht 1938). Avec les technologies de séquençage de l'ADN, on sait aujourd'hui que plusieurs espèces contiennent des introgressions (Won *et coll.* 2005, Kronforst *et coll.* 2006, Vernot et Akey 2014). Les séquences retrouvées dans les génomes de ces espèces proviennent autant d'espèces contemporaines que d'espèces éteintes. On a aussi remarqué que les patrons d'introgression sont souvent asymétriques (Orive et Barton 2002). La sélection est souvent invoquée pour expliquer ces patrons. La recombinaison pouvant contribuer à la ségrégation des incompatibilités, les combinaisons qui augmenteraient le fitness des hybrides seraient alors sélectionnées (Schumer *et coll.* 2018). De la même manière, les allèles plus

adaptés se propageraient d'une espèce vers l'autre (Barton 2001, Whitney *et coll.* 2006). À cet effet, plusieurs rapports indiquant que les séquences introgressées contribuent à l'adaptation des espèces à leur environnement laissent envisager que la reproduction interspécifique a un rôle important dans l'évolution des espèces (Castric *et coll.* 2008, Racimo *et coll.* 2015).

4.4-Introgression sans rétrocroisements

Bien que l'introgression soit souvent liée à une perte d'hétérozygotie suivant des évènements de rétrocroisement, d'autres mécanismes peuvent être à l'origine de pertes d'hétérozygotie. Chez les allopolyploïdes, on observe souvent une diploïdisation du génome où les copies des gènes dupliqués sont perdues avec le temps. Plusieurs espèces contemporaines diploïdes auraient des origines polyploïdes vu le nombre de gènes dupliqués qu'elles possèdent (Ohno 1970, Lagercrantz et Lydiate 1996, Vision *et coll.* 2000). La restructuration du génome peut être liée à plusieurs mécanismes comme des erreurs lors de la méiose et la mitose, la recombinaison entre les chromosomes homéologues ou encore la réactivation d'éléments transposables (Wolfe 2001, Hufton et Panopoulou 2009). Il peut aussi se produire de la recombinaison entre les chromosomes homologues lors de la mitose (Stern 1936). Ces évènements surviennent à la suite de cassures double brin dans l'ADN d'un chromosome suivant sa réplication. La réparation de ces cassures nécessite l'utilisation d'un gabarit, rôle qui peut être rempli par le chromosome homologue (Haynes et Kunz 1981). La résolution de ces évènements peut mener à des enjambements réciproques de régions chromosomiques ou à de la conversion génique pouvant par la suite conduire à des pertes d'hétérozygotie dans les cellules filles issues de la mitose (Lee *et coll.* 2009). Il est donc possible que des régions ressemblant à des introgressions soient formées sans nécessairement passer par des rétrocroisements.

5-Spéciation et hybridation chez les microorganismes

Les microorganismes sont parmi les organismes les plus diversifiés sur terre, représentant environ 9 % de la biodiversité décrite et 35 % de la biodiversité totale estimée (Chapman 2009). Ils sont les premiers organismes à s'être développés sur terre et se sont diversifiés en lignées donnant éventuellement naissance à toutes les espèces vivantes (Woese 1998, Zaremba-Niedzwiedzka *et coll.* 2017). Ils ont une importance capitale dans le maintien de tous les écosystèmes de la planète (Falkowski *et coll.* 2008), même les plus extrêmes (Whitman *et coll.* 1998, Rothschild et Mancinelli 2001). Les microorganismes sont des modèles intéressants au point de vue de l'étude de l'évolution. Les techniques classiques de microbiologie facilitent la manipulation et le maintien de populations microbiennes au laboratoire, ce qui ouvre la porte à l'utilisation d'approches expérimentales qui ne sont pas accessibles à l'étude des macroorganismes. Leur temps de génération de l'ordre des heures ou des minutes permet d'obtenir des milliers de générations et une variation génétique considérable en très peu de temps si on compare avec des organismes multicellulaires. Finalement, leurs génomes sont assez compacts, ce qui permet le séquençage de génomes entiers à un coût abordable, même pour des centaines d'individus. Ces caractéristiques ont ouvert la voie à des approches expérimentales permettant l'étude des processus évolutifs au laboratoire dans des populations expérimentales sur des périodes de temps allant de plusieurs mois à plusieurs décénnies (Lenski *et coll.* 1991), ce qui peut représenter des dizaines de milliers de générations.

Les procaryotes (eubactéries et archées) sont souvent utilisés pour les études d'évolution expérimentale, la bactérie *Escherichia coli* étant un modèle de choix. Cependant, comme ils se reproduisent par mitose, il est difficile d'aborder des sujets comme l'isolement reproducteur ou encore l'hybridation chez les procaryotes. Il y a tout de même des possibilités de

flux de gène entre les populations microbiennes, qui se produit lors de transferts horizontaux de gènes. Malgré les difficultés que pose la définition d'espèces pour les bactéries (Lawrence 2002, Gevers *et coll.* 2005), elles forment tout de même des populations qui maintiennent une certaine indépendance génétique, écologique et géographique (Shapiro et Polz 2015). Il est aussi possible de tracer des parallèles entre isolement reproducteur chez les eucaryotes et interruption ou réduction du transfert horizontal entre deux lignées bactériennes (Bobay et Ochman 2017).

Les eucaryotes unicellulaires sont un groupe formé de protozoaires, de plantes et de champignons. En conditions favorables, ils se reproduisent souvent par mitose comme les procaryotes et peuvent aussi procéder à des transferts horizontaux (Andersson 2005). Cependant, lorsqu'un stress survient, ils ont souvent recours à la reproduction sexuée impliquant méiose et fusion de gamètes (Bernstein et Bernstein 2010). Ceci rend donc possible l'étude de l'isolement reproducteur (pré- ou postzygotique) et ultimement de la spéciation en utilisant les mêmes critères que pour les macroorganismes. Le rôle de l'hybridation dans l'évolution et l'adaptation des champignons est bien connu chez certaines espèces économiquement importantes comme des pathogènes ou des souches industrielles (Arnold 2004, Morales et Dujon 2012, Stukenbrock 2016). Cependant, peu d'information est disponible sur la spéciation par hybridation comparativement aux macroorganismes.

5.1-*Saccharomyces cerevisiae*

Parmi les champignons, *Saccharomyces cerevisiae* est un organisme important autant pour la recherche que pour l'économie. Premier eucaryote dont le génome a été séquencé (Goffeau *et coll.* 1996), c'est un modèle extrêmement utile en biologie qui a permis, entre autres, la description moléculaire du cycle cellulaire et la découverte de la fonction de plusieurs gènes chez les humains (Spellman *et coll.* 1998, Stanchi *et coll.* 2001). *S. cerevisiae* est aussi impliquée dans plusieurs procédés de

fermentations, et ce, depuis plus de 7000 ans (Mortimer 2000, Cavalieri *et coll.* 2003). Ceci fait en sorte que cette levure est souvent considérée comme une espèce domestiquée (Ciani *et coll.* 2004). Des travaux de génomique des populations ont révélé que la structure des populations de *S. cerevisiae* est plus liée à ses diverses utilisations industrielles qu'à sa répartition géographique (Liti *et coll.* 2009). Son évolution semble donc avoir été fortement influencée par son étroite association avec les activités humaines.

5.2-Cycle de vie de la levure

Le cycle de vie des levures a deux phases, une phase haploïde et une phase diploïde. Dans la nature, les levures passeraient la majorité du temps en phase diploïde (Knop 2006). Dans cette phase, la levure peut se reproduire par mitose où une cellule mère bourgeonne pour donner naissance à une cellule fille identique. Lorsqu'un déficit en azote dans le milieu survient, la levure diploïde entre dans un cycle de reproduction sexuée. La cellule diploïde va donc faire un cycle de méiose, donnant naissance à quatre spores haploïdes avec des génomes recombinés. Ces spores sont contenues par une paroi, l'asque, qui retient les spores en une structure tétraédrique (Neiman 2005). Lors du retour à un environnement propice à la croissance, les spores germent et ont deux possibilités : se propager de manière mitotique à l'état haploïde ou participer à un évènement de reproduction sexuée pour former des nouveaux diploïdes. La capacité des spores à germiner et se diviser mitotiquement est utilisée comme un indicateur de la fertilité de leur diploïde parental.

Chez les levures, les deux types sexuels sont appelés mat a et mat α. L'identité sexuelle d'une levure est gouvernée par les allèles non homologues MAT a et MAT α. Ces allèles sont trouvés dans des loci où la transcription est réprimée, ce qui est dû à la structure d'hétérochromatine que forme l'ADN dans ces régions (Nasmyth 1982). Un troisième locus situé entre les deux allèles permet leur expression à la suite d'un

évènement de recombinaison entre les régions flanquantes des allèles, médié par une endonucléase (Lee et Haber 2015). MAT α exprime un facteur de transcription qui régule l'expression des gènes de reproduction sexuée incluant une phéromone, le facteur alpha, le récepteur du facteur a, ainsi qu'un répresseur de l'expression des gènes MAT a spécifiques (Hagen *et coll.* 1993). Suivant l'association des phéromones sexuelles avec leurs récepteurs, deux cellules de types sexuels compatibles vont fusionner par des projections cellulaires (schmoo) pour former un zygote. La fusion des noyaux aura lieu dans le zygote et de ce dernier émergera une nouvelle levure diploïde (Merlini *et coll.* 2013).

La reproduction sexuée peut alors prendre trois formes. La première est un évènement de reproduction entre deux spores d'une même tétrade. Les parois des spores les protègent de l'environnement, mais les maintiennent aussi en association même si l'asque est détérioré (Coluccio et Neiman 2004). La possibilité de croisement avec des cellules d'un asque voisin est donc réduite. Comme chaque tétrade porte deux paires de spores génétiquement identiques, les diploïdes résultant de ces croisements auront donc 33 % de chances d'être homozygotes. La seconde forme est un évènement d'autodiploïdisation. Les levures du genre *Saccharomyces* sont généralement homothalliques, ce qui veut dire qu'elles ont la possibilité de changer leur type sexuel (Klar 1987). Après une division mitotique, la cellule fille peut changer son type sexuel pour ensuite fusionner avec sa cellule mère, générant un diploïde complètement homozygote. La dernière forme est la reproduction entre deux spores de tétrades différentes. C'est dans ce contexte que l'hybridation survient chez les levures.

5.3-Le Saccharomyces sensu stricto et l'hybridation

Mis à part *S. cerevisiae*, le genre *Saccharomyces* (*Saccharomyces sensu stricto*) compte actuellement six autres espèces décrites (Kellis *et coll.* 2003) qui divergent de 5 à 20 % en séquence nucléotidique avec *S. cerevisiae* (Liti *et coll.* 2006). Ces champignons unicellulaires se trouvent

dans une variété d'environnements comme les écorces d'arbres feuillus et les sols qui leur sont associés, les fruits en décomposition et les insectes (Boynton et Greig 2014). Ils présentent aussi des différences de température optimale de croissance, ce qui laisse suggérer que le climat a eu un rôle à jouer au cours de leur évolution (Gonçalves *et coll.* 2011, Salvadó *et coll.* 2011). Cependant, leur métabolisme est très similaire et il n'est pas rare de retrouver deux ou trois de ces espèces en sympatrie sur les mêmes substrats dans certaines régions (Sweeney *et coll.* 2004, Sampaio et Gonçalves 2008, Zhang *et coll.* 2010).

Peu de barrières prézygotiques empêchent les levures du genre *Saccharomyces* de s'hybrider, du moins au laboratoire; il est donc possible d'obtenir des hybrides diploïdes entre toutes les espèces par simple coculture (Naumov *et coll.* 2000). Cependant, des études ont montré qu'il était possible de faire évoluer une discrimination de partenaire de reproduction chez les levures (Leu et Murray 2006). Chez les souches sauvages, il est aussi possible d'observer un isolement temporel lié au temps de germination des spores, forme d'isolement prézygotique non détecté chez les cellules haploïdes en croissance (Murphy *et coll.* 2006, Maclean et Greig 2008).

Les hybrides expérimentaux entre des souches de laboratoire sont viables et montrent de l'hétérosis pour la croissance dans plusieurs conditions (Shapira *et coll.* 2014, Bernardes *et coll.* 2016). Comme les levures passent la majeure partie de leur vie à l'état diploïde, l'hétérosis pourrait alors avantager les hybrides dans une compétition avec les espèces parentales. Cependant, les hybrides interespèces sont complètement stériles (Greig 2008). Le mécanisme majeur d'isolement reproducteur serait la divergence génétique globale entre les génomes parentaux. Cette dernière affecte le bon déroulement de la méiose en déclenchant le mécanisme de réparation des mésappariements (Hunter *et coll.* 1996, Greig *et coll.* 2003). Les protéines impliquées dans ce mécanisme reconnaissent les bases mésappariées pendant la mitose et

réparent les lésions causées par les bases modifiées (Harfe et Jinks-Robertson 2000) . Pendant la méiose, ce mécanisme défait les enjambements chromosomiques contenant trop de mésappariements. Les enjambements sont nécessaires à la bonne ségrégation des chromosomes et l'action de ce mécanisme sur les mésaparriements entraîne la formation de spores aneuploïdes (Rogers *et coll.* 2018). Malgré la contribution majeure de cet isolement à l'infertilité des hybrides interespèces, les autres mécanismes d'isolement reproducteur ont aussi été observés dans différentes espèces. Des expériences ont montré que l'évolution des BDMi pouvait se faire rapidement dans des populations expérimentales de levures (Anderson *et coll.* 2010), la contribution des BDMi à l'isolement reproducteur postzygotique semble toutefois mineure (Greig *et coll.* 2002). Des incompatibilités cytonucléaires causent un arrêt de la sporulation et donc un isolement fort, mais sont limitées à quelques espèces (Chou *et coll.* 2010). Les réarrangements de chromosomes semblent aussi contribuer à l'isolement postzygotique et seraient importants particulièrement quand la divergence est faible entre les souches (Charron *et coll.* 2014, Hou *et coll.* 2014).

Malgré l'isolement reproducteur, plusieurs hybrides industriels sont connus entre les espèces du genre *Saccharomyces* utilisées couramment dans les procédés de fermentation (*S. cerevisiae*, *S. bayanus*, *S. uvarum* et *S. kudriavzevii*) (Dunn et Sherlock 2008, González *et coll.* 2008, Baker *et coll.* 2015). Peu d'hybrides naturels ont été décrits jusqu'à maintenant, ce qui pourrait suggérer que l'émergence des hybrides industriels a possiblement été favorisée par l'environnement stressant des milieux de fermentation ou encore par les activités de propagation des souches par l'homme (Belloch *et coll.* 2008). Les données de génomique des populations suggèrent qu'une fraction mineure des évènements de reproduction sexuée surviendrait entre des spores de tétrades indépendantes (Ruderfer *et coll.* 2006, Tsai *et coll.* 2008), ce qui pourrait aussi expliquer la rareté relative, voire l'absence d'hybrides naturels. Trouver des évènements d'hybridation

chez des espèces dont l'évolution a peu été influencée par les activités humaines pourrait nous aider à mieux évaluer le rôle qu'a l'hybridation dans l'évolution et la spéciation des microorganismes eucaryotes.

5.4-Le système des souches américaines de *Saccharomyces paradoxus*

Parmi les Saccharomyces, *S. paradoxus* n'a jamais été associé avec l'activité humaine, son évolution n'a donc pas été influencée par la domestication. Espèce génétiquement proche de *S. cerevisiae*, elle a divergé il y a environ 5 à 10 millions d'années et sa séquence nucléotidique diverge d'environ 15% (Cliften *et coll.* 2001). C'est une levure mésophile, qui a un optimal de croissance autour de 30 °C, mais qui tolère mieux les faibles températures que *S. cerevisiae*. Elle a été isolée dans plusieurs régions de l'hémisphère nord et, contrairement à son espèce sœur, la structure de ses populations semble être grandement influencée par la géographie (Liti *et coll.* 2009). À l'échelle mondiale, on retrouve trois grands groupes génétiques : asiatique, européen et américain (Naumov *et coll.* 1997). Des souches génétiquement très proches de celles du groupe européen ont été retrouvées en Amérique du Nord (Kuehne *et coll.* 2007) et en Nouvelle-Zélande (Zhang *et coll.* 2010), suggérant que les activités de colonisation humaines ont contribué à la migration de ces souches.

En Amérique du Nord, nos récents travaux ont démontré que, bien que *S. cerevisiae* soit très rare au nord de la frontière canado-américaine, *S. paradoxus* est présente partout dans le sud du Québec jusqu'à la péninsule gaspésienne (Leducq *et coll.* 2014). Cette espèce a déjà été décrite comme étant retrouvée en sympatrie avec *S. cerevisiae* dans les forêts sur l'écorce des chênes et les sols qui leur sont associés. En se basant sur une analyse de séquence multilocus (9 gènes nucléaires), Kuehne *et al.* (2007) ont décrit trois groupes génétiques présents sur le continent nord-américain. Le groupe A (*SpA*) est composé de souches proches génétiquement de celles retrouvées en Europe. Les isolats appartenant à ce groupe sont plutôt rares, ce qui suggère une petite taille

de population. Comme mentionné auparavant, ce groupe correspond aux souches européennes de *S. paradoxus* s'étant établies sur le continent à la suite d'un évènement de migration transocéanique récente suivi d'évènements de dispersion.

Le groupe B (*SpB*) est formé de souches indigènes de l'Amérique du Nord. Leur divergence nucléotidique avec *SpA* est d'environ 4 % à l'échelle du génome (Leducq *et coll.* 2016). On retrouve *SpB* un peu partout dans le sud du Québec, mais aussi à l'ouest du continent, des isolats ayant été échantillonnés jusqu'en Colombie-Britannique et en Oregon (Hyma et Fay 2013, Charron *et coll.* 2014). C'est de loin le groupe le plus communément retrouvé lors des échantillonnages, suggérant une grande taille de populations. En accord avec cette distribution étendue et de grandes populations, ce groupe montre une diversité génétique plus élevée que les deux autres (Leducq *et coll.* 2014).

Le troisième groupe est aussi indigène au continent nord-américain. Les souches du groupe C (*SpC*) ont, jusqu'à maintenant, presque exclusivement été isolées au Québec. La divergence nucléotidique de *SpC* est d'environ 4 % avec *SpA* et 2 % avec *SpB* (Leducq *et coll.* 2014, Leducq *et coll.* 2016). Ce groupe semble avoir une distribution limitée à la vallée du Saint-Laurent et partiellement chevauchante avec *SpB* dans la région entourant la ville de Québec (Leducq *et coll.* 2014). Malgré la faible divergence nucléotidique à l'échelle du génome entre *SpB* et *SpC*, ces deux groupes sont caractérisés par des différences marquées pour certains traits comme la croissance à haute température et la résistance aux cycles de gel-dégel (Leducq *et coll.* 2014). Nos travaux précédents ont montré que ces différences pourraient être liées à des adaptations des levures aux conditions climatiques locales (Leducq *et coll.* 2014). Les données génétiques recueillies suggèrent que les groupes *SpB* et *SpC* auraient été isolés dans différents refuges glaciaires pendant la dernière glaciation. Pendant cette période, les levures se seraient adaptées à différents environnements et auraient accumulé de la divergence génétique (Charron

et coll. 2014, Leducq *et coll.* 2016). Selon ce modèle, la distribution aujourd'hui observée constituerait un évènement de contact secondaire (Charron *et coll.* 2014).

5.4.1-Isolement reproducteur chez *S. paradoxus*

Bien que des souches *SpA* et *SpB* aient pu être isolés sur les mêmes arbres (Charron *et coll.* 2014), les analyses génomiques ne présentent aucune trace de flux de gène. Un isolement reproducteur postzygotique intrinsèque partiel réduit la viabilité des spores hybrides à environ 40 % (Charron *et coll.* 2014). Malgré la dominance de l'isolement reproducteur postzygotique entre les espèces de Saccharomyces, des travaux récents suggèrent que les spores du groupe *SpA* ont tendance à favoriser l'appariement avec d'autres spores de leur groupe plutôt qu'avec *SpB* (Murphy et Zeyl 2011). Cet isolement prézygotique est partiel et varie selon les souches utilisées, ce qui suggère qu'il pourrait être sous sélection dans *SpA* et contribuer au renforcement de l'isolement reproducteur total entre ces groupes. Malgré une divergence génétique entre *SpB* et *SpC* un peu moins élevée que celle de ces groupes avec SpA, l'isolement reproducteur postzygotique entre ces groupes est du même ordre de grandeur (Charron *et coll.* 2014). Ces deux groupes se comportent toutefois différemment de *SpA* pour ce qui est de l'isolement reproducteur intragroupe. Certaines souches de *SpC* sont aussi reproductivement isolées du reste de leur groupe que le sont deux souches de groupes différents. Cet isolement reproducteur semble être expliqué par des réarrangements de chromosomes puisque le polymorphisme nucléotidique chez *SpC* est limité (Charron *et coll.* 2014). Au sein de ce système, l'isolement reproducteur partiel entre *SpB* et *SpC* ainsi que leur distribution chevauchante peuvent donc constituer une opportunité pour les lignées de participer à des évènements d'hybridation.

6-Objectifs

Mes travaux de doctorat s'inscrivent dans un contexte où la majorité des connaissances sur l'isolement reproducteur et la spéciation par l'hybridation a été obtenue par l'étude des macroorganismes. Les microorganismes forment un groupe d'organismes très diversifié, mais les connaissances demeurent insuffisantes à propos des mécanismes qui pourraient contribuer à l'émergence de nouvelles espèces au sein de ce groupe. Les levures, eucaryotes microscopiques capables de reproduction sexuée, représentent une opportunité d'étudier la formation d'espèce hybride en laboratoire. Cette thèse est séparée en trois chapitres qui se sont intéressés à trois aspects différents de l'hybridation chez des microorganismes capables de reproduction sexuée.

Dans le chapitre I, nous nous sommes intéressés à la génomique des populations des levures sauvages *Saccharomyces paradoxus*. Les données disponibles suggèrent que les distributions des groupes *SpB* et *SpC* représenteraient un évènement de contact secondaire. Nous avons identifié des signatures génomiques laissées par un évènement d'hybridation en étudiant le génome de 161 souches représentant la diversité nord-américaine de *S. paradoxus*. Des approches de séquençage complet de génome, de phénotypage à haut débit et de caryotypage nous ont permis de découvrir une lignée cryptique au sein de *SpC*. Le génome de cette lignée contient des introgressions et des réarrangements de chromosomes trouvés chez *SpB*, nous suggérant une origine potentiellement hybride. En ajoutant des tests de croisements aux données déjà recueillies, nous avons tenté de déterminer si cette lignée pouvait être issue d'un évènement de spéciation par hybridation.

La découverte de la lignée *SpC** du chapitre I suggère que l'hybridation entre *SpB* et *SpC* a eu lieu récemment. Cependant, les échantillonnages précédents de levures sauvages ne comptent pas d'hybrides de première génération ou dont la contribution génomique des

parents est proche de 50%. Ceci pourrait être signe que les hybrides contemporains sont désavantagés par un isolement reproducteur postzygotique extrinsèque. Cet isolement reproducteur est peu étudié chez les levures puisque les hybrides interespèces sont typiquement avantagés par l'hétérosis. Cependant, la majorité des études rapportant l'hétérosis se sont faites dans des contextes déconnectés de l'écologie des levures. Soit ces études utilisent des souches de laboratoire ou industrielles ou encore elles utilisent des conditions environnementales peu réalistes dans un contexte naturel.

Au chapitre II, nous explorons les traits des hybrides dans des conditions un peu mieux connectées au le contexte écologique connu. Un criblage phénotypique à grande échelle chez plusieurs hybrides expérimentaux et leurs parents a été réalisé dans plusieurs conditions expérimentales représentant diverses sources de carbones, d'azote ou de stress qui pourraient être présents dans l'environnement des *Saccharomyces*. En comparant les phénotypes des hybrides avec celui des parents, il nous a été possible de déterminer si les hybrides montrent des phénotypes qui pourraient les désavantager par rapport aux parents.
Une des différences principales entre les hybrides de première génération et les espèces hybrides établies se trouve dans leur fertilité. Les nouveaux hybrides ont souvent une fertilité faible comparée aux espèces introgressées. L'amélioration de cette fertilité dans le temps est donc une étape clé à l'établissement des lignées hybrides.

Au cours des travauxx décris dans le chapitre III, nous avons testé l'hypothèse que, chez des organismes ayant des cycles sexués et asexués de reproduction, la fertilité pouvait être recouvrée sans avoir recours aux rétrocroisements. Des hybrides expérimentaux de levure *S. paradoxus* ont été soumis à une évolution en absence de sélection. En suivant les changements de fertilité après l'évolution, il nous a été possible d'observer des améliorations de la fertilité et d'en explorer les causes génétiques.

CHAPITRE I: SPECIATION DRIVEN BY HYBRIDIZATION AND CHROMOSOMAL PLASTICITY IN A WILD YEAST

Jean-Baptiste Leducq[1,4]*, Lou Nielly-Thibault[1]*, Guillaume Charron[1]*, Chris Eberlein[1], Jukka-Pekka Verta[1], Pedram Samani[2], Kayla Sylvester[3], Chris Todd Hittinger[3], Graham Bell[2], Christian R. Landry[1]

*These authors contributed equally to this work.

1- Département de Biologie, Institut de Biologie Intégrative et des Systèmes, PROTEO, Pavillon Charles-Eugène-Marchand, 1030 avenue de la Médecine, Université Laval, Québec, Quebec G1V 0A6, Canada

2- Department of Biology, McGill University, Montreal, Quebec H3A 1B1, Canada

3- Laboratory of Genetics, Genome Center of Wisconsin, DOE Great Lakes Bioenergy Research Center, Wisconsin Energy Institute, J.F. Crow Institute for the Study of Evolution, University of Wisconsin-Madison, Madison, Wisconsin 53706, USA

4- Département des Sciences Biologiques, Pavillon Marie-Victorin, 90 rue Vincent d'Indy, Université de Montréal, Montréal, Quebec H2V 2S9, Canada

Nature Microbiology

2016 - 1: 15003

I.1-RÉSUMÉ

Bien que l'hybridation soit reconnue comme ayant un énorme potentiel pour la génération de biodiversité, les exemples de spéciation par hybridation homoploïde sont limités à quelques espèces d'animaux et de plantes. L'absence d'exemple de ce mécanisme de spéciation chez les eucaryotes unicellulaires pourrait laisser croire que l'hybridation n'a qu'un rôle limité dans la spéciation des microorganismes eucaryotes. Pourtant, des études menées en laboratoire sur les champignons et les levures du genre Saccharomyces ont montré que les hybrides peuvent développer de l'isolement reproducteur très rapidement. Ceci suggère que la spéciation par hybridation est possible dans les populations naturelles de tels organismes. Nous rapportons ici un cas de spéciation par hybridation homoploïde chez *Saccharomyces paradoxus*, une levure naturellement associée aux arbres feuillus. Une analyse de génomique des populations nous a permis de montrer qu'un évènement de contact secondaire entre deux espèces en formation a permis l'établissement d'une lignée hybride. Cette lignée s'est par la suite rétrocroisée avec une des espèces parentales pour former la lignée introgressée que nos analyses ont révélée. Nous montrons également que cette lignée possède un profil phénotypique unique qui pourrait lui procurer une forme d'isolement pré-zygotique avec les souches parentales. La lignée introgressée est également isolée reproductivement des espèces parentales par deux mécanismes d'isolement postzygotique différents soit des réarrangements chromosomiques avec un parent et la divergence génétique entre les chromosomes homéologues avec l'autre. Notre étude constitue un des exemples les mieux documentés de spéciation hybride chez un eucaryote.

I.2-ABSTRACT

Hybridization is recognized as a powerful mechanism of speciation and a driving force in generating biodiversity. However, only few multicellular species, limited to a handful of plants and animals, have been shown to fulfill all the criteria of homoploid hybrid speciation. This lack of evidence could lead to the misconception that speciation by hybridization has a limited role in eukaryotes, particularly in single-celled organisms. Laboratory experiments have revealed that fungi such as budding yeasts can rapidly develop reproductive isolation and novel phenotypes through hybridization, showing that in principle homoploid speciation could occur in nature. Here we report a case of homoploid hybrid speciation in natural populations of the budding yeast *Saccharomyces paradoxus* inhabiting the North American forests. We show that the rapid evolution of chromosome architecture and an ecological context that led to secondary contact between nascent species drove the formation of an incipient hybrid species with a potentially unique ecological niche.

I.3-INTRODUCTION

Hybridization is a major force in evolution because it can prevent population divergence by maintaining gene exchange (Barton et Hewitt 1985). However, hybridization can also lead to speciation by combining genomes that have been evolving independently and that, when combined, provide advantageous phenotypes to the hybrid individuals and populations (Rieseberg *et coll.* 1995, The Heliconius Genome Consortium *et coll.* 2012). The genomic composition of the hybrid individuals may also lead to reproductive isolation with their parental lineages (Lukhtanov *et coll.* 2015). The identification of homoploid hybrid speciation events, which occur without change in chromosome number – contrary to what is observed in allopolyploid and homopolyploid speciation – requires that we show the reproductive isolation of the hybrid with the parental species, the identification of traces of past hybridization in the genome, and that the isolating mechanisms are at least partly derived from hybridization (Schumer *et coll.* 2014). Despite extensive investigations in sexually reproducing microbes, no case of such homoploid speciation has been reported in these eukaryotes (Schumer *et coll.* 2014). Yet, laboratory experiments have shown that hybridization can contribute to rapid species formation in yeast (Greig *et coll.* 2002).

Here we examine the genomics and the ecological bases of species formation in a yeast natural system. The budding yeast *Saccharomyces paradoxus* is the closest known relative of the model yeast *Saccharomyces cerevisiae*. *S. paradoxus* is a free-living saprophyte mostly found in the sap and on the bark of deciduous trees and their associate soil (Sniegowski *et coll.* 2002, Maganti *et coll.* 2012, Hyma et Fay 2013, Charron *et coll.* 2014). *S. paradoxus* has a nearly worldwide distribution but contrary to *S. cerevisiae*, there is no evidence for its domestication by humans (Liti *et coll.* 2009, Hyma et Fay 2013, Boynton et Greig 2014). Accordingly, the biogeography (Liti *et coll.* 2009) and the genome evolution (Bergström *et*

coll. 2014) of this unicellular fungus are mostly influenced by natural processes. Four genetically and phenotypically distinct lineages of *S. paradoxus* have been identified so far and correspond to populations from Europe, Far East Asia, America and North-East America (Koufopanou *et coll.* 2006, Kuehne *et coll.* 2007, Leducq *et coll.* 2014). In North America, three lineages of *S. paradoxus* occur in partial sympatry. One of these lineages, corresponding to the European population (Lineage *SpA*), has a sparse distribution and strains of this lineage are mostly isogenic, due to a recent colonization event (Kuehne *et coll.* 2007). The two other lineages in North America (*SpB* and *SpC*) are indigenous and are incipient species (Charron *et coll.* 2014) differentially distributed along a southwest to northeast gradient (Figures I.1a and I.S1). *SpB* strains display enhanced fitness at high temperature and survival to a freeze-thaw cycle when compared to *SpC* strains, consistent with adaptations to climatic conditions (Leducq *et coll.* 2014). Despite the overlapping distributions and partial post-zygotic reproductive isolation between *SpB* and *SpC* (Charron *et coll.* 2014), no first-generation hybrid has been identified so far, suggesting, along with with the monophyly of these lineages, that speciation has been initiated.

I.4-MATERIALS AND METHOD (ONLINE METHODS)

I.4.1-Strain sampling, genome sequencing and phenotypic characterization.

Strain sampling was described previously for most strains (Charron *et coll.* 2014, Sylvester *et coll.* 2015). Thirty-two additional strains were sampled in Quebec, New-Brunswick, Maine and Massachusetts in 2013 as described previously (Sniegowski *et coll.* 2002, Charron *et coll.* 2014). DNA was extracted following standard protocols (QIAGEN DNAeasy). Whole genome sequencing was performed according to Illumina procedures. An initial set of 24 strains was sequenced at high-coverage (HC) and 137 at low coverage (LC) on two separate lanes of HiSeq 2500 Illumina. The HC

libraries were prepared using the New England Biolabs® (n = 8), Lucigen® (n = 8) and TruSeq Illumina® (n = 8) protocols with no noticeable difference in quality. Reads can be retrieved from NCBI under the BioProject number PRJNA277692 (BioSamples SAMN03389655-SAMN03389678) and number PRJNA277692 (BioSamples SAMN03389659 - SAMN03389817). Reads were mapped onto the *S. paradoxus* reference genome (CBS432 38) using Bowtie2 (Langmead et Salzberg 2012) and duplicated reads removed using Picard (http://broadinstitute.github.io/picard/). Preliminary SNP calling on these HC libraries was performed using the BCFtools from SAMtools (Li *et coll.* 2009) with default parameters. The LC libraries were prepared following the Illumina Nextera XT protocol. All 161 libraries (HC+LC) were used together for an overall SNP and indel calling using FreeBayes (Garrison et Marth 2012) with default parameters and variants were filtered using VCFlib (Garrison et Marth 2012). The global phylogenetic reconstruction was performed with filtered variants in complete deletion using PhyML (Guindon *et coll.* 2010) with the model TN93 and a LRT for branch support. Population structure was examined using STRUCTURE (Falush *et coll.* 2003) with a subset of markers randomly sampled in the genome and an assumed number of clusters between 1 and 6.

I.4.2-Phenotypic analyses

The ability of *SpC* (n=5) and *SpB* strains (n=5) to utilize specific carbon sources and nitrogen sources was initially examined using BIOLOG plates (assay plates PM1, PM2 and PM3, catalog numbers 12111, 12112, 12121) following the manufacturer protocol with the modifications described in the supplementary material (Annexe A section 12, Figure I.S2). A subset of these conditions was then selected to measure growth rates of 182 strains covering *SpA*, *SpB*, *SpC* and *SpC** in order to examine how *SpC** strains compare to the parental *SpC* and *SpB* lineages. The 182 strains were assembled into two arrays of 1536 colonies on solid medium (12 replicates per strain per plate, omnitrays) following the procedures

outlined in Rochette *et coll.* (2015) on rich standard yeast YPD (yeast extract, peptone, dextrose) medium. Ten strains, including two *SpC** strains of type SK, did not grow on these arrays and thus were removed from the downstream analyses. These arrays were then replicated on other YPD plates and incubated in a range of temperatures (10, 15, 20, 25, 20 and 35°C). In another experiment, strains were transferred from YPD plates to YP medium with alternative carbon sources (2% Maltose, 2% Galactose, 2% Mannose, 2% Fructose, 2% Methyl α-D-glucopyranoside, 2% Sucrose or 2% Glucose) or synthetic medium with various sources of nitrogen (5 g/L Isoleucine, 5 g/L Tyrosine, 5 g/L Proline, 5 g/L Glutamine, 5 g/L Asparagine, 5 g/L Lysine, 5 g/L Histidine or 5 g/L Glycine) and grown at 25°C. In each of the temperature and the carbon and nitrogen source experiments, strains were transferred for a second round of selection in selective conditions. All experiments were performed using a BMC-BC S&P Robotic platform with 1536 pin tools (0.5 mm diameter). Colony size was measured by imaging plates and counting pixel intensities of colonies as described in Diss *et coll.* (2013). Freeze-thaw survival was performed as described in Leducq *et coll.* (2014). Data from Leducq et al. was used in addition to another set of 36 strains. Briefly, samples of cells were frozen at -80°C and a control sample was kept on ice. The two samples were then plated on YPD medium and the number of colony-forming units (CFUs) in the frozen sample/control sample was used as a measure of survival rate.

I.4.3-Measurement of reproductive isolation

The detailed procedures were described in Charron *et coll.* (2014) The HO locus was inactivated by complete gene deletion using antibiotic resistance cassettes and homologous recombination, and strains were sporulated to isolate haploid strains. Mating types were tested by performing crosses with control strains. Heterothallic strains expressing

an antibiotic resistance cassette at the HO locus were mated and diploid cells were selected using a pair of antibiotics. The resulting diploid strains were then sporulated and tetrads with 4 apparent spores were dissected and deposited on YPD plates. Spore survival was estimated as the number of colony-forming spores after 72h divided by the total number of spores dissected for a given cross.

I.4.4-Analysis of hybridization and identification of introgressed regions

Genomic islands of high diversity (He) and low differentiation (Fst) were identified from the 24 HC genomes in 500bp discrete windows along the genome. Expected heterozygosity (He) was estimated as He = 2p(1-p) and the fixation index Fst was computed using methods implemented in the R package HIERFSTAT (Goudet 2005). Peaks of high Fst and low He values were identified by computing t-statistics (p-value < 0.0001) between He and Fst estimates in 10kb contiguous regions of 20x500bp regions. The phylogenetic approach used to identify introgressed regions was performed on discrete 2,000kb windows (Figure I.S3) and the phylogeny of these regions were determined based on pairwise evolutionary distance (ED) using mutation probabilities (μ_P) estimated from *Saccharomyces cerevisiae* mutation accumulation experiments (Zhu *et coll.* 2014) (see supplementary material, Annexe A section 17 for details; Figure I.S4). The extent of introgression in the *SpC* population was examined by identifying introgressed regions in the LC strains. We used the *SpA* reference genome (CBS432), 10 *SpB* and 6 *SpC* non introgressed HC genomes as reference and used sites with fixed alleles in each group as a diagnostic for introgression in the overall population (see supplementary material, Annexe A section 18). All strains were assigned to one group or the other based on these diagnostic markers across the entire genome using 5Kb discrete windows. Error rates were estimated using two sequencing libraries obtained from the same homozygous clone.

I.4.5-Chromosomal changes

Potential aneuploidies and chromosomal changes were analyzed by Pulse Field Gel Electrophoresis (PFGE). All samples were treated as described in Charron *et coll.* (2014) and standard genomic DNA from *Saccharomyces cerevisiae* was used as control on all gels. Details on the image analyses are provided in supplementary material (Annexe A section 19). The genomes of the 24 HC strains were assembled using ABySS (Simpson *et coll.* 2009). Several parameters were tested on a subset of strains and led to the selection of K=64 as a final parameter for the assembly (Figure I.S5). Chromosomal rearrangements were detected using scaffolds aligned on the reference S288C genome with methods implemented in MAUVE 48 and through the identification of chimeric scaffolds, i.e. scaffolds that align with two non-contiguous regions of the reference genome. Chromosomal fusions were confirmed by mapping reads on the chimeric scaffolds. One major chromosomal inversion (iVI) was confirmed by PCR in the collection of sequenced strains (Figure I.S6, Table I.1).

I.4.6-Analysis of segregation in F2 hybrids

We examined the segregation of the *SpC* and *SpB*-like genomic regions in the *SpC* × *SpC** F2 hybrids by genotyping 25 pools of haploid strains (3-30 strains per pool) grouped by crosses and by number of spores that survived in a given tetrad (n =1-4). Pools were used to construct 25 Illumina Nextera XT libraries and were sequenced in one lane of Illumina HiSEQ2500. Reads can be retrieved from NCBI under the BioProject number PRJNA277692 (BioSamples SAMN03389818-SAMN03389842). Reads were mapped on the reference genome of *S. paradoxus* CBS432 38 using Bowtie2 39. The depth of coverage (number of reads at a given position) was analyzed along the genome as a measure of ploidy to identify regions that segregate abnormally. We also controlled for coverage along the genome of parental strains to ensure that variations were not inherited

by aneuploidy in natural populations, which was not the case (6% of strains; Figures I.S7 and I.S8). Variant calling was performed on the pooled libraries using Freebayes 41 with options –J and –p 2. Genotypes of the spores were compared directly with that of their parental strains. The inversion on chromosome iVI in the spores was genotyped in the spores from the same crosses by PCR using oligonucleotides flanking the inversion.

I.4.7-History reconstruction and timing of introgression

The time of divergence between *SpC*, *SpB* and *SpC** was estimated using BEAST 49 from the introgressed regions and in regions without traces of introgression using the set of HC genomes. Several groups were formed according to the overall phylogeny and the introgressed regions present: CBS432, *SpA* (5 strains), *SpB* (10 strains), *SpC* (5 strains), PE (1 *SpC** strain), PP (2 *SpC** strains) and IO2 (1 *SpC** strains). As calibration time we assumed that *SpC* and *SpB*'s initial divergence was initiated at the onset of the last glaciation 110,000 years ago (see supplementary material, Annexe A section 28 for the justification).

I.5-RESULTS

I.5.1-Incipient speciation in wild yeast

We sequenced the genomes of 161 strains from the north east of North America to uncover the evolutionary history of this ongoing speciation event, using the European lineage (*SpA*) recently introduced in America as outgroup (see sup information section 1; Table S1). A phylogeny based on 14,974 filtered polymorphic sites confirms that *SpB* and *SpC* form distinct lineages with 2.09±0.01 % nucleotide divergence (Figure I.1b) and we observed no first-generation hybrid among lineages, as supported by overall low heterozygosity (~0.1%; Figure I.S9). The monophyletic *SpB* clade shows population substructure (Figure I.1b-c), consistent with its broad geographic distribution (Figures I.1a and I.S1).

This analysis also revealed a third clade that is sister to the *SpC* lineage, which we call *SpC** (Figure I.1b). *SpC** consists of eleven *SpC* strains (22% of all sequenced *SpC*) showing limited genetic admixture with other *SpC* strains (Figure I.1c) and a narrow geographic distribution (Figures I.1a and I.S1).

Because *SpC** strains are mostly found near the southwest limit of the *SpC* distribution, this new lineage may be locally adapted and thus show increased fitness over *SpC* in the relatively warmer temperatures encountered in this region (Ellison *et coll.* 2011). We measured the growth of 182 strains along a temperature gradient from 10 to 35°C (Figure I.1d) and found that *SpC** strains show a growth advantage over *SpC* (+3-42% colony growth) similar to the *SpB* advantage over *SpC*, particularly at high temperature (+5-36%; $p < 0.001$; Tukey test; Tables S2-S3; Figure I.1e). We previously reported that populations of *S. paradoxus* showed variation for resistance to freeze-thaw cycles that is correlated with their latitude, with higher resistance in the south where these cycles are more frequent (Leducq *et coll.* 2014). We found that like *SpC*, *SpC** is highly sensitive to a freeze-thaw cycle (8.9% and 5.2% survival, respectively), while *SpB* has significantly higher survival (74.4%; $p < 0.05$; Tukey test; Tables S2-S3; Figure I.1f). We also examined other growth conditions that may reflect the large diversity of substrates on which yeasts were isolated (Charron *et coll.* 2014, Leducq *et coll.* 2014, Sylvester *et coll.* 2015). Because *SpC* and *SpB* were recently shown to perform differently in maltose, lyxose and glucosamine (Samani *et coll.* 2015), we first examined growth in various carbon sources that are known to shape sap microbial diversity (Filteau *et coll.* 2011). We also examined growth performance on different nitrogen sources, including several amino acids, which are known to be a major nitrogen source in tree sap (Karley *et coll.* 2002). We found that the three lineages have contrasted metabolic profiles, as suggested by the growth advantage of *SpB* and *SpC* on specific carbon and nitrogen sources, respectively (Tukey test: $p<0.001$; Figure I.1g, Tables I.2-I.3). Overall, *SpC**

strains show a metabolic performance profile intermediate between *SpB* and *SpC*. Assuming that yeasts have a have high dispersal potential like many other eukaryotic microbes (Finlay 2002), these results suggest that the overall performance of *SpC** in specific climatic conditions and substrates is a potential cause for its limited ecological distribution in the middle of the *SpB*-*SpC* range.

The monophyly of *SpC** and its contrasted phenotypes imply limited gene flow with *SpC*, despite these two lineages have overlapping distributions and co-occur on neighboring trees in the region of sympatry (Figure I.S1). This indicates that *SpC** is reproductively isolated from *SpC* and could therefore represent an incipient species. Budding yeasts of the genus Saccharomyces have no obvious pre-zygotic isolation mechanisms in the laboratory (Liti *et coll.* 2006), which allows to measure progeny viability between *SpC**, *SpB* and *SpC* (Figure I.2). Our results indicate that survival of spores from *SpC*×*SpC** crosses is significantly lower than for *SpC*×*SpC* or for *SpC**×*SpC** ($p<0.001$; Tukey test) and on the same order of magnitude as what is observed for *SpB*×*SpC*, and *SpB*×*SpC** (Figure I.2, Tables I.4-I.5), despite the nearly 10-fold difference in terms of molecular divergence (0.27% vs. 2.09%).

I.5.2-Genome dynamics following hybridization

We found that the *SpC** lineage genome is a mosaic of *SpC* and *SpB* genotypes, with genomic islands of *SpB*-like alleles (i.e. fixed within *SpB*) present in *SpC** strains but absent in lineage *SpC* (Figure I.3a). These *SpB*-like regions correspond to 12 large regions (28-123kb) of high diversity within the *SpC*+*SpC** clade ($He > 0.005$; $p < 0.0001$; Figure S3a-c), overlapping with 23 large regions (22-119kb) for which divergence between *SpB* and *SpC*+*SpC** populations is low ($FST < 0.75$; $p < 0.0001$; Figure I.S10d-g) relative to the genome-wide average ($Fst \approx 1$). These *SpB*-like regions most likely result from the introgression of *SpB* into *SpC*, as shown by a window-based phylogenetic analysis (Neafsey *et coll.* 2010) of 24

strains with high-quality genomes (Figures I.S10h and I.S11a) and by a site-wise clustering analysis (Ellison *et coll.* 2011) of the 161 strains (HC and LC combined; Figure I.3a). The monophyly of the *SpC** clade is maintained even when these introgressed regions are removed (Figure I.S11b), suggesting that divergence, although very limited, has accumulated between *SpC* and *SpC** outside of the introgressions. The *SpB*-like regions represent 2.2-5.8% of the *SpC** genome, with 1.6% shared among all *SpC** strains. The common set of introgressed regions could be the remnants of a single ancestral hybridization event between the *SpB* and *SpC* lineages (marked as H0; Figure I.3a) and the polymorphic introgressed regions could result from the ongoing loss of *SpB*-like regions in *SpC** or from secondary introgression events (Table I.6).

We examined chromosome size variation within and between the *SpC* and *SpC** populations (Figure I.S12) and found that the *SpC** karyotype was significantly different from both *SpC* and *SpB* based on chromosome migration profiles ($p < 0.001$; Tukey test; Figure I.S13). Using de novo genome assembly, we found extensive variation in chromosome configuration among lineages (Figure I.3b) that supports the chromosome profile analysis (Figures I.S12 and I.S13). The patterns of inversion and small-scale exchange (≤ 20kb) among telomeric regions that are fixed within lineages confirm the overall *S. paradoxus* phylogeny (Figure I.S14a). One exception is a 42kb inversion within chromosome VI (iVI; Figure I.3b). This inversion is only shared between *SpB* and *SpC** and is absent in *SpC* (Figure I.S14a, Annexe A Table I.S1). Another exception is the fusion between chromosomes VI and the right arm of chromosome XIII (VItXIII) that is absent in *SpC* but fixed in *SpC** (Figures I.3b and I.S14b). This fusion segregates at low frequency in *SpB* (3.3%; referred as *SpB*f strains; Figure I.S14c). A phylogeny based on the iVI region reveals that *SpB*f is the sister group of *SpC**, suggesting that *SpC** inherited iVI and VItXIII from *SpB*f strains (Figure I.S14d). Chromosomal rearrangements are often

under-dominant and are thus unlikely to fix within a population (Walsh 1982), which could explain the low frequency of VItXIII in *SpB*.

I.5.3-Chromosomal changes and reproductive isolation

Given the low level of divergence between *SpC* and *SpC** outside of the introgressed regions, reproductive isolation is unlikely driven by molecular incompatibilities such as mismatch repair (Greig *et coll.* 2003), although those could be important in *SpB*×*SpC* and *SpB*×*SpC** hybrids. We thus hypothesized that chromosomal changes originating from hybridization contribute to the reproductive isolation between *SpC** and *SpC*. Accordingly, these changes would miss-segregate in the progeny of the F1 hybrids and correlate with spore viability (Loidl *et coll.* 1998) (Figure I.4a). We tested this hypothesis by sequencing 325 haploid F2 strains that we obtained from the sporulation of eight *SpC* × *SpC** diploid F1 hybrids (Figure I.4b, Table I.7). We identified large genomic regions (100-200kb) where *SpC** genotypes segregate unevenly in meiotic products with low spore viability (25%; Figure I.4c), corresponding to either *SpB*-like regions (IIL, VIr and XIIr; Figure I.4d) or to regions in linkage with the translocation (LT regions; Figure 4e). We also observed systematic ploidy increase in the translocated arm of chromosome XIII (XIIIr; Figure I.4e). We confirmed that the genomic regions involved in at least two major chromosomal rearrangements that differentiate *SpC** from *SpC*, namely the inversion in chromosome VI (iVI) and the right arm of chromosome XIII (XIIIr), segregate unevenly in the surviving spores and contribute to reduced spore viability, as predicted from their simple Mendelian segregation (Figure I.5a). The iVI inversion also co-localizes with a *SpB*-like region that is preferentially transmitted in surviving spores (Figure I.S15a). Both the frequency of the iVI inversion typical of *SpC** (Figure I.5b) and the ploidy of the XIIIr region (Figure I.5c) significantly decrease with spore survival (1,000 random permutations; $p<0.05$), showing that surviving spores systematically inherit these regions more than expected under

normal segregation. The breakpoint of ploidy change in XIIIr coincides with the fusion point with chromosome VI in *SpC** (Figure I.S15a), indicating that VItXIIIr causes lethal aneuploidy in *SpC* × *SpC** haploid hybrids by the sorting of XIIIr (Loidl *et coll.* 1998) (Figure I.5a). We observed no significant ploidy variation on chromosome VI, suggesting that the over-representation of the iVI inversion in the spores and of the associated *SpB*-like region results from their linkage with XIIIr (LT region; Figure I.S15b). These results show that the chromosomal polymorphism segregating at low frequency within *SpB* (VItXIIIr) was transmitted by hybridization with *SpC* and now causes reproductive barriers between *SpC** and *SpC*. This transmission led to or co-occurred with the formation of the new lineage *SpC** and contributed to reproductive isolation between *SpC** and its parental lineage *SpC*. Because we observed similar patterns of abnormal segregation in other genomic regions (Figure I.5c), we reasoned that additional chromosomal differences between *SpC* and *SpC** also contribute to reproductive isolation (Figure I.S13; see also supplementary material, Annexe A section 27 for discussion). We estimated that taken together, chromosomal differences may lead to lethal aneuploidy in 45% of spores, hence contributing to up to 80% of actual reproductive isolation between *SpC* and *SpC** (Figure I.5a).

I.5.4-Timing of speciation

The distributions of *SpB* and *SpC* mirror those of many post-glacial taxa in this region (Table I.8) (Charron *et coll.* 2014). Therefore, we hypothesized that the present population pattern is the consequence of an allopatric separation of the *SpC* and *SpB* lineages during the last glaciation (110,000-12,000 years ago) with a secondary contact after the glacial retreats (Figure I.6). The pattern of polymorphism revealed by Tajima's D statistics is overall negative for the three groups, which suggests recent population expansion, with much greater effects in *SpC* than *SpC** and *SpB* populations. Because *SpC** strains mostly come from locations where

SpB and *SpC* are found in sympatry, the initial hybridization event most likely occurred in the secondary contact zone between the two post-glacial lineages *SpB* and *SpC* (Figures I.1a and I.S1). The common set of introgressed regions would then be the remnants of a single ancestral hybridization event between the *SpB* and *SpC* lineages (H0). Assuming that *SpB* and *SpC* have originally been separated before the onset of the last glaciation and using polymorphism and divergence data, we found that *SpC* and *SpC** diverged about 10,000 years ago (Figure I.S16, Table I.9). This date is again consistent with the fact that climate warming facilitated secondary contact and hybridization. Although no independent data point can be used to confirm these estimates, these assumptions correspond to 1.72 generations per day on average, which is consistent with all previous studies (Table I.10). Under the same assumptions, we estimated the timing of the introduction of *SpA* into North America about 300 years ago, and the divergence between American and European lineages at about 176,000 years ago (Figure I.S16, Table I.9). These again are consistent with previous scenarios (Kuehne *et coll.* 2007, Zhang *et coll.* 2010) and estimations (Liti *et coll.* 2006) and support the proposed model for the *SpB* and *SpC* divergence and secondary contact (Figure I.6).

I.5-DISCUSSION

We report a fully documented case of homoploid hybrid speciation in a eukaryotic microbe. In the yeast *S. paradoxus*, the incipient species (*SpC**) occupies the contact zone of its parental lineages *SpB* and *SpC* and displays intermediate phenotypes typical of ecological gradients in this region, indicating that a hybrid zone may have been found in the region 10,000 years ago (Hewitt 1988). The unequal contributions of *SpB* and *SpC* to the *SpC** genome suggest several non-exclusive scenarios for the mosaic *SpC** genome formation. First, backcrosses between the newly-formed hybrids and parental lineages could have been more frequent with *SpC*, for instance if *SpC* strains were more abundant than *SpB* in the hybrid zone.

This would have led to the slow erosion of the *SpB* haplotypes in the original hybrid. Such mechanism could be further enhanced by incompatibilities between parental genomic backgrounds (Hou *et coll.* 2015) that purged the *SpB* genome. This is supported by non-fixed *SpB*-like regions, which segregate unevenly in the surviving *SpC*×*SpC** progeny (Figure I.5d). However, backcrosses with *SpB* cannot be ruled out because the *SpB*-like region of chromosome II may have been acquired more recently by crosses with *SpB* (Figure I.S16, Table I.9). Another possibility is that the introgression of *SpB*-specific elements in the *SpC* genome may have conferred a fitness advantage to *SpC** in the hybrid zone and this would have maintained these genomic regions specifically (The Heliconius Genome Consortium *et coll.* 2012, Huerta-Sánchez *et coll.* 2014). Accordingly, rounds of mitotic recombination (Magwene *et coll.* 2011) or repeated selfing within the ancestral hybrid strains (Greig *et coll.* 2002) combined with natural selection could have contributed to the erosion of the *SpB* regions without the need for backcrosses with the parental lineages.

Studies in experimental evolution (Dunn *et coll.* 2013) and on the genomics of industrial and domesticated fungi (Libkind *et coll.* 2011) showed that frequent chromosomal rearrangements and genomic instability of hybrids between species may lead to reproductive isolation in artificial conditions. Extensive chromosomal polymorphism also segregates in natural yeast populations (Liti *et coll.* 2006, Wang *et coll.* 2012), making them a great potential for initiating reproductive barriers and ecological novelties when the ecological opportunities are present. The natural system we report is comparable with *S. cerevisiae* × *S. paradoxus* hybrids generated in the laboratory (Greig *et coll.* 2002). These hybrids were also marked by extensive chromosomal rearrangements and strong reproductive isolation with parental species, but did not have the distinctive ecological divergence that has evolved in *SpC**. The actual extent of reproductive isolation between *SpC** and its parental lineages is

likely underestimated because ecological divergence is a potential factor in generating prezygotic barriers in the wild (Keller et Seehausen 2012). Yet we demonstrated that alone, chromosomal differences could be the main cause of postzygotic reproductive isolation. Because these differences were inherited at the very beginning of hybridization and are fixed within lineages, they were likely the bases for initiating the ongoing genomic divergence and speciation. Our observations provide strong evidence that the processes of homoploid hybrid speciation can actually lead to incipient speciation in natural populations of fungi. These processes, which were shown to play key roles in shaping plant (Rieseberg *et coll.* 1995) and animals (Lukhtanov *et coll.* 2015) natural diversity, are therefore also shaping eukaryotic microbial diversity. The historical and ecological contexts into which this event took place suggest that changes of microbial species distribution caused by climate changes, in conjunction with the limited possibility for pre-zygotic isolation and the plasticity of their genomes, make speciation by hybridization a potential mechanism of diversification in the forest microbiota (Ordonez et Williams 2013).

I.6-AUTHORS' CONTRIBUTIONS

JBL, CRL, LNT and GC planned the experiments. GC, JBL and CE performed experiments. JBL, LNT and JPV performed bioinformatic analyses. PS, KS, CTH and GB provided strains and discussion in the early stages of this study. JBL and CRL drafted the manuscript with contributions from LNT, GC, CE, PS and GB.

I.7-ACKNOWLEDGEMENTS

We thank A. K. Dubé, K. Lambert, R. Nuwal, S. Haughian, A.-E. Chrétien, M. Caouette, I. Kukavica-Ibrulj, R. Levesque and the IBIS sequencing platform (B. Boyle) for technical help; P. Sniegowski, M.-A. Lachance and J. Anderson for providing strains, I. Levade and C. Lemieux for discussion, and N. Aubin-Horth, A. Moses, L. Bernatchez, J. Shapiro,

S. Pavey, F. Rousseau-Brochu, I. Gagnon-Arsenault, A. K. Dubé, A.-M. Dion-Côté, H. Vignaud, M. Nigg and three anonymous reviewers for comments on the manuscript. The data reported in this paper are provided in the supplementary materials and raw sequencing reads were deposited in NCBI (BioProject PRJNA277692).

I.8-FUNDING

NSERC Discovery Grant and HFSP Grant (RGY0073/2010) to C.R.L. FRQS Fellowships to J.-B.L., NSERC USRA Summer scholarships to L.N.T., FRQNT and NSERC PhD Fellowships to G.C. Some of this material (yeast collection) is based upon work supported by the National Science Foundation under Grant No. DEB-1253634 (C.T.H.) and by the DOE Great Lakes Bioenergy Research Center (DOE Office of Science BER DE-FC02-07ER64494). C.T.H. is a Pew Scholar in the Biomedical Sciences, supported by the Pew Charitable Trusts. CRL is a FRQS Junior Investigator and a Canada Research Chair in Evolutionary Cell Biology.

I.9-FIGURES

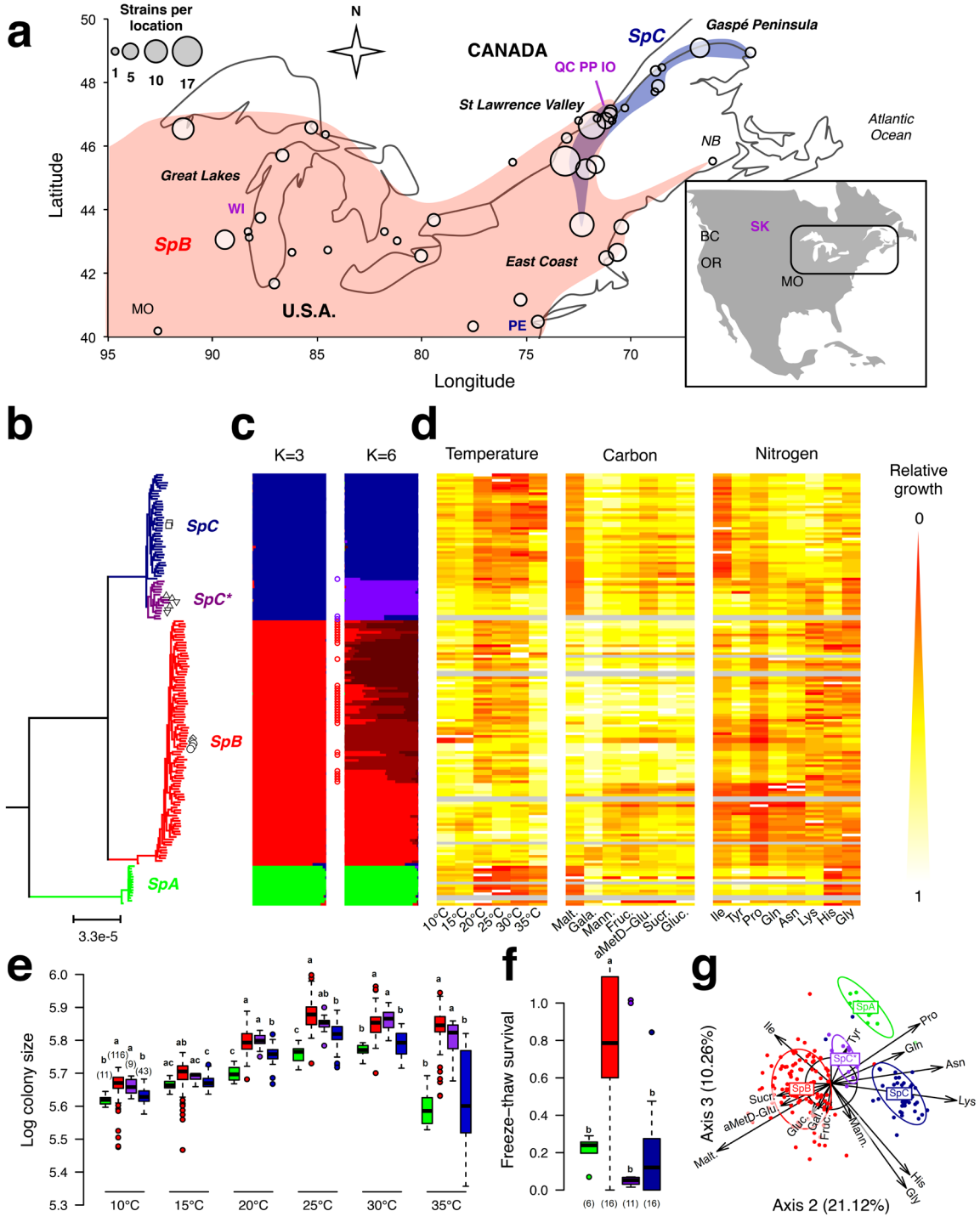

a
Strains per location
1 5 10 17
N
CANADA
SpC
Gaspé Peninsula
QC PP IO
St Lawrence Valley
Atlantic Ocean
NB
Great Lakes
WI
SpB
East Coast
MO
U.S.A.
PE
Latitude
Longitude
SK
BC
OR
MO
b
SpC
SpC*
SpB
SpA
3.3e-5
c
K=3
K=6
d
Temperature
Carbon
Nitrogen
Relative growth
0
1
10°C 15°C 20°C 25°C 30°C 35°C
Malt. Gala. Mann. Fruc. aMetD-Glu. Sucr. Gluc.
Ile Tyr Pro Gln Asn Lys His Gly
e
Log colony size
10°C 15°C 20°C 25°C 30°C 35°C
f
Freeze–thaw survival
(6) (16) (11) (16)
g
Axis 3 (10.26%)
Axis 2 (21.12%)
SpA
SpC*
SpB
SpC
Ile
Tyr
Pro
Gln
Asn
Lys
His
Gly
Sucr.
aMetD-Glu.
Malt.
Gluc.
Gala.
Fruc.
Mann.

Figure I.1: A cryptic *S. paradoxus* lineage revealed by a population structure and a distinct ecological niche from its sympatric close lineages.

(a) Sampling locations (circles) and distribution (inner frame) of *SpB* (red), *SpC* (dark blue; one strain in Pennsylvania) and the cryptic lineage *SpC** (purple). Location abbreviations are listed in Table S1. (b) Eleven *SpC** strains form a monophyletic group within the *SpC* lineage. Phylogenetic tree of 161 strains rooted with *S. cerevisiae* (14,974 SNPs). *SpA* is the European lineage (green). Open symbols indicate replicates and spores from a same strain. (c) STRUCTURE diagrams (6,881 randomly sampled SNPs, assuming K=3 and 6 populations) reveal 33 admixed *SpB* strains between three populations (<80% of assignation; red circles) and only three between *SpC* and *SpC** (PE and SK strains; purple circles). (d) Relative growth (log of colony size; warm colors) measured at different temperatures and in limiting nutrient conditions for seven carbon and eight nitrogen sources for 151 strains. Ten strains (grey) did not pass the selection step. (e) *SpC** shows enhanced growth at high temperature, similarly to *SpB*. Bold letters indicate significant differences among lineages (Tukey test: $p \leq 0.05$). The number of strains per lineage is indicated. Error bars indicate interquartile ranges. (f) Fraction of cells surviving a freeze-thaw cycle. *SpC** and *SpC* show reduced survival as compared to *SpB* ($p \leq 0.05$). (g) Metabolic profiles in limiting nutrient conditions summarized by a principal component analysis (axes 2 and 3 displayed). *SpC** is intermediate between *SpC* (amino-acid preference) and *SpB* (carbon preference).

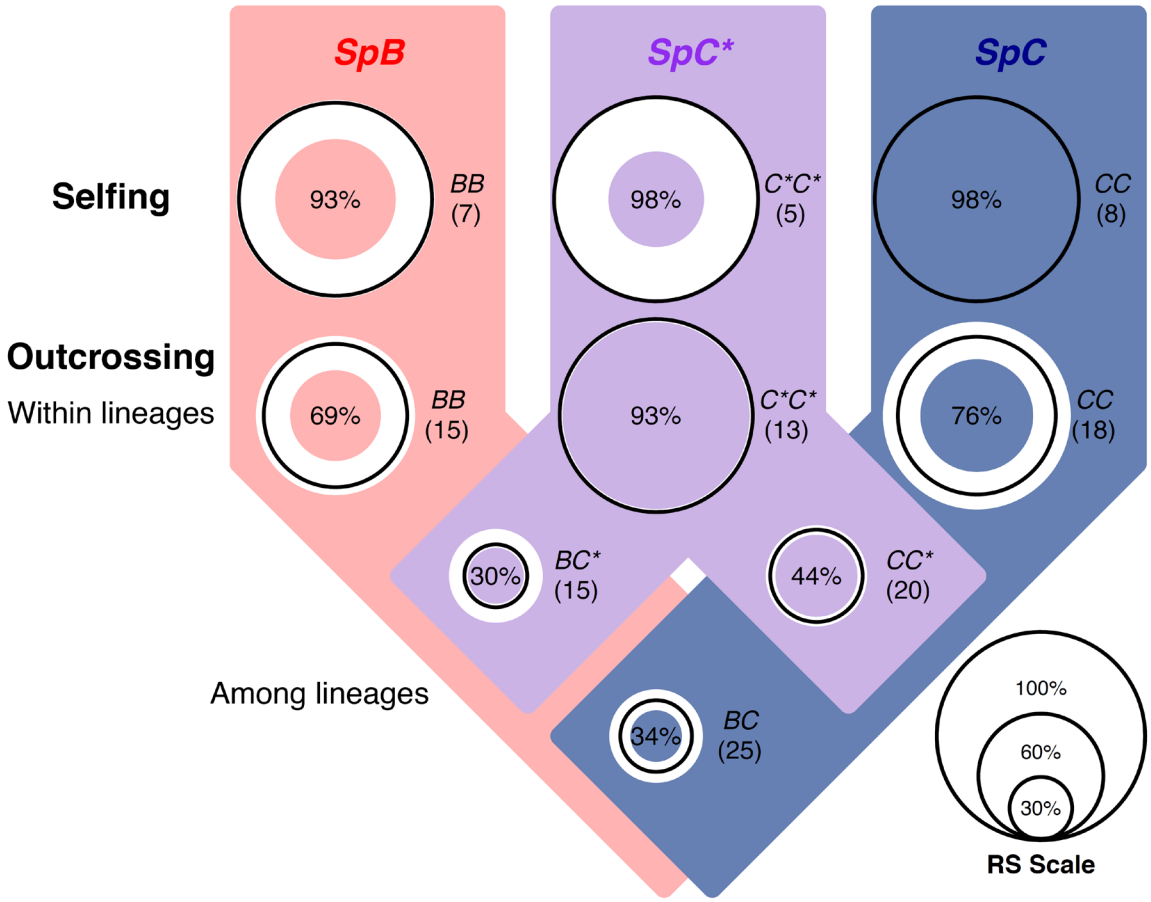


Figure I.2: The *SpC lineage is an incipient species, as revealed by its limited reproductive success with its sister lineages.**
Post-zygotic reproductive success (RS) within and between lineages as estimated by progeny viability. Circle diameter is proportional to median RS (scale on bottom right; 75 percentiles of RS in white). The number of crosses per category is indicated. *SpC** shows partial reproductive isolation with *SpC* and *SpB*, to an extent that is similar to crosses between *SpC* and *SpB*.

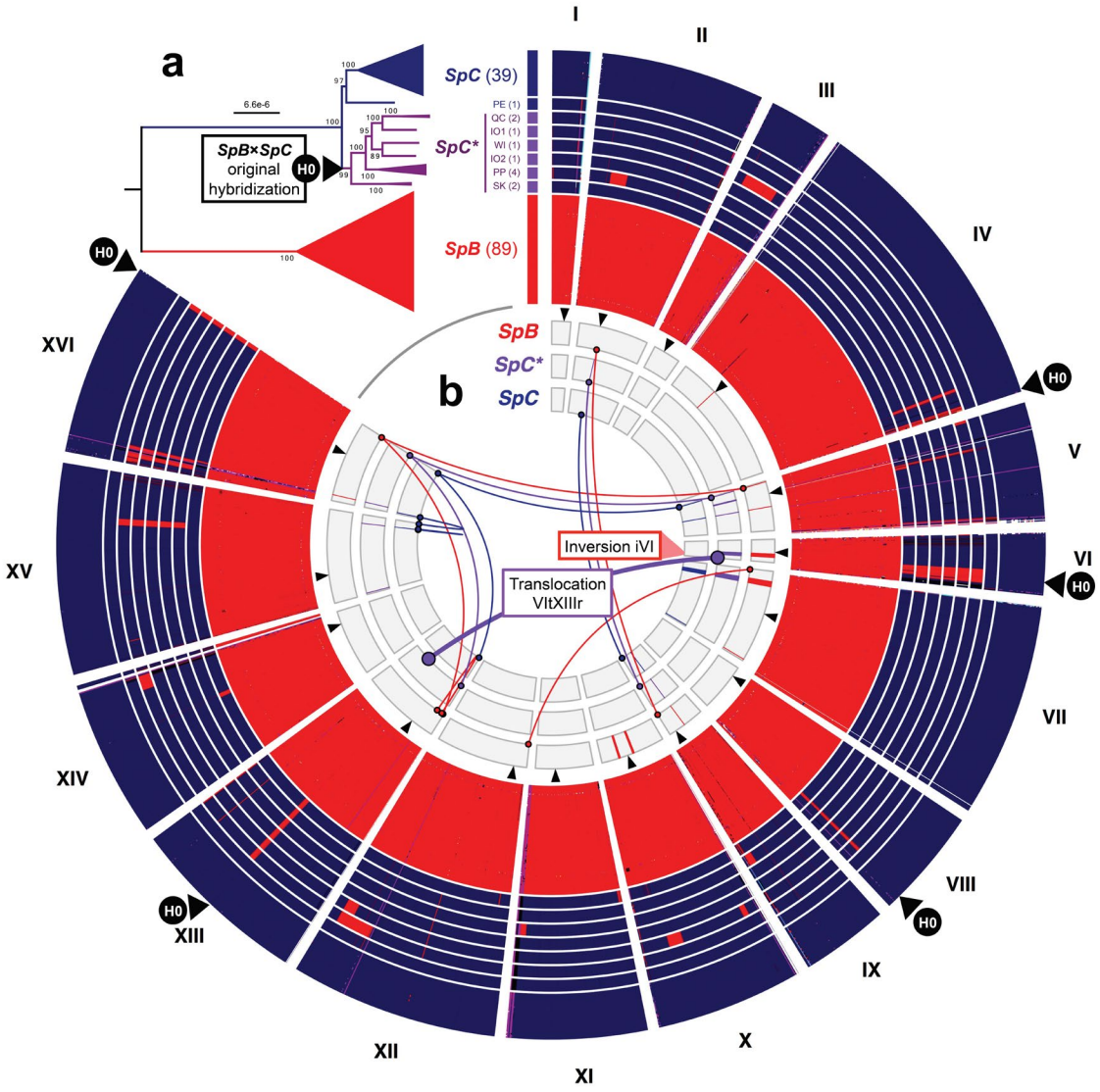

a
6.6e-6
SpB×SpC
original
hybridization
H0
SpC (39)
PE (1)
QC (2)
IO1 (1)
WI (1)
IO2 (1)
PP (4)
SK (2)
SpC*
SpB (89)
b
SpB
SpC*
SpC
Inversion iVI
Translocation
VItXIIIr
I
II
III
IV
V
VI
VII
VIII
IX
X
XI
XII
XIII
XIV
XV
XVI

Figure I.3: The *SpC lineage is a mosaic of *SpC* and *SpB* genomes and results from past hybridization.**

(a) *SpB*-like (red) and *SpC*-like (blue) genotypes (site-wise clustering approach, the *SpA* outgroup was removed) in 5kb windows along the genome of 147 *S. paradoxus* strains reveal that *SpC** strains are a mosaic of the two lineages, with a vast majority of *SpC* genotypes. *SpC** strains share a common set of *SpB*-like regions likely acquired during an original hybridization event between *SpB* and *SpC* (H0). Polymorphic *SpB*-like regions distinguish six *SpC** groups (see Table S6) that result from the ongoing loss of *SpB*-like regions or from secondary introgressions. (b) Lineage-specific karyotypes of 19 strains with high-quality genomes (*SpA* removed). Telomere exchanges (thin lines), translocations (thick lines) and inversions (colored portions in chromosomes) as compared to *S. cerevisiae*, mapped on circular diagrams of the 16 chromosomes (I-XVI). A translocation (VItXIII) between chromosome VI and the right arm of chromosome XIII (240kb) is fixed within *SpC** and absent elsewhere. The 42kb inversion in chromosome VI overlapping with a *SpB*-like region (iVI) was likely transmitted from *SpB* to *SpC** and is absent among the *SpC* high-quality genomes (see also Figure I.S14).

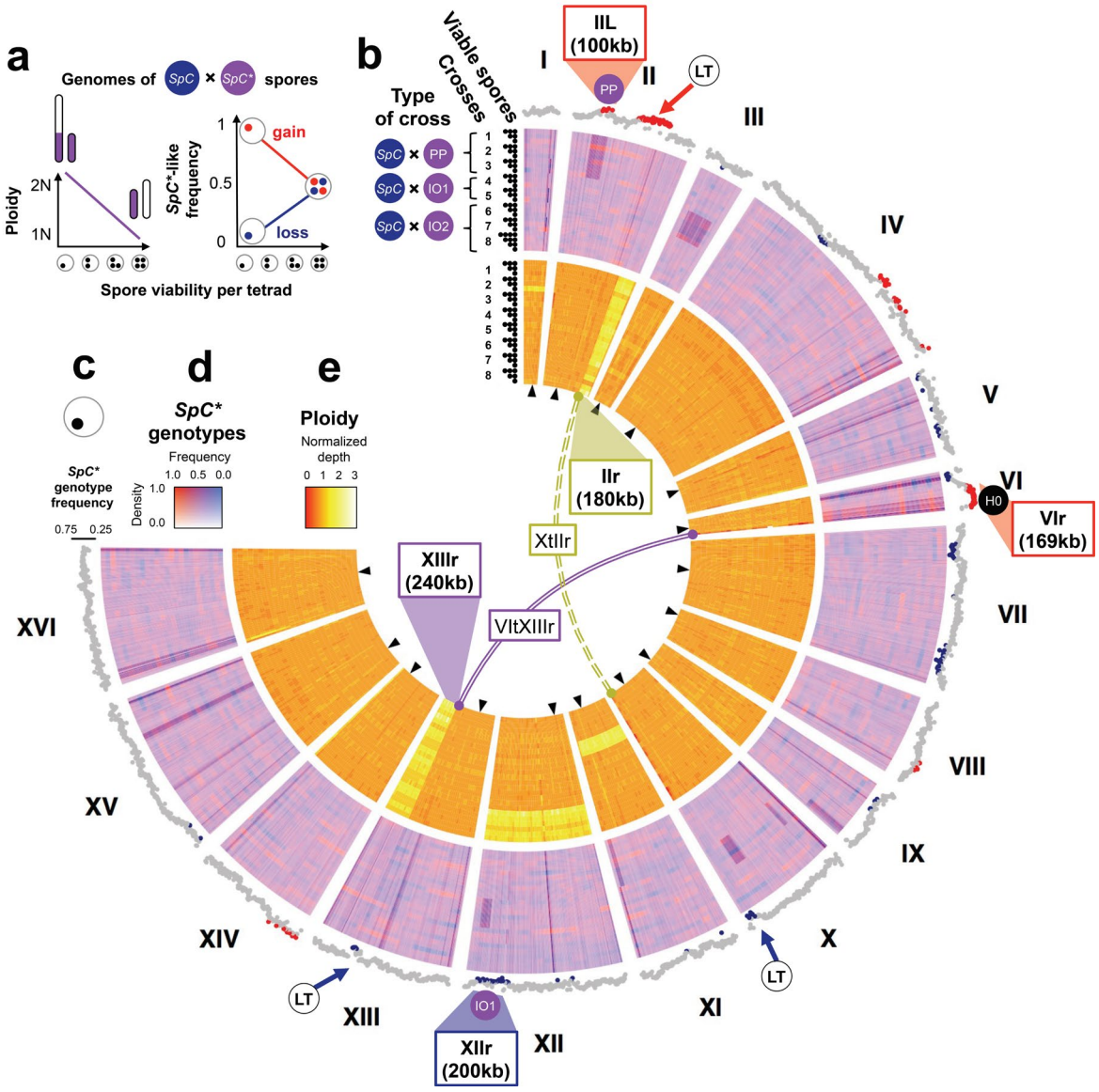

a
Genomes of SpC × SpC* spores
Ploidy
2N
1N
SpC*-like frequency
1
0.5
0
gain
loss
Spore viability per tetrad
b
Viable spores
Crosses
Type of cross
SpC × PP
SpC × IO1
SpC × IO2
IIL (100kb)
PP
LT
I
II
III
IV
V
VI
VII
VIII
IX
X
XI
XII
XIII
XIV
XV
XVI
H0
VIr (169kb)
IIr (180kb)
XtIIr
XIIIr (240kb)
VItXIIIr
IO1
XIIr (200kb)
c
SpC* genotype frequency
0.75 0.25
d
SpC* genotypes
Frequency
1.0 0.5 0.0
Density
1.0
0.0
e
Ploidy
Normalized depth
0 1 2 3

Figure I.4: Chromosomal rearrangements and introgressed regions unevenly segregate in the *SpC×*SpC* hybrid progeny.**

(a), Expected patterns of ploidy increase (left) and *SpB*-like genotype frequency variation (right) in genomic regions involved in decrease of *SpC**×*SpC* spore viability. (b), Genome sequencing of 384 spores from eight *SpC**×*SpC* crosses (*SpC** types indicated) pooled according to the number of surviving spores per tetrad (*n*=1-4). (c), Mean *SpB*-like genotype frequency in tetrads with low viability (*n*=1) in discrete 20kb windows along the genome. Strong segregation biases toward *SpC* (blue) or *SpC** (red) genotypes are observed in some *SpB*-like regions (IIL, VIr, XIIr) and regions linked to chromosomal translocations (LT). (d), Frequency of *SpC** genotypes (red-blue heatmap, 5kb windows) per pool of spores. Dark regions (intensity proportional to divergence between *SpC* and *SpC** parents) correspond to *SpB*-like regions. (e). Heatmap of ploidy variation (warm color scale, 5kb windows) suggests aneuploidy in hybrid progeny caused by chromosomal translocations. The rightmost breakpoint in XIIIr corresponds to the fusion point with chromosome VI (VItXIIIr, purple). A similar pattern in chromosome II (IIr) suggests the contribution of another chromosomal change (XtIIr, yellow; see Figures III.S13, III.S15). The iVI inversion has no visible effect on ploidy. Ploidy was calculated as sequencing coverage per pool normalized by coverage in a control tetrad (*n*=4; removed).

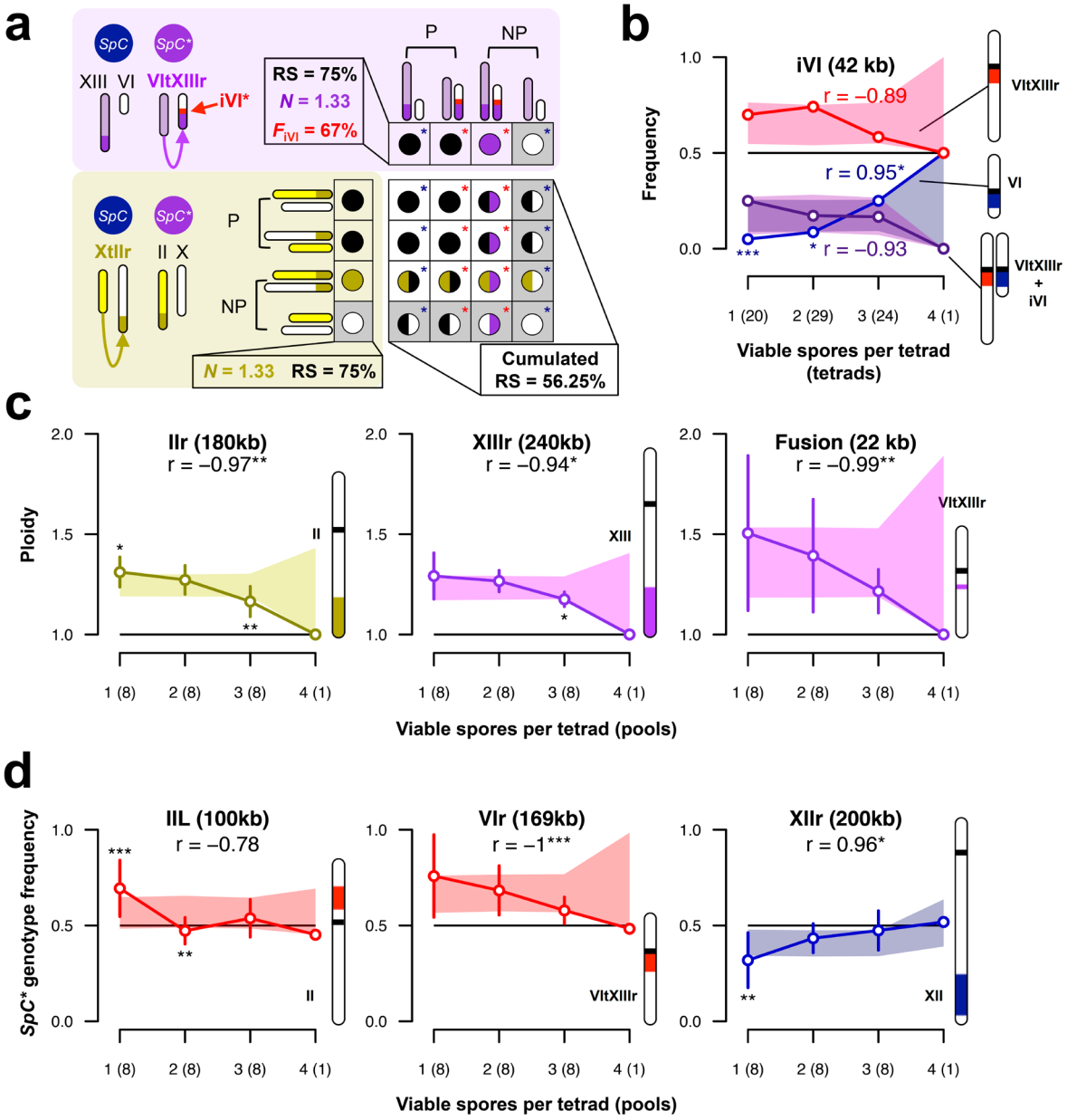

a
SpC
SpC*
XIII VI
VItXIIIr
iVI*
RS = 75%
N = 1.33
F iVI = 67%
P
NP
XtIIr
II X
N = 1.33 RS = 75%
Cumulated
RS = 56.25%
b
iVI (42 kb)
r = −0.89
r = 0.95*
r = −0.93
Frequency
VItXIIIr
VI
VItXIIIr
+
iVI
1 (20) 2 (29) 3 (24) 4 (1)
Viable spores per tetrad
(tetrads)
c
IIr (180kb)
r = −0.97**
XIIIr (240kb)
r = −0.94*
Fusion (22 kb)
r = −0.99**
Ploidy
VItXIIIr
1 (8) 2 (8) 3 (8) 4 (1)
Viable spores per tetrad (pools)
d
IIL (100kb)
r = −0.78
VIr (169kb)
r = −1***
XIIr (200kb)
r = 0.96*
SpC* genotype frequency
Viable spores per tetrad (pools)

Figure I.5: Chromosomal rearrangements and introgressed regions contribute to the decreasing viability of *SpC×*SpC* hybrid progeny.**
(a) Translocation VItXIIIr (purple) and inversion iVI (red) typical of *SpC** and translocation XtIIr (yellow) typical of *SpC*, could theoretically lead to reduced reproductive success (RS), increased ploidy in translocated regions (*N*) and increased frequency of the inversion (F_{iVI}) in surviving spores. The diallelic table represents the expected cumulated effects of translocations on spore survival considering parental (P) or non-parental (NP) associations of chromosomes. Spores (circles) are either diploid (*N*=2; purple and yellow), haploid (*N*=1; black) or have no copy of the translocated region (*N*=0; white), which is presumably lethal (grey frames). Red and blue stars indicate the iVI configurations from *SpC** and *SpC*, respectively. (b) The iVI inversion (*SpC** configuration; red line, frequency assessed by PCR) is more often transmitted than the *SpC* configuration (blue line) in tetrads with few surviving spores (number of tetrad per spore category in parenthesis). The purple line indicates the proportion of surviving spores harboring both configurations. (c) The right arms of chromosomes II (IIr; yellow) and XIII (XIIIr; purple) and the fusion region of VItXIIIr (purple) show systematic ploidy excess (384 *SpC**×*SpC* hybrid spores; see Figure I.4), which is negatively correlated with the number of surviving spores per tetrad. Mean (points) and standard deviation (bars) among crosses per tetrad category (number of pools per spore category in parenthesis). (d), *SpB*-like regions on chromosomes II (IIL) and VI (VIr) are more likely to be transmitted, while the *SpB*-like region of chromosome XII (XIIr) is most likely to be lost in surviving spores and this bias in transmission is negatively correlated with the number of surviving spores per tetrad. The following information is given for panels (b-d): values expected under a 2:2 segregation (black line) and correlation coefficient (*r*). Statistical significance ("*": $p \leq 0.05$; "**": $p \leq 0.01$; "***": $p \leq 0.001$) and 95% statistic range (frame) were estimated after 1,000 random permutations among spore categories within crosses (*n*=8).

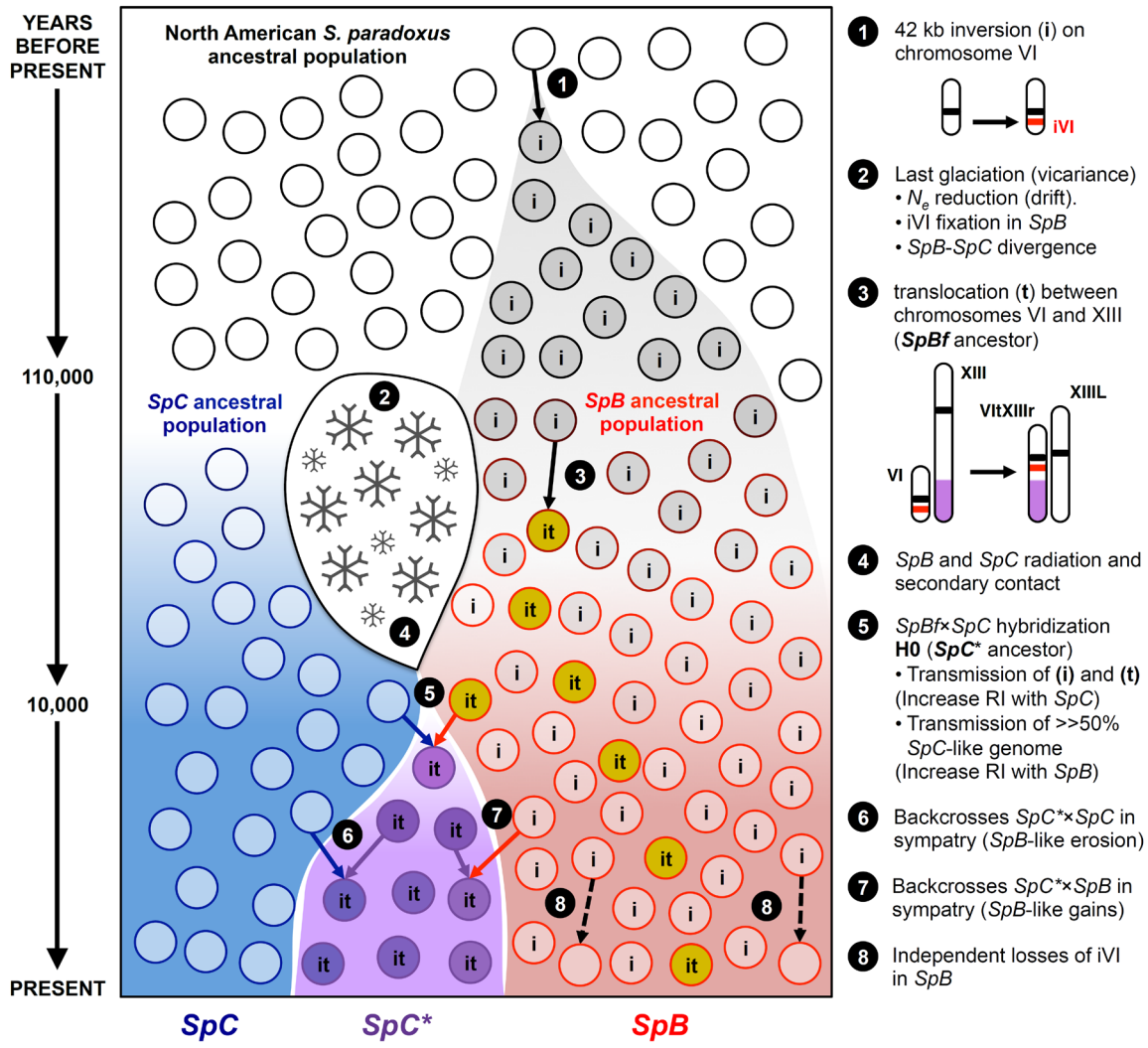


Figure I.6: A biogeographic scenario for the emergence of *S. paradoxus* lineages in North America.

Genomic and phenotypic analyses support the following scenario (see details in points 1-8): an ancestral *S. paradoxus* population (black circles) occupied the North-East American forest before the last glaciation (110,000 years ago). Climate change (ice sheet formation; 2) lead to vicariance of the ancestral population in two populations that progressively diverged for nucleotide polymorphisms (red and blue circles). Chromosomal changes (black arrows; 1, 3 and 8) segregated or eventually became fixed by drift, giving rise to the *SpB* and *SpC* ancestral populations. Secondary contact after the ice sheet retreat 10,000 years ago (4) and hybridization between *SpB* and *SpC* (5) followed by backcrosses (6 and 7) lead to the formation of the new lineage *SpC** (purple circles).

I.7-SUPPLEMENTARY FIGURES

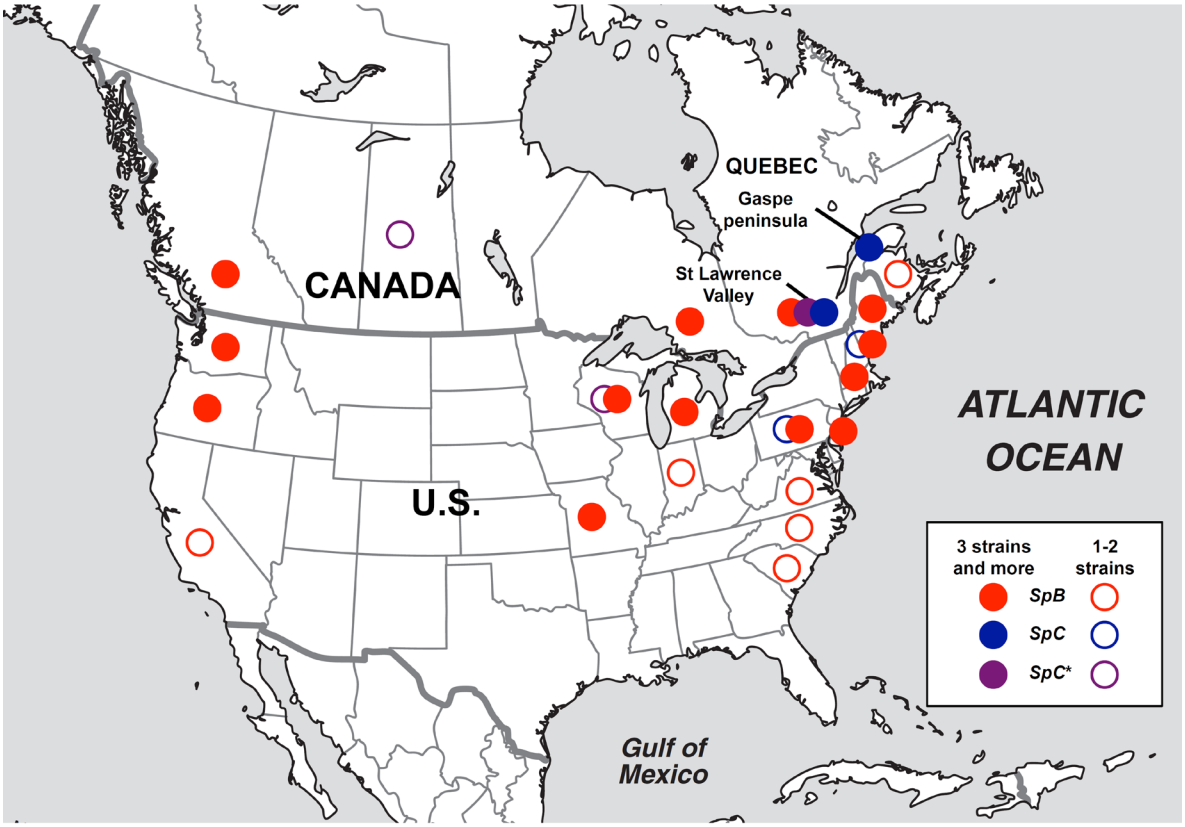


Figure I.S1: Distributions of *S. paradoxus* lineages *SpC* (blue) and the hybrid lineage *SpC (purple) are mostly restricted to Eastern Canada, while *SpB* is widely distributed across North America.**

Occurrence of at least one strain in a U.S. state or a Canadian province is indicated by an open circle. More frequent observations (3 strains and more) are indicated by full circles. Details are given for the St Lawrence Valley and the Gaspe Peninsula (QC, Canada). Sampling data from this study, the literature (Sniegowski *et coll.* 2002, Kuehne *et coll.* 2007, Liti *et coll.* 2009, Hyma et Fay 2013, Charron *et coll.* 2014, Leducq *et coll.* 2014, Sylvester *et coll.* 2015) and unpublished data (Sylvester K, Moriarty RV, Wright S, Langdon Q, Hittinger CT).

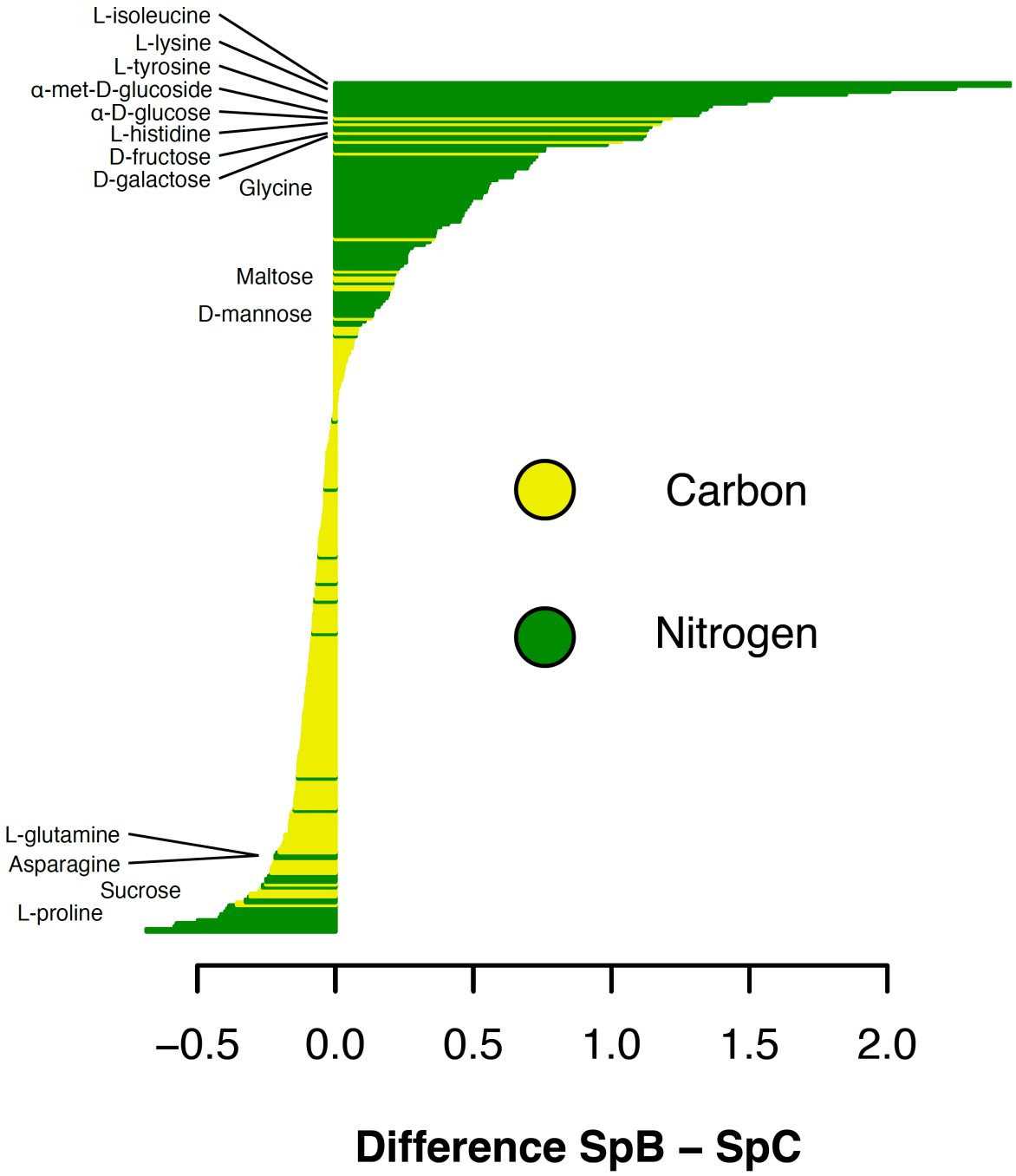


Figure I.S2: Difference in metabolic profiles.
(BIOLOG® plates, absorbance cumulated over 36 hours) between 5 *SpB* and 5 *SpC* strains for 194 carbon (yellow bars) and 95 nitrogen sources (green bars). Seven carbon and eight nitrogen (amino acids) sources for which *SpB* and *SpC* have opposite metabolism profiles are indicated.

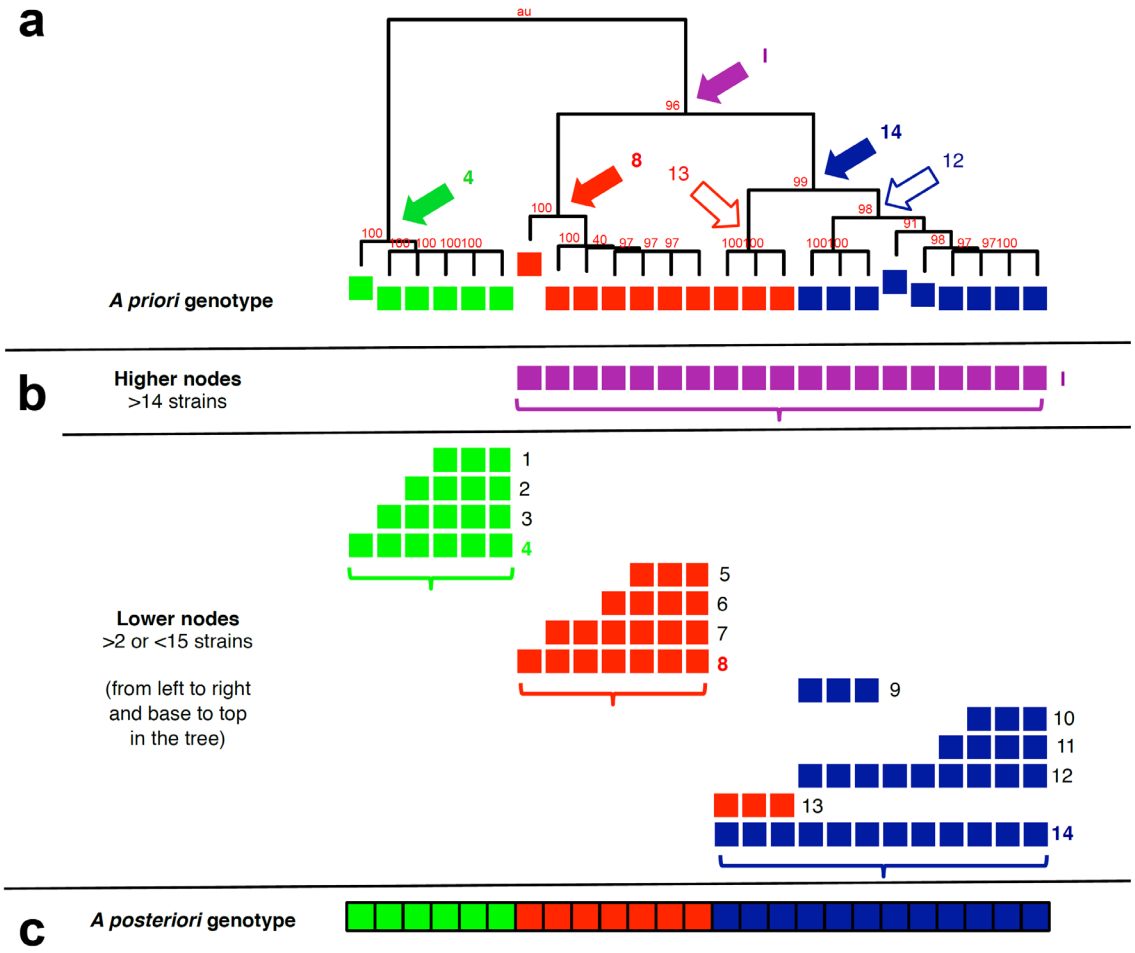

a
au
I
96
14
12
8
13
4
99
100
98
91
100
100
40
97
97
97
98
97
97100
100100
100100
100100
A priori genotype
b
Higher nodes
>14 strains
I
Lower nodes
>2 or <15 strains
(from left to right
and base to top
in the tree)
1
2
3
4
5
6
7
8
9
10
11
12
13
14
c
A posteriori genotype

Figure I.S3: Clustering analysis of 24 HC strains and the reference (see Figure I.S10h)(a) In each 2kb window, an evolutionary tree is established based on the total pairwise E_D («manhattan» distances; function *pvclust*: R package (Suzuki et Shimodaira 2006)). After 1000 bootstrap resampling, only significant clusters ($a \geq 95$) with at least three strains are conserved for the following analyses. The *a priori* strain grouping is based on whole-genome phylogeny (Fig S3h; lineage: *SpA*: green; *SpB*: red; *SpC*: blue). (b) Significant clusters are divided in two categories according to the number (n) of strains they contain: large clusters ($n > 14$; higher nodes) and small clusters ($n \leq 14$; lower nodes). In each category, only the largest non-overlapping clusters (brackets) are conserved and are assigned to a lineage if its *a posteriori* frequency in the cluster is at least half plus one. (c) *A posteriori* genotypes based on lower node assignment. Here, clusters 4 and 8 contain only *SpA* and *SpB* strains, respectively (full green and red arrows in the tree). Cluster 14 contains 3 *SpB* and 9 *SpC*, so *SpC* prevails (full blue arrow in the tree). Higher nodes (here, cluster I, purple arrow in the tree) define the mixed genotype of a strain only if it is assigned to no lower node.

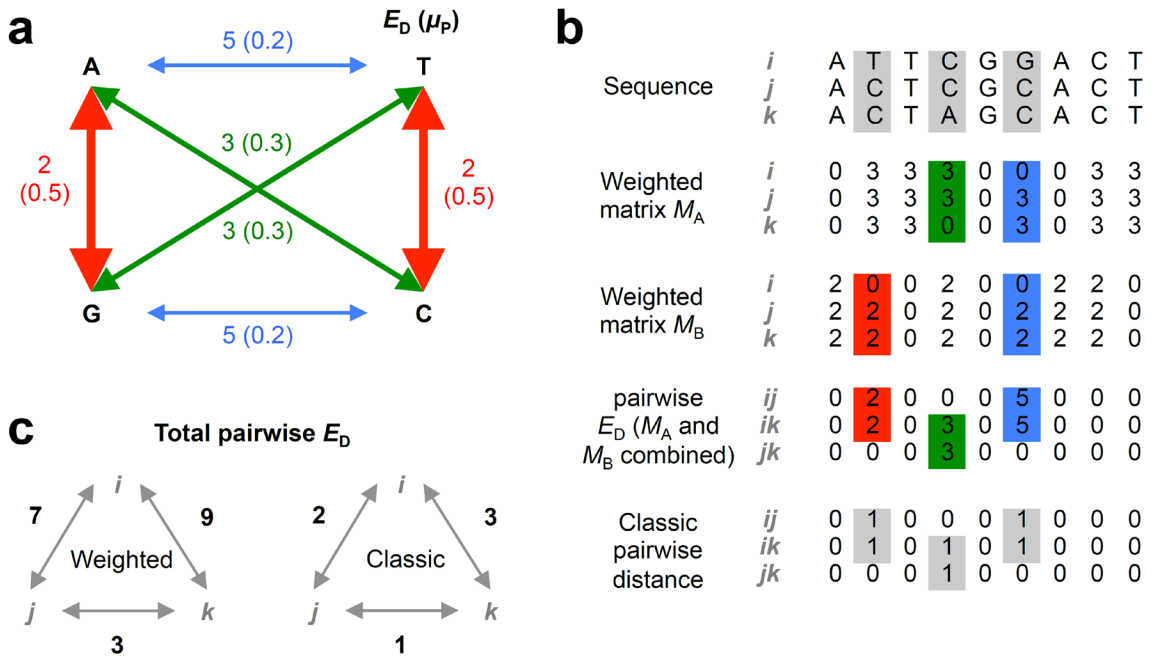


Figure I.S4: Conversion of nucleotide sequences into evolutionary distance matrices for the clustering analysis of 24 HC strains and the reference CBS432 using a sliding phylogeny approach.

(Figure I.S10h).(a) Mutation probabilities (μ_P) estimated from *S. cerevisiae* (Zhu *et coll.* 2014) in evolutionary distances ($E_D \approx 1/\mu_P$). (b) To convert the genotype of three hypothetical individuals (*i*, *j*, *k*) in a E_D matrix containing only the four possible E_D values (0, 2, 3 or 5), nucleotide sequences are first converted in two distinct matrices M_A and M_B. Matrix M_A is weighted 3 For T/C and 0 for A/G positions, and matrix M_B is weighted 2 for A/C and 0 for T/G positions. The pairwise E_D between two individuals *i* (M_{Ai}, M_{Bi}) and *j* (M_{Aj}, M_{Bj}) at a given position is $E_{Dij} = |M_{Ai} - M_{Aj}| + |M_{Bi} - M_{Bj}|$. (c) Comparison of total pairwise E_D between three individuals calculated according to our method (left) and under a classic pairwise distance matrix (right; all mutations are equally weighted).

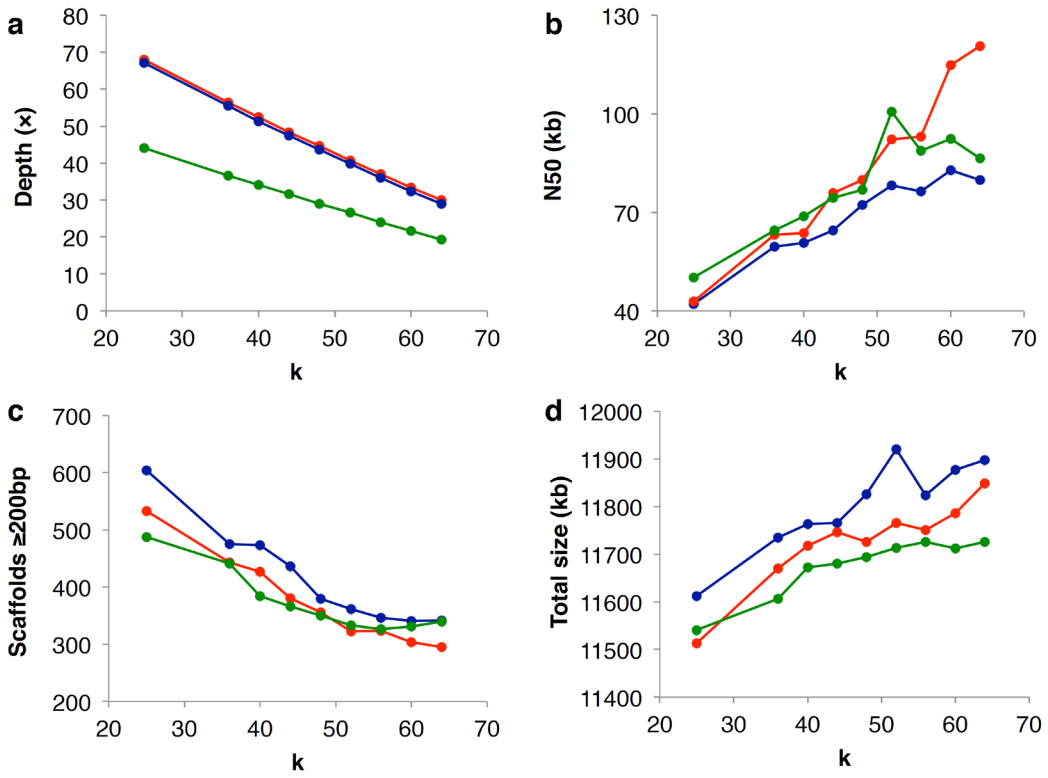


Figure I.S5: Statistics of genome assembly with ABySS for three HC genomes from lineage SpA (green), *SpB* (red) and *SpC* (blue) for k-mer lengths (k) tested in the range 25-64.

(a) Average depth of coverage as a function of *k*. (b) *N50* (in kb) as a function of *k*. (c) Number of scaffolds larger than 199 bp as a function of *k*. (d) Total genome size (scaffold smaller than 200bp removed) as a function of *k*.

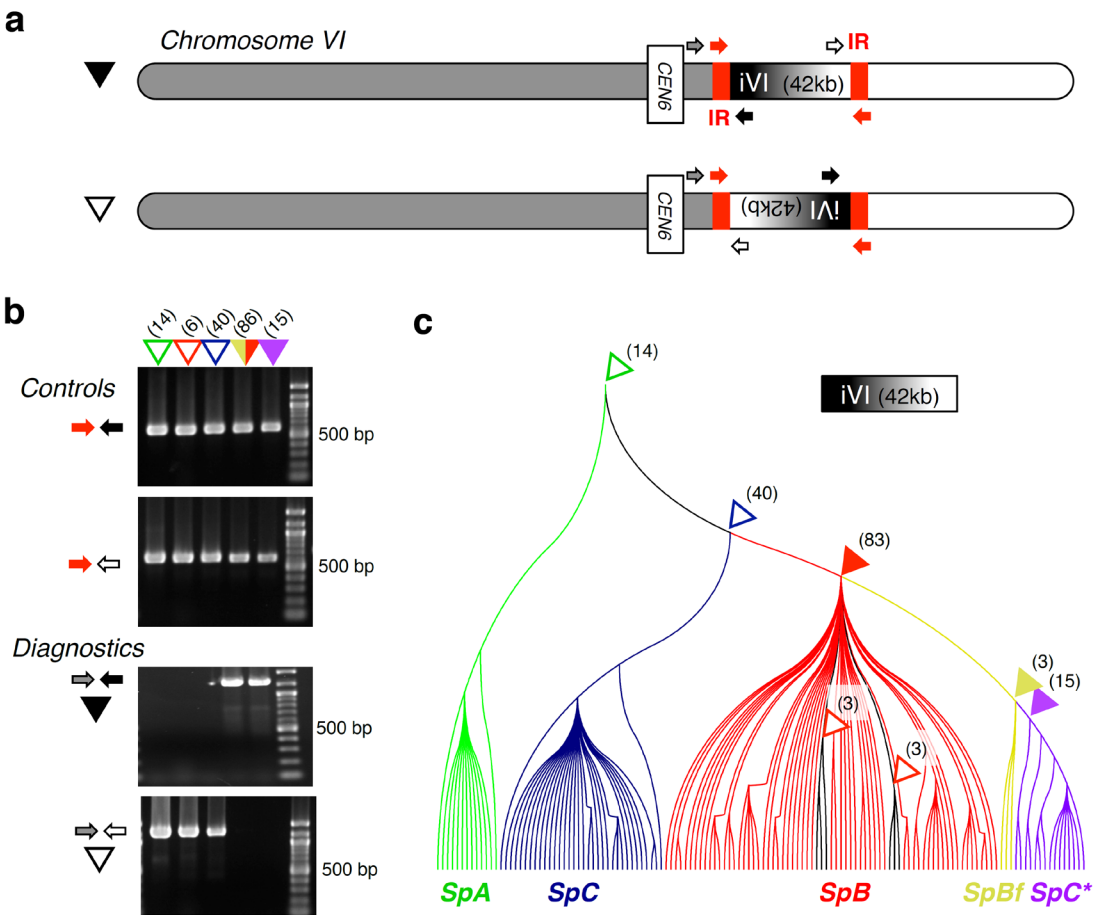

a
Chromosome VI
CEN6
iVI (42kb)
IR
IR
CEN6
iVI (42kb)
b
(14)
(6)
(40)
(86)
(15)
Controls
500 bp
500 bp
Diagnostics
500 bp
500 bp
c
iVI (42kb)
(14)
(40)
(83)
(3)
(15)
(3)
(3)
SpA
SpC
SpB
SpBf
SpC*

Figure I.S6: PCR-based validation and identification of a 42kb inversion on chromosome VI (iVI) in 153 *S. paradoxus* strains.
(a) Position of the iVI region in chromosome VI, flanked by two ~150bp inverted/repeated (IR) regions (red). "Ancestral" configuration (no inversion) observed in *SpA* and *SpC* (top; full diamond); "derived" configuration (inversion) observed in most *SpB* and *SpC** (bottom; empty diamond). Positions of oligonucleotides for diagnostic PCR (Table I.1), shown by arrows. (b) Example of iVI diagnostic in 5 *S. paradoxus* strains from lineage *SpA* (green), *SpB* (red; iVI not present in this case), *SpC* (blue), *SpB* or *SpBf* (red and yellow; when iVI is present) and *SpC** (purple). The number of each case is indicated in parenthesis. The two first panels are positive controls with control oligonucleotides located in the IR region (CLOP72-B12, red). The two last panels are diagnostic for the inversion with the anchor primer located upstream iVI in chromosome VI (CLOP72-B6, grey). The ancestral configuration (no inversion, black diamond) is revealed by amplification with oligonucleotides CLOP72-B2 (black arrow; third panel). The inversion (empty diamond) is revealed by amplification with oligonucleotides CLOP72-B7 (empty arrow; fourth panel). (c) Presence (full diamonds) or absence (empty diamonds) of iVI displayed on the iVI phylogeny in 153 *S. paradoxus* strains after diagnostic PCR (results detailed in Table I.S1). The number of each case is indicated in parenthesis. The iVI inversion was lost at least twice in the *SpB* lineage, as attested by its absence in six strains that from two monophyletic groups (black clusters and red, empty diamonds). Evolutionary tree based on the number of nucleotide difference (1135 positions, complete deletion) in 134 nucleotide sequences (strains with >50% missing positions were removed). Only nodes supported by at least 90% of bootstrap tests (1000 replicates, Neighbor-Joining method) are displayed.

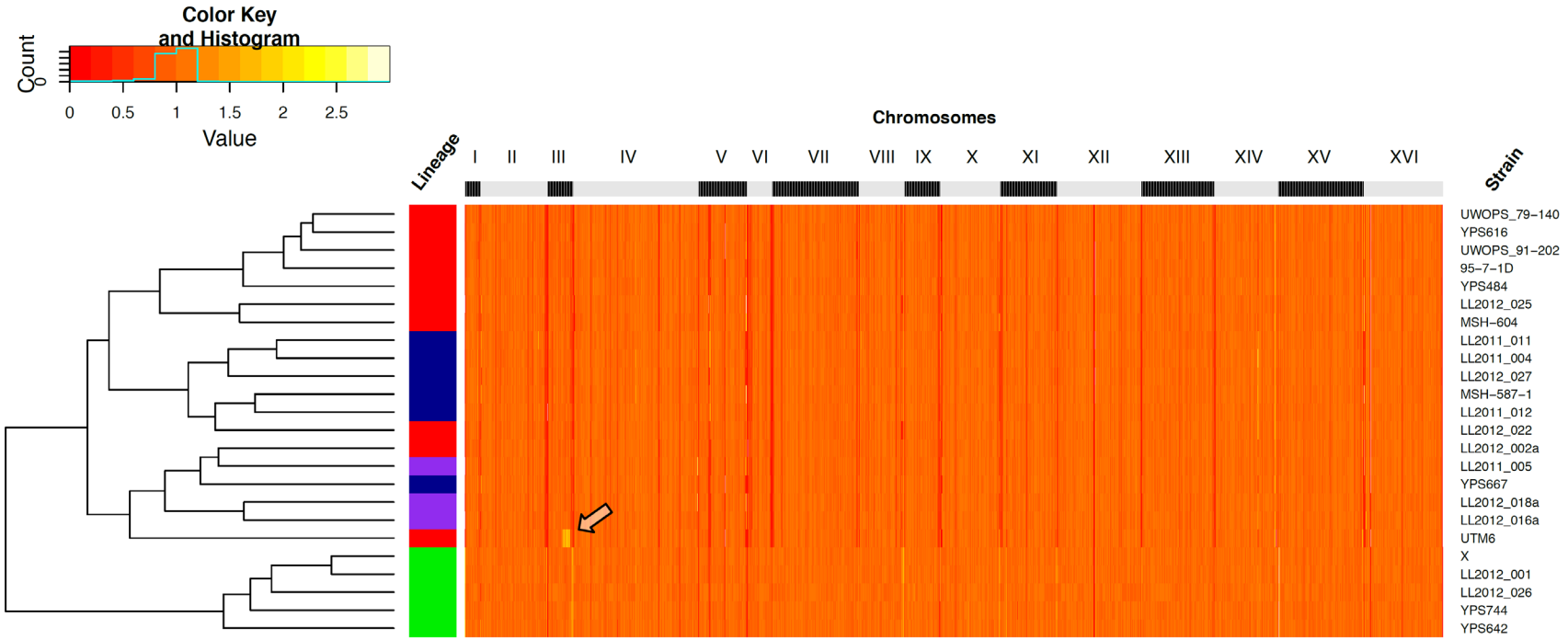


Figure I.S7: Most natural strains of *Saccharomyces paradoxus* are euploid (HC genomes).
Average depth of coverage in discrete 5000 bp windows along the genome reveals few cases of aneuploidy in the 24 high-coverage genome (HC) of *S. paradoxus* strains. For each HC library (TrueSeq Illumina®), sequencing depth (*D*) was normalized by the median calculated over the whole genome and windows containing highly repeated regions such as telomeric regions or rRNA genes on chromosome XII ($D > 3$) were removed (scale on top left). Yellow regions indicate putative duplications (e.g. right arm of chromosome III for strain UTM6: arrow) and red regions indicate putative deletions or removed regions, as compared to the *S. paradoxus* reference genome CBS432 (Liti *et coll.* 2009). Color code for lineages: *SpA*: green; *SpB*: red; *SpC*: blue; *SpC**: purple. The tree indicates coverage similarities among strains.

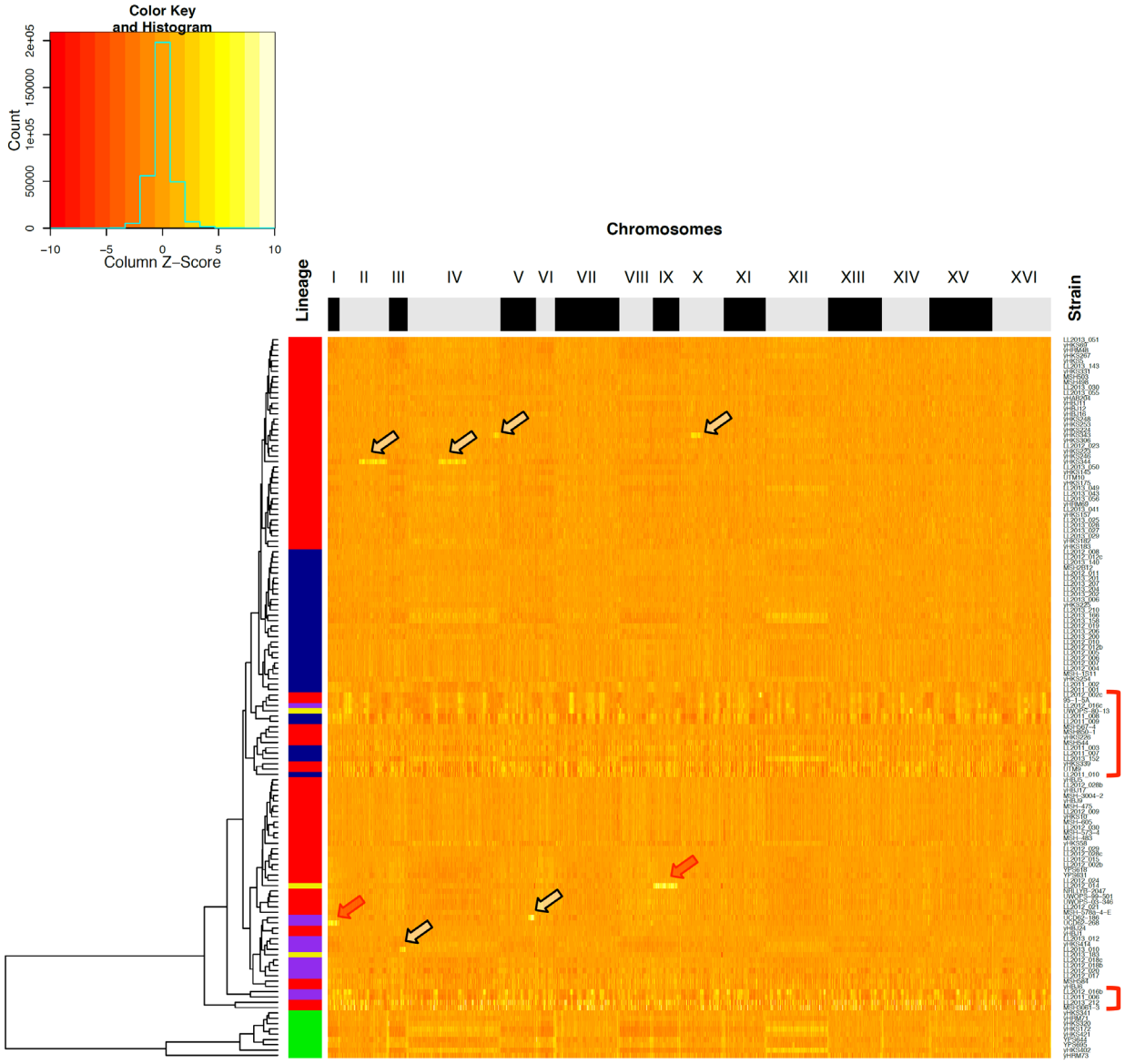

Color Key
and Histogram
Count
0
50000
1e+05
150000
2e+05
−10
−5
0
5
10
Column Z−Score
Chromosomes
Lineage
I
II
III
IV
V
VI
VII
VIII
IX
X
XI
XII
XIII
XIV
XV
XVI
Strain

Figure I.S8: Most natural strains of *Saccharomyces paradoxus* are euploid (LC genomes).

Average depth of coverage in discrete 5000 bp windows along the genome reveals few cases of aneuploidy in the 137 low-coverage (LC) genomes of *S. paradoxus*. For each library, depth of coverage (*D*) was first normalized by a randomly chosen LC library (LL2013_205) and then normalized by the median calculated over the whole genome. Windows containing highly repeated regions such as telomeric regions or rRNA genes on chromosome XII ($D > 3$) were removed (scale on top left). Overall low coverage for some strains (red brackets) leads to variation in coverage along the genome and prevents the identification of ploidy variation. Yellow regions indicate putative segmental (white arrows) or total chromosomal duplications (red arrows) and red regions indicate deletions or removed regions, as compared to the *S. paradoxus* reference genome CBS432 (Liti *et coll.* 2009). Color code for lineages: *SpA*: green; *SpB*: red; *SpC*: blue; *SpC**: purple; *SpBf*: yellow. The tree indicates coverage similarities among strains.

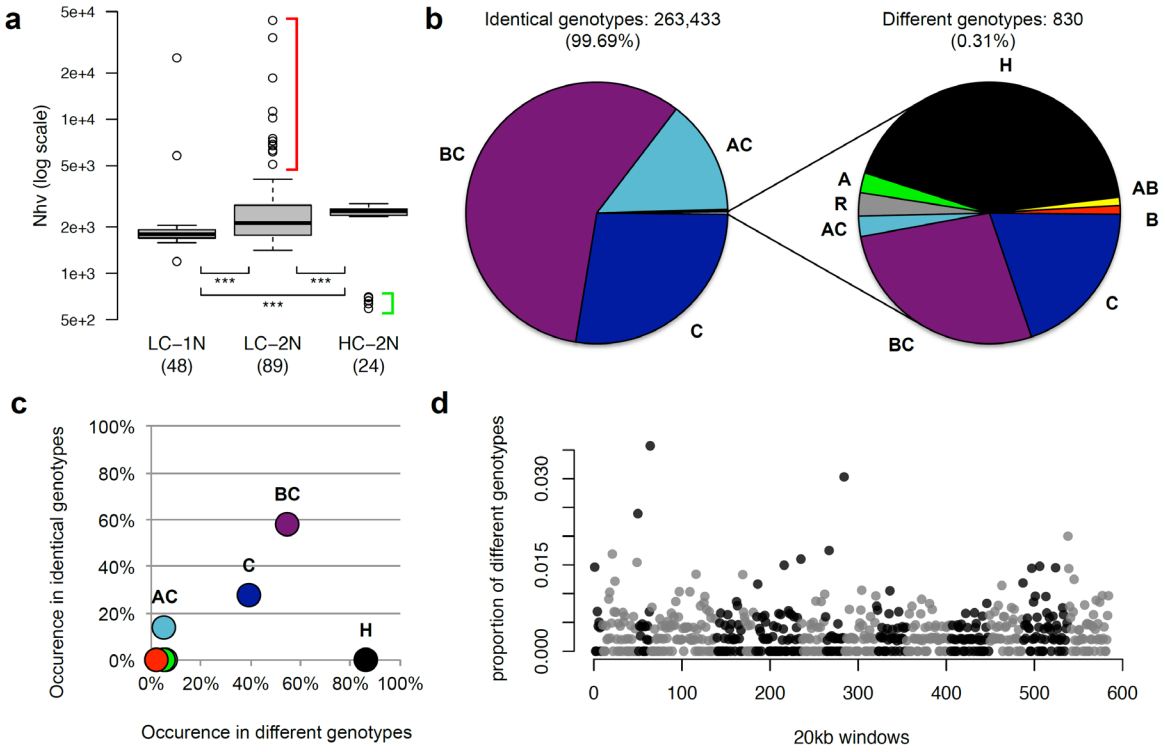

a
Nhv (log scale)
5e+4
2e+4
1e+4
5e+3
2e+3
1e+3
5e+2
***
***
***
LC−1N
(48)
LC−2N
(89)
HC−2N
(24)
b
Identical genotypes: 263,433
(99.69%)
Different genotypes: 830
(0.31%)
BC
AC
C
H
A
R
AC
AB
B
C
BC
c
Occurence in identical genotypes
100%
80%
60%
40%
20%
0%
BC
C
AC
H
0%
20%
40%
60%
80%
100%
Occurence in different genotypes
d
proportion of different genotypes
0.030
0.015
0.000
0
100
200
300
400
500
600
20kb windows

Figure I.S9: Heterozygosity is low in natural diploid *S. paradoxus* strains and SNP calling errors are mostly caused by false heterozygous sites.

(a) The number of heterozygous sites (*Nhv*; log scale) detected in *S. paradoxus* is higher in diploid (2N) than in haploid (1N) strains, as expected ("***": $p < 0.001$; Kruskal-Wallis rank sum test), suggesting low but significant heterozygosity in natural diploid strains ($Nhv \approx 2.5\times10^3$). Few individuals exhibit slightly higher levels of heterozygosity ($Nhv > 4.1\times10^3$; red bracket) that is evenly distributed along the genome (data not shown). *Nhv* is higher in HC than in LC libraries ($Nhv \approx 1.8\times10^3$; $p < 0.001$) suggesting that lower ($D<60\times$) and non-uniform depth of coverage along the genome (i.e. biased by tagmentation-induced coverage variation) in LC libraries may bias the estimation of *Nhv*. Five HC strains corresponding to *SpA* strains have almost null *Nhv* ($Nhv \approx 6\times10^2$; green bracket), which could reflect the isogenicity of *SpA* strains in North America (Kuehne *et coll.* 2007), or results from mapping on *SpA*-like reference genome CBS432 that induces overestimation of *Rh* in non-*SpA* strains, or underestimation of *Rh* in *SpA* strains. Error bars indicate interquartile ranges. (b) Per-site comparison of genotypes between two LC libraries independently obtained from one spore of the strain LL2012_012. The left pie chart represents the repartition of sites that are polymorphic for at least two strains of our dataset, but identical between the two replicated libraries (99.69% of sites). The right pie chart represents the remaining proportion of sites (0.31%) that are different between the replicated libraries and thus represent calling errors. Colors indicate genotype calls: A (green), B (red), C (blue), AB (yellow), AC (cyan), BC (purple), rare genotypes "R" (grey) of heterozygous sites "H" (black). (c) Heterozygous genotypes are overrepresented in sites assigned to different genotypes between the two replicated libraries, indicating that heterozygous sites mainly correspond to SNP calling errors. (d) SNP calling error is evenly distributed along the genome (proportion of different genotypes between replicated libraries per 20kb window; the 16 chromosomes are identified by switching grey and black).

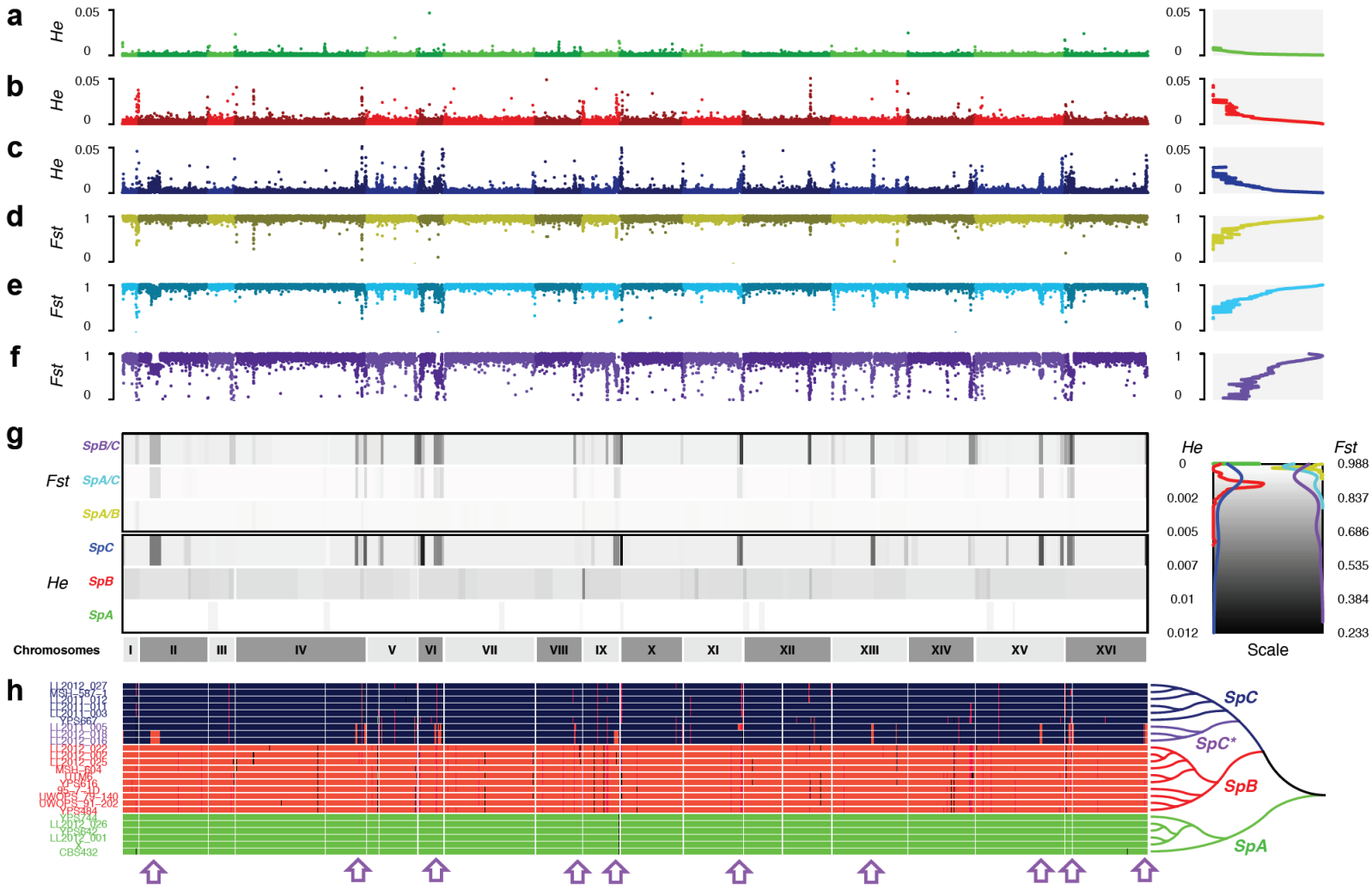

a
b
c
d
e
f
g
h
He
Fst
0.05
0
1
SpB/C
SpA/C
SpA/B
SpC
SpB
SpA
Chromosomes
I
II
III
IV
V
VI
VII
VIII
IX
X
XI
XII
XIII
XIV
XV
XVI
0.002
0.005
0.007
0.01
0.012
0.988
0.837
0.686
0.535
0.384
0.233
Scale
SpC
SpC*
SpB
SpA

Figure I.S10: Large genomic regions of high diversity (He) within the *S. paradoxus* lineage *SpC* and low differentiation (FST) between lineages *SpB* and *SpC* suggest introgressions from *SpB* in three *SpC* (*SpC) strains with HC genomes.**

(a) *He* measured in 5 *SpA* HC libraries in 500bp discrete sliding windows along the genome (distribution in the right panel). Null values indicate no polymorphism within a lineage. Chromosome numbers are indicated (see below panel g). (b) *He* in 10 *SpB* HC libraries. (c) *He* in 9 *SpC* HC libraries. (d) Divergence (F_{ST}) measured between *SpA* and *SpB* HC libraries in 500bp sliding windows along the genome (distribution in the right panel). Null values indicate no differentiation between lineages. (e) F_{ST} between *SpA* and *SpC* HC libraries. (f) F_{ST} between *SpB* and *SpC* HC libraries. (g) The *S. paradoxus* genome was split according to windows of low F_{ST}/high *He* (dark grey) and high F_{ST}/null *He* (white). Different shades of grey (scale and distribution in the right panel) indicate that two contiguous genomic regions (≥ 10kb) have different F_{ST} (top) or *He* (bottom) values at a highly conservative threshold ($p \leq 0.0001$; Welch two-sample *t*-tests). High *He* windows in *SpC* correspond to low F_{ST} between *SpA* and *SpC*, and especially between *SpB* and *SpC*. (h) Clustering analysis among 24 HC strains and the reference CBS432 using a phylogeny approach in 2kb discrete sliding windows along the genome. Details of the analysis are given in Figures I.S3 and I.S4. *SpB* (red) and *SpC* (blue) form well-separated lineages in most windows but in some well-delimited regions (purple arrows) for which three *SpC* strains (names in purple) arbor *SpB*-like regions, corresponding to windows of low F_{ST} between *SpB* and *SpC* and high *He* in *SpC*. The phylogeny at the bottom right is based on the TN93 model of DNA evolution. All nodes are at least 99% supported (approximate Likelihood-Ratio Test).

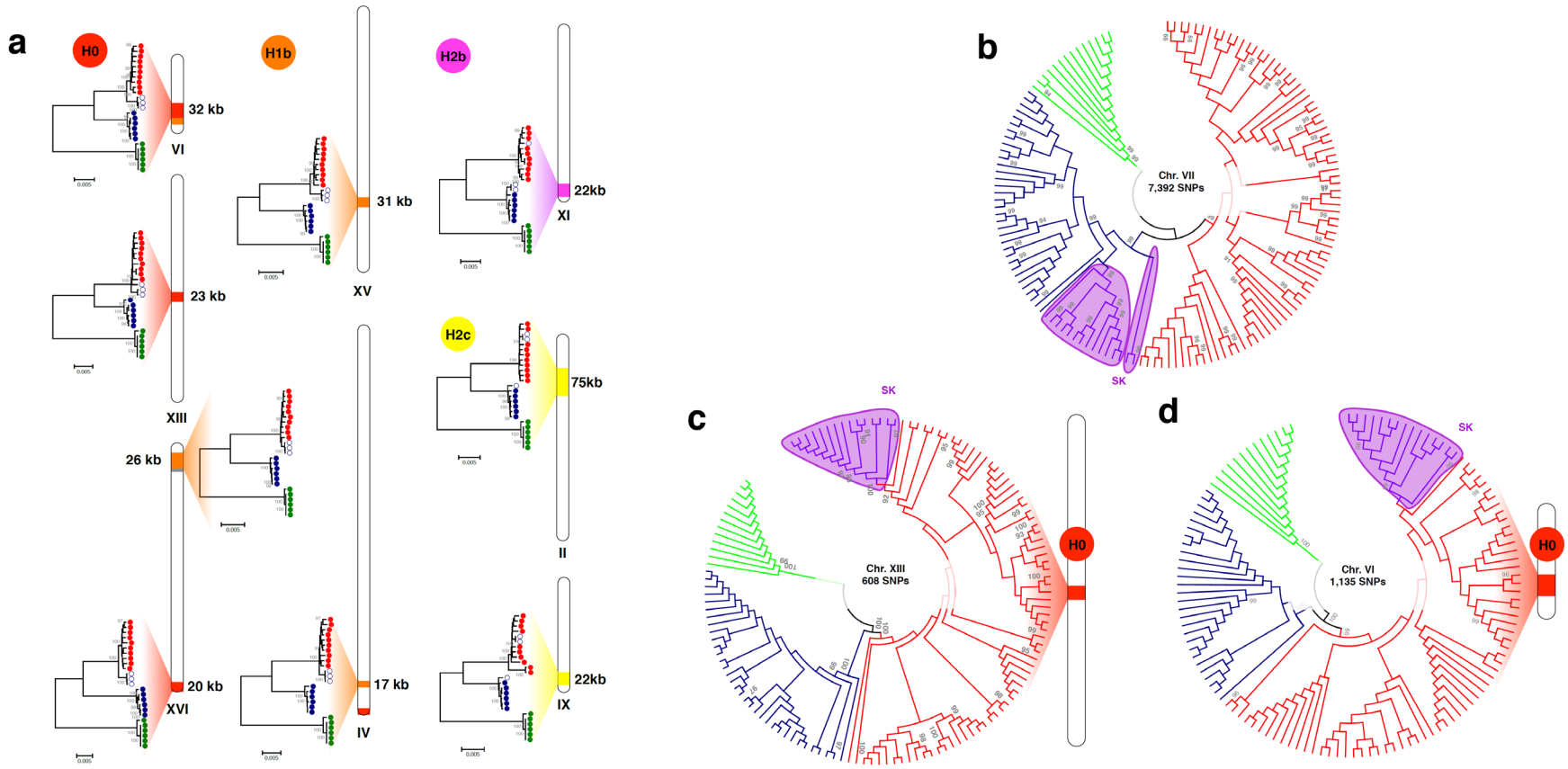

a
H0
32 kb
VI
23 kb
XIII
26 kb
20 kb
XVI
H1b
31 kb
XV
17 kb
IV
H2b
22kb
XI
H2c
75kb
II
22kb
IX
b
Chr. VII
7,392 SNPs
SK
c
SK
Chr. XIII
608 SNPs
H0
d
SK
Chr. VI
1,135 SNPs
H0

Figure I.S11: Detailed phylogeny in *SpB*-like regions.
Detailed phylogeny in *SpB*-like regions.
(a) Details on the phylogenies (24 HC strains) and chromosomal location of *SpB*-like regions. *SpC** strains are identified by open circles. (b) Control phylogeny using chromosome VII based on SNPs (no *SpB*-like regions) in 143 *S. paradoxus* strains (strains with >40% missing positions removed). (c), Phylogeny of the ancestral *SpB*-like region of chromosome XIII in 123 *S. paradoxus* strains (strains with >40% missing positions removed) showing that this region is derived for *SpC** and from strains within the *SpB* lineage. (d), Phylogeny from the ancestral *SpB*-like region of chromosome VI in 134 *S. paradoxus* strains (strains with >40% missing positions removed). *SpC** (without SK for chromosome VII phylogeny) form a monophyletic group (purple frames), regardless of its phylogenetic position (in *SpB* or *SpC*).

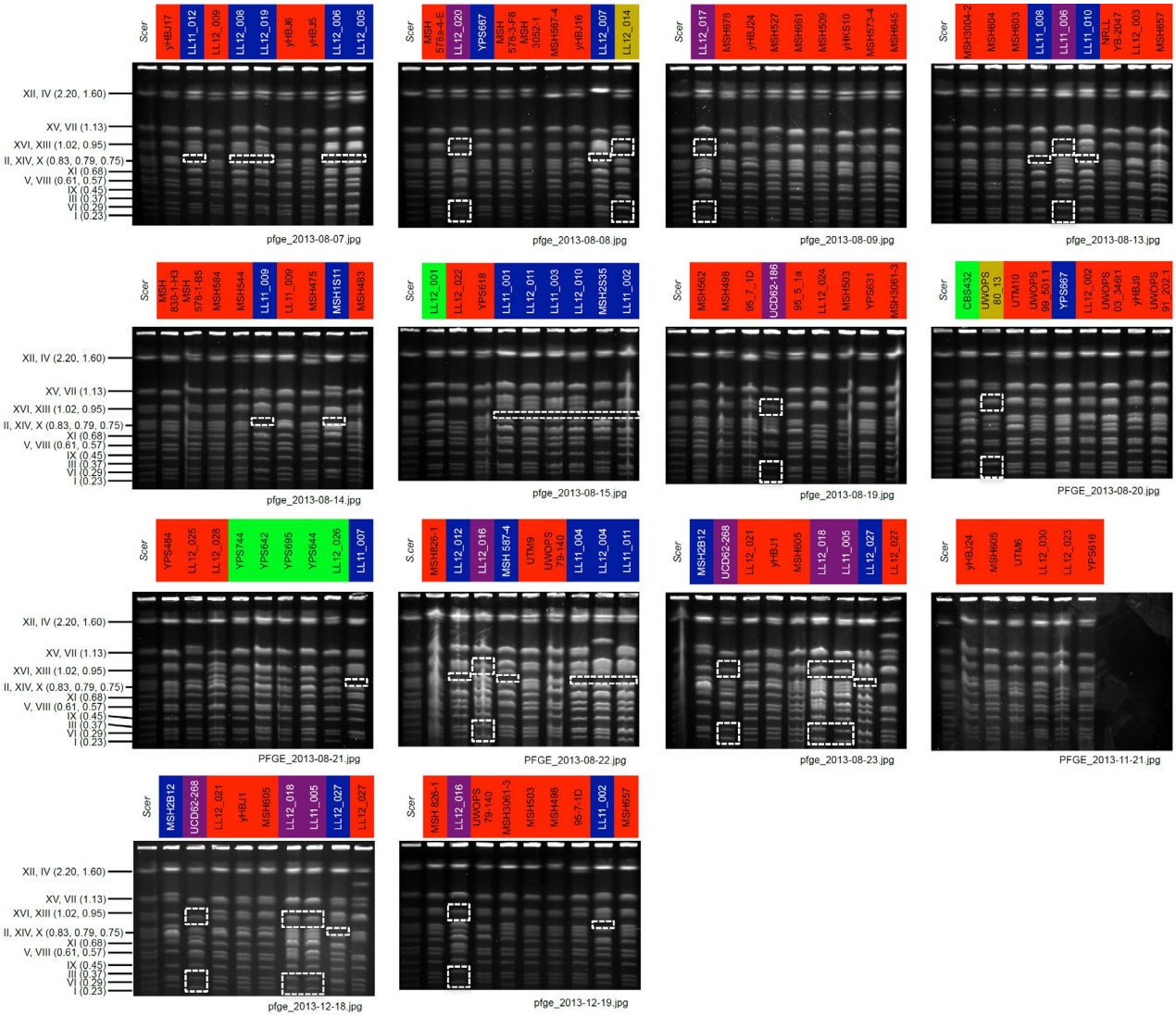


Figure I.S12: Pulse-field gel electrophoresis (PFGE) evidence for the VItXIIIr translocation.

Full PFGE profiles for 116 *S. paradoxus* strains from Charron *et coll.* (2014). The expected locations of chromosomes II (missing in 23 out of 25 *SpC* strains), VI, XIII (both missing in 8 *SpC** strains and 2 *SpBf* strains) are highlighted by dashed line boxes. The names of the strains are listed above each lane (*SpA*: green, *SpB*: red, *SpC**: purple, *SpBf*: yellow and *SpC*: blue). The first lane in each image is *S. cerevisiae* PFGE standard (*Scer*) with identification of each chromosome and their respective length in Mb. PFGE gels were replicated for some strains when the signal was not usable for analysis.

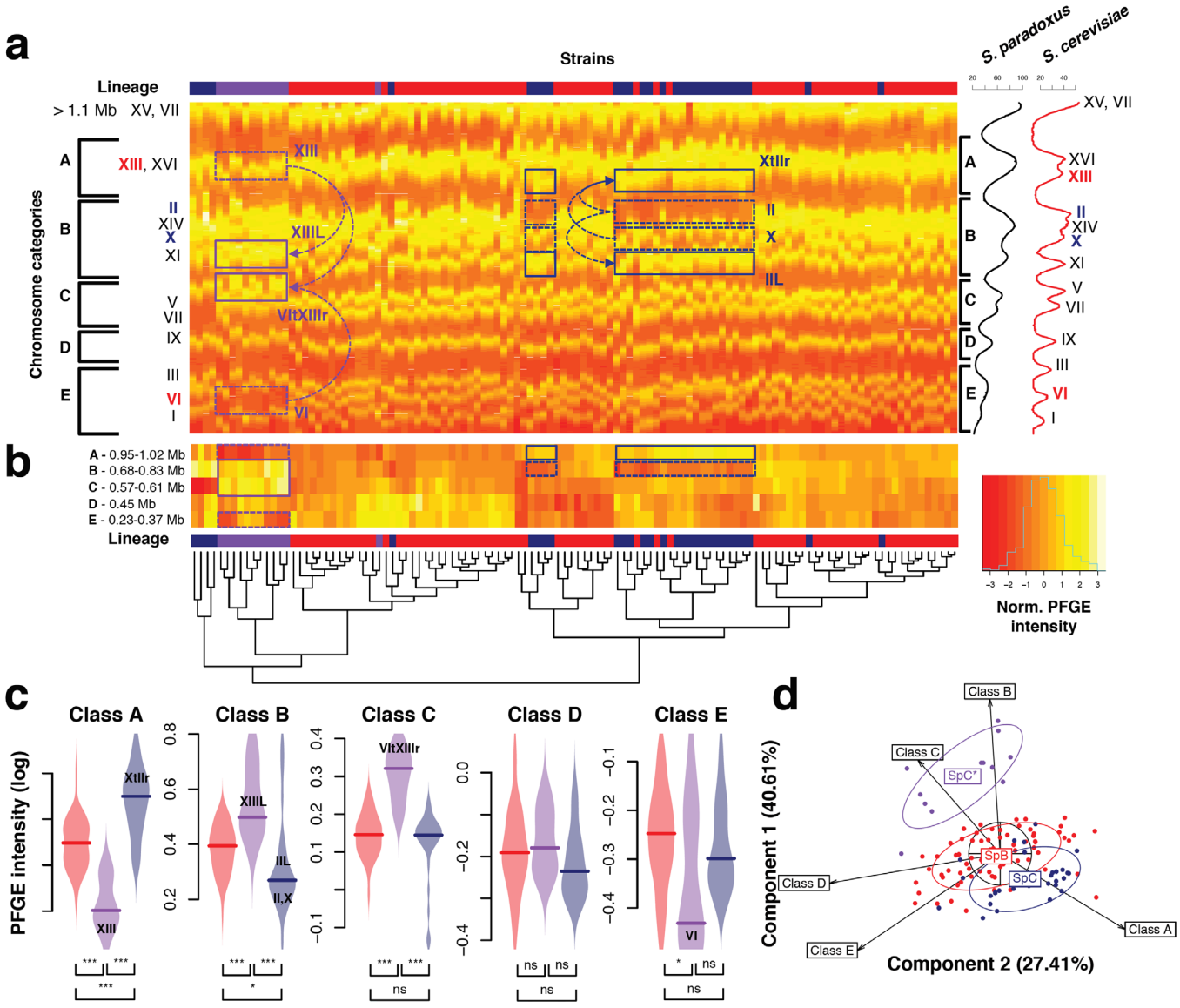

a
Strains
S. paradoxus
S. cerevisiae
Lineage
> 1.1 Mb XV, VII
Chromosome categories
A
B
C
D
E
XIII, XVI
II
XIV
X
XI
V
VII
IX
III
VI
I
XIII
XIIIL
VItXIIIr
VI
XIIIr
II
X
IIL
XV, VII
XVI
XIII
II
XIV
X
XI
V
VII
IX
III
VI
I
b
A - 0.95-1.02 Mb
B - 0.68-0.83 Mb
C - 0.57-0.61 Mb
D - 0.45 Mb
E - 0.23-0.37 Mb
Lineage
Norm. PFGE intensity
c
Class A
Class B
Class C
Class D
Class E
PFGE intensity (log)
XIII
XIIIr
XIIIL
IIL
II,X
VItXIIIr
VI
ns
d
Class B
Class C
SpC*
SpB
SpC
Class D
Class A
Class E
Component 1 (40.61%)
Component 2 (27.41%)

Figure I.S13: *S. paradoxus* lineages have contrasted karyotypic profiles as revealed by PFGE analysis that shows lineage-specific chromosomal changes.

(a), Chromosome profiles based on PFGE profiling (see Figure I.S12) in 116 *S. paradoxus* strains from lineages *SpB* (red), *SpC* (blue) and *SpC** (purple) were converted into pixel intensity profiles and band sizes were normalized according to a standard sample (*S. cerevisiae* S288C chromosomes; red curve on right panel). Unresolved bands corresponding to large chromosomes (> 1.1 Mb) were removed. Yellow indicates the presence of a band on PFGE profile; red indicates no band. Chromosome sizes in *S. paradoxus* were grouped in five size categories (A to E) to decrease chromosome size variation due to experimental noise. Dotted frames indicate different migration pattern for chromosomes VI and XIII in *SpC** (categories A and E, respectively; purple) and chromosomes II and X in *SpC* (category B; blue). Full frames indicate increased intensities corresponding to the likely positions of neo-chromosomes after putative translocations VItXIIIr in *SpC** and XtIIr in *SpC*. (b), Median pixel intensities per strain and per size categories. The tree is based on similarity among strains based upon the five size categories and show strong clustering of strains according to their lineages (*SpB*: red; *SpC*: blue; *SpC**: purple). (c), Violin plots of PFGE intensity distribution in each size category and each lineage. Tukey tests show significant differences between lineages for categories affected by chromosome gains or losses ($p<0.001$: "***"; $p<0.01$: "**"; $p<0.05$: "*"; $p>0.05$; "ns"). (d), Principal component analysis on PFGE intensities combined among all size categories shows that lineages can be discriminated based on their chromosomal profiles (components 1 and 2).

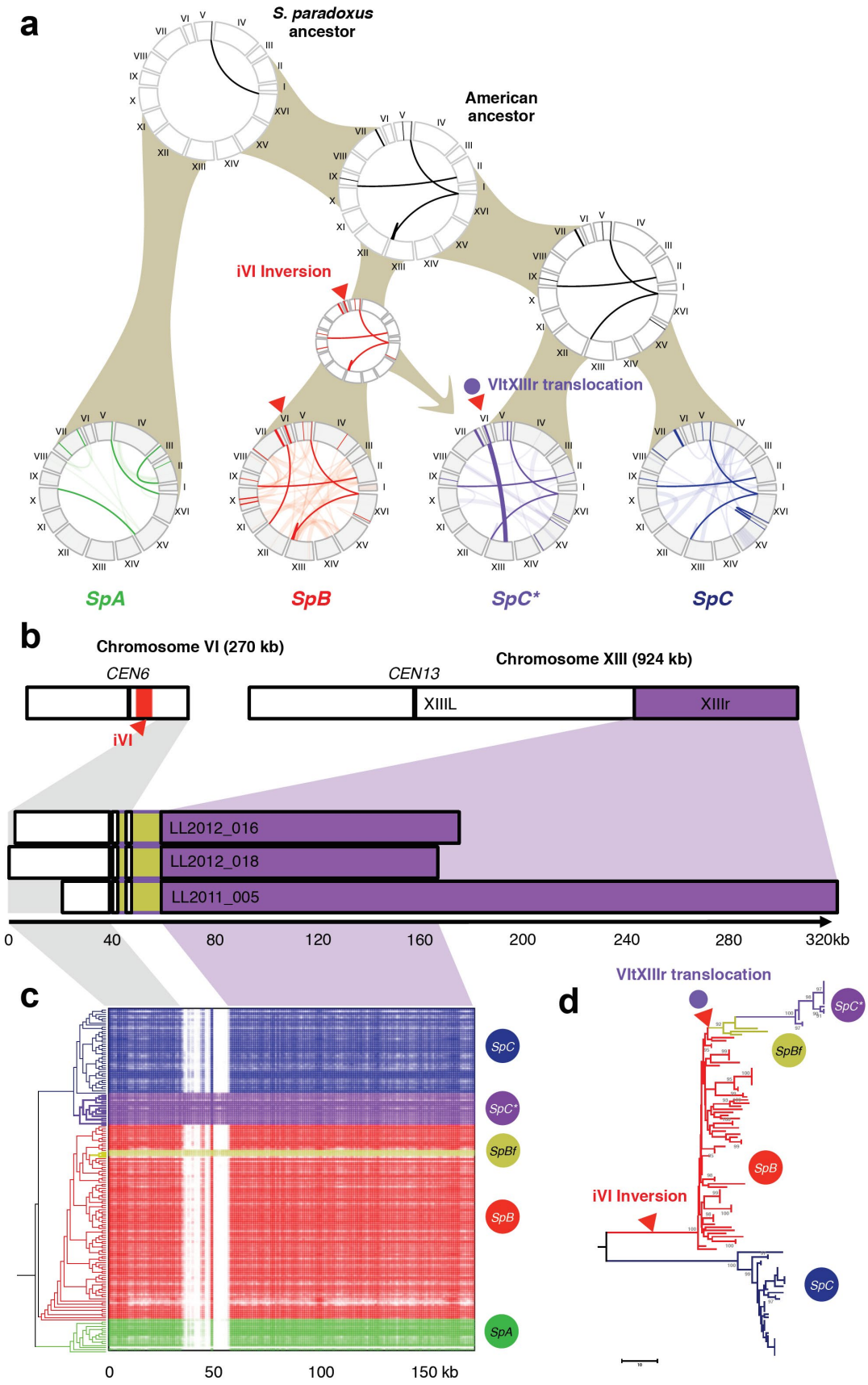

a
S. paradoxus ancestor
American ancestor
iVI Inversion
VltXIIIr translocation
SpA
SpB
SpC*
SpC
b
Chromosome VI (270 kb)
Chromosome XIII (924 kb)
CEN6
CEN13
XIIIL
XIIIr
iVI
LL2012_016
LL2012_018
LL2011_005
0
40
80
120
160
200
240
280
320kb
c
SpC
SpC*
SpBf
SpB
SpA
0
50
100
150 kb
d
VltXIIIr translocation
SpC*
SpBf
SpB
iVI Inversion
SpC

Figure I.S14: Chromosomal rearangements in *SpB* were inherited by *SpC during hybridization.**

(a), Lineage-specific karyotypes of 24 HC strains. Telomere exchange (thin lines), translocation (thick lines) and inversion (colored portions in chromosomes) as compared to *S. cerevisiae*, mapped on circular diagrams of the 16 chromosomes (I-XVI). Fixed rearrangements (dark) within lineages (colors) allow inferring ancestral states (black). A translocation (VItXIIIr) between chromosome VI and the right arm of chromosome XIII (XIIIr, 240kb) is fixed within *SpC** (purple circle) and absent elsewhere. A 42kb inversion in chromosome VI (iVI) was likely transmitted from *SpB* to *SpC** (red triangle) and is absent among the HC *SpC* genomes. (b), Chimeric scaffolds from *de novo* assembly reveal the fusion between chromosome VI and the right arm of chromosome XIII (VItXIIIr) in three HC *SpC** strains (LL2011_005, LL2012_016 and LL1012_018). The three chimeric scaffolds (bottom) were identified from MAUVE alignments onto the *S. cerevisiae* S288c genome and confirmed by blast against the NCBI database and *S. paradoxus* database from the SANGER institute (Liti *et coll.* 2009). Each of the three scaffolds is homologous (>97% of homology) with at least 20kb (40kb in LL2012_018) of the right arm chromosome VI (white, top left) and at least 115kb (260 kb in LL2011_005) of the right arm of chromosome XIII (XIIIr; purple, top right). The two regions are separated by a 22kb fusion region (yellow), for which no close homology (<75%) was identified. Centromere positions are indicated (*CEN6* and *CEN13*). (c), Coverage from mapping on a *SpC** chimeric scaffold (overlapping chromosomes VI and XIII) reveals that only *SpC** and three «*SpBf*» strains (yellow) possess the VItXIIIr translocation. Evolutionary tree from Figure 1b. (d), The iVI inversion corresponds to the ancestral *SpB*-like region of chromosome VI. A phylogeny of iVI (1135 polymorphic positions; 134 individuals) reveals that *SpBf* is at the root of *SpC**, suggesting that *SpC** inherited iVI and VItXIIIr from *SpBf*-like strains.

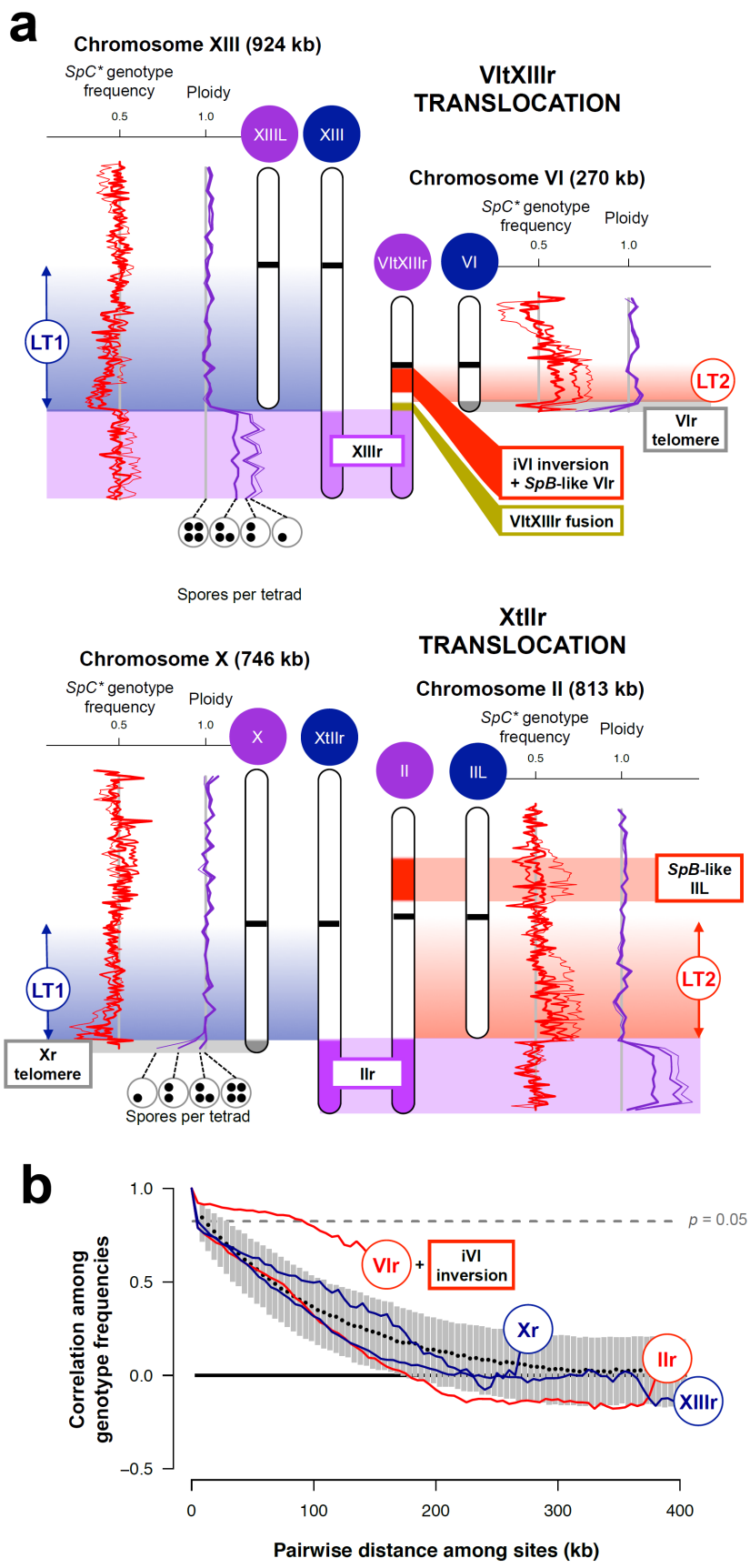

a
Chromosome XIII (924 kb)
SpC* genotype frequency
Ploidy
0.5
1.0
XIIIL
XIII
VltXIIIr TRANSLOCATION
Chromosome VI (270 kb)
VltXIIIr
VI
LT1
LT2
XIIIr
VIr telomere
iVI inversion + SpB-like VIr
VltXIIIr fusion
Spores per tetrad
XtIIr TRANSLOCATION
Chromosome X (746 kb)
Chromosome II (813 kb)
X
XtIIr
II
IIL
SpB-like IIL
Xr telomere
IIr
b
Correlation among genotype frequencies
1.0
0.5
0.0
−0.5
p = 0.05
VIr
+
iVI inversion
Xr
IIr
XIIIr
0
100
200
300
400
Pairwise distance among sites (kb)

Figure I.S15: Two translocations distinguishing strains from lineages *SpC* and *SpC contribute to their reproductive isolation.**
(a), *SpC** genotype frequency (red lines; 5 kb windows) and ploidy (purple line; 20kb windows) in *SpC*×*SpC** hybrid progeny plotted along the maps of chromosomes involved in translocations VItXIIIr (top) and XtIIr (bottom). The VItXIIIr translocation was identified by *de novo* genome assembly (Figure I.3b) and validated by PFGE (Figure I.S13) and re-mapping of reads (Figure I.S14c). Lineage *SpC* (blue circles) harbors ancestral topologies of chromosomes VI and XIII. Lineage *SpC** (purple circle) harbors the translocation. Regions separating centromeres to the translocated region are denoted as LT regions. Ploidy increase in the translocated region (XIIIr; purple frame) indicates that surviving spores inherit this region more than expected under a normal 2:2 segregation. Ploidy decreases with spore viability per tetrad (*n*=1-3; proportional to line width). The inverse pattern in the right telomere of chromosome VI (20kb; grey frame; absent from VItXIIIr) indicates that it is preferentially lost in spores. *SpC** genotypes in the *SpB*-like region of chromosome VI (red frame) and the iVI inversion (pink frame; see Figure I.5b) are preferentially transmitted in surviving spores (LT1). The inverse pattern in chromosome XIII (blue frame) indicates that *SpC* genotypes are more frequently transmitted (LT2). Biases in genotype frequencies in LT regions follow the over-transmission of XIIIr, indicating that they are co-segregating with the translocation due to linkage desequilibrium. The putative XtIIr translocation is supported by PFGE (Figure I.S13) and shows highly similar patterns of segregation bias in ploidy (IIr region; purple) and *SpC** genotype frequency (LT regions) to that of VItXIIIr. The only exception is the *SpB*-like region of chromosome II (IIL) that is over-transmitted in surviving spores but is distant from the LT region, suggesting other mechanisms than chromosomal changes. Lineage *SpC** (purple circle) harbors ancestral topologies of chromosomes II and X. Lineage *SpC* (blue circles) harbors the putative translocation. (b), Co-segregation extent in *SpC*×*SpC** hybrid progeny assessed by association (Pearson correlation test) of *SpC** genotypes frequency between distant regions of the genome (5kb windows). Black points and grey bars represent mean and distribution (interquartile range) of correlation coefficients for each pairwise distance class, respectively. The grey dotted line indicates that correlation coefficients are significant above $r = 0.82$ ($p \leq 0.05$, Bonferroni correction). LT regions (red and blue lines) follow the genomic background, excepted for VIr where the iVI inversion is likely to reduce recombination.

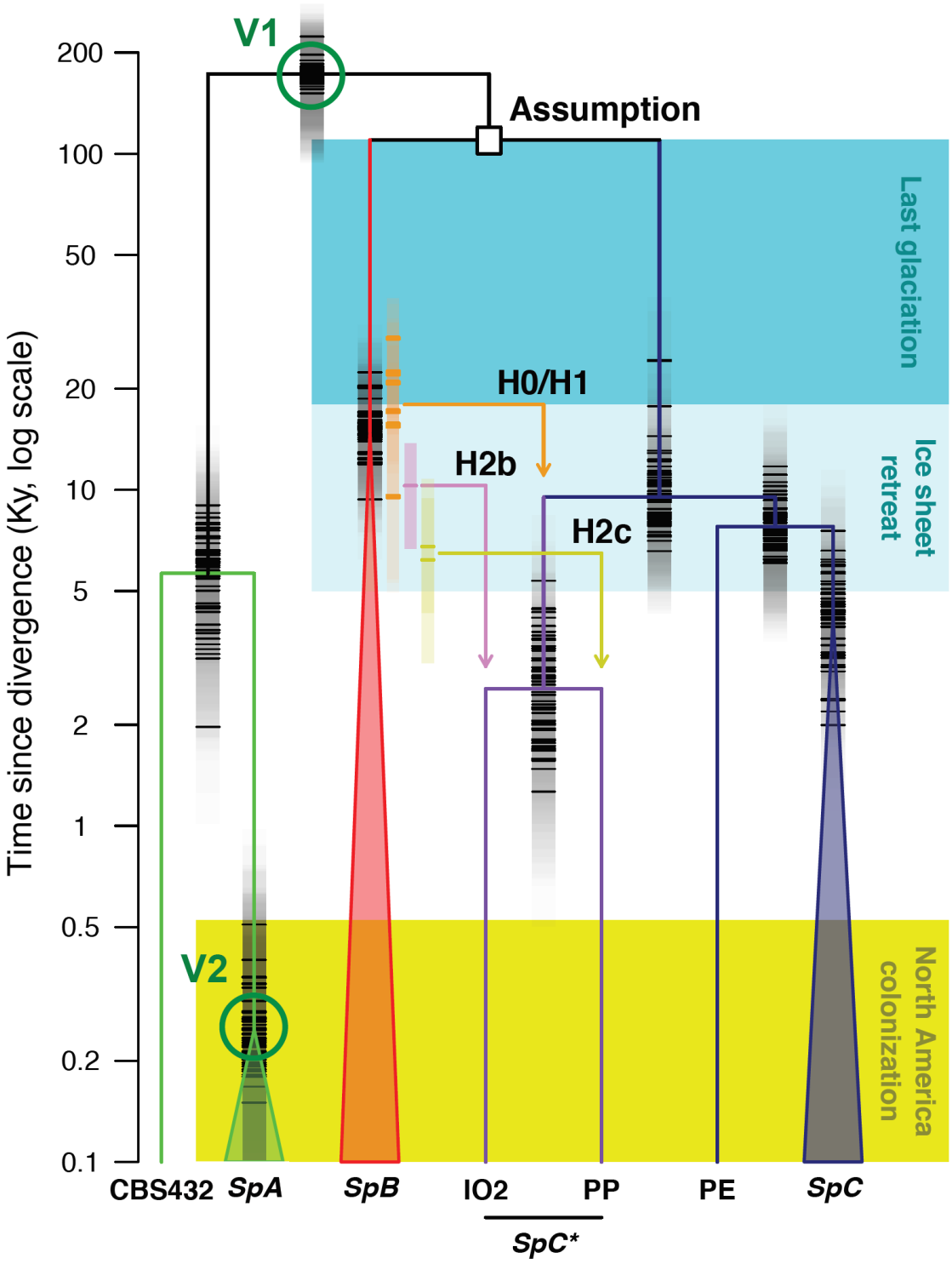

V1
Assumption
Last glaciation
Ice sheet retreat
H0/H1
H2b
H2c
V2
North America colonization
Time since divergence (Ky, log scale)
200
100
50
20
10
5
2
1
0.5
0.2
0.1
CBS432
SpA
SpB
IO2
PP
PE
SpC
SpC*

Figure I.S16: The initial hybridization giving rise to *SpC dates after the last glaciation about 10,000 years ago.**

Divergence times (Kya; log scale) among, and radiation within, evolutionary categories (CBS432, *SpA*, *SpB*, *SpC*, PE, PP, IO2, Table S6) were estimated using BEAST with the 24 HC genomes sliced according to known introgressions in *SpC** strains, and assuming that *SpB* and *SpC* diverged at the beginning of last glaciation (cyan frame), 110,000 years ago. Black bars and grey frames represent the median and standard deviation of time estimated from 53-50kb non-introgressed (control) regions. Colored bars and frames represent the time estimated from nine *SpB*-like regions that are either shared (H0, H1) or not (H2) among all *SpC** types (see Tables I.6 and I.9). Green circles indicate validation time points that are coherent with previous studies, corresponding to the divergence between European (*SpA*) and American (*SpB*+*SpC*) lineages (V1) and to the introduction of the *SpA* lineage in North America (V2). These confirmed that the estimated time of initial divergence between *SpB* and *SpC* is accurate.

I.8-TABLES

Table I.1: List of oligonucleotides for genotyping inversion iVI.
(See Figure I.S6)

Primer name	Sequence	Description
CLOP72-B12	AGTTGGTTTAAGGCGCAAGA	Control in IR region
CLOP72-B2	TCGATCGTGTTGTCGAGGTA	Configuration B/C*
CLOP72-B6	CAATGGTTTGCCCTTATTGG	Anchor in chr6
CLOP72-B7	GCAGGAAAAAGAATCCAAAGC	Configuration A/C

Table I.2: ANOVA (GLM) indicate that lineage is the main factor explaining variance among *S. paradoxus* strains.

For growth-assays under different temperatures (log of colony size), limitation to 15 carbon and nitrogen sources (three main axes of PCA on log of colony size) and fraction of survival to a freeze-thaw cycle. Deviance explained by each factor and their interaction. In bold are the two main factors explaining deviance (Lineage in most cases). Symbols of significance: $p<0.001$: "***"; $p<0.01$: "**"; $p<0.05$: "*"; $p<0.1$: "°"; $p>0.1$; "ns".

		Latitude	Longitude	**Lineage**	Lat*Long	Lat*Lin	Long*Lin	Lat*Long*Lin
Temperature	10°C	**0.01***	0.00°	**0.05*****	0.00	0.00	0.00	0.00
	15°C	0.00	**0.02*****	**0.02****	0.00	0.00	0.00	0.00
	20°C	0.01*	0.01*	**0.14*****	0.01*	**0.02****	0.00	0.01
	25°C	0.00	0.00	**0.22*****	0.00°	**0.03*****	0.01	0.00
	30°C	0.00	0.00	**0.17*****	0.00°	**0.04*****	0.00	0.02**
	35°C	**0.31*****	0.01	**1.54*****	0.05**	0.09**	0.04°	0.11***
Freeze-thaw survival		**3.36*****	0.02	**0.91****	1.52**	0.19	0.03	0.01
Nutrients (PCA)	Axis 1	**106.02*****	**52.94****	1.40	55.52**	45.92*	22.61	24.30
	Axis 2	**79.26*****	32.59***	**365.43*****	2.03*	0.50	2.25	6.06**
	Axis 3	**16.76*****	1.95	**100.25*****	0.70	0.89	1.47	4.72

Table I.3: Tukey test on growth assays under different temperatures (log of colony size) limitation to 15 carbon and nitrogen sources (three main axes of PCA on log of colony size) and fraction of strains surviving to a freeze-thaw cycle.

Differences in mean fitness between lineages are shown. Symbols of significance: $p<0.001$: "***"; $p<0.01$: "**"; $p<0.05$: "*"; $p<0.1$: "°"; $p>0.1$; "ns".

		SpA-SpB	*SpA-SpC*	*SpA-SpC**	*SpB-SpC*	*SpB-SpC**	*SpC-SpC**
Temperature	10°C	-0.045***	-0.012	-0.041°	0.033***	0.004	-0.029
	15°C	-0.030°	-0.010	-0.023	0.020*	0.008	-0.013
	20°C	-0.097***	-0.057***	-0.096***	0.040***	0.002	-0.038*
	25°C	-0.119***	-0.061***	-0.091***	0.057***	0.028	-0.030
	30°C	-0.081***	-0.019	-0.094***	0.062***	-0.013	-0.075***
	35°C	-0.248***	-0.038	-0.208***	0.210***	0.041	-0.170***
Freeze-thaw survival		-0.398	0.085	0.020	0.483***	0.418*	-0.065
Nutrients (PCA)	Axis 1	-0.284	-0.941	-1.305	-0.658	-1.021	-0.363
	Axis 2	2.495***	-1.265***	0.802	-3.761***	-1.693***	2.067***
	Axis 3	3.162***	3.332***	2.001***	0.170	-1.161**	-1.330**

Table I.4: List of heterothallic strains constructed for the controlled crosses performed in this study.

The following information is shown: *S. paradoxus* strain name, sampling location, reference, type of construction performed according to the cassette resistance used to delete the *HO* gene (*KAN*, *NAT*, *HPH*) and the mating type of the haploid strain (*MATa* or *MATα*) and the lineage of the strain.

Strain	Sampling site	Reference	Construction used			Lineage
			ho::KAN	*ho::NAT*	*ho::HPH*	
LL2012_014	St Michel de Bellechasse, QC	Charron et al. (2014)	-	*α*	*a*	*SpBf*
LL2013_183	Oak Point Provincial Parc, NB	This study	*α*	*a, α*	-	*SpBf*
LL2011_005	Cap Chat, QC	Leducq et al. (2014)	-	*α*	*a, α*	*SpC**
LL2011_006	Ile d'Orléans, QC	Leducq et al. (2014)	*α*	-	*a, α*	*SpC**
LL2011_007	Ile d'Orléans, QC	Charron et al. (2014)	*α*	*α*	*a*	*SpC*
LL2011_008	Ile d'Orléans, QC	Charron et al. (2014)	-	*α*	*a*	*SpC*
MSH-587-1	Mont St Hilaire, QC	Leducq et al. (2014)	-	*a, α*	*a*	*SpC*
MSH-604	Mont St Hilaire, QC	Leducq et al. (2014)	-	*a, α*	*a*	*SpB*
yHKS225	Plainfield, NH	Leducq et al. (2014)	-	*a, α*	*α*	*SpC*
yHKS226	Plainfield, NH	Sylvester et al. (2015)	*α*	-	*a, α*	*SpB*
LL2012_021	Pointe Platon, QC	Leducq et al. (2014)	-	*α*	*a, α*	*SpB*
LL2012_016	Pointe Platon, QC	Leducq et al. (2014)	*α*	-	*a, α*	*SpC**
LL2012_018	Pointe Platon, QC	Leducq et al. (2014)	-	*α*	*a*	*SpC**
LL2012_019	Pointe Platon, QC	Charron et al. (2014)	-	*α*	*a*	*SpC*
LL2011_001	Québec, QC	Leducq et al. (2014)	*α*	-	*a, α*	*SpC*
LL2011_012	Station Dushesnay, QC	Charron et al. (2014)	*α*	-	*a*	*SpC*
LL2013_010	Québec, QC	This study	-	*a, α*	*a, α*	*SpC**
LL2013_012	Québec, QC	This study	-	*a, α*	-	*SpC**
LL2012_028	Sherbrooke, QC	Leducq et al. (2014)	-	*a, α*	*a*	*SpB*
LL2012_027	Sherbrooke, QC	Charron et al. (2014)	-	-	*α*	*SpC*
yHBJ1	Iron River, WI	Leducq et al. (2014)	*α*	*a, α*	-	*SpB*
yHKS414	Sheboygan,WI	Sylvester et al. (2015)	-	-	*a, α*	*SpC**

Table I.5: Post-zygotic reproductive success (RS) within and among lineages of *S. paradoxus*.

Crosses were separated in three categories: selfing, outcrossing within lineages and outcrossing among lineages. Statistics are reported in the five columns on the left: lineage(s), number of crosses (*n*), median of RS, 3rd and 1st quartiles. **a,** Results for selfing and statistics from a Tukey test comparing RS values within lineages only indicate no effect of selfing on RS. **b,** Results for outcrossing and statistics from a Tukey test comparing outcrossing RS values among and within lineages indicate mostly significant decrease of RS among lineages ($p<0.001$: "***"; $p<0.01$: "**"; $p<0.05$: "*"; $p<0.1$: "°"; $p>0.1$; "ns").

Lineage	RS				Selfing					Outcrossing within lineages					Tukey test
	n	Median	3rd Q.	1st Q.	*SpA*	*SpB*	*SpBf*	*SpC**	*SpC*	*SpA×SpA*	*SpB×SpB*	*SpBf×SpBf*	*SpC*×SpC**	*SpC×SpC*	
a															
Selfing															Selfing vs. outcrossing within lineage
SpA	1	79.7%	-	-						ns					
SpB	5	92.7%	93.8%	66.7%	ns					ns	ns				
SpBf	2	72.9%	100.0%	45.8%	ns	ns				ns	ns	ns			
*SpC**	5	97.9%	99.0%	43.8%	ns	ns	ns			ns	ns	ns	ns		
SpC	8	97.9%	99.5%	97.4%	ns	ns	ns	ns		ns	**	ns	ns	°	
					Outcrossing within lineages					*Among lineages*					
					SpA×SpA	*SpB×SpB*	*SpBf×SpBf*	*SpC*×SpC**	*SpC×SpC*	*SpA×SpB*	*SpA×SpC*	*SpA×SpC**	*SpB×SpBf*	*SpB×SpC*	*SpB×SpC**	*SpBf×SpC*	*SpBf×SpC**	
b																		
Outcrossing within lineages																		Outcrossing within vs. between lineages
SpA×SpA	3	83.3%	84.6%	78.9%														
SpB×SpB	9	55.2%	71.9%	44.8%	ns													
SpBf×SpBf	1	85.4%	-	-	ns	ns												
SpC×SpC**	13	92.7%	94.8%	89.6%	ns	***	ns											
SpC×SpC	18	75.8%	91.7%	52.6%	ns	ns	ns	°										
Among lineages																		
SpA×SpB	8	43.8%	48.2%	34.9%	*	ns	ns	***	***									
SpA×SpC	6	33.4%	36.5%	27.1%	**	°	°	***	***	ns								
*SpA×SpC**	1	24.0%	-	-	°	ns	ns	**	ns	ns	ns							
SpB×SpBf	5	70.8%	79.2%	39.6%	ns	ns	ns	*	ns	ns	°	ns						
SpB×SpC	19	38.5%	49.0%	26.3%	***	°	ns	***	***	ns	ns	ns	*					
*SpB×SpC**	10	37.5%	45.8%	20.3%	***	°	ns	***	***	ns	ns	ns	*	ns				
SpBf×SpC	6	25.0%	31.3%	24.0%	***	**	*	***	***	ns	ns	ns	**	ns	ns			
*SpBf×SpC**	5	29.2%	39.6%	27.1%	**	ns	°	***	***	ns	ns	ns	°	ns	ns	ns		
*SpC×SpC**	20	44.3%	49.0%	39.1%	*	ns	ns	***	***	ns	ns	ns	ns	ns	ns	ns	ns	

Table I.6: Summary of *SpB*-like regions found in the genomes of *SpC** **strains.**

The following information is shown: the group and the name of *SpC** strains according to *SpB*-like regions identified in their genome (6 groups named according to the sample location: SK, QC, IO1, IO2, WI, PP); the total size of *SpB*-like regions and the proportion of the genome they represent; the size of the main *SpB*-like regions on each chromosome and the name of hypothetical hybridization events (H) during which these regions were acquired or lost by secondary backcrosses between *SpC** and parental lineages *SpB* or *SpC*.

Group (strains)	*SpB*-like regions				
	Total size (% genome)	Chromosomes (size)			
Saskatchewan (SK) (UCD62-186, UCD62-268)	260kb (2.2%)	**H0 (Ancestral)** IV (15kb) V (20kb)	**H1a** V (15kb) XI (25kb) XIV (30kb)		
Québec City (QC) (LL2013-010, LL2012-012)	605kb (5.2%)	VI (55kb) VIII (15kb) XIII (35kb) XVI (30kb)	**H1b** IV (20kb) VI (15kb) XIII (10kb) XV (35kb)	**H2a** XII (70kb)	**H3a** V (20kb) XII (125kb) XIV (75kb) XVI (20kb)
Île d'Orléans I (IO1) (LL2011-006)	680kb (5.8%)		XVI (35kb)		**H3b** III (195 kb) X (120 kb)
Wisconsin (WI) (yHKS414)	290kb (2.5%)			–	
Île d'orléans II (IO2) (LL2011-005)	325kb (2.8%)			**H2b** XI (35kb)	
Pointe Platon (PP) (LL2012_016, LL2012_017, LL2012_018, LL2012_020)	435kb (3.7%)			**H2c** II (105kb) IX (35kb)	

Table I.7: Description of 25 libraries for whole genome sequencing of 384 *S. paradoxus* strains from eight crosses between lineages *SpC* and *SpC*.*

The following information is shown: name of the Illumina Nextera XT® libraries, name of the *SpC** and *SpC* parental strains, type of tetrads from which strains (single spores) were isolated, number of spores pooled in the library and depth of coverage statistics after read mapping onto the CBS432 reference genome.

Library	*SpC**	*SpC*	Spore per tetrad	Spores	Median Depth (1st Q.-3rd Q.)
LIB-A01	LL2011_005	MSH587-1	1	7	103 (58-157)
LIB-A02			2	24	60 (35-90)
LIB-A03			3	3	57 (34-84)
LIB-A04			4	4	109 (66-158)
LIB-B01	LL2011_005	LL2011_001	1	5	90 (51-138)
LIB-B02			2	8	63 (37-96)
LIB-B03			3	6	59 (34-89)
LIB-C01	LL2011_005	LL2012_006	1	9	67 (37-105)
LIB-C02			2	12	53 (29-82)
LIB-C03			3	24	91 (51-145)
LIB-D01	LL2011_006	LL2012_011	1	5	82 (46-125)
LIB-D02			2	18	57 (32-87)
LIB-D03			3	24	74 (41-117)
LIB-E01	LL2011_006	LL2011_011	1	7	78 (45-118)
LIB-E02			2	20	57 (33-87)
LIB-E03			3	15	58 (33-89)
LIB-F01	LL2012_016	LL2011_002	1	4	133 (77-199)
LIB-F02			2	20	42 (24-64)
LIB-F03			3	18	54 (31-82)
LIB-G01	LL2012_016	LL2011_004	1	6	104 (58-165)
LIB-G02			2	14	78 (46-118)
LIB-G03			3	27	52 (30-80)
LIB-H01	LL2012_016	LL2011_009	1	3	116 (68-174)
LIB-H02			2	12	75 (45-112)
LIB-H03			3	30	94 (56-140)

Table I.8: Postglacial taxa in North-East America have biogeographical distributions similar with those of *S. paradoxus* lineages.

For each taxa or group of taxa, the locations of glacial refugia and of post-glacial secondary contact zones, when applicable, are indicated.

	Taxa	Refugium or populations	Secondary contact zone(s)	Reference
Freshwater fishes	*Rhinichthys obtusus, Rhinichtys abtratulus* *Etheostoma nigrum, Etheostoma olmstedi* *Pimephales promelas*	Atlantic, Mississippian	Great Lakes, south Quebec	April *et coll.* (2013)
Crustacea	*Mysis relicta*	Atlantic, Mississippian	Great Lakes	Dooh *et coll.* (2006)
Insect	*Drosophila athabasca*	Atlantic (2) North America	-	Ford et Aquadro (1996)
Conifers	*Picea glauca*	North-West, Mississippian, East-Appalachian	-	de Lafontaine *et coll.* (2010)
	Pinus banksiana	Atlantic	Central Quebec	Godbout *et coll.* (2005)
	Abies balsamea	Labrador, North America (4)	Great Lakes, Quebec	Cinget *et coll.* (2015)

Table I.9: Estimation of time since the radiation within, and divergence time among evolutionary categories (EC) from 24 HC genomes according to *SpB*-like regions identified in eleven (HC+LC) *SpC genomes.**

The following information is shown: the event that is dated (divergence between *SpB* and *SpC* is the reference, assuming that lineages diverged 110,000 years ago); EC that are used for estimation (only one if radiation is estimated); chromosomes that are used for estimation; size of the genomic region used for the estimation; mean and standard deviation of time estimation since the event, estimated in BEAST v1.8.1. PE corresponds to a single *SpC* strain from Pennsylvania and divergent from the core lineage (YPS667). IO2 (LL2011_005) and PP (LL2012_016 and LL2012_018) correspond to two *SpC** groups that are departing for *SpB*-like regions (Table I.6).

Event	EC 1	EC 2	Chromosomes	Genomic region	Time estimation (kilo years)	Nucleotide divergence
Reference regions (no introgression) in chromosomes I, II, VI, IX, XI, XIII, XV and XVI						
Divergence *SpB-C* (assumption)	***SpB***	***SpC, SpC****	**all**	**53 × 50 kb**	**110.0 ± 10.0**	**2,280 ± 0,013%**
Divergence *SpA-SpB* and *SpC*	*SpA, CBS432*	*SpB, SpC, SpC**	all	53 × 50 kb	176.2 ± 36.6	3,718 ± 0,137%
Divergence *SpA*-CBS432	*SpA*	CBS432	all	53 × 50 kb	6.0 ± 2.5	0,123 ± 0.000%
Radiation in *SpA* (w/o CBS432)	*SpA*	-	all	53 × 50 kb	0.3 ± 0.2	0,001 ± 0.000%
Radiation in *SpB*	*SpB*	-	all	53 × 50 kb	15.5 ± 4.0	0,202 ± 0,022%
Radiation in *SpC* (w/o PE)	*SpC* (w/o PE)	-	all	53 × 50 kb	4.4 ± 1.7	0,053 ± 0,015%
Divergence *SpC*-PE	*SpC* (w/o PE)	PE	all	53 × 50 kb	8.2 ± 2.3	0,128 ± 0,001%
Divergence *SpC-SpC**	*SpC*	*SpC** (PP, IO2)	all	53 × 50 kb	10.5 ± 4.3	0,188 ± 0,005%
Radiation in *SpC** (post-H1b)	PP	LL2011_005 (IO2)	all	53 × 50 kb	2.8 ± 1.2	0,035 ± 0,006%
Introgressed regions (SpB-like)						
Divergence of *SpB-SpC** (pre-H0)	*SpB*	*SpC**	XIII	23 kb	10.0 ± 2.8	
			XVI	20 kb	21.1 ± 2.9	
			VI	34 kb	28.5 ± 4.2	
Divergence of *SpB-SpC** (pre-H1)	*SpB*	*SpC**	XVI	26 kb	15.6 ± 3.1	
			XV	31 kb	17.3 ± 2.9	
			IV	18 kb	22.8 ± 4.4	
Divergence of *SpB-SpC** (pre-H2c)	*SpB*	PP	IX	22 kb	6.5 ± 2.0	
			II	75 kb	6.8 ± 1.4	
Divergence of *SpB-SpC** (pre-H2b)	*SpB*	IO2	XI	28 kb	10.4 ± 1.8	

Table I.10: Number of generations (mitotic cell divisions) per day in *S. paradoxus* lineages *SpB* and *SpC*.

Estimated from pairwise nucleotide divergence among 21 HC genomes (*SpC** excluded), using lineage *SpA* as outgroup and assuming $1.67 \pm 0.04.10^{-10}$ substitutions per generation (rate in *S. cerevisiae* (Zhu *et coll.* 2014)) and 110,000±10,000 years since *SpB*-*SpC* divergence (beginning of the last glaciation). Estimations from the literature in different yeast species indicate that our estimation falls in the lower range of prediction for *Saccharomyces* in northern regions (1-8 (Ruderfer *et coll.* 2006)). Longer generation time in *SpC* as compared to *SpB* may reflect less frequent favorable growth conditions in its northern distribution, i.e. larger number of days below freezing point.

Species	Generations per day	Remarks	Reference
S. paradoxus (*SpB*)	1.7 - 2.2		This study
S. paradoxus (*SpC*)	1.3 - 1.7		
S. cerevisiae	8	Tropical region	Fay et Benavides (2005)
S. paradoxus	< 8	Authors propose that 8 generations per day is likely overestimated for *S. paradoxus* in northern regions	Kuehne *et coll.* (2007)
S. cerevisiae, *S. paradoxus*	1 - 8	Authors note that the rate of cell division of yeast in the wild is not well characterized, but may fall in the range mentioned	Ruderfer *et coll.* (2006)
S. cerevisiae	6.6	Laboratory conditions	Anderson *et coll.* (2003)
Schizosaccharomyces pombe	6 - 12	Laboratory conditions (various media)	Moreno *et coll.* (1991), Feng et D'Urso (2001)
Candida albicans	16	Laboratory conditions (rich media)	Setiadi *et coll.* (2006)
Cryptococcus neoformans	9.6	Laboratory conditions (soil)	Farhi *et coll.* (1970)

CHAPITRE II: NO EVIDENCE FOR EXTRINSIC POST-ZYGOTIC ISOLATION IN A WILD *SACCHAROMYCES* YEAST SYSTEM

Guillaume Charron[1] et Christian R. Landry[1]

1- Département de Biologie, Institut de Biologie Intégrative et des Systèmes, PROTEO, Pavillon Charles-Eugène-Marchand, 1030 avenue de la Médecine, Université Laval, Québec, Quebec G1V 0A6, Canada

Biology Letters

13 :20170197

II.1-RÉSUMÉ

Les microorganismes comptent parmis les organismes les plus diversifiés de la planète, mais nous en savons encore peu sur l'évolution des différentes barrières à la reproduction qui maintiennent les lignées isolées génétiquement. Les levures du genre Saccharomyces, eucaryotes unicellulaires capables de reproduction sexuée, développent rapidement des barrières postzygotiques intrinsèques lorsqu'elles évoluent indépendamment. Cependant, le rôle des barrières extrinsèques dans l'initiation de la spéciation chez ces microorganismes a été peu exploré. À la recherche de défaut de croissances chez les hybrides de la levure *Saccharomcyes paradoxus*, nous avons mesuré la croissance d'hybrides F_1 et l'avons comparée à celles de leurs souches parentales dans 32 environnements. Dans plus de 80% des cas, les hybrides montraient des phénotypes de croissance partiellement dominants ou encore surdominants. L'absence de défaut de croissance indique qu'il n'y a pas de barrières postzygotiques extrinsèques qui limitent le flux de gène entre les populations. Nos résultats suggèrent aussi que l'hétérosis observée chez les hybrides leur confère un potentiel de colonisation d'environnements différents de ceux où l'on retrouve les souches parentales.

II.2-ABSTRACT

Although microorganisms account for the largest fraction of Earth's biodiversity, we know little about how their reproductive barriers evolve. Sexual microorganisms such as *Saccharomyces* yeasts rapidly develop strong intrinsic post-zygotic isolation but the role of extrinsic isolation in the early speciation process remains to be investigated. We measured the growth of F_1 hybrids between two incipient species of *Saccharomyces paradoxus* to assess the presence of extrinsic postzygotic isolation across 32 environments. More than 80% of hybrids showed either partial dominance of the best parent or over-dominance for growth, revealing no fitness defects in F_1 hybrids. Extrinsic reproductive isolation therefore likely plays little role in limiting gene flow between incipient yeast species and is not a requirement for speciation.

II.3-BACKGROUND

Restriction to gene flow between populations and incipient species can be mediated through mechanisms that act before or after mating. Studies in microorganisms have mainly focused on post-zygotic intrinsic isolation (hybrid unviability or sterility)(Greig *et coll.* 2002, Kuehne *et coll.* 2007). In fungi, the evolution of reproductive isolation is influenced by life-history traits that would favor pre-zygotic isolation in the highly dispersing basidiomycetes but not in the substrate resident ascomycetes, which are mostly isolated by post-zygotic mechanisms (Giraud et Gourbière 2012). In the ascomycetes *Saccharomyces* genus, adaptation to climatic conditions could be a driver of divergence and speciation (Gonçalves *et coll.* 2011). This ecologically based divergence could contribute to the interruption of gene-flow between populations and eventual incipient species through either selection against migrants that create a geographical barrier or extrinsic post-zygotic isolation (Rundle et Nosil 2005), which arises when the hybrids between two species show inappropriate phenotypes in parental environments, leading to fitness reduction and their elimination (Coyne et Orr 2004).

Although ecological barriers to gene flow have been understudied in fungi (Giraud *et coll.* 2008), hybrid phenotypes have been studied in *Saccharomyces* yeast. Analyses show extensive hybrid vigour in both inter and intra-specific crosses (Shapira *et coll.* 2014, Bernardes *et coll.* 2016). However, extending these results to natural systems is difficult since the crosses tested involved either strains that would otherwise rarely interact in nature, or uncommon conditions in the yeasts environment. Natural systems such as the North-American *Saccharomyces paradoxus* incipient species can be used to examine this mechanism. The three endemic incipient species of *S. paradoxus* in North America (*SpB*, *SpC* and *SpC**, an introgressed incipient species resulting from an ancient hybridization event between *SpB* and *SpC*) thrive on the bark and soil associated with

deciduous trees and are mainly composed of highly homozygous isolates (Figure II.1a, b) (Kuehne *et coll.* 2007, Leducq *et coll.* 2014, Leducq *et coll.* 2016).

The two main lineages have different but partially overlapping distributions, which could have been partially shaped by selection against migrants. Despite this and the presence of partial intrinsic reproductive isolation through reduced hybrid fertility (Figure II.1a, b), no F1 hybrids have been recovered from more than 3,000 samples representing more than 392 strains (Charron *et coll.* 2014 and Unpublished data, Leducq *et coll.* 2016). This implies that either hybrid formation is prevented in nature through pre-zygotic isolation or selection against hybrids (extrinsic reproductive isolation). We tested this last mechanism by measuring and comparing the growth of 21 *SpB/SpC* F1 hybrids, their wild diploid parents and 20 intra-lineage crosses in 32 different growth conditions.

II.4-METHODS

Strains were isolated across their range (Charron *et coll.* 2014). Hybrids were constructed by mating heterothallic haploid strains (Table II.S1). Solid media were selected based on previous experiments (Leducq *et coll.* 2014, Leducq *et coll.* 2016) (Table II.1). Laboratory conditions were assayed to test if the isolation method was biased against hybrids. Growth was assessed using colony sizes. Sixty-two strains (10 *SpB* and 11 *SpC* (wild diploid isolates), 21 *SpB/SpC*, 7 *SpC/SpC* and 13 *SpB/SpB* hybrids) in 7 replicates were randomly arrayed on OmniTray plates using a BM3-BC robot (S&P Robotics, Inc., North York, Canada). The two outer borders were filled with *S. cerevisiae* strain BY4743 to prevent border effects. Photos were taken every 24 hours for 4 days after which they were replicated on fresh media and the process repeated. Images were analyzed using ImageJ 1.44o (National Institutes of Health) (Diss *et coll.* 2013). Colony pixel intensities after two days were $\log_2$-transformed and used for downstream analyses (Figure II.S1, Annexe B).

Resistance to freezing was measured by growing cells overnight at 30°C in a 96 deep-well plate, diluted to a OD_{600nm} of 0.1 and grown at 30°C until they reached an OD_{600nm} of 0.3-0.4. Samples of 200 µL were transferred to PCR plates, one incubated in an ethanol/dry ice bath for 2 h, the other tested for cell viability as a control. Cells were centrifuged in a 5810R Centrifuge (Eppendorf) at 3,000 rpm for 10 minutes and resuspended in 200 µL of sterile water. Volumes of 25 µL were transferred into 225 µL of a 30nM aqueous solution of SYTOX® Green (ThermoFisher). The plate was incubated in the dark for 20 minutes with a resuspension step after 10 minutes. Fluorescence was measured on a Guava EasyCyte 14HT (EMD Millipore) on 5,000 cells in biological replicates (Pearson's r = 0.992) (excitation: 488_{nm}, detection: $525/530_{nm}$)(Figure II.S2, Annexe B).

Using the median colony size values of the wild diploid parental strains for each hybrid, we calculated mid-parent values as m ((P1+P2)/2)) (Shapira *et coll.* 2014), additive genetic deviation as a (|P1 or P2−m|) and dominant genetic deviation as d (hybrid−m). The degree of dominance (d/a) was used to assign mode of inheritance: values between −1 and 1 indicate partial dominance of either the smaller or bigger colony parent; values lower than −1 indicates under-dominance and higher than 1 indicates over-dominance. Values of −1 and 1 indicate full dominance and a value of 0 indicates co-dominance. Analyses were performed with R v3.0 (R Development Core Team 2013) (Figure II.S3, Annexe B).

II.5-RESULTS

SpB/*SpC* hybrids do not show growth defects but rather show over-dominance (47.02%) and partial dominance (35.71%) across conditions (Figure II.2a, b). Examples are shown in Figures II.2c, d and e. The proportions of over-dominant and partially dominant phenotypes (degree of dominance higher than 0) in hybrids is significantly greater than what is observed in intragroup crosses in one case (BC vs BB $p= 2.2\times10^{-16}$, Fisher's exact test, one-tailed) but not the other (BC vs CC p= 0. 9997).

The proportion of under-dominant hybrids is not significantly different between the laboratory media and all other conditions (p= 0.0535, Fisher's exact test, two-tailed). More than 95% (20/21) of parental strains showed median degree of dominance values above 0 in *SpB*/*SpC* crosses (Figure II.S4-S5, Annexe B).

II.6-DISCUSSION

We examined whether F_1 hybrid growth defects could contribute to the reproductive isolation of two incipient yeast species. Most hybrids show no fitness defects over the array of conditions tested but rather show overdominance or intermediate phenotypes. Therefore, it is unlikely that post-zygotic extrinsic reproductive isolation prevents gene flow between the incipient species *SpB* and *SpC*. We also note that heterosis appears to be prevalent in the *SpC* intra-group crosses. Because isolation data suggest that this group displays smaller effective population sizes and population density (Leducq *et coll.* 2014, Leducq *et coll.* 2016), drift could lead to the accumulation of deleterious recessive mutations locally, which would then be masked when outcrossing with other *SpC* strains. The segregation of deleterious mutations in *SpC* would also lead to dominance of the *SpB* phenotype in crosses with *SpC*. The dominance of the *SpB* phenotype over *SpC* in the *SpB*/*SpC* hybrids could therefore partly be explained by the masking of *SpC* deleterious alleles.

Our results are overall in line with the data reported by both Shapira *et coll.* (2014) and Bernardes *et coll.* (2016) which showed growth advantage in hybrids for both intra- (in *S. cerevisiae*/*S. cerevisiae, S.paradoxus*/*S. paradoxus*) and interspecific (*S.cerevisiae*/*S. paradoxus*) hybrid crosses. In natural systems where closely related species are in sympatry, this growth advantage would severely limit the role of extrinsic postzygotic isolation or would even promote hybridization.

We observed that the hybrids have a non-significant tendency of being under-dominant in our laboratory media used for the yeast isolation. However, because the success rate in strain isolation is less than 10%, there is limited opportunity for inter-strain competition during experiments, which would limit these effects in terms of disfavouring hybrids. One limitation of our study is that we still lack a clear understanding of yeast ecology and what their fitness determinants are in nature (Goddard et Greig 2015), which prevents recreating the exact conditions they face in the wild. However, progress was recently made in identifying their preferred habitats, climate and interactions with other microbial species (Kowallik *et coll.* 2015, Kowallik et Greig 2016, Robinson *et coll.* 2016). One environmental parameter that has been shown to play a key role in yeast ecology is high temperature (Gonçalves *et coll.* 2011, Leducq *et coll.* 2014, Sylvester *et coll.* 2015) and we observed the highest median *d/a* value for conditions implicating this parameter. Finally, the observation of frequent heterosis brings the possibility that newly formed *SpB/SpC* hybrids have the potential to occupy new environments unavailable to their parents and would thus have not been sampled and thus be underestimated (Rieseberg *et coll.* 2007). This was seen in *Saccharomyces cerevisiae* where hybrids between oak tree and vineyard lineages have colonized and adapted to the cherry tree environment (Clowers *et coll.* 2015).

Studies on Fungi suggest that barriers to gene flow in ascomycetes are dominated by post-zygotic mechanisms (Giraud *et coll.* 2008). Our investigation of ecologically-based post-zygotic reproductive isolation did not show any defects in hybrids. Since there is only partial post-zygotic intrinsic isolation, other barriers to gene flow should exist. It has been proposed that adaptation to substrate could lead directly to reproductive isolation in ascomycetes (Giraud et Gourbière 2012). Since the two host where we most commonly find the *S. paradoxus* lineages are oak and maple trees, a specialization of each lineage for a different host tree would

then restrict gene flow. There is also some evidences that reinforcement could have been evolving in other *S. paradoxus* yeasts (Murphy et Zeyl 2015). The study of wild incipient species systems such as the one used here will allow a better understanding of the contributions of environment to reproductive barriers and the evolution of reproductive isolation between incipient species.

II.7-AUTHORS' CONTRIBUTIONS

GC and CRL designed the experiments, GC performed the experiments and analysed the results, and drafted the manuscript with help from CRL. All authors gave final approval for publication and agree to be held accountable for the content herein.

II.8-ACKNOWLEDGEMENTS

We thank Landry laboratory members for comments and discussions on this project, Dr. Paul Sniegowski and two anonymous reviewers for their comments on our manuscript.

II.9-FUNDING

This project was funded by a NSERC Discovery grant to CRL and a NSERC AG Bell Scholarship to GC. CRL holds the Canada Research Chair in Evolutionary Cell and Systems Biology.

II.10-FIGURES

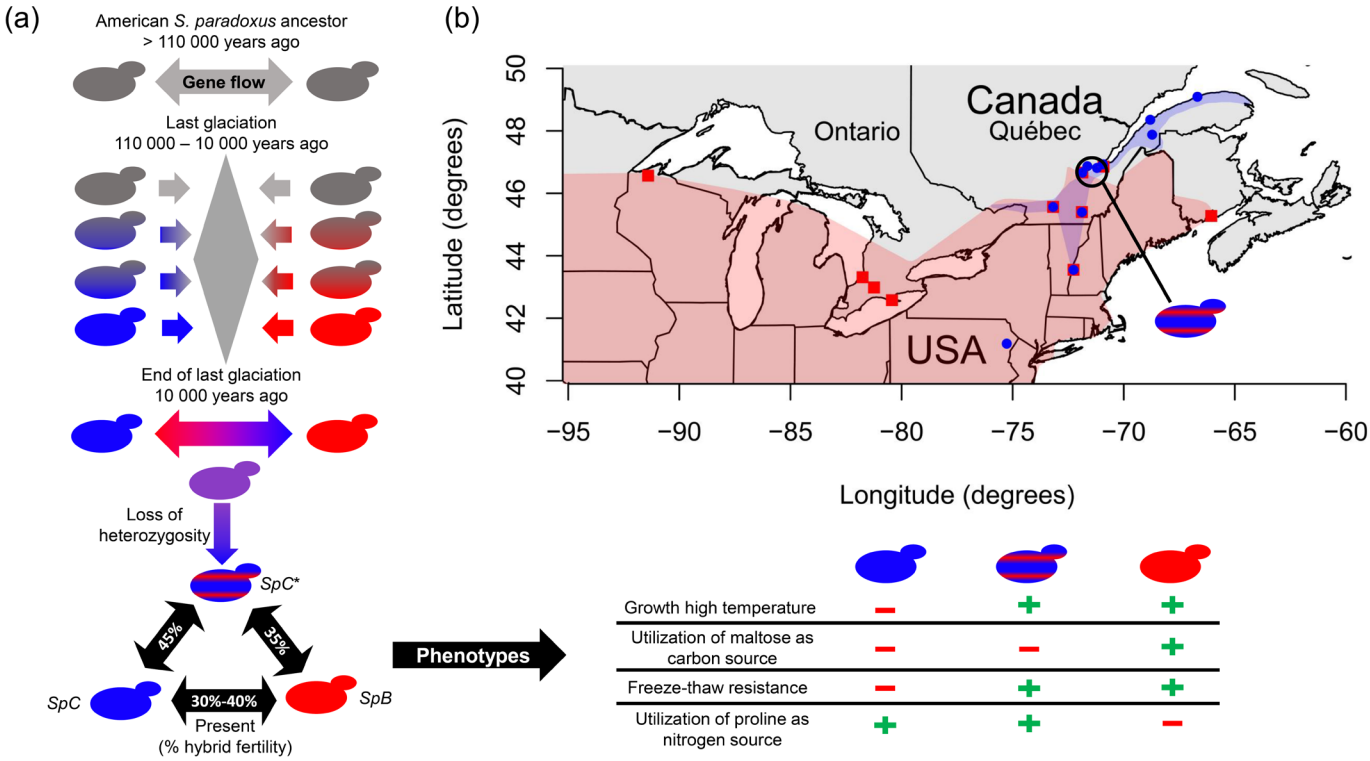


Figure II.1:Incipient speciation in *Saccharomyces paradoxus*.
(a) Biogeographical scenario inferred from genomic and phenotypic data (figure adapted from (Leducq *et coll.* 2016)) and currently known discriminating phenotypes. (b) *SpC* (blue area) is found along in the north east of the *S. paradoxus* distribution whereas *SpB* (red area) can be found in the south of Québec Province and in the northeastern USA. Each of the parental species location is represented on the map by red squares (*SpB* strains) and blue circles (*SpC* strains). The black circle represents the area where the hybrid species *SpC** can be isolated.

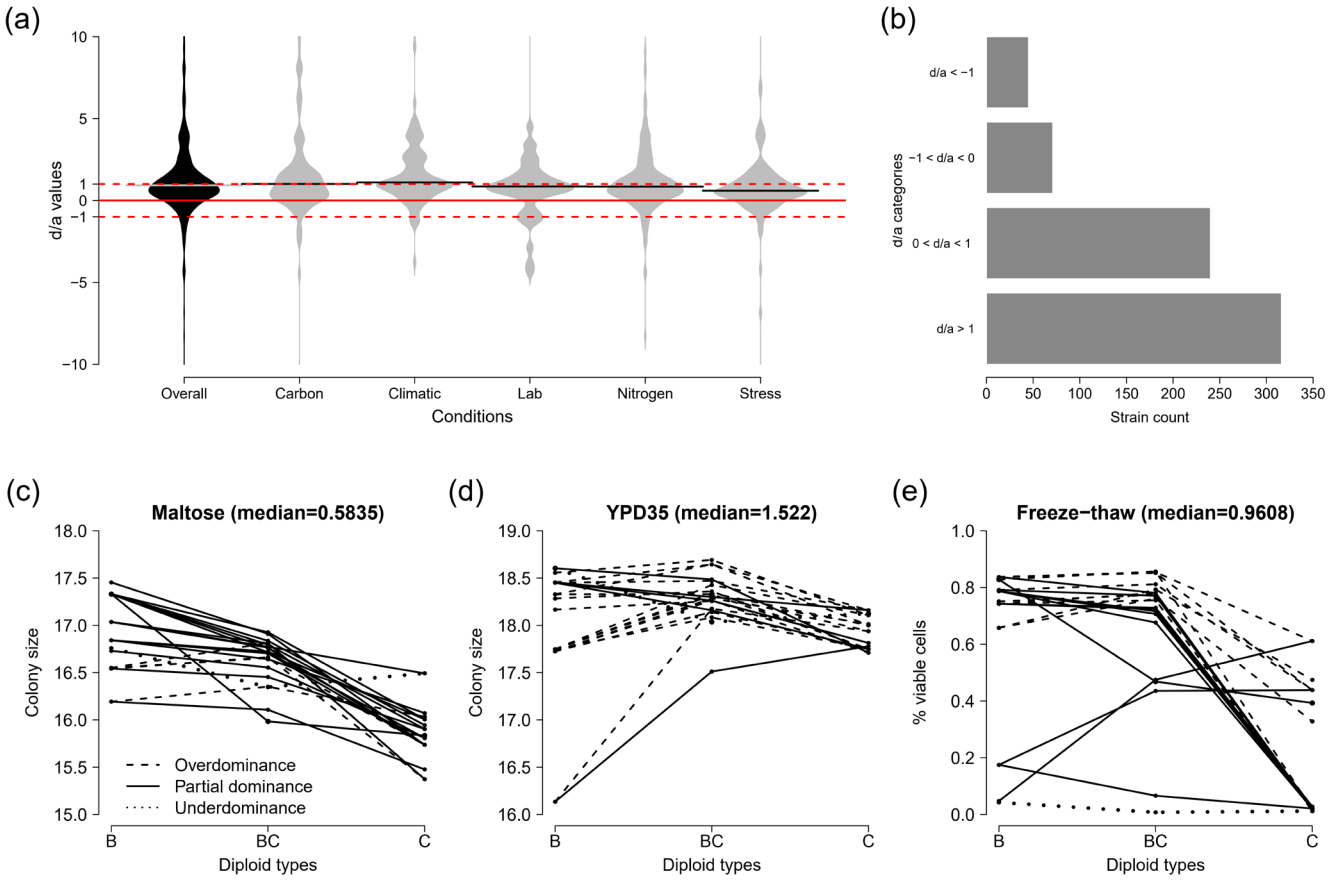


Figure II.2:Hybrid strains show dominance (d/a) values indicating heterosis and intermediate phenotypes.

(a) Frequency distribution of *d/a* values across all conditions (black) and for each condition types (grey). Median values are indicated by horizontal lines. (b) Counts of d/a values by mode of inheritance. Representative conditions showing partial dominance (c), over-dominance (d) and a mix of over- and partial dominance (e) closer to full dominance.

II.11-TABLES

Table II.1:List of media tested in this study.

Abbreviations are used for Yeast Extract (YE), Yeast Nitrogen Base (YNB), Peptones (Pep), Ammonium Sulfate (AS) and D-Glucose (Glc).

Media	Condition type	Composition (+2% Agar)
Ethanol	Carbon	1% YE, 2% Pep, 3% Ethanol
Ethanol 37°C		1% YE, 2% Pep, 3% Ethanol, Incubation 37°C
Fructose		0.174% YNB, 2% Fructose, 0.5% AS
Galactose		0.174% YNB, 2% Galactose, 0.5% AS
Glucose		0.174% YNB, 2% Glc, 0.5% AS
Maltose		0.174% YNB, 2% Maltose, 0.5% AS
Mannose		0.174% YNB, 2% Mannose, 0.5% AS
Sucrose		0.174% YNB, 2% Sucrose, 0.5% AS
α-methyl-D-glucopyranoside		0.174% YNB, 2% α-methyl-D-glucopyranoside, 0.5% AS
SOE		0.1% YE, 0.15% Pep, 0.5% Glc, 0.5% Fructose, 1% Sucrose
25°C	Climatic	1% YE, 2% Pep, 2% Glc, Incubation 25°C
30°C		1% YE, 2% Pep, 2% Glc
35°C		1% YE, 2% Pep, 2% Glc, Incubation 35°C
37°C		1% YE, 2% Pep, 2% Glc, Incubation 37°C
Freeze-Thaw cycle		See methods
Sniegowski I (Solid)	Laboratory	0.3% YE, 0.3% Malt Extract, 0.5% Pep, 1% Sucrose, 5% Ethanol, 1mM HCl, 1mg Chloramphenicol
Sniegowski II		0.174% YNB, 2% Glc, 0.5% AS, 0.134% Complete amino acid drop-out, 4mM HCl
Allantoine	Nitrogen	0.174% YNB, 2% Glc, 0.5% Allantoine
Asparagine		0.174% YNB, 2% Glc, 0.5% Asparagine
Glutamine		0.174% YNB, 2% Glc, 0.5% Glutamine
Glycine		0.174% YNB, 2% Glc, 0.5% Glycine
Histidine		0.174% YNB, 2% Glc, 0.5% Histidine
Isoleucine		0.174% YNB, 2% Glc, 0.5% Isoleucine
Lysine		0.174% YNB, 2% Glc, 0.5% Lysine
Proline		0.174% YNB, 2% Glc, 0.5% Proline
Tyrosine		0.174% YNB, 2% Glc, 0.5% Tyrosine
H2O2	Stress	1% YE, 2% Pep, 2% Glc, 1mM Hydrogen peroxide
NaCl		1% YE, 2% Pep, 2% Glc, 1M NaCl
pH 3		1% YE, 2% Pep, 2% Glc, pH adjusted with HCl
pH 8		1% YE, 2% Pep, 2% Glc, pH adjusted with NaOH
Rapamycine		1% YE, 2% Pep, 2% Glc, 200nM Rapamycine
Sorbitol		1% YE, 2% Pep, 2% Glc, 1M Sorbitol

CHAPITRE III: INTRINSIC PLOIDY INSTABILITY DRIVES FERTILITY RECOVERY IN EXPERIMENTAL HYBRID POPULATIONS

Guillaume Charron[1,2,3†], Souhir Marsit[1,2,3,4†], Mathieu Hénault[1,2,4], Hélène Martin[1,2,3,4] et Christian R. Landry[1,2,3,4*]

Affiliations:
1-Institut de Biologie Intégrative et des Systèmes, 1030 avenue de la Médecine, Université Laval, Québec (Qc), G1V 0A6, Canada.
2-Regroupement Québécois de Recherche sur la Fonction, l'Ingénierie et les Applications des Protéines, (PROTEO), 1045 Avenue de la Médecine, Université Laval, Québec (Qc) G1V 0A6, Canada.
3-Département de Biologie, 1045 Avenue de la Médecine, Université Laval, Québec (Qc) G1V 0A6, Canada.
4-Département de biochimie, microbiologie et bio-informatique, 1045 Avenue de la Médecine, Université Laval, Québec (Qc) G1V 0A6, Canada.

† equal contributions

Soumis à *Nature Communications*
2019-06-24

III.1-RÉSUMÉ

Bien que pouvant bénéficier de l'hétérosis, les hybrides interespèces sont souvent victimes des incompatibilités génétiques. Ces incompatibilités mènent à une fertilité diminuée chez les hybrides de première génération. Les rétrocroisements ou les événements de reproduction entre les hybrides peuvent permettre le rétablissement d'une fertilité normale par la ségrégation des loci incompatibles et la sélection de génotypes plus fertiles. Cependant, la reproduction sexuée a un coût pour les hybrides infertiles. Chez les organismes qui peuvent se reproduire de manière asexuée, les évènements de recombinaison mitotique pourraient aussi contribuer au rétablissement de la fertilité. Afin de tester cette hypothèse, nous avons fait évoluer plus de 600 lignées de levures hybrides pendant environ 700 générations. Ces hybrides ont été générés en croisant des espèces parentales le long d'un gradient de divergence couvrant entre 100 000 et 15 millions d'années de divergence. Afin d'élargir le spectre des mutations observables, nous avons limité l'efficacité de la sélection en soumettant les lignées à de forts goulots d'étranglement sur la taille des populations. En suivant l'évolution de la fertilité au cours du temps, nous avons trouvé que la fertilité de certains hybrides était restaurée entièrement et rapidement. Ces hybrides, autant entre jeunes espèces et espèces bien établies, ont subi une duplication complète du génome. Nos résultats suggèrent donc que l'instabilité de la ploïdie des hybrides peut être un moyen de rétablissement rapide de la fertilité.

III.2-ABSTRACT

Interspecies hybrids often show some advantages over parents but also frequently suffer from reduced fertility, which can sometimes be overcome through sexual reproduction that sorts out genetic incompatibilities. Sex is however inefficient due to the low viability or fertility of hybrid offspring and thus limits their evolutionary potential. Mitotic cell division could be an alternative to fertility recovery in species such as fungi that can also propagate asexually. To test this, we evolved in parallel and under relaxed selection more than 600 diploid yeast inter-specific hybrids that span from 100,000 to 15 M years of divergence. We find that hybrids can recover fertility spontaneously and rapidly through whole-genome duplication. These events occurred in both hybrids between young and well-established species. Our results show that the instability of hybrid ploidy is a spontaneous path to fertility recovery.

III.3-INTRODUCTION

Inter-specific hybridization is common in animals, plants and microorganisms (Mallet 2005, Stukenbrock 2016) and is a potentially frequent source of genetic diversity over short time scales (Lewontin et Birch 1966, Grant et Grant 1994, Arnold et Martin 2009). However, hybrid lineages often suffer from poor fertility that reflects reproductive isolation between parental lineages. The poor fertility of hybrids can prevent their maintenance as independent populations, thus hindering their long term and thus speciation potential. Different molecular mechanisms underlie hybrid infertility, including genetic incompatibilities (nuclear and cytonuclear) (Maheshwari et Barbash 2011) and changes in genome architecture (ploidy number or chromosome rearrangements) (Rieseberg 2001). If the hybrids are to establish as an independent population or species, they need to recover from this low initial fertility. In obligatory sexual species, fertility restoration can be achieved by crosses among hybrid individuals or backcrosses with either parental species, allowing the purge of incompatibilities through recombination. Because of this gene-flow, this process most often leads to the formation of introgressed species (Marsden-Jones 1930) rather than hybrid species (Mallet 2007). In this context, the formation of hybrid species may necessitate other means of isolation from both parental species, which may include geographic or ecological isolation while the recovery of fertility through recombination takes place (Anderson et Stebbins Jr. 1954). Some organisms, however, have access to both sexual and asexual reproduction. In these species, if sexual encounters are rare, periods of asexual reproduction might provide hybrids with alternative mechanisms for fertility recovery, which could facilitate hybrid speciation. Using yeast as an experimental model system, we show that hybrids between closely and distantly related species can recover fertility spontaneously by whole-genome duplication and this, without the need for natural selection. Although rare, these events have

large effects and bring fertility to the levels seen in parental species. Polyploidy is most common in plants (Hilu 1993) but has also been observed in many animals and fungi (Otto et Whitton 2000, Albertin et Marullo 2012), making this mechanism of fertility recovery accessible to many species.

We investigated the evolution of fertility in parallel experimental yeast hybrid populations during mitotic evolution. Using strong population bottlenecking to minimize the efficiency of natural selection allowed for the random accumulation of genetic and genomic changes, providing an estimate of the neutral rate of evolution. We examined whether fertility would increase or decrease with time and, if so, whether it would occur through gradual or punctuated changes (Figure III.1A). We considered hybridization over different levels of parental divergence from intra-population to inter-specific crosses. These crosses represent up to 15 M years of divergence, which is sufficient to achieve almost complete (99%) postzygotic reproductive isolation in budding yeast (Liti *et coll.* 2006, Hittinger 2013). We used a collection of North American natural yeast isolates representing three lineages of the wild species *Saccharomyces paradoxus* and a wild isolate of its bona fide sister species, *S. cerevisiae* (Annexe C, Table III.S1). The *S. paradoxus* lineages (*SpA*, *SpB* and *SpC*) (Kuehne *et coll.* 2007) are incipient species that exhibit up to 4% of genetic divergence (*SpA*-*SpB*) (Leducq *et coll.* 2016, Eberlein *et coll.* 2019) and up to 60% of reduction in hybrid fertility compared to within lineage crosses (Charron *et coll.* 2014). These species and populations occur in partially overlapping geographical ranges, even for the most distant pair, *S. paradoxus* and *S. cerevisiae*, making hybridization possible in a natural context. Including this sister species extends nucleotide divergence between parental strains up to 15%. We mated two *SpB* strains to two other *SpB* strains and to two strains of the diverged lineages and species, producing 4 different types of crosses in duplicates that we classified in

terms of divergence: Very Low (VL_{div}) = *SpB*×*SpB*; Low (L_{div}) = *SpB*×*SpC*; Moderate (M_{div}) = *SpB*×*SpA* and High (H_{div}) = *SpB*×*S. cerevisiae* (Annexe C, Table III.S2 and III.S3). Ninety-six diploid hybrid lines from independent mating events were generated for all but the VL_{div} crosses, for which 48 hybrids were generated, for a total of 672 independent lines (96 lines × 3 types of crosses × 2 pairs of strains + 2 × 48 VL_{div} crosses) (Figure III.1A, B). We randomly selected and streaked colonies on plates every 3 days for 35 passages with an estimated number of mitotic generations of ~22 per passage (Fig III.1C, Figure III.S1).

III.4-MATERIAL AND METHODS

III.4.1-Strain construction

We used the *ade2-Δ* marker to help with visual identification of respiration deficient colonies, a strategy used in past mutation accumulation experiments (Joseph et Hall 2004, Zhu *et coll.* 2014). As described in the main text, this marker did not faithfully indicate inefficient respiration the strain backgrounds that we used. The heterothallic *S. paradoxus* strains were generated previously (Charron *et coll.* 2014, Leducq *et coll.* 2016, Annexe C, Table III.S1). The *ADE2* and *HO* loci of the two wild *S. cerevisiae* strains were deleted following the method described by Güldener *et coll.* (1996). The *ADE2* locus of the *S. paradoxus* strains were replaced by homologous recombination with resistance cassette following the same procedure as for *HO* in *S. paradoxus (Charron et coll. 2014)*. Oligonucleotides with overhangs (Annexe C, Table III.S4) specific to each lineage were used to generate the deletion cassettes from pFA-hphNT1 (Janke *et coll.* 2004) to prevent recombination with the cassettes already present at the *HO* locus (*KANMX* and *NATMX* cassettes).

III.4.2-Experimental crosses

Two crosses were made for each of the divergence levels (L1, M1, H1, VL1 and L2, M2, H2, VL2) (Annexe C, Table III.S3). All incubation steps

were performed at room temperature (RT). Haploids to be crossed were precultured overnight in 5 mL of YPD (1% yeast extract, 2% tryptone and 2% D-glucose). Precultures were then diluted at OD_{600nm} of 1.0 in 500µL aliquots. The aliquots from two strains to be crossed were mixed together and 5µL were used to inoculate 200µL of fresh YPD medium in 96 replicates so all strains would derive from independent mating events and would be truly independent hybrids. Cells were given 6 hours to mate after which 5µL of the mating cultures were spotted on a diploid selection medium (YPD, 100µg/mL G418, 10µg/mL Nourseothricin). From each of the 96 spots, 1 colony was picked as a founding line for the evolution experiment, resulting in 96 independent lines for each of the six interlineage crosses (48 lines for the two intra lineage crosses).

III.4.3-Evolution experiment

Each of the independent lines (single colonies) were streaked on one third of a YPD agar plate. To facilitate the detection of potential lines mixing during the experiment, each Petri was streaked with three different crosses (series L1, M1, H1 and L2, M2, H2). Crosses VL1 and VL2 were streaked on two different sets of Petri dishes, with three lineages per Petri. The 192 plates were split into 3 sets of 64 (lines 1-64 and lineages 65-96). Plates were incubated at room temperature for 3 days after which a new single colony was streaked as a progenitor for the new generation. Each set was rotated between three manipulators at each replication step. The criteria for the new colony were to be 1) the closest to a predesigned mark on the Petri dish, allowing for unbiased colony selection, 2) a single colony and 3) big enough to allow for both replication on a new medium and the inoculation of a liquid culture to generate a frozen stock. If the colony closest to the mark did not meet criteria 2 and 3, the second closest colony was then examined, and the process was repeated until all criteria were met. Every 3 passages, the colonies were both streaked and used to inoculate the wells of a 96 wells plate containing 150µL of fresh YPD

medium. After a 24 h incubation at room temperature, 75 µL of 80% glycerol was added and the plates were placed in a −80°C freezer for archiving. The lines were maintained on plates for a total 35 passages.

III.4.4-Estimation of generation time

To evaluate the generation time on plates and thus estimate the total number of mitotic divisions during the experiment, 3 lines from each cross were randomly selected. Strains for T_{ini} and T_{end} were thawed (48 total strains tested), streaked on YPD solid medium and let grow for 3 days. Strains were then replicated on fresh YPD solid medium following the same protocol as for the evolution experiment. After 3 days of incubation, the colony closest to the predesigned mark was extracted from the media using a sterile scalpel. The agar block with the colony was put in a sterile 1 mL Eppendorf tube and the colony was resuspended in 500µL of sterile water. Optical density at 600nm (OD_{600nm}) of the resuspensions was estimated using a TECAN Infinite 200 plate reader (TECAN, Männedorf, Switzerland). These resuspensions were diluted in 200µL of sterile water to obtain OD_{600nm} values of about 0.05 (500 cells/µL). The dilutions were then analyzed with a Guava® easyCyte HT (Millipore Sigma, Burlington, USA) flow cytometer to estimate actual cell numbers. The estimated number of cells/µL were used to calculate the initial number of cells in the volume used in the dilution and then in the initial 500µL. The $\log_2$ of this number represents the number of cell doubling during the colony growth for 1 passage of the experiment assuming the colony was formed from a single cell (Figure III.S1).

III.4.5-Sporulation protocol

Strains were thawed and 2µL of the stocks were spotted on a fresh YPD medium and incubated for 3 days. A small number of cells was used to inoculate 4 mL of fresh YPD media and incubated for another day. From those precultures, a new 4 mL culture was inoculated at 0.6 OD_{600nm} in fresh YPD and grown for 3 hours. Cultures were then centrifuged at 250 g

and the YPD was replaced with 4mL of YEPA medium (1% yeast extract, 2% tryptone and 2% potassium acetate). Cultures were incubated for 24 h after which they were centrifuged again at 250 g, washed once with sterile deionized water and put into 4 mL of SP medium (0.3% potassium acetate 0.02% D-Raffinose). After 3-5 days of incubation, the strains were dissected as in Charron *et coll.* 2014 with a SporePlay™ dissection microscope (Singer Instruments, Somerset, UK) on YPD plates and incubated for 5 days. Pictures of the plates were taken after the incubation and fertility was determined as the number of spores forming a colony visible to the naked eye after 5 days.

III.4.6-Mitochondrial DNA genotyping

Two mitochondrial loci were genotyped for presence or absence by PCR. Total DNA extractions were performed using the method described by Looke *et coll.* (2011). The two PCR assays target loci in the *RNL* and *ATP6* mitochondrial genes, respectively, as described in Leducq *et coll.* (2017). Multiplex PCR with both primer pairs (Annexe C, Table III.S4) was performed with the following cycle: 3 min at 94°C; 40 times the following cycle: 30 s at 94°C, 30 s at 57.5°C, and 50 s at 72°C; and 10 min at 72°C. A PCR targeting the ITS1-5.8S-ITS2 locus was performed on the same DNA samples as positive controls following the method described in Montrocher *et coll.* (1998).

III.4.7-Colony pigmentation analysis

The stocks of evolved lines at P0, P1, P16 and P35 were printed on OmniTray plates (Thermo Fisher Scientific, Waltham, USA) of YPD solid medium using a BM5-BC-48 colony processing robot (S&P Robotics Inc., Toronto, Canada) and incubated at room temperature for 4 days. Plate images were analyzed with the gitter (Wagih et Parts 2014) package in R and positions which had no growth, had visible contamination or which were flagged as non-circular or overlapping were filtered out. Downstream analyses were performed using custom scripts in

Python 3.7.1. A square of 400 pixels was extracted for each position on the plate and the HSV (hue, saturation, value) were extracted using the package pillow v3.5.0. For each position, the top 10% pixels according to color hue values were retained. Given that hue values allow to discriminate between foreground (colony) and background (media) pixels (Figure III.S2A), this filtering aimed to minimize the contamination from background pixels. The average of saturation values for each position was used for the analysis. The association between colony saturation and sporulation capacity was tested using a linear model with logistic link function using the statsmodels package v.0.9.0.

III.4.8-Colony growth on complex medium with glycerol

The colonies grown on YPD solid medium for the coloration analysis were replicated on OmniTray plates of YPG solid medium (1% yeast extract, 2% tryptone, 3% glycerol and 2% agar) using a BM5-BC-48 colony processing robot and incubated at room temperature for 6 days. Plate images analysis and filtering were performed as described for the colony coloration analysis. A value of one was added to the pixel counts per colony output by gitter prior to conversion in $\log_2$. The association between colony size and sporulation ability was tested using a linear model with logistic link function using the statsmodels package v.0.9.0 (Seabold et Perktold 2010).

III.4.9-Fertility assessment in the evolved lines

Fertility of the evolved lineages (VL_{div}, M_{div} and H_{div}) was measured on 24 randomly chosen lines per cross (14 which lived through P35 and 10 that were lost before the end of the experiment). To ensure we had the same numbers of diploids to compare with the other lines, more strains were chosen for the two L_{div} crosses, 36 lines were used (24 diploids and 12 triploids, 26 strains that lived until P35 and 10 lost before). For each line, we measured fertility at three different time points: 1) immediately after mating (T_{ini}), 2) at the halfway point for the given lineage (T_{mid}) and 3)

at the last passage of the given line (T_{end}). This means that T_{mid} and T_{end} do not always refer to 352 and 770 mitotic generations (Table III.S5).

III.4.10-Autodiploidization of spores

In order to generate fully homozygous strains that should have fully recovered fertility, we performed autodiploidization experiments. After the dissections, some spores from the L_{div} and M_{div} crosses were typed for their mating type locus and their antibiotic resistance markers. When possible, four spores were kept as frozen stocks (two of each mating type and resistance combination). A subset of the spores expressing the G418 resistance was selected to undergo autodiploidization. This was performed by transformation of the spores with the plasmid pHS3 containing *S. cerevisiae HO* gene with its endogenous promoter and the CloNAT resistance cassette (pHS3 was a gift from John McCusker, Addgene plasmid # 81038). Transformants were selected on fresh selection medium (YPD, 200µg/mL G418, 100µg/mL CloNAT) to be sporulated and dissected.

III.4.11-Intra-tetrad crosses (ITC)

Intra-tetrad mating was conducted in order to generate hybrids with rapid loss of heterozygosity and to test whether fertility recovery was possible. Strains from the L1 and the L2 crosses were sporulated as described above, but the dissection steps were modified to allow the mating of pairs of spores from the same tetrad (leaving 2 spores instead of one at the designed dissection spot). The plates were incubated for 5 days. In this manner, pairs of spores of opposite mating types could mate to generate new diploids while pairs of identical mating types would divide mitotically as haploids. To ensure that the yeasts recovered were the result of a mating event, we selected diploids with resistance to both G418 and CloNAT by replica plating the colonies on a fresh selective medium (YPD, 200µg/mL G418, 100µg/mL CloNAT). From the surviving diploids, one colony was randomly selected to be sporulated again. This process was repeated 12 times (ITC 1 to ITC 12) for 16 lines (2 replicates from lines L1-

6, L2-8, L2-36 and L2-63 and 8 other unique line). This number of meiosis was expected to generate extensive LOH as, for a single given heterozygous locus, less than 1% of the population will have maintained heterozygosity (Knop 2006). The ITC 0 (initial hybrids), ITC 1, ITC 6 and ITC 12 strains were sporulated and dissected as described above.

III.4.12-Determination of ploidy

Measurement the cell DNA content was performed using flow cytometry with the SYTOX™ green staining assay (Thermo Fisher, Waltham, USA). Cells were first thawed from glycerol stocks on solid YPD in omnitray plates (room temperature, 3 days) including controls. The parental strain *SpB* (MSH604) was used as control on both its haploid and diploid (wild strain) state. Liquid YPD cultures of 1 ml in 96 deepwell (2ml) plates were inoculated and incubated for 24h at room temperature. Cells were subsequently prepared as in Gerstein *et coll.* (2006) but stained with a final SYTOX™ green concentration of 0.6 µM for a minimum of 1 h at room temperature in the dark. The volume of cells was adjusted to be around a cell concentration less than 500 cells/µL. Five thousand cells for each sample were analysed on a Guava® easyCyte 8HT flow cytometer using a sample tray for 96-well microplates. Cells were excited with the blue laser at 488 nm and fluorescence was collected with a green fluorescence detection channel (peak at 512 nm). The distributions of the green fluorescence values were processed to find the two main density peaks, which correspond to the two cell populations, respectively in G1 and G2 phases. The data was analysed using R version 3.4.1 (R Development Core Team 2013).

III.4.13-Mating type DNA genotyping

The MAT locus was genotyped for the presence of MATa, MATα or both copies by PCR in the stock of the haploid *SpC* parental strains (LL11_004 and LL11_009). Genomic DNA was extracted following standard

protocols (QIAGEN DNAeasy, Hilden, Germany) from overnight cultures issued from the 5 isolated colonies from the *SpC* haploid stocks showing 2n ploidy and as controls 3 haploid *SpC* strains from the same stocks, two triploid L_{div} lines and a diploid *SpC* control strain (LL11_004). The PCR assay target a region in the active mating type locus that differentiate the two mating type sequences. PCR with three primers (Table III.S4) was performed with the following cycle: 5 min at 94°C; 35 times the following cycle: 30 s at 94°C, 30 s at 56°C, and 50 s at 72°C; and 10 min at 72°C.

III.4.14-Genotyping by sequencing

We performed genotyping-by-sequencing (GBS) to investigate the genomic composition of the triploid and tetraploid hybrids. We sampled 77 strains in total: 8 diploids and 16 triploids from the L_{div} crosses at T_{ini}; all 8 tetraploids and as controls 2 diploids from each cross L_{div}, M_{div} and H_{div} at T_{ini} and T_{end} as well as all parental strains. DNA was extracted from overnight cultures issued from one isolated colony following standard protocols (QIAGEN DNAeasy, Hilden, Germany). As controls, we prepared artificial hybrid genomes by mixing DNA of parental strains with different proportion from each 0.5/0.5, 0.66/0.33 or 0.33/0.66. DNA was quantified using Accuclear® Ultrahigh sensitivity dsDNA Quantification kit (Biotium, Fremont, USA) in a Spark® microplate reader (TECAN, Männedorf, Switzerland). DNA concentration was normalized to 10 ng/µl and subsequently used for library preparation.

Libraries for Ion Proton GbS were prepared using the procedure described by Masher *et al.*, 2013 (Mascher *et coll.* 2013) at the Plateforme d'Analyses Génomiques of the Institut de Biologie Intégrative et des Systèmes (IBIS, Université Laval, Québec, Canada) with the following modifications: ApeKI endonuclease enzyme and ApeKI barcodes were used instead of the PstI/MspI combination and a blue Pippin (SAGE science, Beverly, USA) was used to size libraries (150-300bp) before PCR amplification. Libraries were prepared for sequencing using an Ion Chef™,

Hi-Q reagents and PI™ chip kit V3 (Thermo Fisher, Waltham, USA) and the sequencing was performed for 300 flows. A single fastq file was obtained and demultiplexed using Radtags tool from STACKS with default options (Catchen *et coll.* 2013) which generated separated fastq files for each sample. Reads were mapped onto the *S. paradoxus* reference genome (CBS432) (Yue *et coll.* 2017) for all crosses and onto the *S. cerevisiae* reference genome (YPS 128) (Yue *et coll.* 2017) for the H_{div} crosses using Bowtie2 (Langmead et Salzberg 2012). Read coverage for each position in the genome was also determined using SAMtools depth. Single nucleotide polymorphism (SNP) calling on these GBS libraries was performed using the BCFtools from SAMtools (Li *et coll.* 2009) with default parameters. The data was subsequently analyzed using R version 3.4.1 (R Development Core Team 2013). To measure the allelic frequency, we filtered for SNPs with more than 10X coverage and corresponding to positions in the genome covered in both parental strains to be able to identify the parental origin of the alleles in hybrids.

III.4.15-Whole-genome sequencing

We performed whole genome sequencing to investigate the genomic composition of the 8 tetraploid hybrids at T_{ini} and T_{end} (for 16 total) as well as the 6 corresponding haploid parental strains. Genomic DNA was extracted from overnight cultures issued from one isolated colony following standard protocols (QIAGEN DNAeasy, Hilden, Germany). Libraries were prepared with the Illumina Nextera kit (Illumina, San Diego, USA) following the manufacturer's protocol and modifications from by Baym and colleagues (Baym et al., 2013). Pooled libraries were sequenced on different lanes of HiSeqX (150PE, Illumina, San Diego, USA) at the Genome Quebec Innovation Center (Montréal, Canada). The 22 genomes were sequenced with an average genome-wide coverage of 90X. Raw sequences are accessible at NCBI (bio project ID PRJNA515073).

III.4.16-Read mapping and variant calling

Raw reads were mapped on the reference genome of MSH604 (Eberlein *et coll.* 2019), one of the two used *SpB* parental strains (Eberlein *et coll.* 2019), *S. paradoxus* reference genome (CBS432) (Yue *et coll.* 2017) for all crosses and onto the *S. cerevisiae* reference genome (YPS 128) (Yue *et coll.* 2017) for the H_{div} crosses using Bowtie2 v2.1.0 (Langmead et Salzberg 2012) with default settings and the local alignment option. Raw reads were also mapped using the same method on a generated *S. paradoxus* MAT locus reference sequence for copy number variation and allele frequency of MATa and MATα specific sequences. The MAT locus reference were created by extracting the sequences of MATa and MATα with +/-1kb flanked regions from *S. paradoxus* reference genome (NRRL Y-17217) (Kellis *et coll.* 2003). Sequences were sorted and indexed with Samtools v1.8 (Li *et coll.* 2009). Coverage and variant-calling were performed as for the GBS analysis described above.

III.4.17-Statistical analyses

Survival curves were produced and analyzed using the R packages *survival* (Therneau 2015) and *survminer* (Kassambara A. 2018). The analysis of the correlation between sporulation and genetic divergence was performed using the glm R function to perform a logistic regression with the formula: Sporulation T_{end} ~ Genetic Divergence. Statistical analyses and figure creation for fertility data were done using custom scripts in R version 3.3.2 (R Development Core Team 2013). Figures and statistical analyses for mtDNA loss and sporulation capacity were performed using custom scripts in Python (version 3.6.3).

III.4.18-ade2-Δ colony coloration phenotype

Although we used the *ade2-Δ* marker as a visual aid to track loss of mitochondrial DNA, we still passaged strains that seem to have lost

mitochondrial DNA. During the experiment, some colonies from all crosses suddenly turned whitish or light orange. This paler coloration did correlate with the absence of sporulation in the strains. The only thing that changed during the evolution experiment is the yeast extract (EMD millipore, Burlington, USA), for which the lot number changed. Further testing suggests that the pink/red coloration of the *ade2-Δ* mutants is media dependent. On one of the yeast extract lot used, white colonies appear red and show slower growth while on the other, most colonies are white and show normal growth. The slower growth is common to all *ade2-Δ* strains (Figure III.S3).

III.5-RESULTS

III.5.1-Hybrid survival through serial bottlenecks

Not all lines could be propagated through the entire experiment. After 770 mitotic generations, 77.9% of the lines (524 out of the 672 initial lines) were still propagated using standard conditions. About 50% (n= 72) of the extinct lines were lost within the first 250 mitotic generations. This suggests that the loss of these lines is mostly due to genomic instability that arises rapidly after hybridization rather than spontaneous mutations, which would happen at a much slower pace (Lynch *et coll.* 2008). The L_{div}, H_{div} and one of the M_{div} crosses (M1) had a significantly lower proportion of surviving lines (averages of 69.3% for L_{div}, 70.3% for M_{div} and value of 79.2% for M1, Fig. 1D, Figure III.S4) compared to VL_{div} and M2 (average of 95.8% for VL_{div} and value of 91.7% for M2, Figure III.1D, $P < 0.01$, Log-rank test, Figure III.1D, Annexe C, Table III.S6). This suggests that hybrids from divergent parents may suffer from exacerbated genomic instability that lead to the rapid collapse of populations when faced with serial bottlenecks. The cause of line extinction remains to be investigated in detail, but the data suggests that it may be because of the frequent segregation of highly deleterious variants generated by genome instability. Indeed, the within-species control crosses of VL_{div}, which are expected to

be stable because of their low heterozygosity (Nishant *et coll.* 2010) and thus represent a measure of experimental noise, show little line extinction. In general, the extent of line loss correlates with genetic divergence of the parental strains, with the exception of the L_{div} crosses, which shows elevated line extinction. The L_{div} crosses is also the one that shows the greater level of ploidy instability (Annexe C, "Ploidy evolves following hybridization" section), suggesting that these hybrids are generally less stable. Therefore, these results suggest that yeast hybrids, even when maintained by mitotic division only, segregate very unstable clones at high frequency. This is also supported by the fact that the rate of loss decreases with time, which we would expect if unstable genotypes are eliminated through replication.

III.5.2-Hybridization rapidly leads to sterility

We measured two components of strain fertility, first by testing the ability to sporulate and then by measuring spore viability for sporulating strains. We induced sporulation (meiosis) and counted meiotic progeny survival in 214 randomly selected lines at three time points roughly corresponding to the initial (T_{ini}, right after hybridization, before first passage), middle (T_{mid}, 352 generations) and terminal (T_{end}, 770 generations) time points. For lines that did not survive until the end (58 out of 214), these timepoints were determined as the last available frozen stock (Annexe C, Table III.S5). Initial spore survival values were consistent with previous estimates for these types of crosses (averages of < 1% for H_{div}, 27.7% for M_{div}, 34.2% for L_{div} and 60.3% for VL_{div}, Figure III.S5) (Charron *et coll.* 2014). However, there were significant differences between the biological replicates (one way ANOVA $F(7,552)= 171.8$) in the VL_{div} (averages of 47.2% and 73.4, $P < 0.01$, Tukey HSD) and M_{div} crosses (averages of 36.4% and 18.1%, $P < 0.01$, Tukey HSD). These differences are probably due to strain specific genetic variation or even genomic architecture leading to variable levels of postzygotic isolation (Charron *et*

coll. 2014). Unexpectedly, spore viability could not be assessed for all the lines at T_{end} because 17% (n=37) of the tested lines lost their ability to enter sporulation. We found that the probability of successful sporulation at the end points considered is negatively correlated with parental divergence (Fig. 1E, r=−.76, P < 0.01, logistic regression). The loss of sporulation ability in yeast is multifactorial (Neiman 2011) but we hypothesized that it could be caused by mitochondrial malfunctions. Indeed, functional aerobic respiration is necessary for sporulation and thus requires the maintenance of functional mitochondrial DNA (mtDNA) (Küenzi *et coll.* 1974). The genotyping of two mitochondrial loci revealed a strong association between the loss of ability to sporulate and the absence of at least one mitochondrial marker (Fisher's exact test, odds ratio > 77, P = 1.43×10^{-5}, Figure III.S6C). This observation suggests that mtDNA instability could contribute to reproductive isolation among closely related yeast populations by leading to sterility, as shown for more distant species (Chou *et coll.* 2010), and that this effect increases with genetic distance. There are also rare cases in which lines lost their sporulation ability while both mitochondrial markers were detected (Figure III.S6), suggesting that the loss of mtDNA integrity is not the only cause of sporulation inability. As the Saccharomyces Genome Database (SGD) reports that there are more than 200 genes that, when knocked out in *S. cerevisiae*, lead to an absence of sporulation (Cherry *et coll.* 2012), any loss of function mutation happening in one of those genes could explain this phenotype. Although the ADE2 gene deletion strategy was used as a color reporter to avoid replicating "petite" mutants (Joseph et Hall 2004, Zhu *et coll.* 2014), it seems that partial or complete mtDNA loss in the hybrid backgrounds did not yield the expected "petite" phenotype. Despite the fact that lines do not show typical "petite" phenotypes on glucose, we found a significant association between the absence of sporulation capacity and poor growth on complex media containing glycerol as main carbon source (Figure III.S2B, P=3.97×10^{-10}, logistic regression), which is consistent with the

reduced of aerobic respiration. Moreover, we also found that sporulation ability is significantly associated with a change in red pigment saturation for colonies growing on the standard media used for the evolution experiment (Figure III.S2B, P=0.029, logistic regression). Although the increase of saturation with mtDNA loss was contrary to the phenotype expected from ADE2 deletants, this pigmentation change seems to be related with mitochondrial function, as it is well correlated with the ability to grow on glycerol (Figure III.S2B, $P=1.336x10^{-26}$, Pearson's r=0.547). These results suggest that the failure of mtDNA maintenance or function plays a major role in causing rapid and irreversible sterility in experimental yeast hybrids through the loss of sporulation ability.

III.5.3-Fertility evolves through time

To investigate whether the spore survival component of fertility improved over the experiment, we calculated a fertility recovery score (FRS) as the difference in spore viability between T_{end} and T_{ini} (Figure III.2A, B). As a point of comparison for fertility restoration with sexual reproduction, we performed 12 meiotic generations of intra-tetrad crosses (ITC) in randomly chosen diploid L_{div} lines and calculated FRS, all of which were positive (Figure III.2C, Figure III.S7). This is in stark contrast with mitotic lines, in which we found no statistically significant bias in FRS values as the distributions were unimodal and centered on 0 (Figure III.2B, Figure III.S8), showing that spore viability is as likely to increase as it is to decrease. To make sure that low spore viability was not due to the intrinsic inability of strains to produce viable spores, due to dominant mutations for instance, we performed autodiploidization on a random set of 16 haploid spores from the L_{div} and M_{div} crosses and this, at the three timepoints. In most cases, fertility was restored to more than 85% upon selfing (Figure III.2C), showing that infertility mostly derives from the presence of two divergent genomes in the same cell.

Although no general trend towards the recovery of spore viability was observed during the experiment, 23 individual lines presented statistically significant difference in their fraction of viable spores between T_{ini} and T_{end} (Fisher's exact test, fdr corrected, Table III.S7). Eleven of those 23 lines (5 VL_{div}, 5 L_{div} and 1 M_{div}) showed a decreased fertility and the remaining 12 (7 L_{div}, 4 M_{div} and 1 H_{div}) showed improvement. Among the lines that showed improvement, seven (4 L_{div}, 2 M_{div} and 1 H_{div}) displayed spectacular recovery in spore viability within 352 mitotic generations: their FRS values were close to that of ITC lines (Figure III.2B), and their spore viability was similar to what is typically observed for their non-hybrid diploid parents (Leducq *et coll.* 2016). One line presented a low FRS value, which is explained by the fact that its fertility returned to its initial value by T_{end} (Figure III.2B). As infertility of *Saccharomyces* hybrids is mainly due to anti-recombination caused by the mismatch repair machinery acting on homeologous chromosome pairs, leading to chromosome missegregation (Greig *et coll.* 2003, Rogers *et coll.* 2018), there could be at least three main explanations for these sudden increases in fertility. All of these mechanisms rely on the loss of heterozygosity across the genome and re-establishment of correct chromosome pairing during meiosis.

The first potential mechanism would be an endoreduplication event, i.e. spontaneous chromosome doubling following a failed cell division during mitosis (Harari *et coll.* 2018). Such an event would lead to the production of identical homologues that would restore correct chromosome segregation. The second mechanism would be damage to a copy of the MAT locus that would convert the diploid hybrid into behaving as a gamete. Two such diploid gametes could then mate, generating a fertile tetraploid hybrid. This path to fertility recovery was recently observed in hybrid species of the *Zygosaccharomyces* genus (Ortiz-Merino *et coll.* 2017, Braun-Galleani *et coll.* 2018), so it is in principle an accessible path to fertility recovery. However, this would need rare events to co-occur in the same colony and to produce two diploid gametes of opposite mating type.

The last potential mechanism would be that strains could have sporulated during the experiment and spores of opposite mating types could have mated. This would be the equivalent of our ITC lines where sometimes a single cross between two spores can bring fertility back to high values (Figure III.S7). This third option is very unlikely because sporulation happens under very specific environmental conditions, principally nitrogen starvation. The frequency of streaking to fresh media during the experiment would prevent such depletion to happen. In addition, this scenario would often lead to spores that are aneuploid, making fertility recovery unlikely even after mating. All these mechanisms would generate strains with increased spore viability, but the lines are expected to show a change from diploidy to tetraploidy in the first two scenarios, allowing to differentiate these mechanisms. Mitotic loss of heterozygosity could also be involved but would not be expected to lead to such dramatic recovery of fertility (Dutta *et coll.* 2017). We tested these potential mechanisms by measuring the total cellular DNA content of the lines to infer ploidy, genome-wide genotyping and whole-genome sequencing of some of the strains.

III.5.4-Ploidy evolves following hybridization

We measured ploidy in the 672 lines at the three time points T_{ini}, T_{mid} and T_{end} using DNA staining and flow cytometry. Surprisingly, these analyses revealed that some lines already deviate from diploidy after hybridization. While almost all the hybrids are diploid, both independent L_{div} hybrids show frequent triploidy (average at T_{ini} of 54% triploid lines) (Figure III.3A, B). It appears that this triploidization is a major driver of low initial spore viability in the L_{div} hybrids (both crosses), with an average reduction of 45% compared to diploids (20.7% compared to 37.6% considering all time points, Figure III.S9). This triploidy could either be a consequence of aneuploidies that led to an overall DNA content equivalent to triploidy, or as a consequence of whole genome duplication of one of the

parental genomes. We examined the genotype of hybrids using genotyping-by-sequencing (GBS) and found that at T_{ini}, all triploid hybrids were composed of two copies of the *SpC* genome and one copy of the *SpB* genome (Figure III.3C), suggesting that the change in ploidy predates mating. This would be possible if some *SpC* haploid cells were in fact diploid. We indeed observed a small fraction of diploid clones in both parental *SpC* haploid stocks (Figure III.S10). These clones were identified pseudo-haploids, i.e. diploid but competent for mating (Figure III.S11). Triploidy appears to be frequent in the L_{div} cross and was observed in the two biological replicate matings performed with independent strains of *SpC* and *SpB* (Figure III.1). It however does not seem to occur in all crosses between these two species (Figures III.S12, and III.S13) and is variable among replicates (Annexe C, Table III.S8) suggesting some stochastic effects within and between strains. These results suggest that the stochastic duplication of the *SpC* genome happened prior to mating and that some *SpC* haploid strains may be prone to spontaneous genome doubling.

Consistent with the hypothesis above based on spore viability recovery by polyploidization, the 7 lines that displayed significant spectacular fertility recovery and one L line that showed significant fertility improvement, but of lower magnitude, during evolution became tetraploid (Figure III.4). This is also true even for hybrids between *S. paradoxus* and *S. cerevisiae*, which were initially completely sterile (Figure III.4). One of the two tetraploid M1 lines (M1_40) returned to diploidy at T_{end} but this is more likely to be due to segregating ploidy and colony heterogeneity at T_{mid} rather than return to diploidy (for more details see Annexe C, Supplementary text). GBS analysis revealed that tetraploid hybrids have equal copies of parental genomes, since allele frequencies across 171 markers among the 16 chromosomes are around 50% (Figure III.3C, Figures III.S14, III.S15 and III.S16).

As stated above, whole genome doubling could occur either by endoreduplication which is a consequence of cytokinesis failure (Edgar et Orr-Weaver 2001) or by means of damage to one copy of the MAT locus in the hybrid (Ortiz-Merino *et coll.* 2017, Harari *et coll.* 2018) (Figure III.S17). This damage to the MAT locus could cause hybrid cells to behave as a haploid, switch mating type and hence autotetraploidize. In this experiment, mating type switching may not occur using the standard process because the necessary HO gene was deleted. The main way by which autotetraploidization could occur by mating in our study is to have two hybrids with damage to the opposite MAT loci that are in the same colony and are close enough to mate with each other (Figure III.S17), which is a very unlikely event. To investigate this, we sequenced the genome of the 8 tetraploid lines (5 L_{div}, 2 M_{div} and 1 H_{div}) at T_{ini} and T_{end}. We indeed found that the frequency of parental alleles across the genomes are roughly 50%, showing that the strains are not aneuploids that would have DNA content equivalent to tetraploidy (Figure III.3D, Figures III.S18 and III.S19). One exception is observed for the M1_40 line that show allele frequency corresponding to a triploid state at T_{end} (Figure III.3D, Figure III.S18) while GBS data show a tetraploid state at T_{mid} and a diploid state at T_{end} (Figure III.S18). A different colony was isolated each time from these time-points. This is again consistent with segregating ploidy and colony heterogeneity which are probably due to extreme genomic instability in this particular line (for more details see Annexe C supplementary text). Next, we investigated the total or partial chromosome loss or LOH of the MAT locus region (Figure III.S17). We identified only two tetraploids with aneuploidy on chromosome III containing the MAT locus (Figures III.S20, III.S21 and III.S22). However, these aneuploidies affect only one copy of the mating type (Annexe C Supplementary text and Figure III.S22). We thus find no evidence of damage to the MAT locus of tetraploid hybrids that could have caused mating between diploid hybrids. Thus, these results suggest that endoreduplication is the most likely mechanism of

whole genome doubling. However, we cannot exclude that damage to the MAT loci occurring by the loss of chromosome III could be undetectable because this chromosome could have been regained following tetraploid formation. Chromosome III was indeed found to be the most unstable chromosome among *S. cerevisiae* diploids, triploids and *S. cerevisiae-S. bayanus* hybrids (Kumaran *et coll.* 2013). Furthermore, transient dynamics of aneuploidy were repeatedly observed in yeast laboratory evolution experiments under stressful conditions (Yona *et coll.* 2012, Gilchrist et Stelkens 2019). Such successive aneuploidies could happen in our hybrids lines because aneuploidy is prevalent (Figure III.3C, Figure III.S19) and tetraploids are notorious for showing genome instability (Selmecki *et coll.* 2015).

III.6-DISCUSSION

We measure the rate of recovery from reproductive isolation in yeast hybrids using experimental evolution. We propagated parallel hybrids with serial bottlenecks to measure the neutral rate of recovery. We eliminated the potential confounding effect that natural selection on growth rate could have on genome dynamics and consequently on components of fertility.

One of the most striking changes to fertility we observed is a complete loss of the ability to sporulate within thirty generations, effectively reducing fertility to 0. While decrease in sporulation efficiency was observed in similar experiments performed on homozygous strains (Zeyl *et coll.* 2005, Hill et Otto 2007), such complete loss of the ability to sporulate in strains that initially were able to sporulate at high efficiencies was, to our knowledge, not reported before. Our results show that a greater genetic distance between the founding parents lead to higher probability of losing the ability to sporulate, suggesting that genome instability or genetic incompatibilities could cause this decrease of fertility.

This, added to the results indicating that those strains also lost some or all of their mitochondrial DNA and growth defects on non fermentable sources, indicate that genetic incompatibilities or instabilities involving the mitochondrial genome may be responsible for these fertility losses.

We find that reproductive isolation as assessed by spore viability does not have a global directional fate during the mitotic propagation of hybrid lines. The different scenarios presented in figure 1A were almost all observed. Overall, spore viability of hybrids evolved like a neutrally evolving quantitative trait, with incremental gains and losses, with no overall particular direction. However, we did observe spectacular punctuated improvements in spore viability for some lineages. From the 23 lines that had statistically significant differences between the initial hybridization and the end of the experiment, fertility decreased for 11 of them. Reductions in spore viability was observed in all crosses except the H_{div} crosses, which were already completely infertile at the beginning of the evolution and could not be reduced further. One explanation for such decrease includes the accumulation of genomic rearrangements, which would lead to incorrect segregation of chromosomes during meiosis to various degrees depending on the extent of the rearrangements (Liti *et coll.* 2006, Charron *et coll.* 2014). The segregation of recessive lethal alleles following a *de novo* mutation could also be implicated and would lead to strains exhibiting halved fertilities, which were rarely observed, but more complex patterns of fertility decrease are also possible given the potential for genetic interactions in these heterogeneous genetic backgrounds.

Punctual and almost complete restoration of fertility was observed at low frequency in all but the lowest parental divergence crosses. We show that these lines experienced a genome duplication, most likely caused by the doubling of all chromosomes. It was shown before that artificially induced chromosome doubling in infertile hybrids between *S. paradoxus* and *S.*

cerevisiae hybrids could restore fertility (Greig *et coll.* 2002). Our results show that this happens spontaneously, without the need for natural selection and in between species that can naturally hybridize. Finally, a more gradual form of statistically significant recovery was also observed in about six lines which were not subjected to genome duplication. For four of those lines, fertility almost doubled compared to their ancestral lines. The other two have improvement of 3 and almost 5 times their initial fertilities. As mitotic recombination often leads to gene conversion events spanning tens of thousands of base pairs (Lee *et coll.* 2009), accumulation of a large number of mitotic loss of heterozygosity events might allow such recoveries. More in depth genomic analyses will therefore be needed to understand the basis of these recoveries. This would suggest that the contribution of mitotic recombination to the recovery of fertility in hybrid lineages could be a slow process compared to meiotic recombination and whole-genome duplication.

It is important to note that our results also show that there is some variation for fertility and ploidy 1) within the colonies that were replicated and frozen, 2) among lines within a cross and 3) between biological replicates of the same divergence category. This variation could be due to the interactions of the parental genetic backgrounds, generating instability in the hybrid lines, and their elevated instability that starts in the zygotes after the initial matings. Therefore, taking into account the initial genetic background of experimentally evolved hybrids and considering multiple independent hybridization events is crucial for the understanding of their possible evolutionary fates.

Sexual reproduction is thought to be extremely rare in yeast (Tsai *et coll.* 2008) and thus may not always explain successful hybridization and introgression in the wild. The slow rate of improvement of spore viability by mitotic growth alone, without change in ploidy, could therefore contribute

to improved fertility because mitotic proliferation is much more frequent than sexual reproduction. This could be even more likely for instance if natural selection could accelerate the loss of heterozygosity, although this remains to be tested. Our data shows that hybrids can recover fertility by becoming allopolyploids in less than 400 cell divisions, which could represent less than 250 days in nature (Leducq *et coll.* 2016). In principle, fertility recovery could therefore happen before sexual reproduction occurs. The rate of fertility restoration by genome doubling observed in our experiments might be an underestimate of the rate that would occur in yeast hybrids that contain the HO gene. In the event of damage to one of the mating type loci, the wild type HO would allow the mating type switch and self mating in hybrids, leading to a duplicated genome. How frequently whole-genome duplication occurs in nature remains to be examined but it could contribute to important events. For instance, it was shown recently that the whole-genome duplication (Marcet-Houben et Gabaldon 2015) in an ancestor of the *Saccharomyces* genus originated from inter-species hybridization. The success of such an event would have been unlikely if the F_1 hybrids were completely sterile and if sexual reproduction would have been needed to recover fertility. Our results suggest that whole-genome duplication could have happened spontaneously and neutrally, thus restoring fertility at the same time. Fertility recovery without sex is likely to apply to multicellular eukaryotes as well because somatic chromosome doubling in diploid tissues or zygotes can lead to the emergence of polyploids, which may display both restored fertility and reproductive isolation with parental species (Ramsey et Schemske 1998). While polyploidy has been shown to be prevalent in plants (Grant 1981), polyploidy is not common in animals (Muller 1925, Orr 1990). However, the animal lineages containing stable polyploids species often have access to asexual reproductive strategies such as parthenogenesis which improve tolerance to polyploidy and could enable hybrid fertility restoration without sex (Mable 2004). As ploidy changes contribute to restore fertility in

partially fertile hybrids, we also find that ploidy instability generates triploids in some of our crosses, which contributes to poor fertility. Ploidy changes are therefore a double-edged sword, causing both reproductive isolation and fertility recovery.

III.7-AUTHORS' CONTRIBUTION

Conceptualization, C.R.L., G.C., M.H., S.M.; Data curation, G.C, S.M and M.H; Funding acquisition, C.R.L.; Experimental work, formal analysis and interpretation: Cell propagation by G.C., S.M., M.H., Fertility by G.C., Ploidy and GBS by S.M. Genome sequencing by S.M. and M.H, Mitochondrial DNA work by M.H, R scripts support for ploidy/GBS by H.M. Project supervision, C.R.L.; Writing - original draft, G.C. and S.M.; Writing – review & editing, all authors

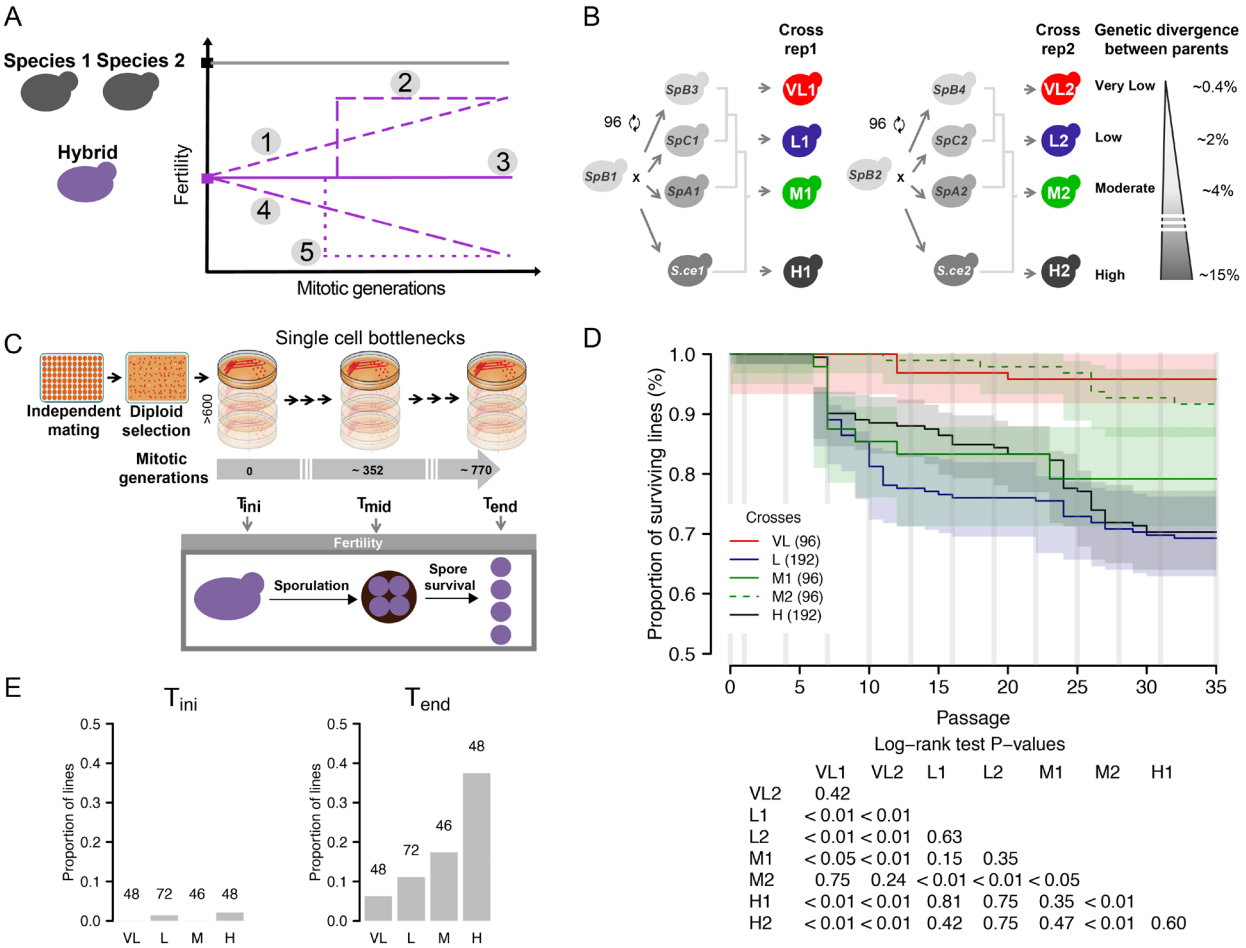
III.8-ACKNOWLEDGEMENTS

Acknowledgments: We thank the members of the Landry lab for discussions and J. Hallin, A. K. Dubé, C. Eberlein, A. Fijarczyk, N. Aubin-Horth, J.B. Anserson, L. Kohn, S. Otto, C. Mérot and A.M. Dion-Coté for comments on the manuscript. We thank C. Mérot for help with the library construction protocols and A. Fijarczyk for help with genome data analysis.

III.9-FUNDING

NSERC Discovery grant and Canada Research Chair to C.R.L, FRQNT scholarship to G.C and M.H, NSERC Alexander Graham-Bell scholarship to G.C. and FRQS post-doctoral fellowship to S.M.

III.10-FIGURES



	VL1	VL2	L1	L2	M1	M2	H1
VL2	0.42						
L1	< 0.01	< 0.01					
L2	< 0.01	< 0.01	0.63				
M1	< 0.05	< 0.01	0.15	0.35			
M2	0.75	0.24	< 0.01	< 0.01	< 0.05		
H1	< 0.01	< 0.01	0.81	0.75	0.35	< 0.01	
H2	< 0.01	< 0.01	0.42	0.75	0.47	< 0.01	0.60

Figure III.1: Neutral evolution of yeast hybrids shows the effect of genetic divergence on the evolution of viability and sporulation ability.

(A) Evolution of fertility during mitotic proliferation of hybrids between species (S1 and S2): Potential scenarios: (1) Gradual recovery over time, (2) rapid sudden recovery, (3) no significant recovery, (4) decline over time, (5) rapid and sudden decline. (B) Crosses were performed among S. paradoxus lineages (VL_{div}, L_{div} and M_{div}) and with S. cerevisiae (H_{div}). Each type of cross involves two biological replicates, i.e. involving found independent strains, and each individual cross was performed independently to represent independent hybridization events. (C) The 672 hybrids were evolved in conditions of weak selection to examine the neutral spontaneous changes of fertility. Mitotic propagation was performed through repeated bottlenecks of single cells. Fertility was measured by estimating spore viability after meiosis at T_{ini}, T_{mid} and T_{end}. (D) Survival rates vary among lines. Timing of each glycerol stock indicated by vertical grey lines. Logrank test pairwise comparisons FDR corrected P-values are shown. (E) Fraction of lines that lost their sporulation capacity at T_{ini}, T_{mid} and T_{end}. The number of strains tested per cross type is indicated over the corresponding bars.

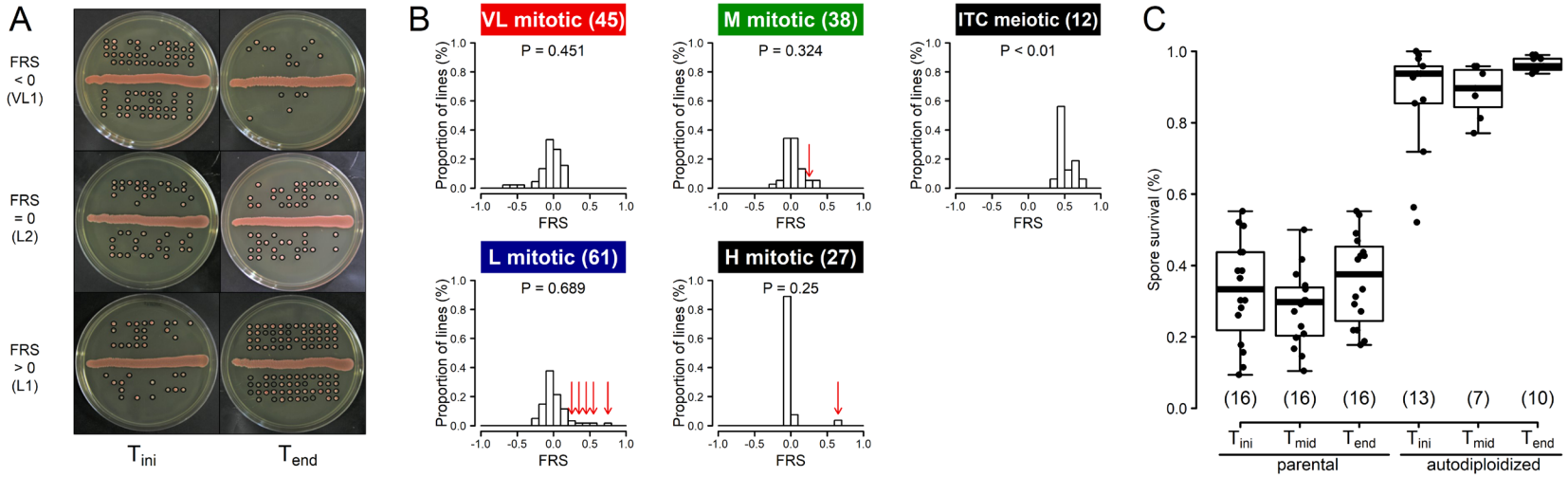


Figure III.2: Hybrids gain and lose fertility at similar rates while sexual reproduction systematically improves fertility.

(A) Example of spore survival from lines showing a reduction of fertility (top), no change of fertility (middle) and recovery of fertility (bottom). Spores considered as viable are circled in black. Spores from sporulation at the initial time point are shown on the left and from the final time point on the right/ (B) Fertility recovery scores (FRS), which measure the change in spore viability, show no clear directionality in mitotically propagated lines (VL_{div}, L_{div}, M_{div} and H_{div}). P-values given for exact binomial tests. FRS for intra-tetrad crosses (ITC) indicate large and systematic increase in fertility. Bins containing strains with high fertility by the midpoint of the experiment highlighted with red arrows. (C) We performed autodiploidization on a random set of 16 haploid viable spores dissected from the L_{div} and M_{div} crosses at T_{ini}, T_{mid} and T_{end}. Fertility of initial hybrids and evolved lines (parental) increases to above 80% after spore autodiploidization. Numbers in parentheses represent sample sizes.

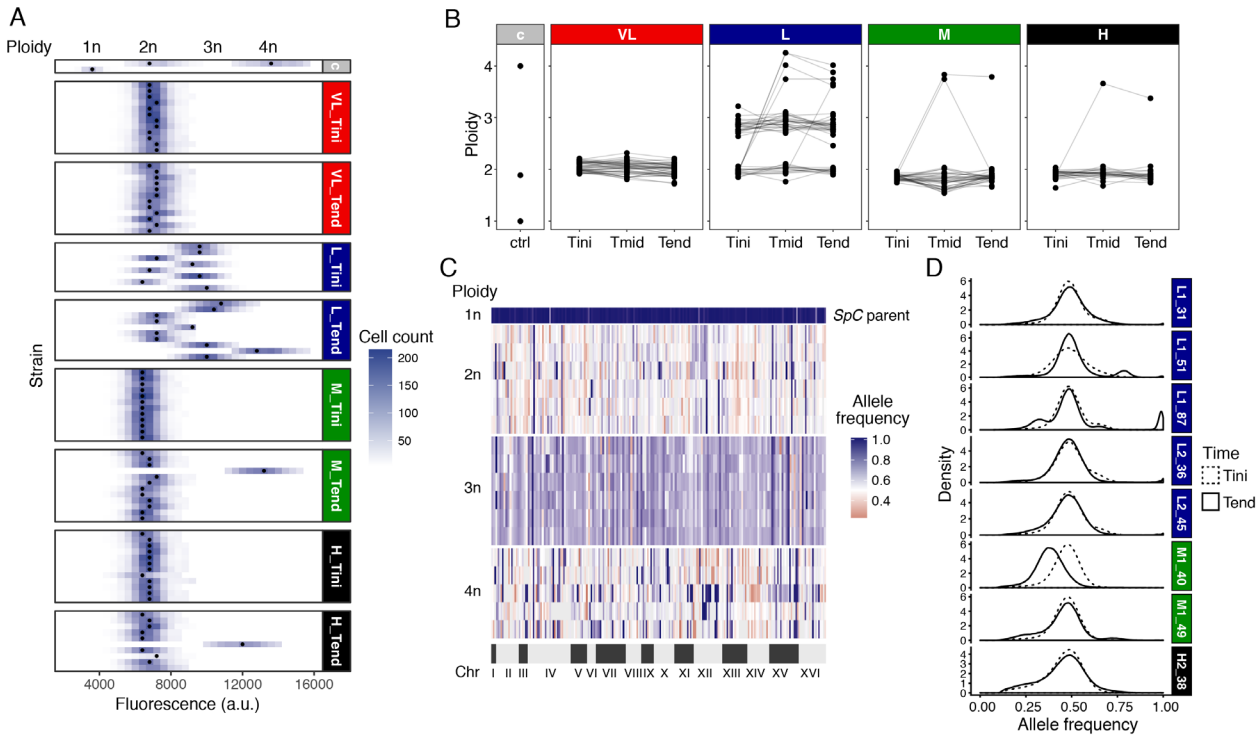


Figure III.3: Ploidy varies among hybrids and evolves through time.
(A) Ploidy of a subset of hybrid lines after hybridization (T_{ini}) and after ~770 (T_{end}) of mitotic generations. The c grey panel corresponds to controls. (B) Ploidy at the three tested timepoints (T_{ini}, T_{mid} and T_{end}). Connected dots represent independent lines. The c grey panel corresponds to controls. (C) Frequencies of 171 markers across the genome show around 50% of *SpC* alleles in the diploid and tetraploid strains and 66% in triploid strains.

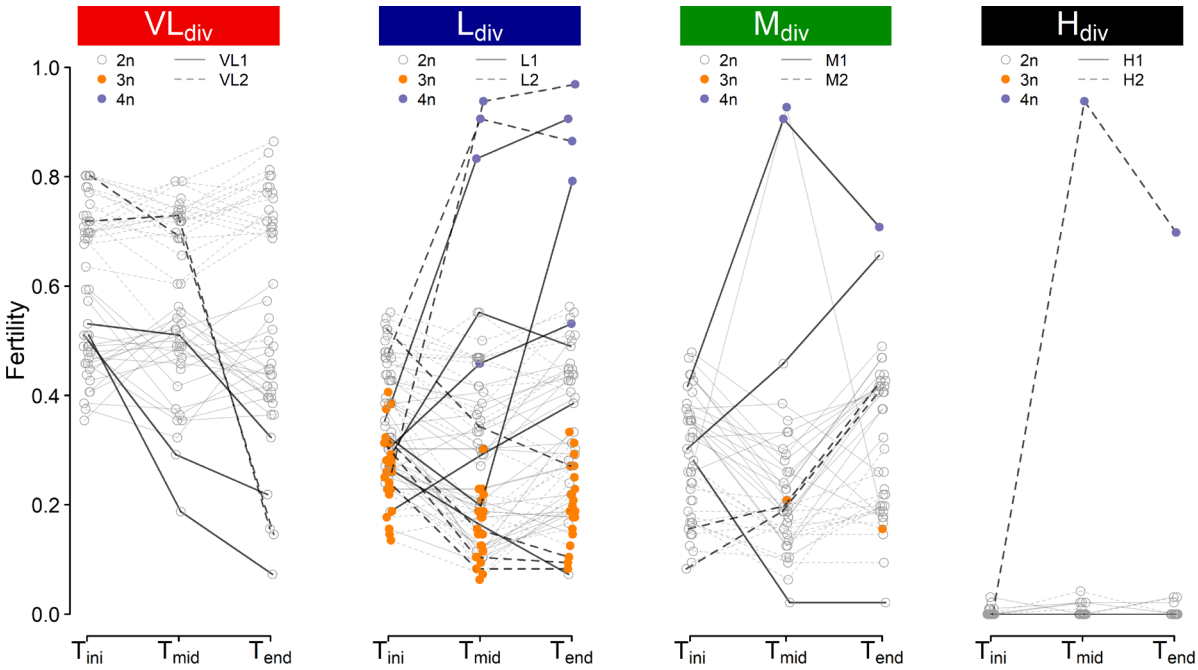


Figure III.4: Tetraploidization leads to sudden fertility recovery over time.

Fertility trajectory at the three timepoints (T_{ini}, T_{mid} and T_{end}). Each connected set of dots represents an independent line. The colors correspond to ploidy. Dotted and full lines represent the two independent crosses within each genetic divergence class. Bolder lines correspond to the lines with significantly different proportions of viable spores between T_{ini} and T_{end}. In the case of VL1 and VL2, the two pairs of strains show different level of fertility, which is frequent in yeast intra-species crosses.

III.11-SUPPLEMENTARY FIGURES

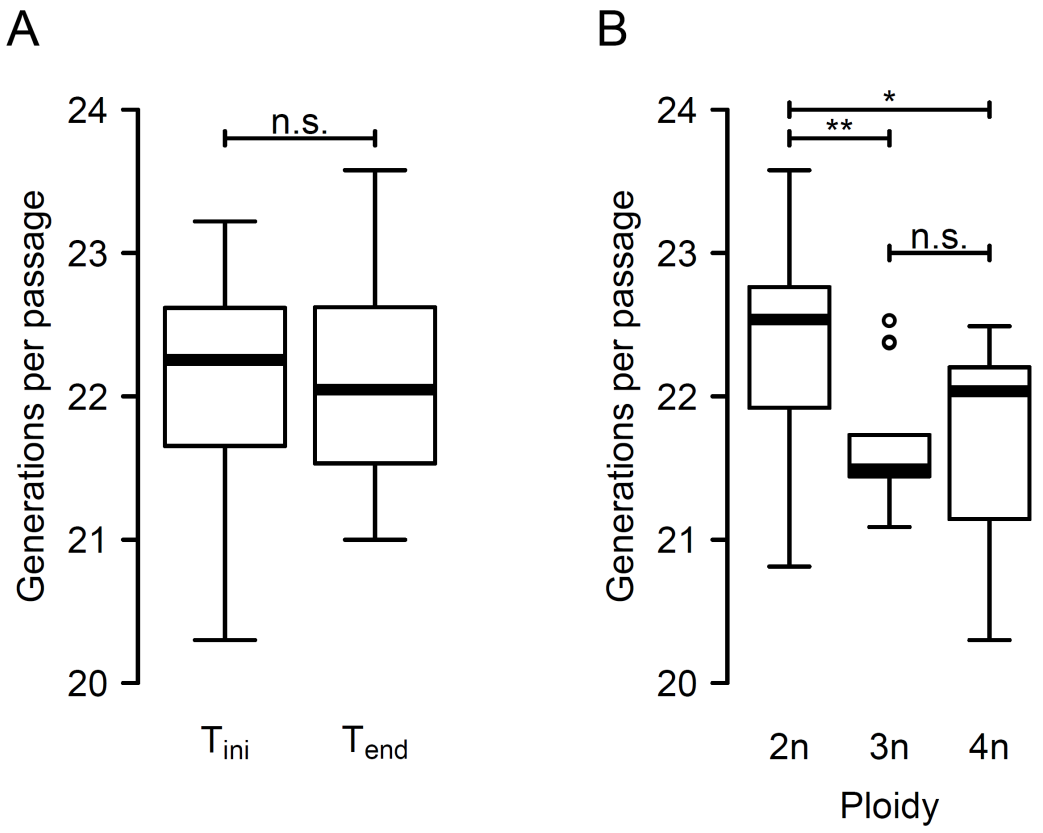


Figure III.S1: Number of mitotic generations per passage is stable through time but differ between hybrids with different ploidies.
Number of mitotic generations estimated as the $\log_2$ of the number of cells in colonies after a growth cycle by following the experimental procedure of the evolution experiment. This was performed for a subset of 40 lines. The number of cells was estimated by flow cytometry of a colony resuspended in water. (A) There is no statistically significant difference in the number of mitotic generations when considering the timepoints (T_{ini} or T_{end}) of the experiment (Welch two sample t-test, P = 0.902). (B) but there is a significant effect of ploidy (one-way ANOVA ($F(2,60) = 9.256$, Tukey HSD, "**" $P < 0.01$, "*" $P < 0.05$, "n.s." non-significant), showing that changes to higher ploidy was not favored by selection during the experiment.

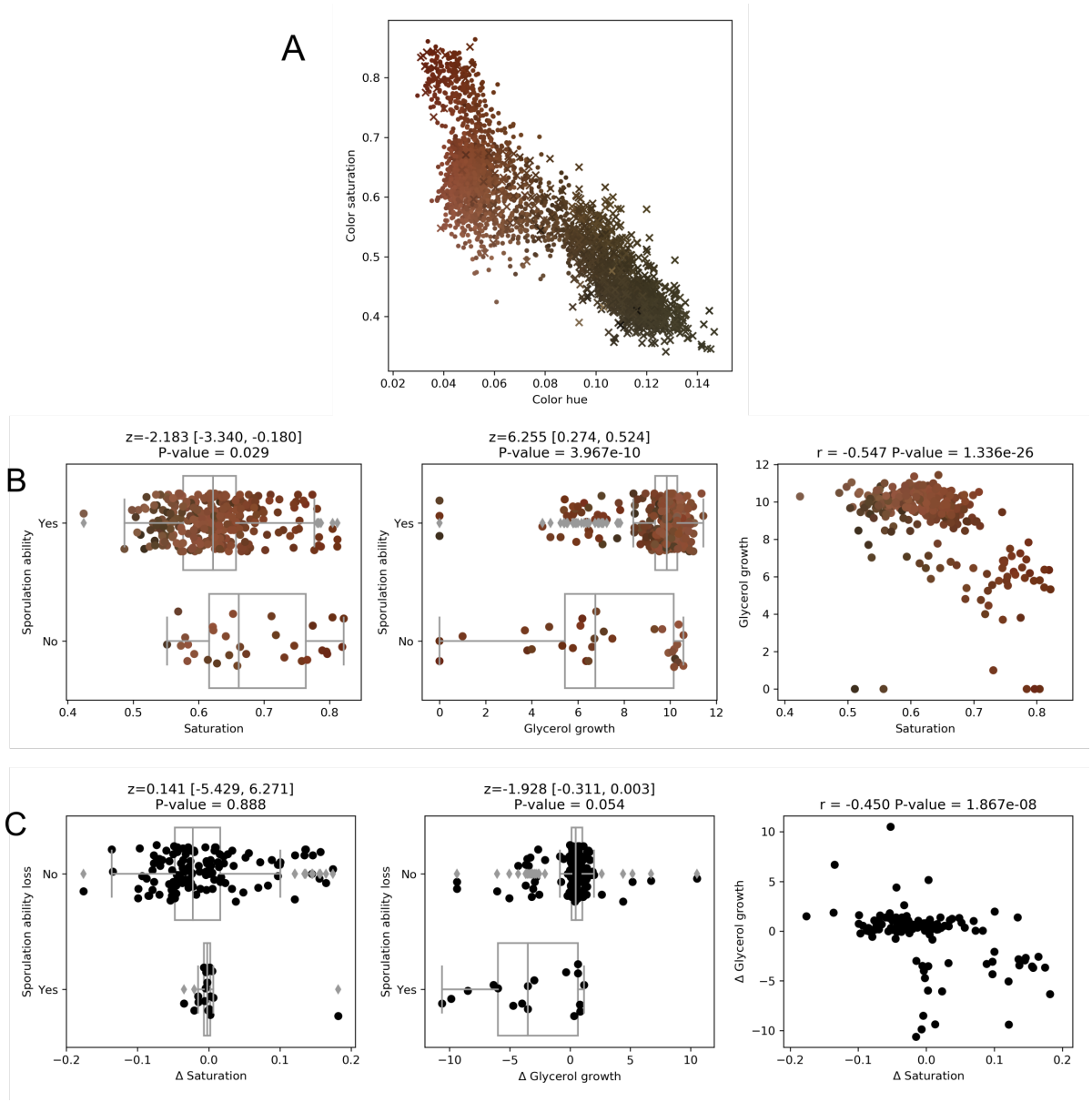

A
Color saturation
Color hue
B
z=-2.183 [-3.340, -0.180]
P-value = 0.029
Sporulation ability
Yes
No
Saturation
z=6.255 [0.274, 0.524]
P-value = 3.967e-10
Glycerol growth
r = -0.547 P-value = 1.336e-26
C
z=0.141 [-5.429, 6.271]
P-value = 0.888
Sporulation ability loss
Δ Saturation
z=-1.928 [-0.311, 0.003]
P-value = 0.054
Δ Glycerol growth
r = -0.450 P-value = 1.867e-08
Δ Glycerol growth
Δ Saturation

Figure III.S2: Sporulation inability is associated with increase in red pigment saturation and decrease of growth on glycerol.
(A) Raw data of the colony image analysis shows that color hue discriminates between background and colonies, while color saturation captures the variation in red pigmentation intensity. Dots: data points kept for the analysis, cross symbols: data points filtered out (see Methods). (B) Sporulation ability is significantly associated with color saturation (left) and growth on glycerol (center). P-values and 95% confidence intervals for logistic regressions are shown. Saturation and growth on glycerol are correlated (right). P-value for Pearson's correlation is shown. (C) Sporulation ability loss with time shows no association with change in color saturation (left) and a marginally significant association with change in growth on glycerol (center). P-values and 95% confidence intervals for logistic regression are shown. Variations of saturation and growth on glycerol are correlated (right). P-value for Pearson's correlation is shown.

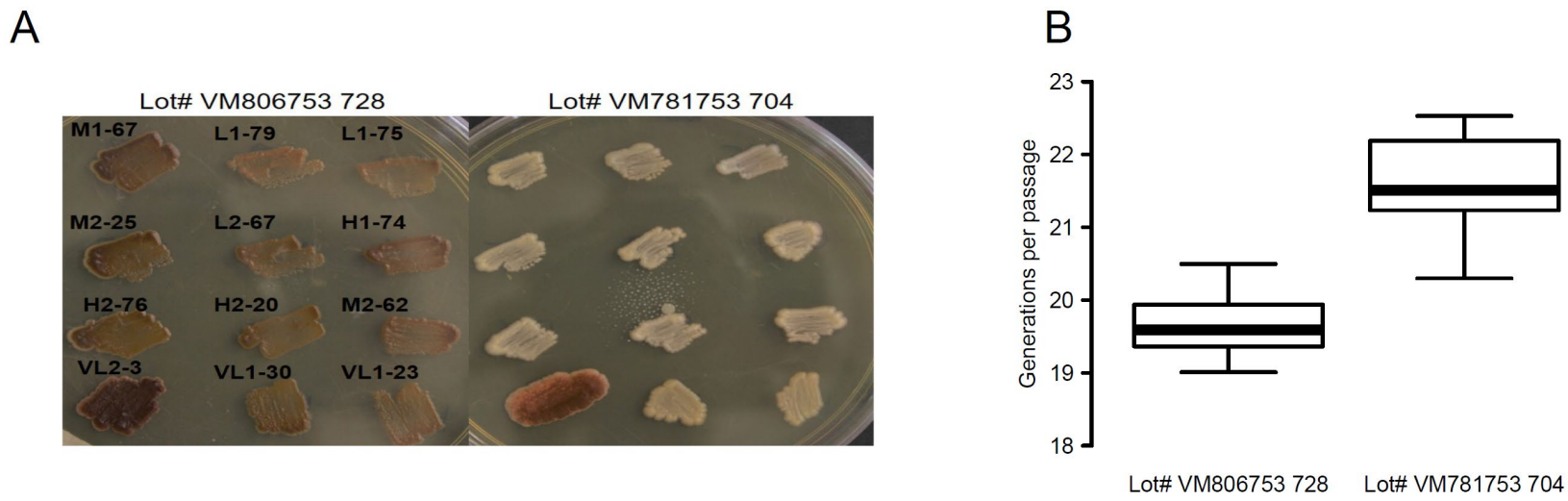


Figure III.S3: The ade2Δ coloration phenotype and growth rate is media dependent.
(A) Sporulation negative strains appear to be red when grown on YPD prepared with yeast extract from one lot (panel A left). Using the same ingredients but changing the lot number of the yeast extract yielded a very different coloration for all tested lines but one (panel A right). Evolution line numbers indicated above the cell patches. Plates were prepared and inoculated on the same day. Photos were taken after 5 days of growth. (B) Generations per passage on the two different media measured by cytometry on a subset of triploids and tetraploids that were sporulation positive.

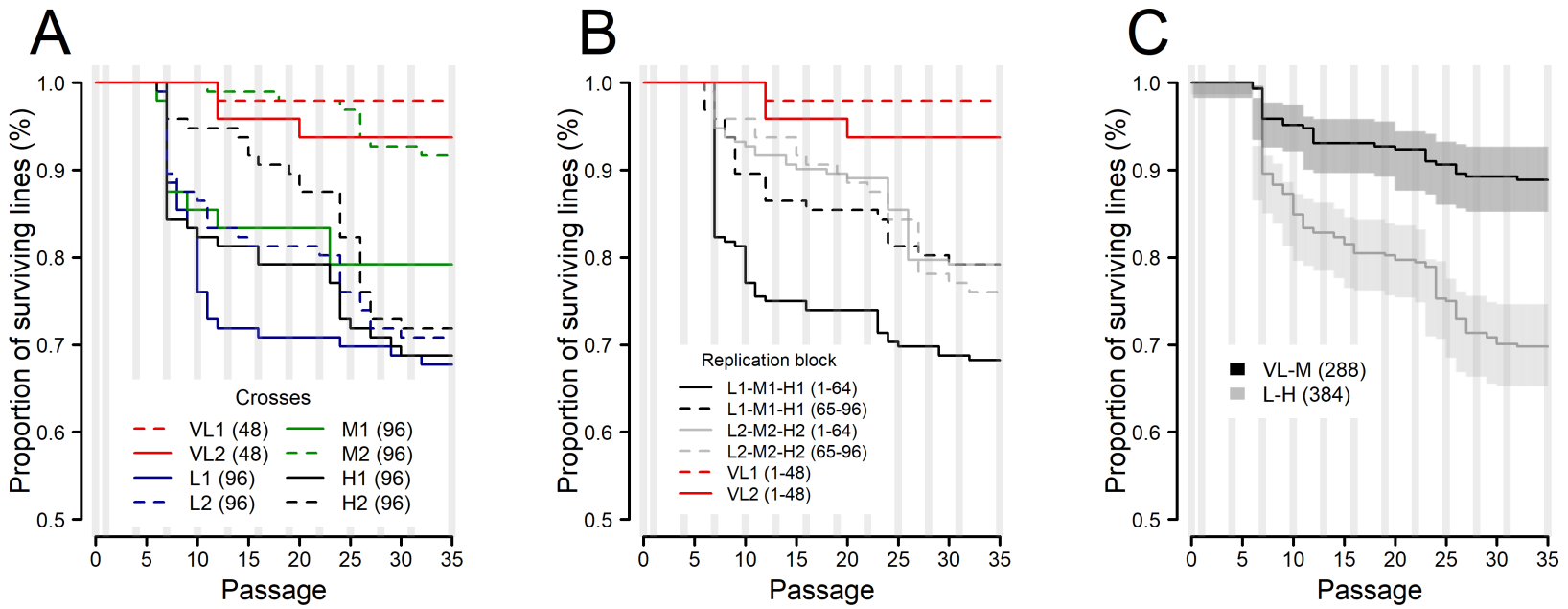


Figure III.S4: Survival of the evolution lines.
Survival of the lines grouped by (A) individual crosses, (B) replication sets used during the evolution experiment and (C) crosses showing high (VL-M) and low (L-H) survival.

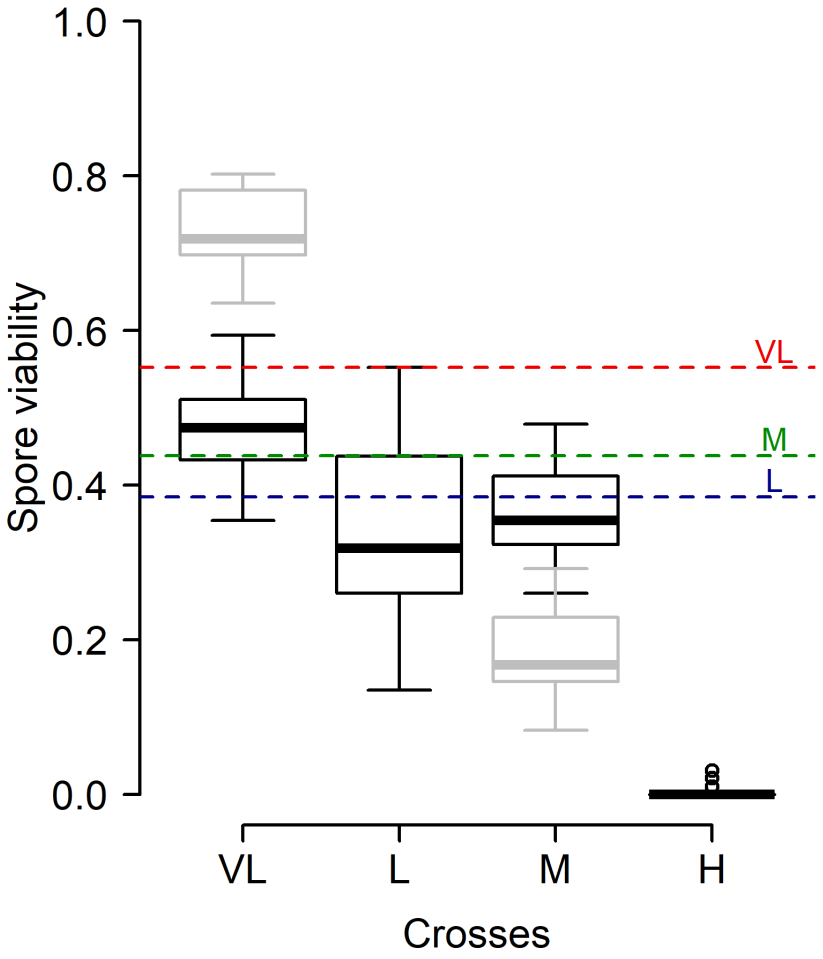


Figure III.S5: Ancestral lines show expected spore viabilities for their cross types.

Boxplot of spore viabilities for the different hybrid lines at T_{ini}. A single box represents the average for the two crosses when their mean is not different. In case of significant difference in mean Black and grey boxes represent the first and second crosses respectively. Dashed colored lines represent median spore viabilities values for each of the cross types from Leducq *et coll.* 2016.

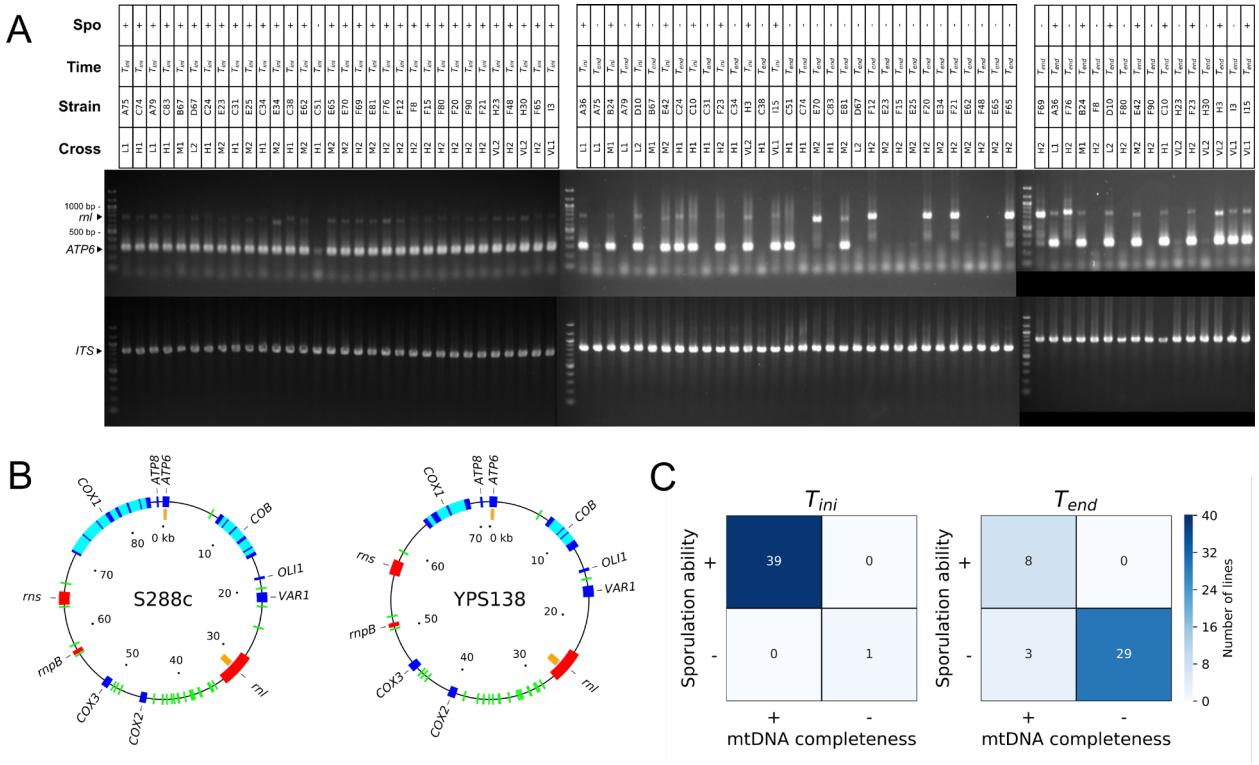


Figure III.S6: Partial or complete mtDNA deletion is associated with the loss of sporulation ability after 770 mitotic generations.

(A) PCR assays for the rnl and ATP6 mitochondrial loci and for the ITS nuclear locus. Presence or absence of these loci was assayed for 32 lines which lost the ability to sporulate at the end of the mitotic evolution experiment and 8 lines that maintained the ability. PCRs were performed on DNA extracted from stocks from T_{ini} and T_{end} timepoints. The rnl amplicon is expected to vary in size around 700 bp, while the ATP6 amplicon is expected to be 285 bp long. (B) Map of S. cerevisiae (S288c) and S. paradoxus (YPS138) mtDNAs. The rnl and ATP6 amplicons are shown with orange shapes. Protein-coding genes are shown in blue, protein-coding gene introns in cyan, RNA-coding genes in red and tRNAs in green. The genome annotations used are from Yue et al. 2017 2. (C) Partial or complete loss of mtDNA is associated with loss of sporulation ability. For T_{ini} and T_{end}, contingency tables show the counts of lines according to mtDNA completeness (+: both markers present, -: at least one marker absent) versus sporulation capacity. The association is significant for T_{end} (Fisher's exact test, odds ratio > 77, P-value=1.43x10-5).

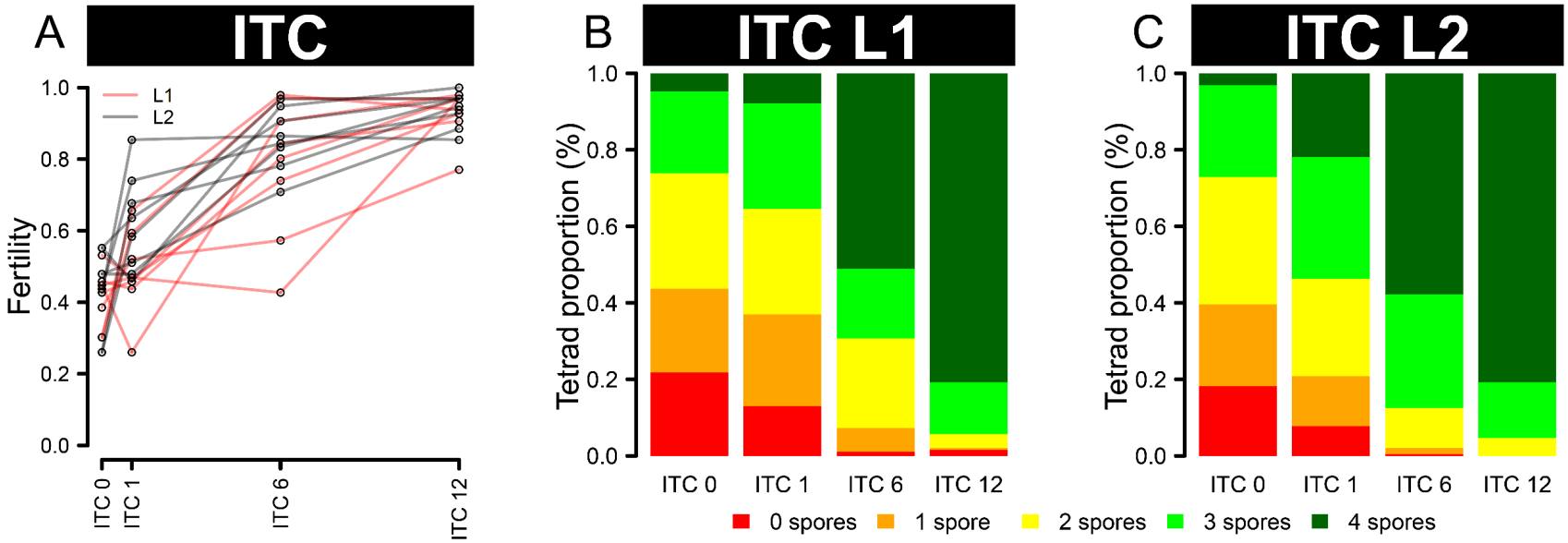


Figure III.S7: Intra-tetrad mating restores hybrid fertility.
(A) Fertility trajectories of 16 hybrids through 12 sporulations followed by intra-tetrad mating events. Line colors indicate the SpB×SpC cross identity. (B) and (C) Combined proportions of tetrad types per cross type after hybridization (ITC 0) and after 1, 6 and 12 intra-tetrad cross rounds.

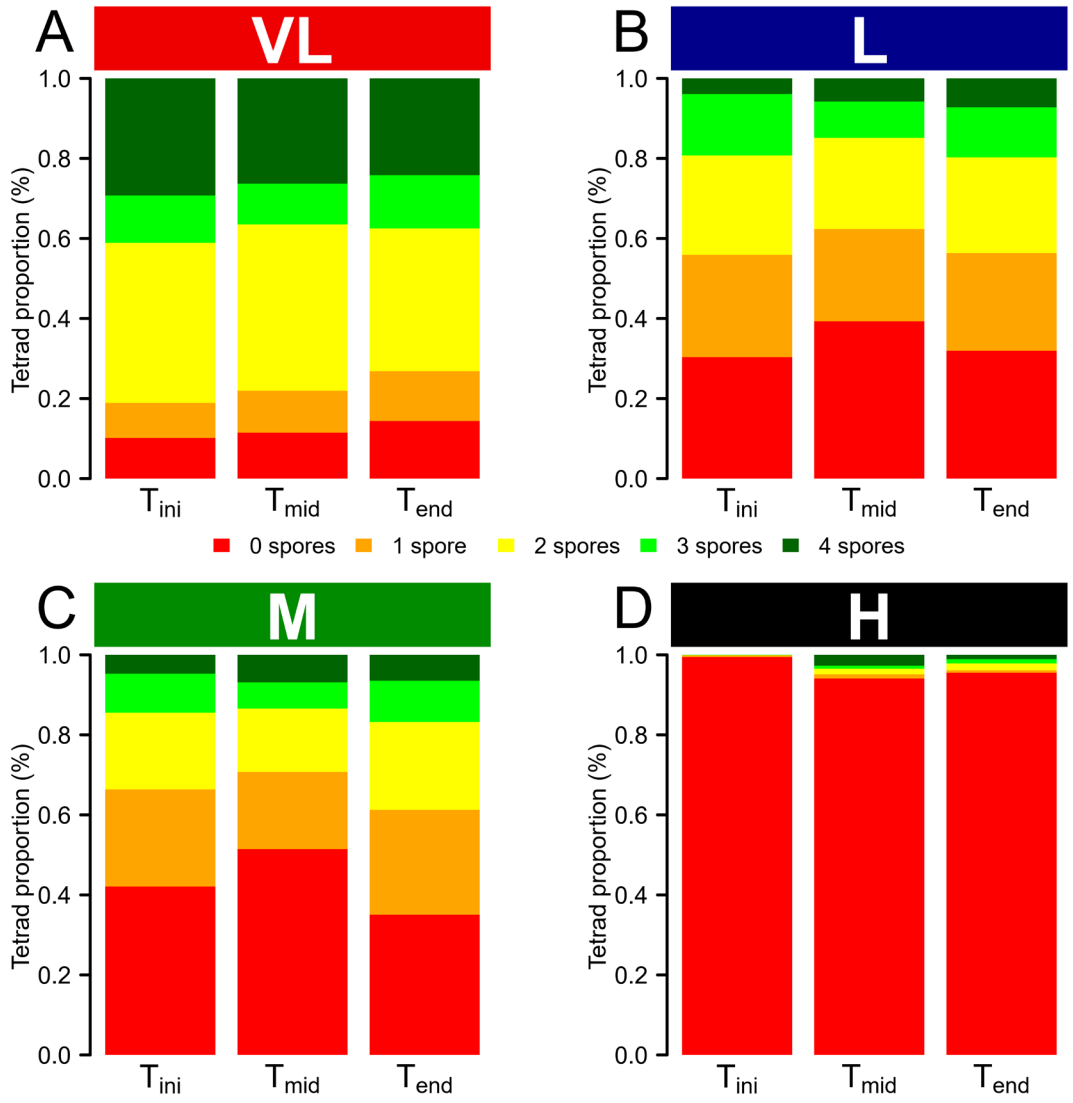


Figure III.S8: The proportions of tetrad types do not change during evolution.

The proportions of possible tetrad types per time point tested for the (A) VL_{div}, (B) L_{div}, (C) M_{div} and (D) H_{div} crosses. The data of the two independent crosses for each cross type were merged for this analysis.

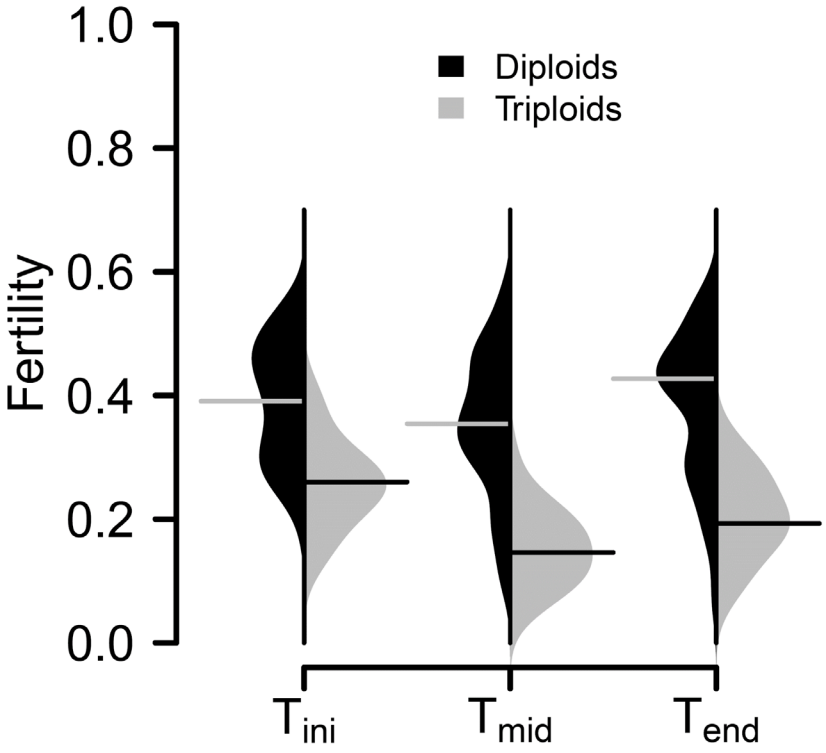


Figure III.S9: Triploid hybrids have a lowered fertility.
Distributions of fertility values (fraction of viable spores) for the diploid (black) and triploid (grey) individuals from the L lines at each of the three tested timepoints.

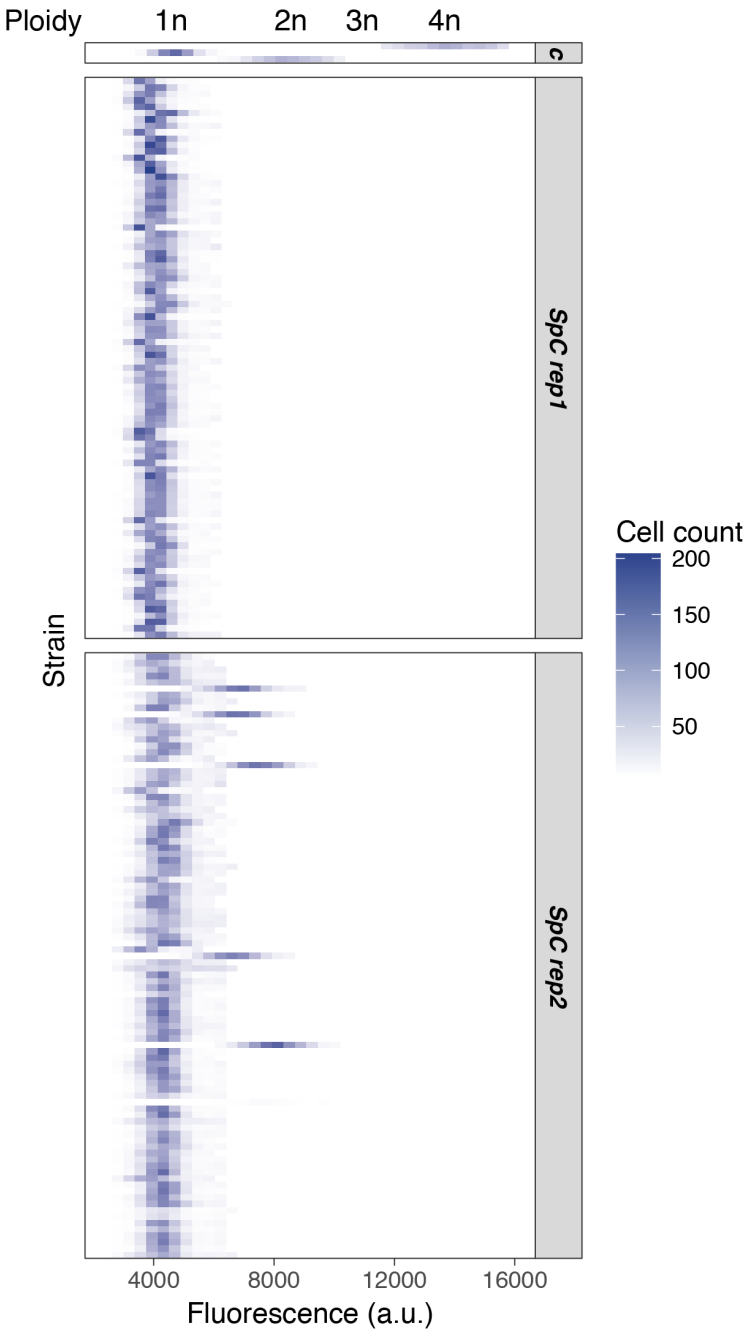


Figure III.S10: The parental SpC haploid stocks contain a small fraction of diploids.

Ploidy of 94 isolated colonies from the parental SpC haploid stocks (LL2011_004 and LL2011_009) using flow cytometry repeated at two independent times. The c panel at the top indicates controls.

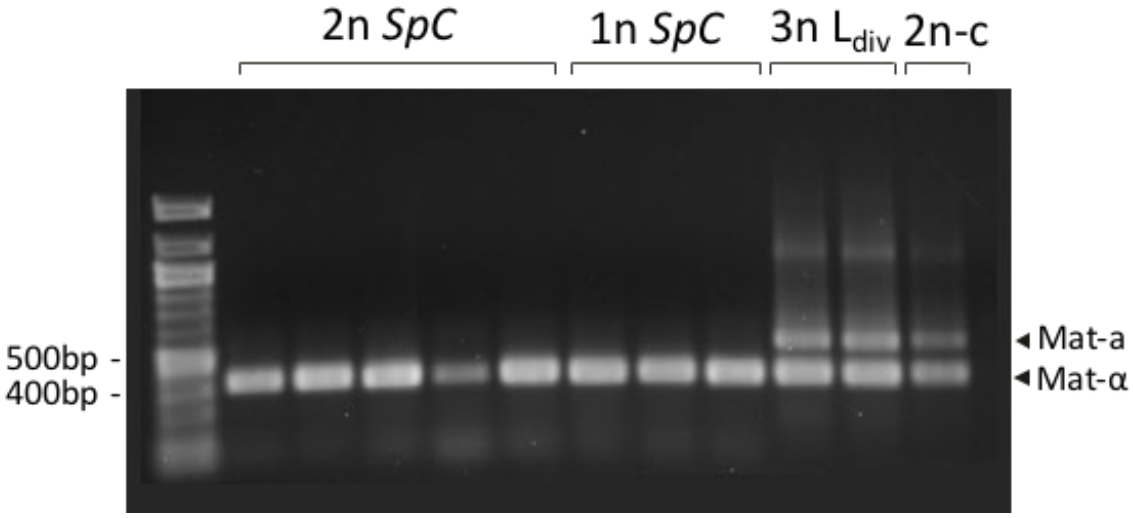


Figure III.S11: The small fraction of diploids in the *SpC* haploid stocks are pseudo-haploids.
PCR of the mating type locus on genomic DNA of isolated colonies. The 5 diploids (2n SpC) identified in the SpC haploid stocks are shown along with 3 haploid SpC (1n SpC) from the same stocks, two triploid L_{div} lines (3n L_{div}) and a diploid control (2n-c).

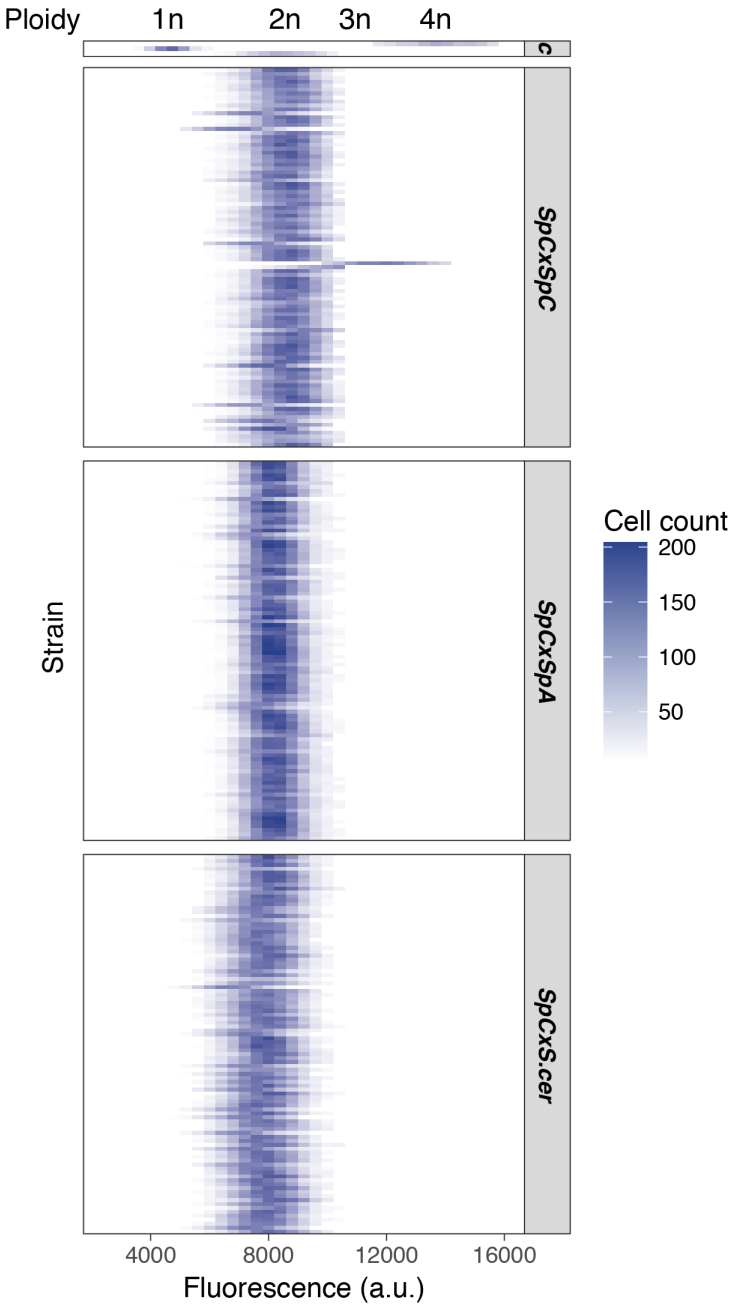


Figure III.S12: Triploidy is rare in hybrids from crosses between *SpC* and other lineages and species.

Ploidy of 94 independent replicates of (*SpC* x *SpC*), (*SpC* x *SpA*) and (*SpC* x *S. cerevisiae*) crosses using flow cytometry. The c panel corresponds to controls.

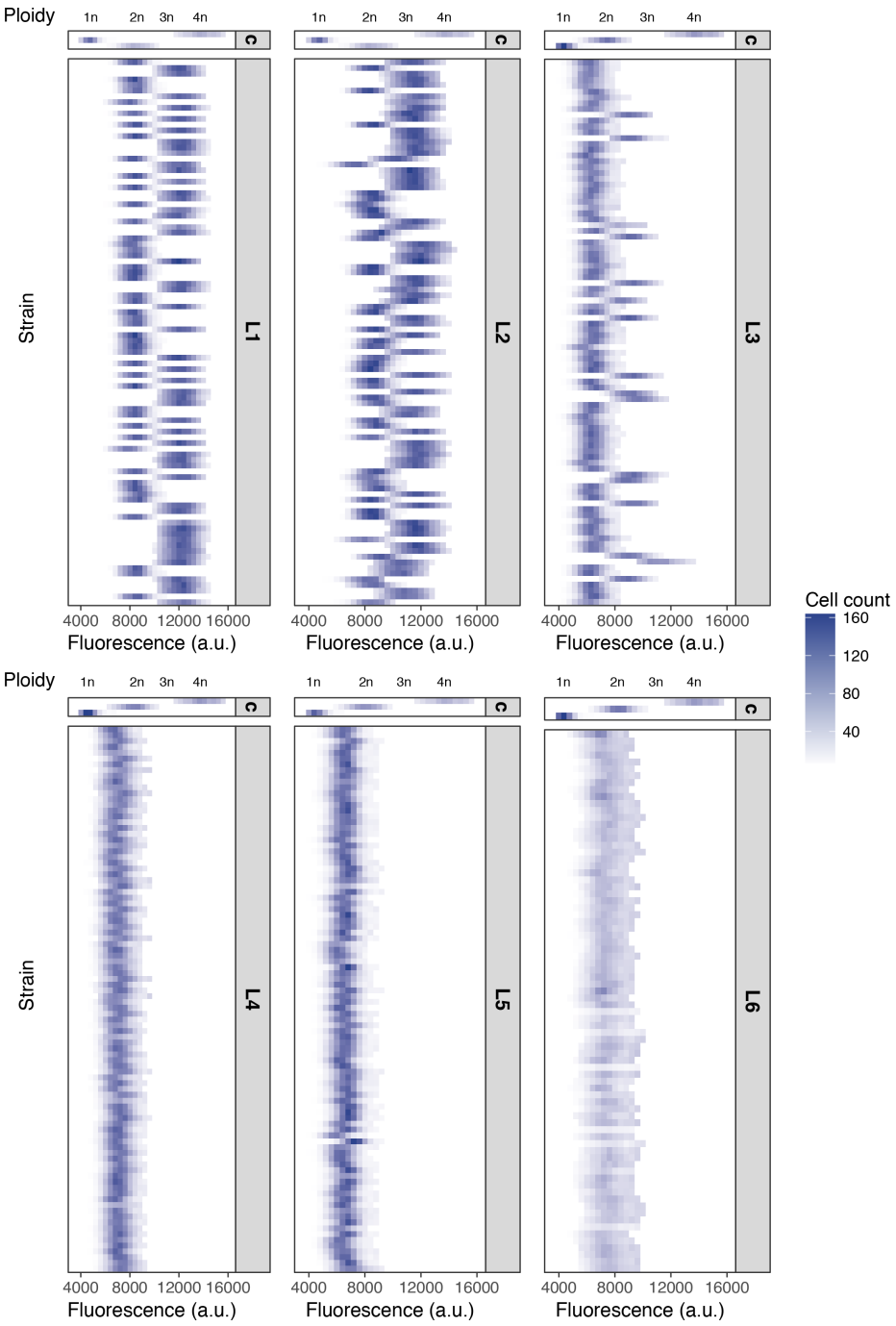


Figure III.S13: Triploidy is observed in crosses between different *SpB* and *SpC* strains but not in all crosses.

Ploidy of 94 independent replicates of different *SpB*×*SpC* crosses ($L_{div}1$, $L_{div}2$, $L_{div}3$, $L_{div}4$, $L_{div}5$ and $L_{div}6$). The c panel corresponds to controls. The strains used for each cross are listed in the Table III.S1 and III.S3. $L_{div}5$ and L_{div6}).

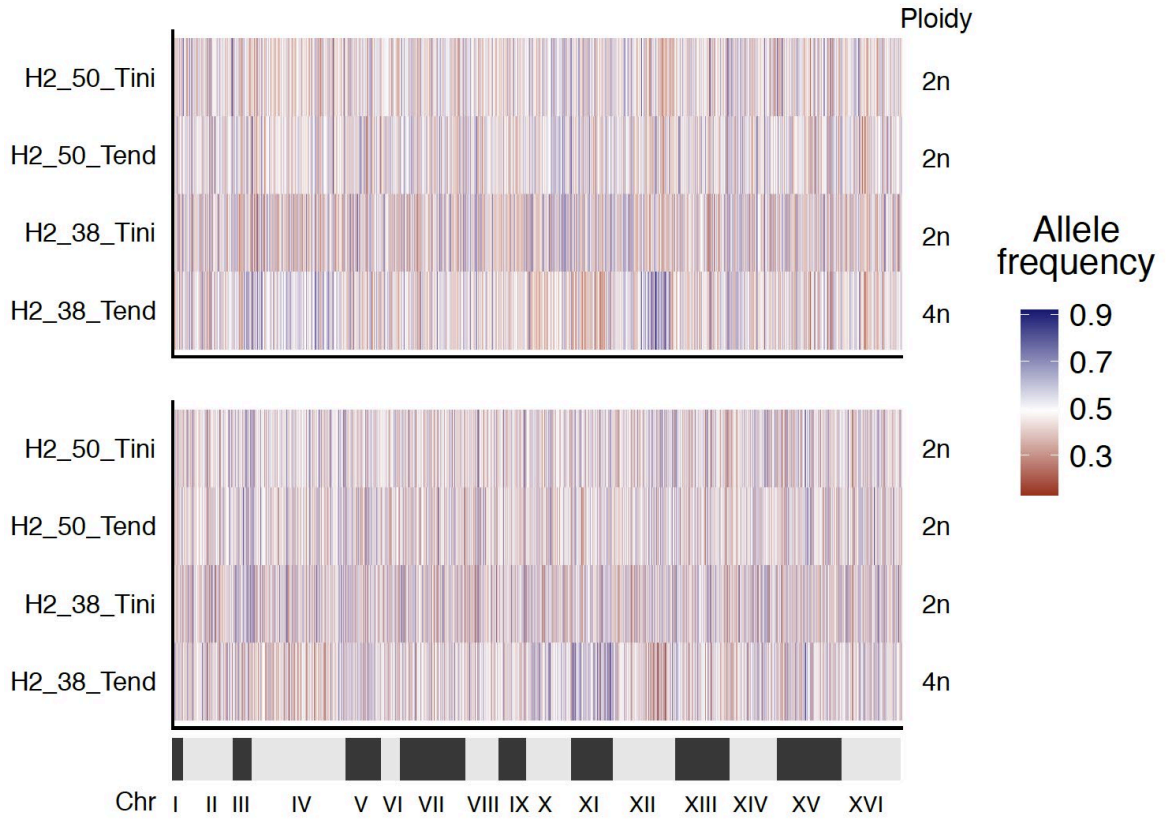


Figure III.S14: The tetraploid hybrids between S. paradoxus and S. cerevisiae result from a whole genome duplication of both parental genomes.

Allele frequencies along the 16 chromosomes of the tetraploid H2_38 at T_{end} are around 50%, similar to what is seen for the H2_38 at T_{ini} and the diploid hybrid H2_50 at T_{ini} and T_{end}. The heatmap on the top represents the allele frequency of reads generated by GBS after mapping on S. cerevisiae reference genome (YPS128) (53078 markers) and the one on the bottom after mapping on S. paradoxus reference genome (CBS432) (56280 markers), showing consistent results independently from the reference genome used.

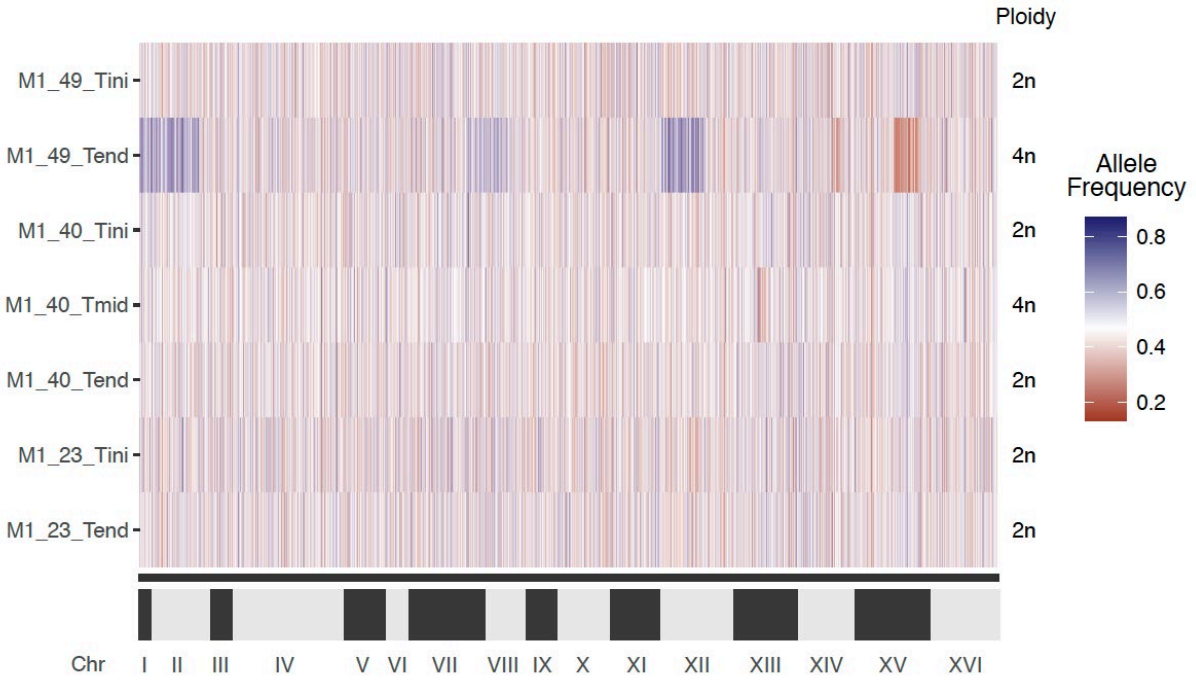


Figure III.S15: The tetraploid hybrids between *SpB* and *SpA* result from a whole genome duplication of both parental genomes.
Allele frequencies along the 16 chromosomes of the tetraploid M1_40 at T_{mid} and M1_49 at T_{end} are around 50%, similar to what is seen for the M1_49 at T_{ini} and M1_40 at T_{ini} and T_{end} and the diploid hybrid M1_23 at T_{ini} and T_{end}. The heatmap represent allele frequencies after mapping reads generated by GBS on the S. paradoxus reference genome (CBS432) (16020 markers).

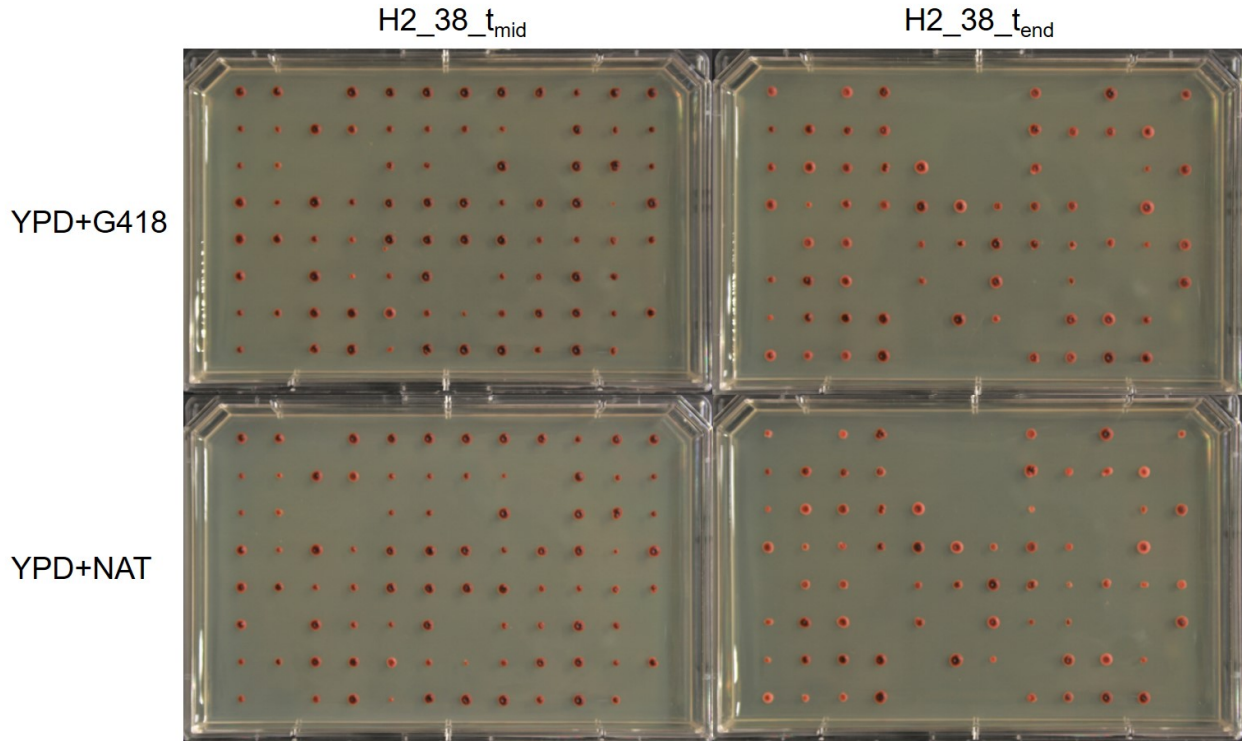


Figure III.S16: The S. paradoxus × S. cerevisiae (H2) hybrids segregate a copy of each parental genome in its diploid spores.

Spores dissected at T_{mid} and T_{end} from the tetraploid line H2_38 were plated on selective media to assess the segregation, during meiosis, of the selection cassettes that were introduced at the HO locus. The S. cerevisiae genome harbor the NAT resistance while S. paradoxus genome harbor the G418 resistance at the HO locus (Chr IV). Empty spaces on plates represent spores that did not form colonies on the original dissection plate. As all the viable spores inherited both resistances, it is likely that the spores contain full non-recombined S. cerevisiae and a S. paradoxus haplotypes.

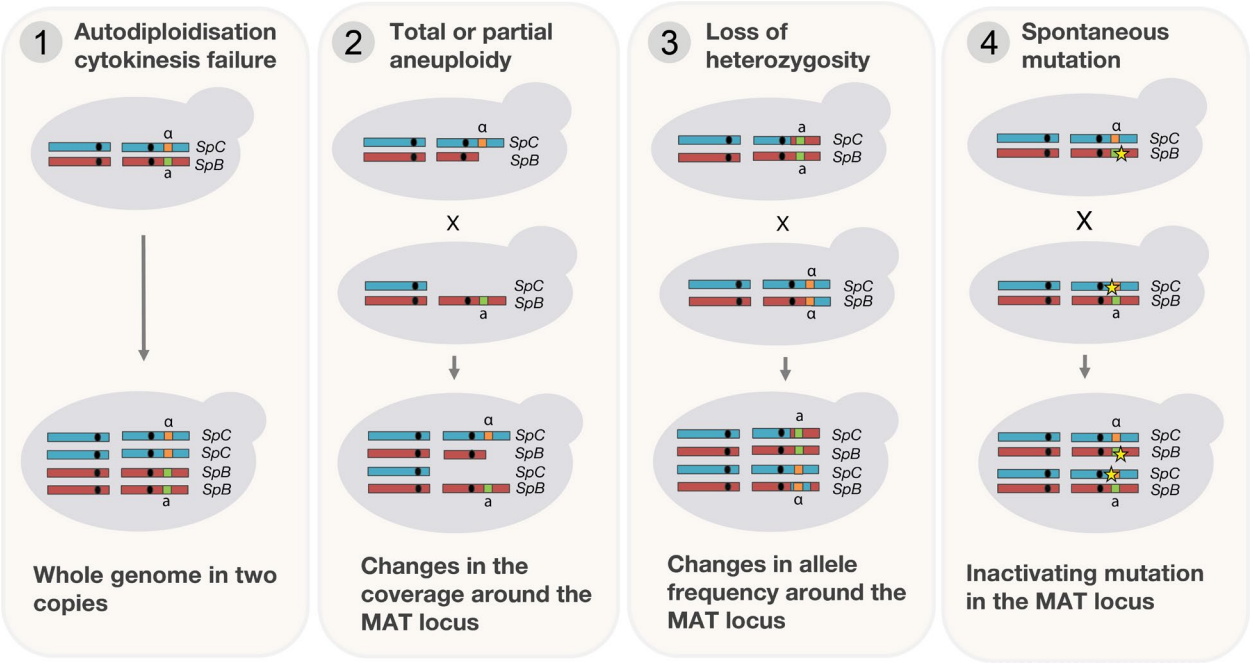


Figure III.S17: Major potential mechanisms of whole-genome doubling.

Whole genome doubling could occur by (1) autodiploidization, which is a consequence of cytokinesis failure. It could also be caused by means of damage to one copy of the MAT locus. The MAT locus damage could be caused by (2) partial or complete chromosome loss of the chromosome containing the MAT locus (chromosome III), (3) loss of heterozygosity around the MAT locus, (4) or an inactivating mutation in one copy of the MAT locus.

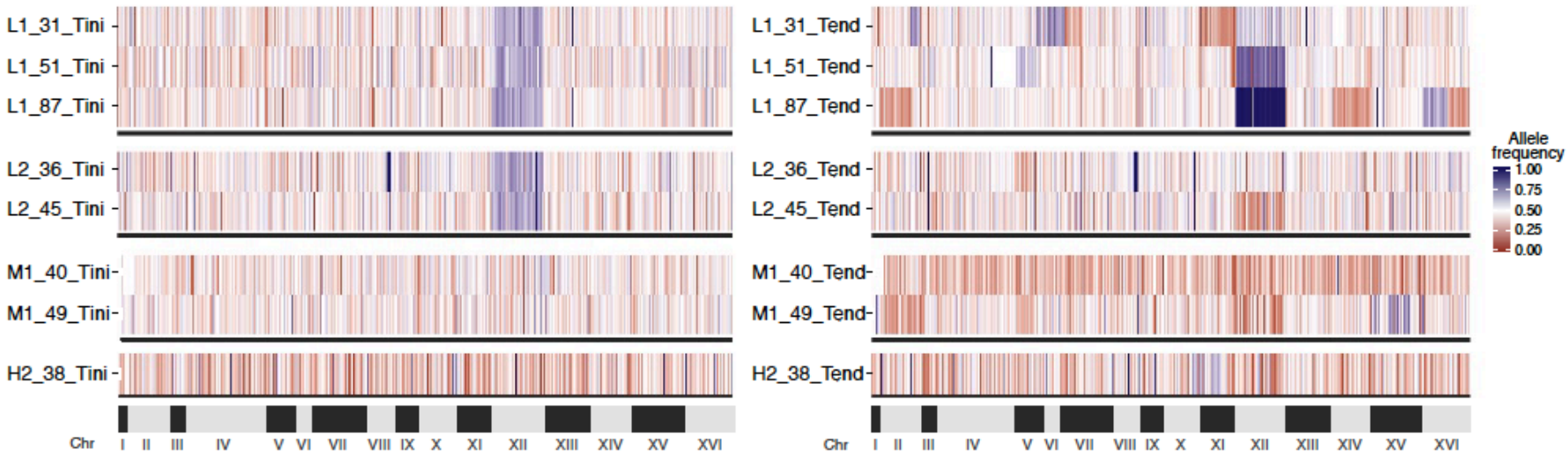


Figure III.S18: The tetraploid hybrids result from a whole genome duplication of both parental genomes.

Allele frequencies along the 16 chromosomes of the tetraploid hybrids at T_{end} are around 50%, similar to what is seen for the same diploid hybrids at T_{ini}. The heatmap represents allele frequencies after mapping reads generated by whole genome sequencing on the S. paradoxus reference genome (MSH604) (276,003 markers for L1 lines, 280,756 markers for L2 lines, 471,547 markers for M1 lines and 981,580 markers for the H2_38).

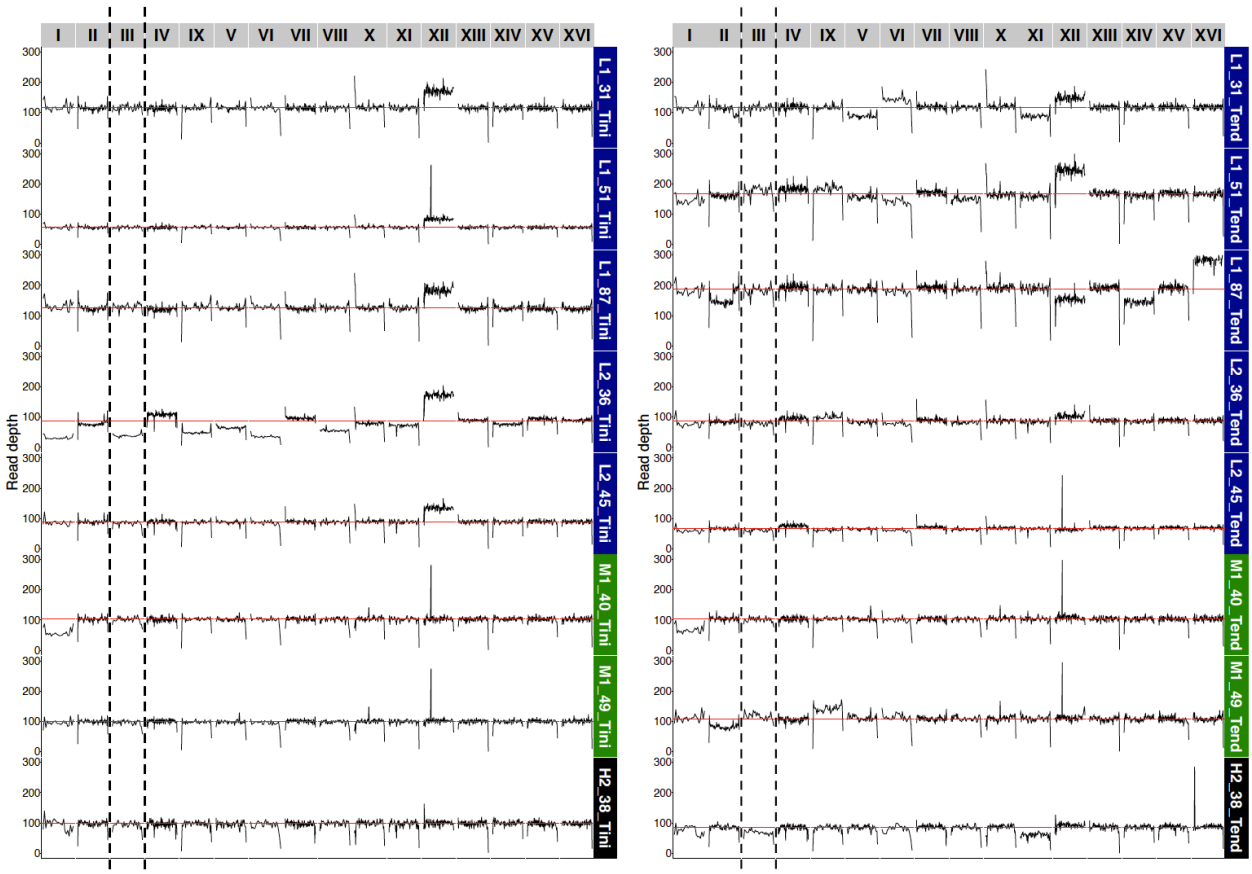


Figure III.S19: The loss of chromosome III is not the molecular mechanism leading to whole genome doubling.
Sequencing read depth for bins of 10kb on the 16 chromosomes of the 8 tetraploid lines at T_{ini} (the left panel) and T_{end} (the right panel). The red line represents the average sequencing read depth of the whole genome. Several aneuploidies are detected in almost all hybrids at T_{ini} and T_{end}.

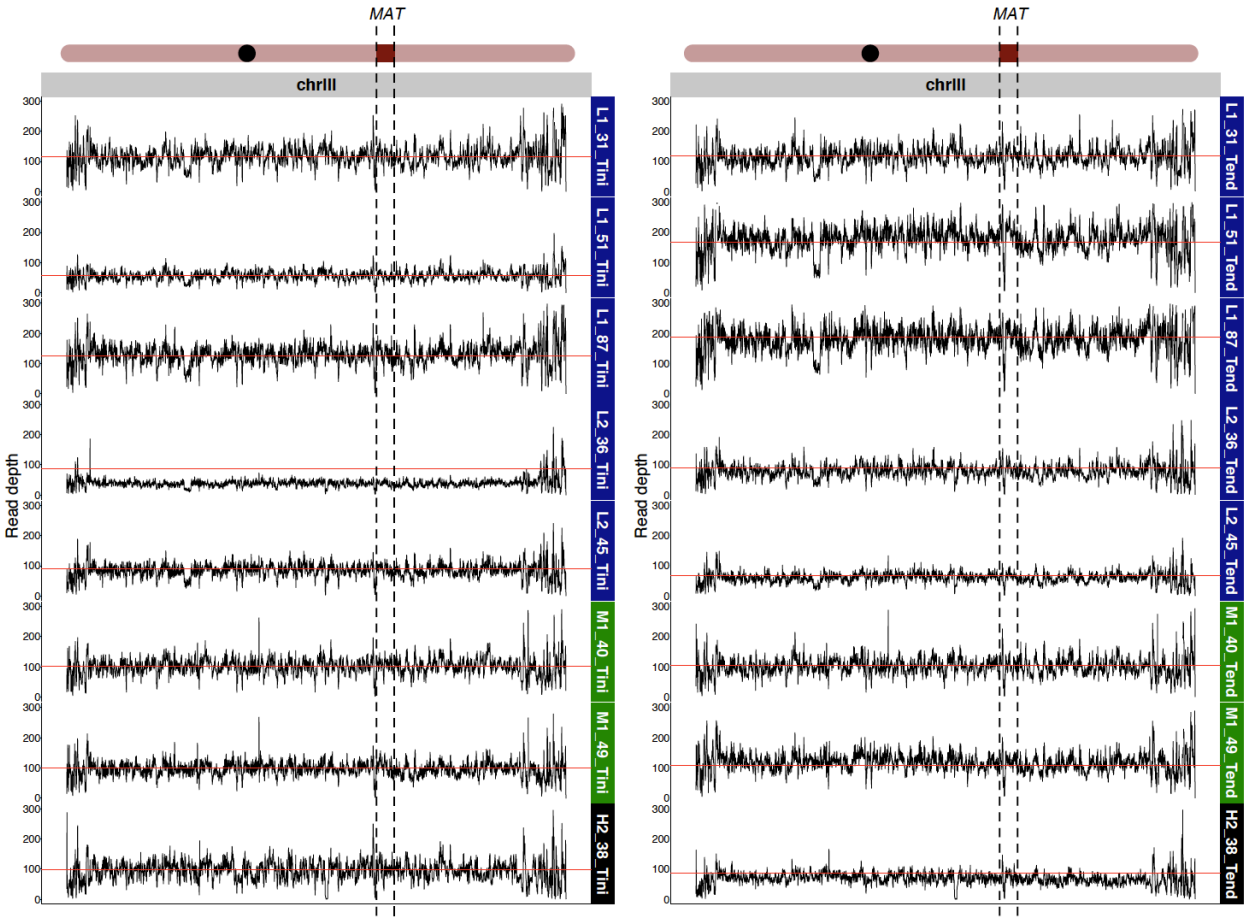


Figure III.S20: The partial loss of Chromosome III is unlikely to be the molecular mechanism leading to whole genome doubling.
Sequencing read depth for each bin of 100bp on chromosome III for the 8 tetraploid lines at T_{ini} (the left panel) and T_{end} (the right panel). The red lines represent the average sequencing read depth of the whole genome. The active mating type locus (MAT) is indicated by dashed lines.

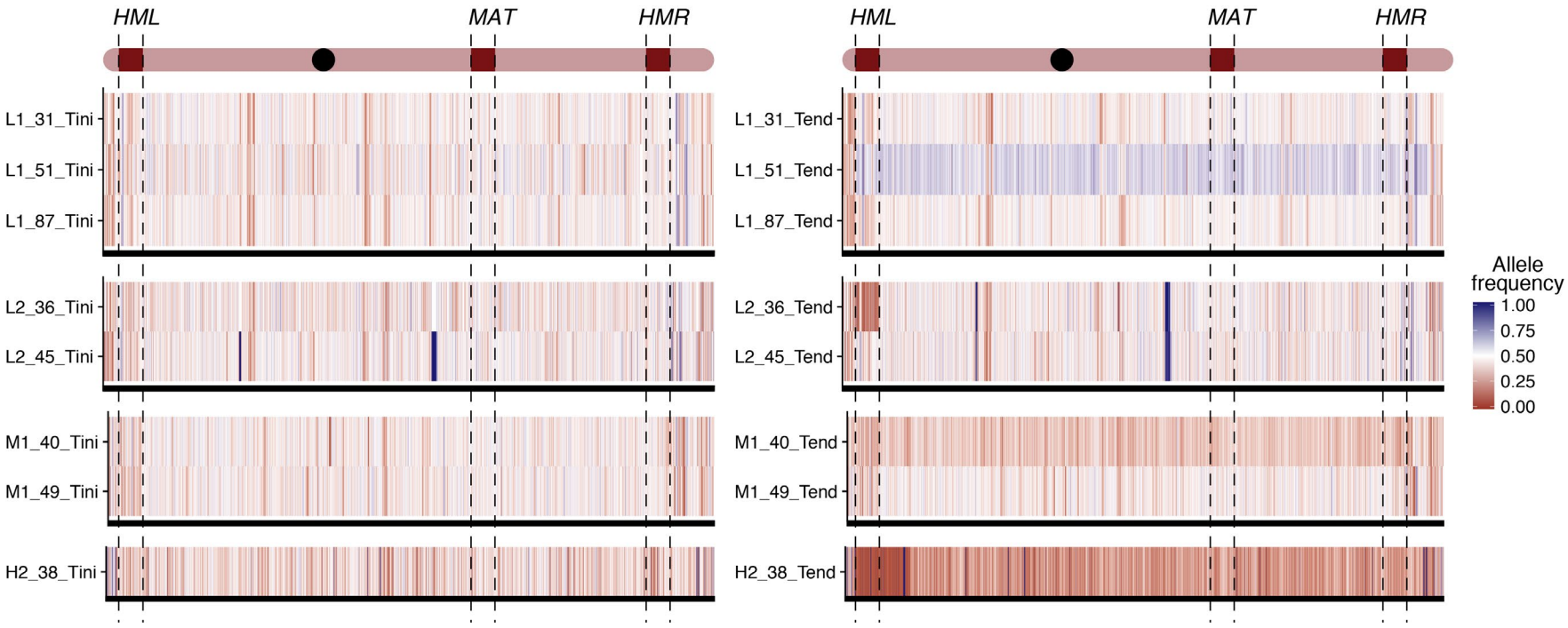


Figure III.S21: Loss of heterozygosity around the MAT locus is most likely not the molecular mechanism causing whole genome doubling.

Allele frequencies along the chromosome III for the 8 tetraploid lines at T_{ini} (left panel) and T_{end} (right panel). The active mating type locus (MAT) and the silent mating type loci (HML and HMR) are indicated by dashed lines. The heatmaps represent allele frequency after mapping on S. paradoxus reference genome (6,611 markers for L1 lines, 5,986 markers for L2 lines, 11,794 markers for M1 lines and 20,537 markers for the H2_38).

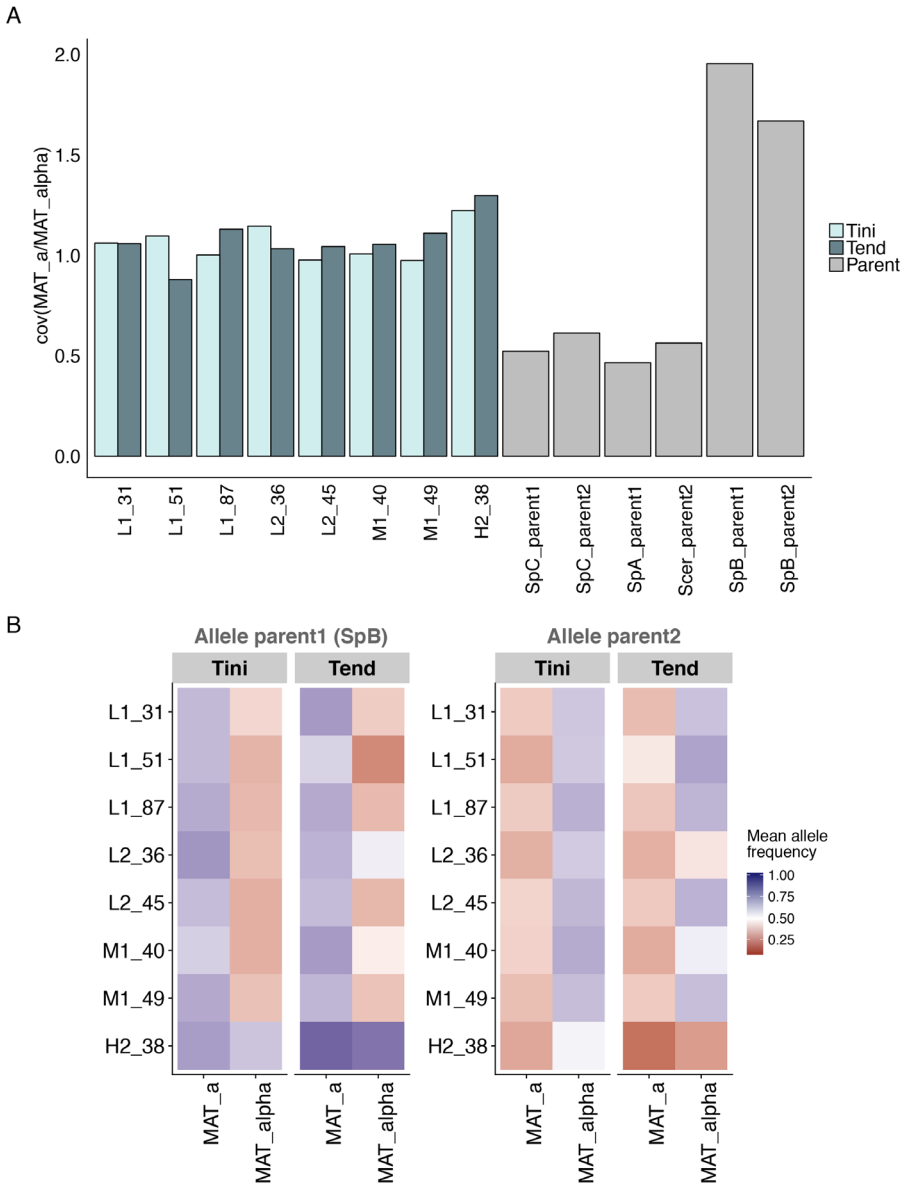

A
cov(MAT_a/MAT_alpha)
2.0
1.5
1.0
0.5
0.0
L1_31
L1_51
L1_87
L2_36
L2_45
M1_40
M1_49
H2_38
SpC_parent1
SpC_parent2
SpA_parent1
Scer_parent2
SpB_parent1
SpB_parent2
Tini
Tend
Parent
B
Allele parent1 (SpB)
Allele parent2
Tini
Tend
Tini
Tend
L1_31
L1_51
L1_87
L2_36
L2_45
M1_40
M1_49
H2_38
MAT_a
MAT_alpha
Mean allele frequency
1.00
0.75
0.50
0.25

Figure III.S22: Copy number variation and allele frequency of MATa and MATα sequences of tetraploid hybrids show no double damage in the MAT locus.

(A) The bar plots represent the ratio of MATa and MATα average read depth for the 8 tetraploid lines at T_{ini} and T_{end} as well as the 6 haploid parental strains. The average sequencing read depth of MATa and MATα were calculated for a sequence of 1.3 kb containing SNPs that differentiates the two copies. (B) The heatmaps show the average allele frequencies (AF) of the MATa and MATα sequences for the 8 tetraploid lines at T_{ini} and T_{end}. After mapping reads on S. paradoxus MATa and MATα reference sequences, the average AF were calculated for a sequence of 1.3 kb containing SNPs that differentiate the two mating type copies. The panel on the left shows the average allele frequencies of both mating type sequences corresponding to the haploid SpB parent (which is MATa) alleles and the panel on the right those corresponding to the haploid SpC (for L1 and L2 lines), SpA (for M1 lines) or S. cerevisiae (for H2_38) parents (which are MATα) alleles.

CONCLUSION GÉNÉRALE

1-Résumé des résultats principaux

L'objectif principal de cette thèse était d'étudier l'hybridation et sa contribution à l'émergence de nouvelles espèces. Cette thématique avait déjà été explorée chez les plantes et les animaux, mais les connaissances sur le rôle de l'hybridation dans la spéciation chez les microorganismes sont beaucoup moins étendues. Bien que la domestication de certains microorganismes par l'homme ait sélectionné des hybrides pour certaines de leurs caractéristiques, on ignorait si l'hybridation était commune dans les populations naturelles (Sipiczki 2008). Nous avons premièrement utilisé une approche de génomique des populations à la recherche d'hybrides ou encore d'indices génomiques d'hybridation.

Au chapitre I, nous avons exploré la génomique des populations naturelles de la levure *Saccharomyces paradoxus*. La structure génétique de cette espèce ainsi que sa biogéographie présentent un contexte qui pourrait être propice à des évènements d'hybridation. Nous avons trouvé une lignée au sein du groupe *SpC* qui était jusque-là cryptique avec des approches de séquençage de marqueurs spécifiques. Nous avons trouvé que cette lignée, *SpC* introgressée (*SpC**), possédait un génome mosaïque. Le génome est composé à 95% d'un génotype très similaire à *SpC* et le 5% restant est quant à lui similaire à *SpB*. Un criblage phénotypique des souches nous a permis de montrer que ces souches présentent des différences phénotypiques marquées, *SpB* ayant une meilleure croissance à haute température et une meilleure résistance au stress gel-dégel que *SpC*. Les données de divergence génétique entre les génomes de *SpC** et de ses deux parents supposés suggèrent un évènement d'hybridation vers la fin de la dernière glaciation. Comme le nombre de chromosomes ne semble pas avoir été altéré et que nous répondions au critère de la présence de trace génomique d'hybridation, nous avons émis l'hypothèse que cette lignée représente un évènement de spéciation par hybridation homoploïde.

Les résultats obtenus à la suite de croisements expérimentaux nous ont indiqué que l'isolement reproducteur entre *SpC** et les deux lignées parentales était aussi fort que celui qui est présent entre *SpB* et *SpC*. Ce résultat nous permettait donc de répondre au critère de l'isolement reproducteur entre la lignée hybride et les lignées parentales. En faisant des analyses de ségrégation de marqueurs, nous avons pu relier l'isolement reproducteur entre *SpC* et *SpC** à une translocation fixée dans la lignée *SpC** et présente à une faible fréquence chez *SpB*. Nos résultats suggèrent que l'hybridation chez les microorganismes déjà reconnue pour être impliquée dans l'émergence de pathogènes fongiques des animaux, des végétaux et des humains (Brasier *et coll.* 1999, Boekhout *et coll.* 2001, Stukenbrock 2016) et de souches utilisées dans les bioprocédés (González *et coll.* 2006, Dunn et Sherlock 2008, González *et coll.* 2008), contribue aussi à la formation d'espèces microbiennes non influencées par les activités humaines.

Avec les données recueillies au chapitre I, la question de l'hybridation contemporaine dans l'environnement s'est soulevée. Nous savons que des évènements d'hybridation et de rétrocroisements ont eu lieu chez les lignées nord-américaines de *S. paradoxus*, mais aucune trace d'hybridation contemporaine n'a été trouvée. Une hypothèse pouvant expliquer la rareté des hybrides est qu'ils sont contre-sélectionnés dans l'environnement puisqu'ils ne sont pas en mesure de compétitionner avec leurs parents.

Au cours des travaux dérits au chapitre II, nous avons exploré la possibilité que les hybrides de premières générations présentent un isolement reproducteur postzytgotique extrinsèque. Peu de recherches avaient été menées sur ce genre d'isolement reproducteur chez les microorganismes. Nous avons utilisé une approche de criblage phénotypique à haut débit qui nous a permis de comparer la croissance d'hybrides et de leurs parents dans plusieurs conditions environnementales. Nos résultats suggèrent plutôt que les hybrides

présentent généralement une croissance égale ou supérieure à celle de leur meilleur parent. Ces signes d'hétérosis dans la majorité des environnements expérimentaux testés suggèrent donc qu'une sélection contre les génotypes hybrides est donc peu probable. Ce résultat nous a permis d'écarter l'hypothèse de l'isolement reproducteur extrinsèque pour expliquer l'absence d'hybrides récents entre *SpB* et *SpC*.

Une des observations que l'on peut faire sur la fertilité des souches introgressées est que cette dernière est aussi élevée que celle de souches non hybrides. Cependant, les hybrides expérimentaux de première génération présentent une fertilité diminuée. La fertilité des hybrides est donc recouvrée dans le temps. Les rétrocroisements peuvent être responsables de l'amélioration de la fertilité puisqu'ils diminuent l'hétérozygotie du génome et par conséquent ont le potentiel de réduire le nombre de loci incompatibles dans le génome (Schumer *et coll.* 2018). Cependant, comme les hybrides initiaux sont partiellement fertiles, la reproduction sexuée leur impose un coût de fitness. Dans des espèces ayant accès à la reproduction asexuée, il se pourrait que le recouvrement de la fertilité puisse se faire sans l'intervention de rétrocroisements.

Au chapitre III, nous avons tenté de mesurer le potentiel de recouvrement de la fertilité chez des hybrides se reproduisant uniquement de manière asexuée. Pour ce faire, nous avons créé des hybrides expérimentaux le long d'un gradient de divergence génétique. Nous avons fait évoluer les hybrides sur plus de 700 générations mitotiques en limitant l'effet de la sélection par de forts goulots d'étranglement sur la taille des populations. La fertilité de nos hybrides a été mesurée pour déterminer comment elle évoluait au cours du temps. Nous avons observé autant de recouvrement que de dégradation de la fertilité à l'issue de l'expérience, mais ces changements étaient le plus souvent de faible magnitude. Cependant, nous avons aussi pu observer de rares cas où le recouvrement de la fertilité était complet. L'étude de la ploïdie des individus testés a révélé que dans tous ces cas, le contenu cellulaire en ADN a doublé,

indiquant une duplication du génome. Nos résultats indiquent qu'au cours de leur évolution, les hybrides peuvent recouvrer leur fertilité par des évènements stochastiques de duplication du génome. Ceci pourrait s'avérer important pour la recherche en évolution chez la levure puisque l'origine de la duplication complète du génome chez un ancêtre des Saccharomyces est attribuée à un évènement d'allopolyploïdisation chez un ancêtre de ce groupe (Marcet-Houben et Gabaldon 2015).

Notre travail a répondu aux objectifs initiaux de la thèse. Au moment d'entamer le travail décrit dans cette thèse, les connaissances de l'hybridation et de son rôle chez les populations de microorganismes étaient limitées à peu d'exemples focalisés sur des espèces économiquement importantes telles des champignons pathogènes (Nelson 1963, Stukenbrock 2016) ou encore des souches industrielles domestiquées (Masneuf *et coll.* 1998, González *et coll.* 2008). Dans les populations naturelles, peu d'hybrides étaient alors rapportés (Sipiczki 2008), probablement parce que les cas d'hybridation interspécifique sont rares en nature vu l'isolement reproducteur post-zygotique très fort chez les Saccharomyces (Greig 2008). L'utilisation des populations naturelles de *S. paradoxus* a permis de porter un regard nouveau sur l'hybridation puisqu'on a étudié des lignées qui, contrairement aux espèces appartenant au *Saccharomyces sensu stricto*, sont seulement partiellement isolées reproductivement. Cet isolement partiel couplé aux changements environnementaux de la dernière glaciation a probablement favorisé l'hybridation. De plus, nous avons utilisé ce système en laboratoire pour explorer la possibilité d'un isolement extrinsèque présent entre *SpB* et *SpC*. Malgré la littérature déjà abondante sur l'hétérosis chez les hybrides de levures Saccharomyces (Shapira *et coll.* 2014, Bernardes *et coll.* 2016), nous avons contribué à l'acquisition de nouvelles connaissances puisque les hybrides utilisés représentent des lignées avec un historique connu d'hybridation. Nos résultats suggèrent que des barrières au flux de gène, autres que les barrières post-zygotiques extrinsèques doivent être

présentes pour expliquer l'absence d'hybrides dans nos échantillonnages. Finalement, en utilisant nos levures dans un contexte d'évolution expérimentale, nous avons pu explorer l'évolution des hybrides. D'autres études ont utilisé une approche similaire chez la levure *S. cerevisiae* pour étudier l'évolution de l'architecture du génome après l'hybridation (Dutta *et coll.* 2017, Smukowski Heil *et coll.* 2017). Nous nous sommes plutôt penchés sur l'évolution de la fertilité chez des lignées n'ayant pas accès à la reproduction sexuée. Cette étude de grande envergure autant pour le nombre de lignées d'évolution que pour le gradient de divergence testé a généré beaucoup d'informations qui restent encore à être analysées. Les données que l'on pourra en tirer ont un grand potentiel dans le développement de nouvelles connaissances sur l'évolution des hybrides en général. Plusieurs questions demeurent en suspens, mais notre système peut nous aider à y répondre.

2-Discussion et Perspectives

2.1-Spéciation par hybridation chez *S. paradoxus*

Nous avons établi au chapitre 1 que notre système d'étude répondait aux trois critères de la spéciation par hybridation homoploïde qui sont : 1) La présence de trace d'hybridation dans le génome, 2) Un isolement reproducteur entre l'hybride et les lignées parentales et 3) un lien génétique entre l'hybridation et l'isolement reproducteur (Schumer *et coll.* 2014). Les deux premiers critères sont souvent plus faciles à vérifier expérimentalement que le troisième. Les méthodes de détection basées sur l'utilisation des SNP entre les lignées, comme les statistiques *F3* et *F4* ou encore la statistique *D* nous informent sur la présence des introgressions et les espèces impliquées (Reich *et coll.* 2009, Green *et coll.* 2010). Cependant, ces méthodes ne permettent pas d'obtenir de l'information sur la relation entre les introgressions et les évènements de spéciation.

Pour répondre à cette problématique, Hibbins et Hahn (2019) ont développé un modèle utilisant le modèle du coalescent de réseau

multiespèces (multispecies network coalescent) (Degnan 2018). Les prédictions du modèle ont permis aux auteurs de proposer deux nouvelles statistiques, D_1 et D_2 (Hibbins et Hahn 2019). La statistique D_2 permet de déterminer la direction de l'introgression. La statistique D_1, quant à elle, permet de déterminer le moment relatif de l'introgression par rapport au moment de la spéciation. Ceci est fait en comparant les temps de coalescence des gènes dans les introgressions à celui des gènes à l'extérieur de celles-ci. Si l'hybridation et la spéciation se sont faites simultanément comme le requiert le troisième critère, ces deux temps ne devraient pas être différents. Les auteurs ont utilisé les données de séquençages du chapitre I afin de démontrer l'application de leur modèle à un cas réel. En mesurant D1 dans notre système, ils ont obtenu une valeur significativement différente de zéro, indiquant que l'introgression de *SpB* dans *SpC** s'est faite après la spéciation entre *SpC* et *SpC** réfutant ainsi notre conclusion de spéciation par hybridation. Les auteurs nuancent toutefois leur conclusion en disant qu'il est possible que SpC* ait une origine hybride qui a par la suite été introgressée par *SpB*, mais qu'il est impossible de différencier les deux modèles. Notre étude avait donc des limitations dans l'analyse des données vu l'absence de statistique permettant de vérifier l'hypothèse de la spéciation par hybridation homoploide. Peut-être aurions-nous pu poser des conclusions un peu plus prudentes sur la relation entre *SpC*, *SpB* et *SpC** si nous avions pu appliquer ces statistiques à nos analyses génomiques.

Il est difficile de savoir exactement si l'évènement d'hybridation initial a mené à un isolement reproducteur immédiat. Notre hypothèse de la spéciation par hybridation était toutefois soutenue par la présence de la translocation VItXIIIr (figure I.5a) qui jouait un rôle dans l'isolement reproducteur en déséquilibrant le contenu chromosomique des produits de méiose. Le rôle des réarrangements de chromosomes dans l'initiation de l'isolement reproducteur est mis en doute puisqu'ils affectent le fitness de l'hybride et sont considérés comme sous-dominants (Walsh 1982, Lande

1985). Cependant, chez les levures, l'homothallisme permet à n'importe quelle mutation (non létale chez l'haploïde) de se fixer rapidement. Les réarrangements de chromosomes sont déjà connus comme source d'isolement entre les espèces *S. paradoxus* et *S. cariocanus* (Fischer *et coll.* 2000), chez les populations naturelles de *S. cerevisiae* (Hou *et coll.* 2014) et pour des caryotypes complètement artificiels (Luo *et coll.* 2018). Pour vérifier l'effet de la translocation sur la fertilité, il serait possible de créer des hybrides avec ou sans translocation que nous pourrions croiser avec les souches parentales. Ceci pourrait se faire grâce à une approche basée sur CRISPR-Cas9, déjà utilisée pour former des *S. cerevisiae* artificielles avec tous les nombres de chromosomes possibles entre 1 et 15 (Shao *et coll.* 2018). Une réduction significative de la fertilité des hybrides portant cette translocation lorsque comparé à des hybrides ne la portant pas pourrait supporter notre hypothèse.

Finalement, les critères proposés par Schumer *et coll.* (2014) ont ouvert un débat quant à la caractérisation de ces évènements. Nieto Feliner *et coll.* (2017) suggèrent que le critère selon lequel l'hybridation doit mener à l'isolement reproducteur est peut-être trop sévère et pourrait rendre le cadre conceptuel de la spéciation homoploïde trop étroit. Ils proposent que l'origine de l'isolement reproducteur importe peu tant que la nouvelle lignée émergeant par hybridation devient isolée reproductivement et est stabilisée dans une niche différente (Nieto Feliner *et coll.* 2017). Ces critères un peu plus souples permettraient donc de considérer *SpC** comme une espèce hybride.

2.2-Introgressions chez *SpC**

Les introgressions ont souvent un grand potentiel adaptatif puisqu'elles génèrent de nouvelles combinaisons génétiques et chromosomiques sur lesquelles la sélection naturelle peut ensuite agir (Hedrick 2013, Tigano et Friesen 2016, Suarez-Gonzalez *et coll.* 2018). C'est une question qui reste en suspens dans notre système: est-ce que les

introgressions ont contribué à l'adaptation et au maintien des populations de *SpC** dans l'environnement? Trois informations nous indiquent qu'une partie des introgressions pourrait être adaptative. Premièrement, nous avons montré par le passé que les souches *SpB* et *SpC* semblaient localement adaptées aux conditions climatiques locales (Leducq *et coll.* 2014). Il y a donc un groupe d'allèles potentiellement adaptés qui diffèrent entre *SpB* et *SpC*. Des travaux récents par d'autres membres du laboratoire ont aussi montré que certains gènes ont divergé par l'action de la sélection positive et ces gènes pourraient contribuer à leur adaptation ou leur spécialisation à leur environnement (Eberlein *et coll.* 2017). Deuxièmement, les génomes des souches de la lignée *SpC** présentent des introgressions fixées (H0, Figure I.3) alors que d'autres semblent toujours ségréguer dans la population. La conservation de ces régions introgressées à travers les individus pourrait être un indicateur de l'action de la sélection naturelle sur certains des gènes contenus dans ces régions. De ces 105 gènes, il ressort un enrichissement de terme d'ontologie lié au métabolisme des acides aminés (Hénault *et coll.* 2019), métabolisme qui semble différent entre *SpB* et *SpC* (Leducq *et coll.* 2014, Filteau *et coll.* 2017). Troisièmement, le site où nous avons le plus échantillonné d'individus *SpC** est la Pointe-Platon. Les populations dominantes de ce site sont *SpB* (39.1%) et *SpC** (42.4%) qui représentent plus de 80% des *S. paradoxus* isolées. Ceci pourrait suggérer que les gènes portant les allèles *SpB* dans *SpC** contribuent à un fitness plus élevé par rapport à *SpC* dans l'environnement de la Pointe-Platon. Afin de cibler des gènes potentiellement responsables des différences phénotypiques entre *SpC* et *SpC**, il serait possible d'utiliser une approche de type QTL (Paterson *et coll.* 1988) en faisant des croisements entre *SpC* et *SpC**. En mesurant la croissance des ségrégants dans des environnements où *SpC** croît mieux que *SpC*, nous pourrions trouver des régions génomiques impliquées dans ces différences. Il serait aussi possible de remplacer les allèles des souches *SpC* par leur équivalent *SpB* trouvé chez *SpC**, ce qui permettrait d'évaluer

l'effet individuel des gènes sur le fitness ou encore de procéder en évaluant le fitness d'hémizygotes réciproques.

2.3-Isolement reproducteur entre *SpB* et *SpC*

Les résultats du chapitre II suggèrent que *SpB* et *SpC* ne sont pas isolés par un isolement reproducteur extrinsèque. Nous avons alors émis l'hypothèse que d'autres mécanismes d'isolement reproducteur pourraient s'additionner à l'isolement post-zygotique intrinsèque déjà présent. Une de ces hypothèses, l'isolement écologique, pourrait venir de la spécialisation d'une ou des deux lignées pour des substrats en particulier. Les résultats d'Eberlein et al. 2017 suggèrent la présence de gènes sous évolution asymétrique entre *SpB* et *SpC*, ce qui pourrait être le résultat d'une spécialisation pour des environnements différents. Un échantillonnage à grande échelle des environnements où l'on retrouve *SpB* et *SpC* en sympatrie pourrait être un moyen de trouver une association entre les levures et différents substrats comme les chênes et les érables ou encore les sols et les écorces d'arbres. Les résultats du chapitre III nous ont aussi mis sur la piste d'autres mécanismes génétiques d'isolement reproducteur. Les souches triploïdes obtenues lors des croisements entre *SpB* et *SpC* pourraient résulter d'une instabilité de la ploïdie chez les hybrides ou encore chez les individus d'une des lignées, qui agirait comme une source supplémentaire d'isolement reproducteur post-zygotique. Cependant, nos expérimentations ont montré que nos stocks d'haploïdes contenaient des cellules pseudo-haploïdes. Ces cellules qui contiennent un génome haploïde doublé et un seul type sexuel, se comportent donc comme des haploïdes et sont capables de réagir aux phéromones et fusionner avec des haploïdes. Jusqu'ici, la présence des pseudo-haploïdes semble anecdotique (Figure III.S10), mais si cette instabilité est généralisée chez les haploïdes *SpC*, elle pourrait contribuer à l'isolement reproducteur en générant des hybrides triploïdes. Nous avons aussi observé une perte rapide de la capacité de sporulation de certaines lignées hybrides au cours de

l'évolution. Nos observations suggèrent que ces lignées ont perdu une partie du contenu de leur génome mitochondrial, impliqué dans la respiration qui elle, est nécessaire à la sporulation (Küenzi *et coll.* 1974). Chez les Saccharomyces, les hybrides sont généralement hétéroplasmiques après leur formation, ils contiennent donc les mitochondries de leurs deux parents. Après 10 à 20 générations mitotiques, un mitotype se fixera dans les cellules (Birky *et coll.* 1978). Pendant cette transition, il peut se produire de la recombinaison entre les génomes mitochondriaux, générant de nouveaux génotypes mitochondriaux (Thomas et Wilkie 1968, Leducq *et coll.* 2017). Le nombre de lignées affectées par cette perte de fonction est corrélé avec la divergence génétique entre les parents d'un croisement. La perte de la sporulation chez les hybrides pourrait donc être causée par la fixation d'incompatibilités mitonucléaires dans les générations suivant l'hybridation. Si ces incompatibilités sont létales, les levures ayant perdu une partie de leur génome mitochondrial sont les seules que nous pouvions alors transférer lors de l'évolution. La recombinaison mitochondriale pourrait aussi être un moyen de résoudre des incompatibilités mitonucléaires en évitant la fixation d'un génotype mitochondrial parental. Le génome des souches utilisées dans notre expérience d'évolution a aussi été séquencé, ce qui nous sera utile afin de trouver l'origine génétique de cette possible incompatibilité.

2.4-Écologie de *S. paradoxus*

Au chapitre II, nous avons montré que les hybrides F_1 entre *SpB* et *SpC* ne montrent pas d'isolement reproducteur extrinsèque dans la majorité des conditions expérimentales testées. Nous avions choisi ces conditions pour nous renseigner sur des traits qui pouvaient être écologiquement pertinents dans leur environnement naturel comme l'utilisation de certaines sources de carbone ou d'azote. Cependant, beaucoup d'information reste manquante à propos de l'écologie des *Saccharomyces* (Boynton et Greig 2014). Il est donc difficile de faire le lien

direct entre la capacité des levures à utiliser un substrat au laboratoire (dans des conditions optimales de croissance) et les réels déterminants du fitness des levures dans leur milieu naturel. Une étude récente faite en sol européen a montré une association entre la litière de feuilles de chêne et *S. paradoxus* (Kowallik et Greig 2016). La litière, plus que l'écorce des arbres, serait donc l'habitat de préférence de cette espèce. Les auteurs ne rapportent pas d'association avec les érables, ce qui est probablement dû à l'absence de cette espèce dans la région échantillonnée. De plus, des méthodes permettant de suivre les levures lors d'expériences en jardin commun (common garden experiment) et de mesurer leur fitness dans le sol ont déjà été testées (Boynton *et coll.* 2017) et des recensements dans le temps indiquent que les mêmes génotypes persistent d'année en année (Xia *et coll.* 2017). La construction de collections marquées avec des code-barres génétiques (Bleuven *et coll.* 2018) et la découverte de mutants naturels sélectionnables (Anderson *et coll.* 2018) permettent de suivre les levures dans des microcosmes au laboratoire. Ces nouvelles ressources pourront faciliter l'étude de l'écologie de *S. paradoxus*. Appliquées aux hybrides, ces techniques pourraient nous aider à confirmer les observations faites au laboratoire comme la présence d'hétérosis.

2.5-Évolution des hybrides

Le chapitre III nous a permis de générer une quantité considérable d'hybrides artificiels qui pourront servir à en apprendre plus sur l'évolution des hybrides et de leur génome. Nous avons exploré l'évolution de la fertilité et de la ploïdie des lignées évoluées, mais plusieurs autres informations nous sont accessibles. Nous avons testé un gradient de divergence de 0.5% jusqu'à 15% et que chaque point dans ce gradient compte au minimum 96 lignées indépendantes. Comme nous avons soumis nos hybrides à une sélection très limitée, nous devrions avoir un portrait général des changements pouvant se produire dans un génome hybride. Nous avons aussi des possibilités d'étudier ces changements à

fine échelle puisque nous avons gardé des stocks de nos lignées à chaque tranche d'environ 70 générations mitotiques. Mes collègues s'affairent déjà à analyser les séquences génomiques d'une partie des lignées initiales et finales. Dutta et Lin (2017) présentent une étude similaire à la nôtre et rapportent la dynamique du génome des hybrides. Ils rapportent que leurs hybrides montrent des pertes d'hétérozygotie très variables pouvant aller jusqu'à plus du tiers des sites hétérozygotes. Leur approche est limitée à peu de lignées hybrides d'évolution (12) et une divergence minimale entre les parents (60,000 SNP ou environ 0.5% de divergence). Nos lignées d'évolution VL_{div} constituent en quelque sorte une réplication indépendante de leur expérience. Nous pourrons en plus tester s'il y a un effet de la divergence sur la fréquence et l'étendue des pertes d'hétérozygotie. Les études passées suggèrent que les hybrides interspécifiques subissent beaucoup d'évènements de réarrangements de chromosomes et de variation du nombre de copies de certains gènes (Mason et Batley 2015, Smukowski Heil *et coll.* 2017). Ces changements semblent associés à une activité des éléments transposables (Dunham *et coll.* 2002, Gresham *et coll.* 2008), ce que nos lignées nous permettront aussi d'étudier plus en détail.

3-Conclusion

Peloria (monstre), le nom donné par Linné au premier organisme hybride résume bien ce que l'on pensait de l'hybridation à une époque où l'on croyait à la Création comme origine pour toutes les espèces (Gustafsson 1979). On finit par accepter l'existence de l'hybridation, mais comme une erreur menant à un cul-de-sac évolutif. L'avènement de la génomique a révélé que le génome de beaucoup d'espèces contient des traces d'hybridation passée et que les introgressions pouvaient être avantageuses (Hedrick 2013, Racimo *et coll.* 2015, Suarez-Gonzalez *et coll.* 2018). On sait que l'hybridation est commune chez les animaux, les plantes (Mallet 2007) et maintenant qu'elle contribue aussi à la diversité

des microorganismes naturels. En obtenant les séquences génomiques de plusieurs espèces, on a pu constater que l'hybridation contemporaine ou ancienne ne représente pas des anecdotes dans l'histoire des espèces (Payseur et Rieseberg 2016). En amassant plus de génomes complets de microorganismes, peut-être pourrions-nous observer que ces organismes ont recours à l'hybridation plus souvent que les animaux ou les plantes. Une des difficultés auxquelles la détection des évènements d'hybridation fait face est l'extinction des lignées parentales, puisque sans elles il est difficile de détecter les hybrides. Avec les macroorganismes, il est possible de suivre les extinctions dans le registre fossile et même de récupérer des échantillons d'ADN ancien et même d'étudier des hybrides éteints (Slon *et coll.* 2018). Ceci est beaucoup plus ardu, voire impossible, à faire avec les microbes et rend l'étude de l'hybridation chez ces organismes centrée sur le passé récent et le présent. Les microorganismes seraient peut-être donc des modèles idéaux pour étudier l'initiation des évènements de spéciation par hybridation. (Eberlein *et coll.* 2019) ont trouvé une nouvelle lignée de *S. paradoxus* qui semble être un hybride récent entre *SpB* et *SpC**, révélant que l'hybridation continue de manière contemporaine en Amérique du Nord. Ceci pourrait suggérer que l'hybridation chez les microorganismes se fait de manière continue en nature et que les hybrides peuvent ensuite continuer de générer de nouvelles lignées par rétrocroisements.

Il est fort probable qu'il y ait une augmentation des évènements d'hybridation dans le contexte actuel de changements climatiques. Face à ceux-ci, les espèces n'auront d'autres choix que de migrer ou de s'adapter (Aitken *et coll.* 2008). D'un côté, les migrations augmenteront le nombre de rencontres interspécifiques entre espèces proches (Garroway *et coll.* 2010) et donc le nombre d'hybrides potentiels. De l'autre côté, les hybrides et les lignées introgressées pourraient alors mener à l'extinction des lignées non hybrides par exclusion compétitive (Muhlfeld *et coll.* 2014), menant à un enrichissement en hybrides.

Une des questions que soulève cette problématique à propos de l'hybridation est si cette dernière a un effet net positif ou encore négatif sur la biodiversité. Des termes comme « contamination génétique », « pollution génétique » et « agression génétique » sont souvent utilisés pour décrire l'hybridation et l'introgression chez les animaux (Rhymer et Simberloff 1996). Ces termes à connotation négative sont dus au fait que les hybrides peuvent venir à remplacer une espèce et aussi éliminer de la population les génomes parentaux « purs » (Muhlfeld *et coll.* 2014). Cependant, l'hétérosis et la ségrégation transgressive peuvent conférer un avantage aux hybrides dans certaines situations comme une résistance accrue au stress (Rieseberg *et coll.* 2007, Ma *et coll.* 2010) ou encore, contribuer à la sauvegarde évolutive de lignées génétiques (Stelkens *et coll.* 2014, Whiteley *et coll.* 2015). Il est donc important de bien comprendre la formation et l'évolution des hybrides pour savoir dans quelles circonstances ils sont une menace ou encore un atout pour la biodiversité présente et future (Allendorf *et coll.* 2001, Fitzpatrick *et coll.* 2015, Jackiw *et coll.* 2015).

BIBLIOGRAPHIE

Abbott, R., D. Albach, S. Ansell, J. W. Arntzen, S. J. E. Baird, N. Bierne, J. Boughman, A. Brelsford, C. A. Buerkle, R. Buggs, R. K. Butlin, U. Dieckmann, F. Eroukhmanoff, A. Grill, S. H. Cahan, J. S. Hermansen, G. Hewitt, A. G. Hudson, C. Jiggins, J. Jones, B. Keller, T. Marczewski, J. Mallet, P. Martinez-Rodriguez, M. Möst, S. Mullen, R. Nichols, A. W. Nolte, C. Parisod, K. Pfennig, A. M. Rice, M. G. Ritchie, B. Seifert, C. M. Smadja, R. Stelkens, J. M. Szymura, R. Väinölä, J. B. W. Wolf and D. Zinner (2013). "Hybridization and speciation." Journal of Evolutionary Biology **26**(2): 229-246.

Adams, K. L., Y. L. Qiu, M. Stoutemyer and J. D. Palmer (2002). "Punctuated evolution of mitochondrial gene content: high and variable rates of mitochondrial gene loss and transfer to the nucleus during angiosperm evolution." Proc Natl Acad Sci U S A **99**(15): 9905-9912.

Aitken, S. N., S. Yeaman, J. A. Holliday, T. Wang and S. Curtis-McLane (2008). "Adaptation, migration or extirpation: climate change outcomes for tree populations." Evolutionary Applications **1**(1): 95-111.

Albertin, W. and P. Marullo (2012). "Polyploidy in fungi: evolution after whole-genome duplication." Proc Biol Sci **279**(1738): 2497-2509.

Alberts, B., A. Johnson, J. Lewis, D. Morgan, M. Raff, K. Roberts and P. Walter (2014). Molecular Biology of the Cell, Sixth Edition, Taylor & Francis Group.

Allendorf, F. W., R. F. Leary, P. Spruell and J. K. Wenburg (2001). "The problems with hybrids: setting conservation guidelines." Trends in Ecology & Evolution **16**(11): 613-622.

Anderson, E. and L. Hubricht (1938). "Hybridization in Tradescantia. III. The Evidence for Introgressive Hybridization." American Journal of Botany **25**(6): 396-402.

Anderson, E. and G. L. Stebbins Jr. (1954). "Hybridization as an evolutionary stimulus." Evolution **8**(4): 378-388.

Anderson, J. B., J. Funt, D. A. Thompson, S. Prabhu, A. Socha, C. Sirjusingh, J. R. Dettman, L. Parreiras, D. S. Guttman, A. Regev and L. M. Kohn (2010). "Determinants of divergent adaptation and Dobzhansky-Muller interaction in experimental yeast populations." Curr Biol **20**(15): 1383-1388.

Anderson, J. B., D. Kasimer, W. Xia, N. C. H. Schröder, P. Cichowicz, S. Lioniello, R. Chakrabarti, E. Mohan and L. M. Kohn (2018). "Persistence of Resident and Transplanted Genotypes of the Undomesticated Yeast Saccharomyces paradoxus in Forest Soil." mSphere **3**(3): e00211-00218.

Anderson, J. B., C. Sirjusingh, A. B. Parsons, C. Boone, C. Wickens, L. E. Cowen and L. M. Kohn (2003). "Mode of selection and experimental evolution of antifungal drug resistance in Saccharomyces cerevisiae." Genetics **163**(4): 1287-1298.

Andersson, J. O. (2005). "Lateral gene transfer in eukaryotes." Cellular and Molecular Life Sciences CMLS **62**(11): 1182-1197.

April, J., R. H. Hanner, A. M. Dion-Cote and L. Bernatchez (2013). "Glacial cycles as an allopatric speciation pump in north-eastern American freshwater fishes." Mol Ecol **22**(2): 409-422.

Arnold, M. L. (2004). "Natural hybridization and the evolution of domesticated, pest and disease organisms." Molecular Ecology **13**(5): 997-1007.

Arnold, M. L. and N. H. Martin (2009). "Adaptation by introgression." J Biol **8**(9): 82.

Baker, E., B. Wang, N. Bellora, D. Peris, A. B. Hulfachor, J. A. Koshalek, M. Adams, D. Libkind and C. T. Hittinger (2015). "The genome sequence of Saccharomyces eubayanus and the domestication of lager-brewing yeasts." Molecular biology and evolution **32**(11): 2818-2831.

Barton, N. H. (2001). "The role of hybridization in evolution." Molecular ecology **10**(3): 551-568.

Barton, N. H. and G. M. Hewitt (1985). "Analysis of Hybrid Zones." Annual Review of Ecology and Systematics **16**(1): 113-148.

Bateson, W. (1909). Heredity and variation in modern lights, pp. 85–101 in Darwin and Modern Science, edited by Seward AC, Cambridge University Press, Cambridge, UK.

Bell, G. (1982). The Masterpiece of Nature:: The Evolution and Genetics of Sexuality, CUP Archive.

Belloch, C., S. Orlic, E. Barrio and A. Querol (2008). "Fermentative stress adaptation of hybrids within the Saccharomyces sensu stricto complex." International Journal of Food Microbiology **122**(1): 188-195.

Bergström, A., J. T. Simpson, F. Salinas, B. Barré, L. Parts, A. Zia, A. N. Nguyen Ba, A. M. Moses, E. J. Louis, V. Mustonen, J. Warringer, R. Durbin and G. Liti (2014). "A High-Definition View of Functional Genetic Variation from Natural Yeast Genomes." Molecular Biology and Evolution **31**(4): 872-888.

Bernardes, J. P., R. B. Stelkens and D. Greig (2016). "Heterosis in hybrids within and between yeast species." Journal of Evolutionary Biology **30**(3): 538-548.

Bernstein, H. and C. Bernstein (2010). "Evolutionary Origin of Recombination during Meiosis." BioScience **60**(7): 498-505.

Birky, C. W., R. L. Strausberg, J. L. Forster and P. S. Perlman (1978). "Vegetative segregation of mitochondria in yeast: Estimating parameters using a random model." Molecular and General Genetics MGG **158**(3): 251-261.

Bleuven, C., A. K. Dubé, G. Q. Nguyen, I. Gagnon-Arsenault, H. Martin and C. R. Landry (2018). "A collection of barcoded natural isolates of Saccharomyces paradoxus to study microbial evolutionary ecology." MicrobiologyOpen **0**(0): e773.

Bobay, L.-M. and H. Ochman (2017). "Biological Species Are Universal across Life's Domains." Genome Biology and Evolution **9**(3): 491-501.

Boekhout, T., B. Theelen, M. Diaz, J. W. Fell, W. C. J. Hop, E. C. A. Abeln, F. Dromer and W. Meyer (2001). "Hybrid genotypes in the pathogenic yeast Cryptococcus neoformans." Microbiology **147**(4): 891-907.

Bordenstein, S. R., F. P. O'Hara and J. H. Werren (2001). "Wolbachia-induced incompatibility precedes other hybrid incompatibilities in Nasonia." Nature **409**(6821): 707-710.

Boynton, P. J. and D. Greig (2014). "The ecology and evolution of non-domesticated Saccharomyces species." Yeast (Chichester, England) **31**(12): 449-462.

Boynton, P. J., R. Stelkens, V. Kowallik and D. Greig (2017). "Measuring microbial fitness in a field reciprocal transplant experiment." Mol Ecol Resour **17**(3): 370-380.

Brasier, C. M., D. E. L. Cooke and J. M. Duncan (1999). "Origin of a new Phytophthora pathogen through interspecific hybridization." Proceedings of the National Academy of Sciences **96**(10): 5878-5883.

Braun-Galleani, S., R. A. Ortiz-Merino, Q. Wu, Y. Xu and K. H. Wolfe (2018). "Zygosaccharomyces pseudobailii, another yeast interspecies hybrid that regained fertility by damaging one of its MAT loci." FEMS Yeast Res **18**(7).

Bruce, A. B. (1910). "The mendelian theory of heredity and the augmentation of vigor." Science **32**(827): 627-628.

Buerkle, C. A., R. J. Morris, M. A. Asmussen and L. H. Rieseberg (2000). "The likelihood of homoploid hybrid speciation." Heredity (Edinb) **84 (Pt 4)**: 441-451.

Butlin, R. (1987). "Speciation by reinforcement." Trends Ecol Evol **2**(1): 8-13.

Castric, V., J. Bechsgaard, M. H. Schierup and X. Vekemans (2008). "Repeated adaptive introgression at a gene under multiallelic balancing selection." PLoS Genetics **4**(8): e1000168.

Catchen, J., P. A. Hohenlohe, S. Bassham, A. Amores and W. A. Cresko (2013). "Stacks: an analysis tool set for population genomics." Mol Ecol **22**(11): 3124-3140.

Cavalieri, D., P. E. McGovern, D. L. Hartl, R. Mortimer and M. Polsinelli (2003). "Evidence for S. cerevisiae fermentation in ancient wine." Journal of molecular evolution **57**(1): S226-S232.

Chandley, A. C., R. C. Jones, H. M. Dott, W. R. Allen and R. V. Short (1974). "Meiosis in interspecific equine hybrids. I. The male mule (Equus asinus X E. caballus) and hinny (E. caballus X E. asinus)." Cytogenet Cell Genet **13**(4): 330-341.

Chapman, A. D. (2009). "Numbers of living species in Australia and the world."

Charlesworth, D. and J. H. Willis (2009). "The genetics of inbreeding depression." Nature Reviews Genetics **10**: 783.

Charron, G., J. B. Leducq, C. Bertin, A. K. Dube and C. R. Landry (2014). "Exploring the northern limit of the distribution of Saccharomyces cerevisiae and Saccharomyces paradoxus in North America." FEMS Yeast Res **14**(2): 281-288.

Charron, G., J. B. Leducq and C. R. Landry (2014). "Chromosomal variation segregates within incipient species and correlates with reproductive isolation." Molecular ecology **23**(17): 4362-4372.

Charron, G., J. B. Leducq and C. R. Landry (2014). "Chromosomal variation segregates within incipient species and correlates with reproductive isolation." Mol Ecol **23**(17): 4362-4372.

Chenoweth, S. F. and M. W. Blows (2006). "Dissecting the complex genetic basis of mate choice." Nature Reviews Genetics **7**: 681.

Cherry, J. M., E. L. Hong, C. Amundsen, R. Balakrishnan, G. Binkley, E. T. Chan, K. R. Christie, M. C. Costanzo, S. S. Dwight, S. R. Engel, D. G. Fisk, J. E. Hirschman, B. C. Hitz, K. Karra, C. J. Krieger, S. R. Miyasato, R. S. Nash, J. Park, M. S. Skrzypek, M. Simison, S. Weng and E. D. Wong (2012). "Saccharomyces Genome Database: the genomics resource of budding yeast." Nucleic Acids Res **40**(Database issue): D700-705.

Chou, J.-Y., Y.-S. Hung, K.-H. Lin, H.-Y. Lee and J.-Y. Leu (2010). "Multiple Molecular Mechanisms Cause Reproductive Isolation between Three Yeast Species." PLOS Biology **8**(7): e1000432.

Chou, J. Y., Y. S. Hung, K. H. Lin, H. Y. Lee and J. Y. Leu (2010). "Multiple molecular mechanisms cause reproductive isolation between three yeast species." PLoS Biol **8**(7): e1000432.

Ciani, M., I. Mannazzu, P. Marinangeli, F. Clementi and A. Martini (2004). "Contribution of winery-resident Saccharomyces cerevisiae strains to spontaneous grape must fermentation." Antonie van Leeuwenhoek **85**(2): 159-164.

Cinget, B., S. Gerardi, J. Beaulieu and J. Bousquet (2015). "Less pollen-mediated gene flow for more signatures of glacial lineages: congruent evidence from balsam fir cpDNA and mtDNA for multiple refugia in eastern and central North America." PLoS One **10**(4): e0122815.

Cliften, P. F., L. W. Hillier, L. Fulton, T. Graves, T. Miner, W. R. Gish, R. H. Waterston and M. Johnston (2001). "Surveying Saccharomyces genomes to identify functional elements by comparative DNA sequence analysis." Genome Res **11**(7): 1175-1186.

Clowers, K. J., J. L. Will and A. P. Gasch (2015). "A unique ecological niche fosters hybridization of oak-tree and vineyard isolates of Saccharomyces cerevisiae." Molecular Ecology **24**(23): 5886-5898.

Coluccio, A. and A. M. Neiman (2004). "Interspore bridges: a new feature of the Saccharomyces cerevisiae spore wall." Microbiology **150**(10): 3189-3196.

Coyne, J. A. and H. A. Orr (2004). Speciation, Sinauer.

Crow, J. F. (1948). "Alternative Hypotheses of Hybrid Vigor." Genetics **33**(5): 477-487.

Cutter, A. D. (2012). "The polymorphic prelude to Bateson-Dobzhansky-Muller incompatibilities." Trends Ecol Evol **27**(4): 209-218.

Dai, C.-F. and T. Fan (1992). Sexual Reproduction of Corals in Northern and Southen Taiwan. Proceedings of the Seventh International Coral Reef Symposium.

Dai, C. F., T. Y. Fan and J. K. Yu (2000). "Reproductive isolation and genetic differentiation of a scleractinian coral Mycedium elephantotus." Marine Ecology Progress Series **201**: 179-187.

Danecek, P., A. Auton, G. Abecasis, C. A. Albers, E. Banks, M. A. DePristo, R. E. Handsaker, G. Lunter, G. T. Marth, S. T. Sherry, G. McVean, R. Durbin and G. Genomes Project Analysis (2011). "The variant call format and VCFtools." Bioinformatics (Oxford, England) **27**(15): 2156-2158.

Darling, A. C., B. Mau, F. R. Blattner and N. T. Perna (2004). "Mauve: multiple alignment of conserved genomic sequence with rearrangements." Genome Res **14**(7): 1394-1403.

Davenport, C. B. (1908). "Degeneration, albinism and inbreeding." Science **28**(718): 454-455.

de Lafontaine, G., J. Turgeon and S. Payette (2010). "Phylogeography of white spruce (Picea glauca) in eastern North America reveals contrasting ecological trajectories." Journal of Biogeography **37**(4): 741-751.

De Storme, N. and A. Mason (2014). "Plant speciation through chromosome instability and ploidy change: Cellular mechanisms, molecular factors and evolutionary relevance." Current Plant Biology **1**: 10-33.

Degnan, J. H. (2018). "Modeling Hybridization Under the Network Multispecies Coalescent." Systematic Biology **67**(5): 786-799.

Diss, G., Alexandre K. Dubé, J. Boutin, I. Gagnon-Arsenault and Christian R. Landry (2013). "A Systematic Approach for the Genetic

Dissection of Protein Complexes in Living Cells." Cell Reports **3**(6): 2155-2167.

Dobzhansky, T. (1936). "Studies on hybrid sterility. II. Localization of sterility factors in Drosophila pseudoobscura hybrids." Genetics **21**(2): 113.

Dobzhansky, T. (1940). "Speciation as a Stage in Evolutionary Divergence." The American Naturalist **74**(753): 312-321.

Dobzhansky, T. (1951). Genetics and the origin of species. New York, Columbia University Press.

Dobzhansky, T. and C. Tan (1937). "Studies on hybrid sterility III." Zeitschrift für Induktive Abstammungs-und Vererbungslehre **72**(1): 88-114.

Dooh, R. T., S. J. Adamowicz and P. D. Hebert (2006). "Comparative phylogeography of two North American 'glacial relict' crustaceans." Mol Ecol **15**(14): 4459-4475.

Drummond, A. J., S. Y. Ho, M. J. Phillips and A. Rambaut (2006). "Relaxed phylogenetics and dating with confidence." PLoS Biol **4**(5): e88.

Drummond, A. J., M. A. Suchard, D. Xie and A. Rambaut (2012). "Bayesian phylogenetics with BEAUti and the BEAST 1.7." Mol Biol Evol **29**(8): 1969-1973.

Dufour, L. (1844). "Anatomie générale des Dipteres." Ann. Sci. Nat **1**: 244-264.

Dunham, M. J., H. Badrane, T. Ferea, J. Adams, P. O. Brown, F. Rosenzweig and D. Botstein (2002). "Characteristic genome rearrangements in experimental evolution of *Saccharomyces cerevisiae*." Proceedings of the National Academy of Sciences **99**(25): 16144-16149.

Dunn, B., T. Paulish, A. Stanbery, J. Piotrowski, G. Koniges, E. Kroll, E. J. Louis, G. Liti, G. Sherlock and F. Rosenzweig (2013). "Recurrent Rearrangement during Adaptive Evolution in an Interspecific Yeast Hybrid Suggests a Model for Rapid Introgression." PLOS Genetics **9**(3): e1003366.

Dunn, B. and G. Sherlock (2008). "Reconstruction of the genome origins and evolution of the hybrid lager yeast Saccharomyces pastorianus." Genome Research.

Dutta, A., G. Lin, A. V. Pankajam, P. Chakraborty, N. Bhat, L. M. Steinmetz and K. T. Nishant (2017). "Genome Dynamics of Hybrid Saccharomyces cerevisiae During Vegetative and Meiotic Divisions." G3 (Bethesda) **7**(11): 3669-3679.

Dyke, A. S. (2004). An outline of North American deglaciation with emphasis on central and northern Canada. Developments in Quaternary Sciences, Elsevier. **2:** 373-424.

East, E. (1908). Inbreeding in corn. Reports of the Connecticut agricultural experiment station for yeasrs 1907-1908. New Haven, CT, Connecticut Agricultural Experiment Station**:** 419-428.

Eberlein, C., M. Henault, A. Fijarczyk, G. Charron, M. Bouvier, L. M. Kohn, J. B. Anderson and C. R. Landry (2019). "Hybridization is a recurrent evolutionary stimulus in wild yeast speciation." Nat Commun **10**(1): 923.

Eberlein, C., M. Hénault, A. Fijarczyk, G. Charron, M. Bouvier, L. M. Kohn, J. B. Anderson and C. R. Landry (2019). "Hybridization is a recurrent evolutionary stimulus in wild yeast speciation." Nature Communications **10**(1): 923.

Eberlein, C., L. Nielly-Thibault, H. Maaroufi, A. K. Dube, J. B. Leducq, G. Charron and C. R. Landry (2017). "The Rapid Evolution of an Ohnolog Contributes to the Ecological Specialization of Incipient Yeast Species." Mol Biol Evol **34**(9): 2173-2186.

Edgar, B. A. and T. L. Orr-Weaver (2001). "Endoreplication cell cycles: more for less." Cell **105**(3): 297-306.

Ellison, C. E., C. Hall, D. Kowbel, J. Welch, R. B. Brem, N. L. Glass and J. W. Taylor (2011). "Population genomics and local adaptation in wild isolates of a model microbial eukaryote." Proceedings of the National Academy of Sciences **108**(7): 2831-2836.

Evanno, G., S. Regnaut and J. Goudet (2005). "Detecting the number of clusters of individuals using the software STRUCTURE: a simulation study." Mol Ecol **14**(8): 2611-2620.

Falkowski, P. G., T. Fenchel and E. F. Delong (2008). "The microbial engines that drive Earth's biogeochemical cycles." science **320**(5879): 1034-1039.

Falush, D., M. Stephens and J. K. Pritchard (2003). "Inference of Population Structure Using Multilocus Genotype Data: Linked Loci and Correlated Allele Frequencies." Genetics **164**(4): 1567-1587.

Fan, T.-Y. and C.-F. Dai (1998). "Sexual Reproduction of the Scleractinian Coral Merulina Ampliata in Southern Taiwan." Bulletin of Marine Science **62**(3): 897-904.

Farhi, F., G. S. Bulmer and J. R. Tacker (1970). "Cryptococcus neoformans IV. The Not-So-Encapsulated Yeast." Infect Immun **1**(6): 526-531.

Fay, J. C. and J. A. Benavides (2005). "Evidence for domesticated and wild populations of Saccharomyces cerevisiae." PLoS Genet **1**(1): 66-71.

Feder, J. L., C. A. Chilcote and G. L. Bush (1988). "Genetic differentiation between sympatric host races of the apple maggot fly Rhagoletis pomonella." Nature **336**(6194): 61-64.

Felsenstein, J. (1981). "Skepticism towards Santa Rosalia, or why are there so few kinds of animals?" Evolution **35**(1): 124-138.

Feng, W. and G. D'Urso (2001). "Schizosaccharomyces pombe cells lacking the amino-terminal catalytic domains of DNA polymerase epsilon are viable but require the DNA damage checkpoint control." Mol Cell Biol **21**(14): 4495-4504.

Filteau, M., G. Charron and C. R. Landry (2017). "Identification of the fitness determinants of budding yeast on a natural substrate." ISME J **11**(4): 959-971.

Filteau, M., L. Lagacé, G. LaPointe and D. Roy (2011). "Correlation of maple sap composition with bacterial and fungal communities determined by multiplex automated ribosomal intergenic spacer analysis (MARISA)." Food Microbiology **28**(5): 980-989.

Finlay, B. J. (2002). "Global Dispersal of Free-Living Microbial Eukaryote Species." Science **296**(5570): 1061-1063.

Fischer, G., S. A. James, I. N. Roberts, S. G. Oliver and E. J. Louis (2000). "Chromosomal evolution in Saccharomyces." Nature **405**: 451.

Fitzpatrick, B. M., M. E. Ryan, J. R. Johnson, J. Corush and E. T. Carter (2015). "Hybridization and the species problem in conservation." Current Zoology **61**(1): 206-216.

Ford, M. J. and C. F. Aquadro (1996). "Selection on X-linked genes during speciation in the Drosophila athabasca complex." Genetics **144**(2): 689-703.

Futuyma, D. J. and G. C. Mayer (1980). "Non-Allopatric Speciation in Animals." Systematic Biology **29**(3): 254-271.

Garrison, E. and G. Marth (2012). "Haplotype-based variant detection from short-read sequencing." arXiv preprint arXiv:1207.3907.

Garroway, C. J., J. Bowman, T. J. cascaden, G. L. Holloway, C. G. Mahan, J. R. Malcolm, M. A. Steele, G. Turner and P. J. Wilson (2010). "Climate change induced hybridization in flying squirrels." Global Change Biology **16**(1): 113-121.

Gavrilets, S., H. Li and M. D. Vose (2000). "PATTERNS OF PARAPATRIC SPECIATION." Evolution **54**(4): 1126-1134.

Gerstein, A. C., H.-J. E. Chun, A. Grant and S. P. Otto (2006). "Genomic Convergence toward Diploidy in Saccharomyces cerevisiae." PLOS Genetics **2**(9): e145.

Gevers, D., F. M. Cohan, J. G. Lawrence, B. G. Spratt, T. Coenye, E. J. Feil, E. Stackebrandt, Y. V. de Peer, P. Vandamme, F. L. Thompson and J. Swings (2005). "Re-evaluating prokaryotic species." Nature Reviews Microbiology **3**: 733.

Giaever, G., A. M. Chu, L. Ni, C. Connelly, L. Riles, S. Veronneau, S. Dow, A. Lucau-Danila, K. Anderson, B. Andre, A. P. Arkin, A. Astromoff, M. El-Bakkoury, R. Bangham, R. Benito, S. Brachat, S. Campanaro, M. Curtiss, K. Davis, A. Deutschbauer, K. D. Entian, P. Flaherty, F. Foury, D. J. Garfinkel, M. Gerstein, D. Gotte, U. Guldener, J. H. Hegemann, S. Hempel, Z. Herman, D. F. Jaramillo, D. E. Kelly, S. L. Kelly, P. Kotter, D. LaBonte, D. C. Lamb, N. Lan, H. Liang, H. Liao, L. Liu, C. Luo, M. Lussier, R. Mao, P. Menard, S. L. Ooi, J. L. Revuelta, C. J. Roberts, M. Rose, P. Ross-Macdonald, B. Scherens, G. Schimmack, B. Shafer, D. D. Shoemaker, S. Sookhai-Mahadeo, R. K. Storms, J. N. Strathern, G. Valle, M. Voet, G. Volckaert, C. Y. Wang, T. R. Ward, J. Wilhelmy, E. A. Winzeler, Y. Yang, G. Yen, E. Youngman, K. Yu, H. Bussey, J. D. Boeke, M. Snyder, P. Philippsen, R. W. Davis and M. Johnston (2002). "Functional profiling of the Saccharomyces cerevisiae genome." Nature **418**(6896): 387-391.

Gilchrist, C. and R. Stelkens (2019). "Aneuploidy in Yeast: Segregation Error or Adaptation Mechanism?" Yeast.

Giraud, T. and S. Gourbière (2012). "The tempo and modes of evolution of reproductive isolation in fungi." Heredity **109**: 204.

Giraud, T., G. Refrégier, M. Le Gac, D. M. de Vienne and M. E. Hood (2008). "Speciation in fungi." Fungal Genetics and Biology **45**(6): 791-802.

Godbout, J., J. P. Jaramillo-Correa, J. Beaulieu and J. Bousquet (2005). "A mitochondrial DNA minisatellite reveals the postglacial history of jack pine (Pinus banksiana), a broad-range North American conifer." Mol Ecol **14**(11): 3497-3512.

Goddard, M. R. and D. Greig (2015). "Saccharomyces cerevisiae: a nomadic yeast with no niche?" FEMS Yeast Research **15**(3): fov009-fov009.

Goffeau, A., B. G. Barrell, H. Bussey, R. W. Davis, B. Dujon, H. Feldmann, F. Galibert, J. D. Hoheisel, C. Jacq, M. Johnston, E. J. Louis, H. W. Mewes, Y. Murakami, P. Philippsen, H. Tettelin and S. G. Oliver (1996). "Life with 6000 Genes." Science **274**(5287): 546-567.

Gonçalves, P., E. Valério, C. Correia, J. M. de Almeida and J. P. Sampaio (2011). "Evidence for divergent evolution of growth temperature preference in sympatric Saccharomyces species." PloS one **6**(6): e20739.

González, S. S., E. Barrio, J. Gafner and A. Querol (2006). "Natural hybrids from Saccharomyces cerevisiae, Saccharomyces bayanus and Saccharomyces kudriavzevii in wine fermentations." FEMS Yeast Research **6**(8): 1221-1234.

González, S. S., E. Barrio and A. Querol (2008). "Molecular characterization of new natural hybrids of Saccharomyces cerevisiae and S. kudriavzevii in brewing." Applied and environmental microbiology **74**(8): 2314-2320.

Gosset, C. C. and N. Bierne (2013). "Differential introgression from a sister species explains high FST outlier loci within a mussel species." Journal of Evolutionary Biology **26**(1): 14-26.

Goudet, J. (2005). "Hierfstat, a package for R to compute and test hierarchical F-statistics." Molecular Ecology Notes **5**(1): 184-186.

Grant, P. R. and B. R. Grant (1994). "Phenotypic and Genetic Effects of Hybridization in Darwin's Finches." Evolution **48**(2): 297-316.

Grant, V. (1981). Plant speciation, Columbia University Press.

Grant, V. (1994). "Modes and origins of mechanical and ethological isolation in angiosperms." Proceedings of the National Academy of Sciences **91**(1): 3-10.

Green, R. E., J. Krause, A. W. Briggs, T. Maricic, U. Stenzel, M. Kircher, N. Patterson, H. Li, W. Zhai, M. H. Fritz, N. F. Hansen, E. Y. Durand, A. S. Malaspinas, J. D. Jensen, T. Marques-Bonet, C. Alkan, K. Prufer, M. Meyer, H. A. Burbano, J. M. Good, R. Schultz, A. Aximu-Petri, A. Butthof, B. Hober, B. Hoffner, M. Siegemund, A. Weihmann, C. Nusbaum, E. S. Lander, C. Russ, N. Novod, J. Affourtit, M. Egholm, C. Verna, P. Rudan, D. Brajkovic, Z. Kucan, I. Gusic, V. B. Doronichev, L. V. Golovanova, C. Lalueza-Fox, M. de la Rasilla, J. Fortea, A. Rosas, R. W. Schmitz, P. L. F. Johnson, E. E. Eichler, D. Falush, E. Birney, J. C. Mullikin, M. Slatkin, R. Nielsen, J. Kelso, M. Lachmann, D. Reich and S. Paabo (2010). "A draft sequence of the Neandertal genome." Science **328**(5979): 710-722.

Greig, D. (2008). "Reproductive isolation in Saccharomyces." Heredity **102**: 39.

Greig, D., R. H. Borts, E. J. Louis and M. Travisano (2002). "Epistasis and hybrid sterility in Saccharomyces." Proc Biol Sci **269**(1496): 1167-1171.

Greig, D., E. J. Louis, R. H. Borts and M. Travisano (2002). "Hybrid Speciation in Experimental Populations of Yeast." Science **298**(5599): 1773-1775.

Greig, D., M. Travisano, E. J. Louis and R. H. Borts (2003). "A role for the mismatch repair system during incipient speciation in Saccharomyces." Journal of Evolutionary Biology **16**(3): 429-437.

Greig, D., M. Travisano, E. J. Louis and R. H. Borts (2003). "A role for the mismatch repair system during incipient speciation in Saccharomyces." J Evol Biol **16**(3): 429-437.

Gresham, D., M. M. Desai, C. M. Tucker, H. T. Jenq, D. A. Pai, A. Ward, C. G. DeSevo, D. Botstein and M. J. Dunham (2008). "The Repertoire and Dynamics of Evolutionary Adaptations to Controlled Nutrient-Limited Environments in Yeast." PLOS Genetics **4**(12): e1000303.

Guindon, S., J.-F. Dufayard, V. Lefort, M. Anisimova, W. Hordijk and O. Gascuel (2010). "New Algorithms and Methods to Estimate Maximum-Likelihood Phylogenies: Assessing the Performance of PhyML 3.0." Systematic Biology **59**(3): 307-321.

Güldener, U., S. Heck, T. Fiedler, J. Beinhauer and J. H. Hegemann (1996). "A new efficient gene disruption cassette for repeated use in budding yeast." Nucleic acids research **24**(13): 2519-2524.

Gustafsson, Å. (1979). "Linnaeus' Peloria: The history of a monster." Theoretical and Applied Genetics **54**(6): 241-248.

Hagen, D. C., L. Bruhn, C. A. Westby and G. F. Sprague (1993). "Transcription of alpha-specific genes in Saccharomyces cerevisiae: DNA sequence requirements for activity of the coregulator alpha 1." Molecular and Cellular Biology **13**(11): 6866-6875.

Harari, Y., Y. Ram and M. Kupiec (2018). "Frequent ploidy changes in growing yeast cultures." Curr Genet **64**(5): 1001-1004.

Harfe, B. D. and S. Jinks-Robertson (2000). "DNA mismatch repair and genetic instability." Annual review of genetics **34**(1): 359-399.

Harlan, J. R. and J. M. J. deWet (1975). "On Ö. Winge and a Prayer: The origins of polyploidy." The Botanical Review **41**(4): 361-390.

Harrison, R. G. (1990). "Hybrid zones: windows on evolutionary process." Oxford surveys in evolutionary biology **7**: 69-128.

Hatfield, T. and D. Schluter (1999). "Ecological speciation in sticklebacks: Environment-dependent hybrid fitness." Evolution **53**(3): 866-873.

Hawthorne, D. J. and S. Via (2001). "Genetic linkage of ecological specialization and reproductive isolation in pea aphids." Nature **412**: 904.

Haynes, R. H. and B. A. Kunz (1981). DNA Repair and Mutagenesis in Yeast.

Hedrick, P. W. (2013). "Adaptive introgression in animals: examples and comparison to new mutation and standing variation as sources of adaptive variation." Molecular ecology **22**(18): 4606-4618.

Hénault, M., C. Eberlein, G. Charron, É. Durand, L. Nielly-Thibault, H. Martin and C. R. Landry (2019). Yeast Population Genomics Goes Wild: The Case of Saccharomyces paradoxus. Population Genomics: Microorganisms. M. F. Polz and O. P. Rajora. Cham, Springer International Publishing**:** 207-230.

Hewitt, G. M. (1988). "Hybrid zones-natural laboratories for evolutionary studies." Trends in Ecology & Evolution **3**(7): 158-167.

Hewitt, G. M. (2008). "Some genetic consequences of ice ages, and their role in divergence and speciation." Biological Journal of the Linnean Society **58**(3): 247-276.

Hibbins, M. S. and M. W. Hahn (2019). "The Timing and Direction of Introgression Under the Multispecies Network Coalescent." Genetics: genetics.301831.302018.

Hill, J. A. and S. P. Otto (2007). "The role of pleiotropy in the maintenance of sex in yeast." Genetics **175**(3): 1419-1427.

Hilu, K. W. (1993). "Polyploidy and the Evolution of Domesticated Plants." American Journal of Botany **80**(12): 1494-1499.

Hittinger, C. T. (2013). "Saccharomyces diversity and evolution: a budding model genus." Trends Genet **29**(5): 309-317.

Hopkins, R. (2013). "Reinforcement in plants." New Phytologist **197**(4): 1095-1103.

Hou, J., A. Friedrich, J. de Montigny and J. Schacherer (2014). "Chromosomal rearrangements as a major mechanism in the onset of reproductive isolation in Saccharomyces cerevisiae." Current Biology **24**(10): 1153-1159.

Hou, J., A. Friedrich, J.-S. Gounot and J. Schacherer (2015). "Comprehensive survey of condition-specific reproductive isolation reveals genetic incompatibility in yeast." Nature Communications **6**: 7214.

Huerta-Sánchez, E., X. Jin, Asan, Z. Bianba, B. M. Peter, N. Vinckenbosch, Y. Liang, X. Yi, M. He, M. Somel, P. Ni, B. Wang, X. Ou, Huasang, J. Luosang, Z. X. P. Cuo, K. Li, G. Gao, Y. Yin, W. Wang, X. Zhang, X. Xu, H. Yang, Y. Li, J. Wang, J. Wang and R. Nielsen (2014). "Altitude adaptation in Tibetans caused by introgression of Denisovan-like DNA." Nature **512**: 194.

Hufton, A. L. and G. Panopoulou (2009). "Polyploidy and genome restructuring: a variety of outcomes." Current opinion in genetics & development **19**(6): 600-606.

Hunter, N., S. R. Chambers, E. J. Louis and R. H. Borts (1996). "The mismatch repair system contributes to meiotic sterility in an interspecific yeast hybrid." The EMBO Journal **15**(7): 1726-1733.

Hyma, K. E. and J. C. Fay (2013). "Mixing of vineyard and oak-tree ecotypes of Saccharomyces cerevisiae in North American vineyards." Molecular Ecology **22**(11): 2917-2930.

Jackiw, R. N., G. Mandil and H. A. Hager (2015). "A framework to guide the conservation of species hybrids based on ethical and ecological considerations." Conservation Biology **29**(4): 1040-1051.

Janke, C., M. M. Magiera, N. Rathfelder, C. Taxis, S. Reber, H. Maekawa, A. Moreno-Borchart, G. Doenges, E. Schwob, E. Schiebel and M. Knop (2004). "A versatile toolbox for PCR-based tagging of yeast genes: new fluorescent proteins, more markers and promoter substitution cassettes." Yeast **21**(11): 947-962.

Joseph, S. B. and D. W. Hall (2004). "Spontaneous mutations in diploid Saccharomyces cerevisiae: more beneficial than expected." Genetics **168**(4): 1817-1825.

Joseph, S. B. and D. W. Hall (2004). "Spontaneous mutations in diploid Saccharomyces cerevisiae: more beneficial than expected." Genetics **168**(4): 1817-1825.

Kaeppler, S. (2012). "Heterosis: Many Genes, Many Mechanisms-End the Search for an Undiscovered Unifying Theory." ISRN Botany **2012**: 12.

Karley, A. J., A. E. Douglas and W. E. Parker (2002). "Amino acid composition and nutritional quality of potato leaf phloem sap for aphids." Journal of Experimental Biology **205**(19): 3009-3018.

Kassambara A., K. M. (2018). "survminer: Drawing Survival Curves using 'ggplot2'."

Keeble, F. and C. Pellew (1910). "The mode of inheritance of stature and of time of flowering in peas (Pisum sativum)." Journal of Genetics **1**(1): 47-56.

Keller, I. and O. Seehausen (2012). "Thermal adaptation and ecological speciation." Molecular Ecology **21**(4): 782-799.

Kellis, M., N. Patterson, M. Endrizzi, B. Birren and E. S. Lander (2003). "Sequencing and comparison of yeast species to identify genes and regulatory elements." Nature **423**(6937): 241.

Kellis, M., N. Patterson, M. Endrizzi, B. Birren and E. S. Lander (2003). "Sequencing and comparison of yeast species to identify genes and regulatory elements." Nature **423**(6937): 241-254.

Khachane, A. N., K. N. Timmis and V. A. Martins dos Santos (2007). "Dynamics of reductive genome evolution in mitochondria and obligate intracellular microbes." Mol Biol Evol **24**(2): 449-456.

Kim, J. M., S. Vanguri, J. D. Boeke, A. Gabriel and D. F. Voytas (1998). "Transposable elements and genome organization: a comprehensive survey of retrotransposons revealed by the complete Saccharomyces cerevisiae genome sequence." Genome Res **8**(5): 464-478.

Kirkpatrick, M. and M. J. Ryan (1991). "The evolution of mating preferences and the paradox of the lek." Nature **350**(6313): 33-38.

Klar, A. (1987). "The mother-daughter mating type switching asymmetry of budding yeast is not conferred by the segregation of parental HO gene DNA strands." Genes & development **1**(10): 1059-1064.

Klar, A. J., J. B. Hicks and J. N. Strathern (1982). "Directionality of yeast mating-type interconversion." Cell **28**(3): 551-561.

Knop, M. (2006). "Evolution of the hemiascomycete yeasts: on life styles and the importance of inbreeding." Bioessays **28**(7): 696-708.

Koufopanou, V., J. Hughes, G. Bell and A. Burt (2006). "The spatial scale of genetic differentiation in a model organism: the wild yeast Saccharomyces paradoxus." Philosophical Transactions of the Royal Society B: Biological Sciences **361**(1475): 1941-1946.

Kowallik, V. and D. Greig (2016). "A systematic forest survey showing an association of Saccharomyces paradoxus with oak leaf litter." Environmental Microbiology Reports **8**(5): 833-841.

Kowallik, V. and D. Greig (2016). "A systematic forest survey showing an association of Saccharomyces paradoxus with oak leaf litter." Environ Microbiol Rep **8**(5): 833-841.

Kowallik, V., E. Miller and D. Greig (2015). "The interaction of Saccharomyces paradoxus with its natural competitors on oak bark." Molecular Ecology **24**(7): 1596-1610.

Kronforst, M. R., L. G. Young, L. M. Blume and L. E. Gilbert (2006). "Multilocus analyses of admixture and introgression among hybridizing Heliconius butterflies." Evolution **60**(6): 1254-1268.

Kuehne, H. A., H. A. Murphy, C. A. Francis and P. D. Sniegowski (2007). "Allopatric divergence, secondary contact, and genetic isolation in wild yeast populations." Curr Biol **17**(5): 407-411.

Kuehne, H. A., H. A. Murphy, C. A. Francis and P. D. Sniegowski (2007). "Allopatric divergence, secondary contact, and genetic isolation in wild yeast populations." Current Biology **17**(5): 407-411.

Küenzi, M. T., M. A. Tingle and H. O. Halvorson (1974). "Sporulation of Saccharomyces cerevisiae in the absence of a functional mitochondrial genome." Journal of bacteriology **117**(1): 80-88.

Kumaran, R., S. Y. Yang and J. Y. Leu (2013). "Characterization of chromosome stability in diploid, polyploid and hybrid yeast cells." PLoS One **8**(7): e68094.

Lagercrantz, U. and D. J. Lydiate (1996). "Comparative genome mapping in Brassica." Genetics **144**(4): 1903-1910.

Lande, R. (1985). "The fixation of chromosomal rearrangements in a subdivided population with local extinction and colonization." Heredity (Edinb) **54 (Pt 3)**: 323-332.

Langmead, B. and S. L. Salzberg (2012). "Fast gapped-read alignment with Bowtie 2." Nature methods **9**(4): 357-359.

Langmead, B. and S. L. Salzberg (2012). "Fast gapped-read alignment with Bowtie 2." Nat Methods **9**(4): 357-359.

Lawrence, J. G. (2002). "Gene Transfer in Bacteria: Speciation without Species?" Theoretical Population Biology **61**(4): 449-460.

Leducq, J.-B., G. Charron, P. Samani, A. K. Dubé, K. Sylvester, B. James, P. Almeida, J. P. Sampaio, C. T. Hittinger and G. Bell (2014). "Local

climatic adaptation in a widespread microorganism." Proceedings of the Royal Society of London B: Biological Sciences **281**(1777): 20132472.

Leducq, J.-B., L. Nielly-Thibault, G. Charron, C. Eberlein, J.-P. Verta, P. Samani, K. Sylvester, C. T. Hittinger, G. Bell and C. R. Landry (2016). "Speciation driven by hybridization and chromosomal plasticity in a wild yeast." Nature Microbiology **1**: 15003.

Leducq, J. B., M. Hénault, G. Charron, L. Nielly-Thibault, Y. Terrat, H. L. Fiumera, B. J. Shapiro and C. R. Landry (2017). "Mitochondrial Recombination and Introgression during Speciation by Hybridization." Mol Biol Evol **34**(8): 1947-1959.

Leducq, J. B., L. Nielly-Thibault, G. Charron, C. Eberlein, J. P. Verta, P. Samani, K. Sylvester, C. T. Hittinger, G. Bell and C. R. Landry (2016). "Speciation driven by hybridization and chromosomal plasticity in a wild yeast." Nat Microbiol **1**: 15003.

Lee, C.-S. and J. E. Haber (2015). "Mating-type Gene Switching in Saccharomyces cerevisiae." Microbiology Spectrum **3**(2).

Lee, P. S., P. W. Greenwell, M. Dominska, M. Gawel, M. Hamilton and T. D. Petes (2009). "A Fine-Structure Map of Spontaneous Mitotic Crossovers in the Yeast Saccharomyces cerevisiae." PLOS Genetics **5**(3): e1000410.

Lee, P. S., P. W. Greenwell, M. Dominska, M. Gawel, M. Hamilton and T. D. Petes (2009). "A fine-structure map of spontaneous mitotic crossovers in the yeast Saccharomyces cerevisiae." PLoS Genet **5**(3): e1000410.

Lenski, R. E., M. R. Rose, S. C. Simpson and S. C. Tadler (1991). "Long-Term Experimental Evolution in Escherichia coli. I. Adaptation and Divergence During 2,000 Generations." The American Naturalist **138**(6): 1315-1341.

Leu, J.-Y. and A. W. Murray (2006). "Experimental evolution of mating discrimination in budding yeast." Current Biology **16**(3): 280-286.

Lewontin, R. C. and L. C. Birch (1966). "Hybridization as a Source of Variation for Adaptation to New Environments." Evolution **20**(3): 315-336.

Li, H., B. Handsaker, A. Wysoker, T. Fennell, J. Ruan, N. Homer, G. Marth, G. Abecasis, R. Durbin and S. Genome Project Data Processing (2009). "The Sequence Alignment/Map format and SAMtools." Bioinformatics **25**(16): 2078-2079.

Libkind, D., C. T. Hittinger, E. Valério, C. Gonçalves, J. Dover, M. Johnston, P. Gonçalves and J. P. Sampaio (2011). "Microbe domestication and the identification of the wild genetic stock of lager-brewing yeast." Proceedings of the National Academy of Sciences **108**(35): 14539-14544.

Liti, G., D. B. Barton and E. J. Louis (2006). "Sequence diversity, reproductive isolation and species concepts in Saccharomyces." Genetics **174**(2): 839-850.

Liti, G., D. B. Barton and E. J. Louis (2006). "Sequence diversity, reproductive isolation and species concepts in Saccharomyces." Genetics.

Liti, G., D. M. Carter, A. M. Moses, J. Warringer, L. Parts, S. A. James, R. P. Davey, I. N. Roberts, A. Burt, V. Koufopanou, I. J. Tsai, C. M. Bergman,

D. Bensasson, M. J. O'Kelly, A. van Oudenaarden, D. B. Barton, E. Bailes, A. N. Nguyen, M. Jones, M. A. Quail, I. Goodhead, S. Sims, F. Smith, A. Blomberg, R. Durbin and E. J. Louis (2009). "Population genomics of domestic and wild yeasts." Nature **458**(7236): 337-341.

Loidl, J., Q. W. Jin and M. Jantsch (1998). "Meiotic pairing and segregation of translocation quadrivalents in yeast." Chromosoma **107**(4): 247-254.

Looke, M., K. Kristjuhan and A. Kristjuhan (2011). "Extraction of genomic DNA from yeasts for PCR-based applications." Biotechniques **50**(5): 325-328.

Lõoke, M., K. Kristjuhan and A. Kristjuhan (2011). "Extraction of genomic DNA from yeasts for PCR-based applications." Biotechniques **50**(5): 325.

Lukhtanov, V. A., N. A. Shapoval, A. Anokhin Boris, F. Saifitdinova Alsu and G. Kuznetsova Valentina (2015). "Homoploid hybrid speciation and genome evolution via chromosome sorting." Proceedings of the Royal Society B: Biological Sciences **282**(1807): 20150157.

Luo, J., X. Sun, B. P. Cormack and J. D. Boeke (2018). "Karyotype engineering by chromosome fusion leads to reproductive isolation in yeast." Nature **560**(7718): 392-396.

Lynch, M., W. Sung, K. Morris, N. Coffey, C. R. Landry, E. B. Dopman, W. J. Dickinson, K. Okamoto, S. Kulkarni, D. L. Hartl and W. K. Thomas (2008). "A genome-wide view of the spectrum of spontaneous mutations in yeast." Proc Natl Acad Sci U S A **105**(27): 9272-9277.

Ma, F., C. Zhao, R. Milne, M. Ji, L. Chen and J. Liu (2010). "Enhanced drought-tolerance in the homoploid hybrid species Pinus densata: implication for its habitat divergence from two progenitors." New Phytologist **185**(1): 204-216.

Mable, B. K. (2004). "'Why polyploidy is rarer in animals than in plants': myths and mechanisms." Biological Journal of the Linnean Society **82**(4): 453-466.

Maclean, C. J. and D. Greig (2008). "Prezygotic reproductive isolation between Saccharomyces cerevisiae and Saccharomyces paradoxus." BMC evolutionary biology **8**: 1-1.

Maganti, H., D. Bartfai and J. Xu (2012). "Ecological structuring of yeasts associated with trees around Hamilton, Ontario, Canada." FEMS Yeast Research **12**(1): 9-19.

Magwene, P. M., Ö. Kayıkçı, J. A. Granek, J. M. Reininga, Z. Scholl and D. Murray (2011). "Outcrossing, mitotic recombination, and life-history trade-offs shape genome evolution in Saccharomyces cerevisiae." Proceedings of the National Academy of Sciences **108**(5): 1987-1992.

Maheshwari, S. and D. A. Barbash (2011). "The genetics of hybrid incompatibilities." Annu Rev Genet **45**: 331-355.

Majerus, M., P. O'Donald and J. Weir (1982). "Evidence for preferential mating in Adalia bipunctata." Heredity **49**: 37.

Mallet, J. (2005). "Hybridization as an invasion of the genome." Trends Ecol Evol **20**(5): 229-237.

Mallet, J. (2007). "Hybrid speciation." Nature **446**(7133): 279-283.

Mallet, J. (2007). "Hybrid speciation." Nature **446**: 279.

Mani, G. S. and B. C. Clarke (1990). "Mutational order: a major stochastic process in evolution." Proceedings of the Royal Society of London. B. Biological Sciences **240**(1297): 29-37.

Marcet-Houben, M. and T. Gabaldon (2015). "Beyond the Whole-Genome Duplication: Phylogenetic Evidence for an Ancient Interspecies Hybridization in the Baker's Yeast Lineage." PLoS Biol **13**(8): e1002220.

Marsden-Jones, E. M. (1930). "The genetics of Geum intermedium willd. Haud ehrh., And its back-crosses." Journal of Genetics **23**(3): 377-395.

Marshall, D. C. and J. R. Cooley (2000). "Reproductive character displacement and speciation in periodical cicadas, with description of new species, 13-year Magicicada neotredecem." Evolution **54**(4): 1313-1325.

Martin, H., P. Touzet, M. Dufay, C. Godé, E. Schmitt, E. Lahiani, L. F. Delph and F. Van Rossum (2017). "Lineages of Silene nutans developed rapid, strong, asymmetric postzygotic reproductive isolation in allopatry." Evolution **71**(6): 1519-1531.

Mascher, M., T. A. Richmond, D. J. Gerhardt, A. Himmelbach, L. Clissold, D. Sampath, S. Ayling, B. Steuernagel, M. Pfeifer, M. D'Ascenzo, E. D. Akhunov, P. E. Hedley, A. M. Gonzales, P. L. Morrell, B. Kilian, F. R. Blattner, U. Scholz, K. F. Mayer, A. J. Flavell, G. J. Muehlbauer, R. Waugh, J. A. Jeddeloh and N. Stein (2013). "Barley whole exome capture: a

tool for genomic research in the genus Hordeum and beyond." Plant J **76**(3): 494-505.

Masly, J. P. (2012). "170 years of “lock-and-key”: genital morphology and reproductive isolation." International Journal of Evolutionary Biology **2012**.

Masneuf, I., J. Hansen, C. Groth, J. Piskur and D. Dubourdieu (1998). "New Hybrids between *Saccharomyces* Sensu Stricto Yeast Species Found among Wine and Cider Production Strains." Applied and Environmental Microbiology **64**(10): 3887-3892.

Mason, A. S. and J. Batley (2015). "Creating new interspecific hybrid and polyploid crops." Trends Biotechnol **33**(8): 436-441.

Masterson, J. (1994). "Stomatal Size in Fossil Plants: Evidence for Polyploidy in Majority of Angiosperms." Science **264**(5157): 421-424.

Matute, D. R., I. A. Butler, D. A. Turissini and J. A. Coyne (2010). "A test of the snowball theory for the rate of evolution of hybrid incompatibilities." Science **329**(5998): 1518-1521.

Mayr, E. (1963). Animal Species and Evolution.

Mayr, E. (1982). The Growth of Biological Thought: Diversity, Evolution, and Inheritance, Belknap Press.

McPhail, J. D. (1984). "Ecology and evolution of sympatric sticklebacks (Gasterosteus): morphological and genetic evidence for a species pair in

Enos Lake, British Columbia." Canadian Journal of Zoology **62**(7): 1402-1408.

McPhail, J. D. (1992). "Ecology and evolution of sympatric sticklebacks (Gasterosteus): evidence for a species-pair in Paxton Lake, Texada Island, British Columbia." Canadian Journal of Zoology **70**(2): 361-369.

Merlini, L., O. Dudin and S. G. Martin (2013). "Mate and fuse: how yeast cells do it." Open biology **3**(3): 130008-130008.

Montrocher, R., M. C. Verner, J. Briolay, C. Gautier and R. Marmeisse (1998). "Phylogenetic analysis of the Saccharomyces cerevisiae group based on polymorphisms of rDNA spacer sequences." Int J Syst Bacteriol **48 Pt 1**: 295-303.

Morales, L. and B. Dujon (2012). "Evolutionary Role of Interspecies Hybridization and Genetic Exchanges in Yeasts." Microbiology and Molecular Biology Reviews **76**(4): 721-739.

Moreno, S., A. Klar and P. Nurse (1991). "Molecular genetic analysis of fission yeast Schizosaccharomyces pombe." Methods Enzymol **194**: 795-823.

Mortimer, R. K. (2000). "Evolution and variation of the yeast." Genome Res **10**: 891.

Moyle, L. C. and T. Nakazato (2010). "Hybrid Incompatibility “Snowballs” Between *Solanum* Species." Science **329**(5998): 1521-1523.

Muhlfeld, C. C., R. P. Kovach, L. A. Jones, R. Al-Chokhachy, M. C. Boyer, R. F. Leary, W. H. Lowe, G. Luikart and F. W. Allendorf (2014). "Invasive hybridization in a threatened species is accelerated by climate change." Nature Climate Change **4**: 620.

Muller, H. (1925). "Why polyploidy is rarer in animals than in plants." The American Naturalist **59**(663): 346-353.

Muller, H. (1942). Isolating mechanisms, evolution, and temperature. Biol. Symp.

Muller, H. J. (1925). "Why Polyploidy is Rarer in Animals Than in Plants." The American Naturalist **59**(663): 346-353.

Murphy, H. A., H. A. Kuehne, C. A. Francis and P. D. Sniegowski (2006). "Mate choice assays and mating propensity differences in natural yeast populations." Biology Letters **2**(4): 553-556.

Murphy, H. A. and C. W. Zeyl (2011). "PREZYGOTIC ISOLATION BETWEEN SACCHAROMYCES CEREVISIAE AND SACCHAROMYCES PARADOXUS THROUGH DIFFERENCES IN MATING SPEED AND GERMINATION TIMING." Evolution **66**(4): 1196-1209.

Murphy, H. A. and C. W. Zeyl (2015). "A Potential Case of Reinforcement in a Facultatively Sexual Unicellular Eukaryote." The American Naturalist **186**(2): 312-319.

Nagel, L. and D. Schluter (1998). "BODY SIZE, NATURAL SELECTION, AND SPECIATION IN STICKLEBACKS." Evolution **52**(1): 209-218.

Nasmyth, K. A. (1982). "Molecular Genetics of Yeast Mating Type." Annual Review of Genetics **16**(1): 439-500.

Naumov, G., E. Naumova and P. Sniegowski (1997). "Differentiation of European and Far East Asian populations of Saccharomyces paradoxus by allozyme analysis." International Journal of Systematic and Evolutionary Microbiology **47**(2): 341-344.

Naumov, G. I., S. A. James, E. S. Naumova, E. J. Louis and I. N. Roberts (2000). "Three new species in the Saccharomyces sensu stricto complex: Saccharomyces cariocanus, Saccharomyces kudriavzevii and Saccharomyces mikatae." International journal of systematic and evolutionary microbiology **50**(5): 1931-1942.

Naumov, G. I., E. S. Naumova and P. D. Sniegowski (1998). "Saccharomyces paradoxus and Saccharomyces cerevisiae are associated with exudates of North American oaks." Can J Microbiol **44**(11): 1045-1050.

Neafsey, D. E., B. M. Barker, T. J. Sharpton, J. E. Stajich, D. J. Park, E. Whiston, C. Y. Hung, C. McMahan, J. White, S. Sykes, D. Heiman, S. Young, Q. Zeng, A. Abouelleil, L. Aftuck, D. Bessette, A. Brown, M. FitzGerald, A. Lui, J. P. Macdonald, M. Priest, M. J. Orbach, J. N. Galgiani, T. N. Kirkland, G. T. Cole, B. W. Birren, M. R. Henn, J. W. Taylor and S. D. Rounsley (2010). "Population genomic sequencing of Coccidioides fungi reveals recent hybridization and transposon control." Genome Res **20**(7): 938-946.

Neiman, A. M. (2005). "Ascospore formation in the yeast Saccharomyces cerevisiae." Microbiology and molecular biology reviews : MMBR **69**(4): 565-584.

Neiman, A. M. (2011). "Sporulation in the budding yeast Saccharomyces cerevisiae." Genetics **189**(3): 737-765.

Nelson, R. R. (1963). "Interspecific Hybridization in the Fungi." Mycologia **55**(1): 104-123.

Nieto Feliner, G., I. Álvarez, J. Fuertes-Aguilar, M. Heuertz, I. Marques, F. Moharrek, R. Piñeiro, R. Riina, J. A. Rosselló, P. S. Soltis and I. Villa-Machío (2017). "Is homoploid hybrid speciation that rare? An empiricist's view." Heredity **118**: 513.

Nishant, K. T., W. Wei, E. Mancera, J. L. Argueso, A. Schlattl, N. Delhomme, X. Ma, C. D. Bustamante, J. O. Korbel, Z. Gu, L. M. Steinmetz and E. Alani (2010). "The baker's yeast diploid genome is remarkably stable in vegetative growth and meiosis." PLoS Genet **6**(9): e1001109.

Noor, M. A. F., K. L. Grams, L. A. Bertucci and J. Reiland (2001). "Chromosomal inversions and the reproductive isolation of species." Proceedings of the National Academy of Sciences **98**(21): 12084-12088.

Nosil, P. (2012). Ecological speciation, Oxford University Press.

Ohno, S. (1970). "The Enormous Diversity in Genome Sizes of Fish as a Reflection of Nature's Extensive Experiments with Gene Duplication." Transactions of the American Fisheries Society **99**(1): 120-130.

Ordonez, A. and J. W. Williams (2013). "Climatic and biotic velocities for woody taxa distributions over the last 16 000 years in eastern North America." Ecology Letters **16**(6): 773-781.

Orive, M. E. and N. H. Barton (2002). "Associations between cytoplasmic and nuclear loci in hybridizing populations." Genetics **162**(3): 1469-1485.

Orr, H. A. (1990). ""Why Polyploidy is Rarer in Animals Than in Plants" Revisited." The American Naturalist **136**(6): 759-770.

Orr, H. A. (1995). "The population genetics of speciation: the evolution of hybrid incompatibilities." Genetics **139**(4): 1805-1813.

Orr, H. A. and M. Turelli (2001). "The evolution of postzygotic isolation: accumulating Dobzhansky-Muller incompatibilities." Evolution **55**(6): 1085-1094.

Ortiz-Merino, R. A., N. Kuanyshev, S. Braun-Galleani, K. P. Byrne, D. Porro, P. Branduardi and K. H. Wolfe (2017). "Evolutionary restoration of fertility in an interspecies hybrid yeast, by whole-genome duplication after a failed mating-type switch." PLoS Biol **15**(5): e2002128.

Otto, S. P. and J. Whitton (2000). "Polyploid incidence and evolution." Annu Rev Genet **34**: 401-437.

Otto, S. P. and J. Whitton (2000). "Polyploid incidence and evolution." Annual review of genetics **34**(1): 401-437.

Paterson, A. H., E. S. Lander, J. D. Hewitt, S. Peterson, S. E. Lincoln and S. D. Tanksley (1988). "Resolution of quantitative traits into Mendelian

factors by using a complete linkage map of restriction fragment length polymorphisms." Nature **335**(6192): 721-726.

Paulson, D. R. (1974). "Reproductive isolation in Damselflies." Systematic Biology **23**(1): 40-49.

Payseur, B. A. and L. H. Rieseberg (2016). "A genomic perspective on hybridization and speciation." Molecular Ecology **25**(11): 2337-2360.

Pritchard, J. K., M. Stephens and P. Donnelly (2000). "Inference of Population Structure Using Multilocus Genotype Data." Genetics **155**(2): 945-959.

Prochazka, E., F. Franko, S. Polakova and P. Sulo (2012). "A complete sequence of Saccharomyces paradoxus mitochondrial genome that restores the respiration in S. cerevisiae." FEMS Yeast Res **12**(7): 819-830.

Quinn, T. P., M. J. Unwin and M. T. Kinnison (2007). "Evolution of temporal isolation in the wild: Genetic divergence in timing of migration and breeding by introduced chinook salmon populations." Evolution **54**(4): 1372-1385.

R Development Core Team (2013). "R: A language and environment for statistical computing."

R Development Core Team (2013). R: A language and environment for statistical computing. Vienna, Austria, R Foundation for Statistical Computing).

Racimo, F., S. Sankararaman, R. Nielsen and E. Huerta-Sánchez (2015). "Evidence for archaic adaptive introgression in humans." Nature Reviews Genetics **16**(6): 359.

Ramsey, J. and D. W. Schemske (1998). "Pathways, mechanisms, and rates of polyploid formation in flowering plants." Annual review of ecology and systematics **29**(1): 467-501.

Rand, D. M., R. A. Haney and A. J. Fry (2004). "Cytonuclear coevolution: the genomics of cooperation." Trends Ecol Evol **19**(12): 645-653.

Reich, D., K. Thangaraj, N. Patterson, A. L. Price and L. Singh (2009). "Reconstructing Indian population history." Nature **461**(7263): 489-494.

Rhymer, J. M. and D. Simberloff (1996). "Extinction by hybridization and introgression." Annual review of ecology and systematics **27**(1): 83-109.

Rice, W. R. and E. E. Hostert (1993). "LABORATORY EXPERIMENTS ON SPECIATION: WHAT HAVE WE LEARNED IN 40 YEARS?" Evolution **47**(6): 1637-1653.

Rice, W. R. and G. W. Salt (1990). "The Evolution of Reproductive Isolation as a Correlated Character Under Sympatric Conditions: Experimental Evidence." Evolution **44**(5): 1140-1152.

Rieseberg, L. H. (1997). "Hybrid Origins of Plant Species." Annual Review of Ecology and Systematics **28**(1): 359-389.

Rieseberg, L. H. (2001). "Chromosomal rearrangements and speciation." Trends Ecol Evol **16**(7): 351-358.

Rieseberg, L. H., S.-C. Kim, R. A. Randell, K. D. Whitney, B. L. Gross, C. Lexer and K. Clay (2007). "Hybridization and the colonization of novel habitats by annual sunflowers." Genetica **129**(2): 149-165.

Rieseberg, L. H., C. Van Fossen and A. M. Desrochers (1995). "Hybrid speciation accompanied by genomic reorganization in wild sunflowers." Nature **375**: 313.

Robinson, H. A., A. Pinharanda and D. Bensasson (2016). "Summer temperature can predict the distribution of wild yeast populations." Ecology and Evolution **6**(4): 1236-1250.

Rochette, S., G. Diss, M. Filteau, J. B. Leducq, A. K. Dube and C. R. Landry (2015). "Genome-wide protein-protein interaction screening by protein-fragment complementation assay (PCA) in living cells." J Vis Exp(97).

Rogers, D. W., E. McConnell, J. Ono and D. Greig (2018). "Spore-autonomous fluorescent protein expression identifies meiotic chromosome mis-segregation as the principal cause of hybrid sterility in yeast." PLOS Biology **16**(11): e2005066.

Rogers, D. W., E. McConnell, J. Ono and D. Greig (2018). "Spore-autonomous fluorescent protein expression identifies meiotic chromosome mis-segregation as the principal cause of hybrid sterility in yeast." PLoS Biol **16**(11): e2005066.

Rothschild, L. J. and R. L. Mancinelli (2001). "Life in extreme environments." Nature **409**: 1092.

Ruderfer, D. M., S. C. Pratt, H. S. Seidel and L. Kruglyak (2006). "Population genomic analysis of outcrossing and recombination in yeast." Nature Genetics **38**: 1077.

Rundle, H. D. (2002). "A test of ecologically dependent postmating isolation between sympatric sticklebacks." Evolution **56**(2): 322-329.

Rundle, H. D. and P. Nosil (2005). "Ecological speciation." Ecology Letters **8**(3): 336-352.

Ryan, M. T. and N. J. Hoogenraad (2007). "Mitochondrial-nuclear communications." Annu Rev Biochem **76**: 701-722.

Salvadó, Z., F. N. Arroyo-López, J. M. Guillamón, G. Salazar, A. Querol and E. Barrio (2011). "Temperature Adaptation Markedly Determines Evolution within the Genus Saccharomyces." Applied and Environmental Microbiology **77**(7): 2292-2302.

Samani, P., E. Low-Decarie, K. McKelvey, T. Bell, A. Burt, V. Koufopanou, C. R. Landry and G. Bell (2015). "Metabolic variation in natural populations of wild yeast." Ecology and Evolution **5**(3): 722-732.

Sampaio, J. P. and P. Gonçalves (2008). "Natural populations of Saccharomyces kudriavzevii in Portugal are associated with oak bark and are sympatric with S. cerevisiae and S. paradoxus." Applied and environmental microbiology **74**(7): 2144-2152.

Scannell, D. R., O. A. Zill, A. Rokas, C. Payen, M. J. Dunham, M. B. Eisen, J. Rine, M. Johnston and C. T. Hittinger (2011). "The Awesome Power of Yeast Evolutionary Genetics: New Genome Sequences and Strain

Resources for the Saccharomyces sensu stricto Genus." G3: Genes|Genomes|Genetics **1**(1): 11-25.

Schluter, D. (2001). "Ecology and the origin of species." Trends in ecology & evolution **16**(7): 372-380.

Schneider, C. A., W. S. Rasband and K. W. Eliceiri (2012). "NIH Image to ImageJ: 25 years of image analysis." Nat Methods **9**(7): 671-675.

Schumer, M., G. G. Rosenthal and P. Andolfatto (2014). "How common is homoploid hybrid speciation?" Evolution **68**(6): 1553-1560.

Schumer, M., C. Xu, D. L. Powell, A. Durvasula, L. Skov, C. Holland, J. C. Blazier, S. Sankararaman, P. Andolfatto, G. G. Rosenthal and M. Przeworski (2018). "Natural selection interacts with recombination to shape the evolution of hybrid genomes." Science.

Schwartz, D. and W. J. Laughner (1969). "A Molecular Basis for Heterosis." Science **166**(3905): 626-627.

Seabold, S. and J. Perktold (2010). "Statsmodels: Econometric and statistical modeling with python.". Proceedings of the 9th Python in Science Conference.

Seehausen, O. (2006). "Conservation: losing biodiversity by reverse speciation." Curr Biol **16**(9): R334-337.

Seehausen, O., J. J. M. v. Alphen and F. Witte (1997). "Cichlid Fish Diversity Threatened by Eutrophication That Curbs Sexual Selection." Science **277**(5333): 1808-1811.

Seehausen, O., R. K. Butlin, I. Keller, C. E. Wagner, J. W. Boughman, P. A. Hohenlohe, C. L. Peichel, G.-P. Saetre, C. Bank, Å. Brännström, A. Brelsford, C. S. Clarkson, F. Eroukhmanoff, J. L. Feder, M. C. Fischer, A. D. Foote, P. Franchini, C. D. Jiggins, F. C. Jones, A. K. Lindholm, K. Lucek, M. E. Maan, D. A. Marques, S. H. Martin, B. Matthews, J. I. Meier, M. Möst, M. W. Nachman, E. Nonaka, D. J. Rennison, J. Schwarzer, E. T. Watson, A. M. Westram and A. Widmer (2014). "Genomics and the origin of species." Nature Reviews Genetics **15**: 176.

Selmecki, A. M., Y. E. Maruvka, P. A. Richmond, M. Guillet, N. Shoresh, A. L. Sorenson, S. De, R. Kishony, F. Michor, R. Dowell and D. Pellman (2015). "Polyploidy can drive rapid adaptation in yeast." Nature **519**(7543): 349-352.

Servedio, M. R. and M. A. F. Noor (2003). "The Role of Reinforcement in Speciation: Theory and Data." Annual Review of Ecology, Evolution, and Systematics **34**(1): 339-364.

Setiadi, E. R., T. Doedt, F. Cottier, C. Noffz and J. F. Ernst (2006). "Transcriptional response of Candida albicans to hypoxia: linkage of oxygen sensing and Efg1p-regulatory networks." J Mol Biol **361**(3): 399-411.

Shao, Y., N. Lu, Z. Wu, C. Cai, S. Wang, L.-L. Zhang, F. Zhou, S. Xiao, L. Liu, X. Zeng, H. Zheng, C. Yang, Z. Zhao, G. Zhao, J.-Q. Zhou, X. Xue and Z. Qin (2018). "Creating a functional single-chromosome yeast." Nature **560**(7718): 331-335.

Shapira, R., T. Levy, S. Shaked, E. Fridman and L. David (2014). "Extensive heterosis in growth of yeast hybrids is explained by a combination of genetic models." Heredity **113**(4): 316-326.

Shapiro, B. J. and M. F. Polz (2015). "Microbial Speciation." Cold Spring Harbor perspectives in biology **7**(10): a018143-a018143.

Shull, G. H. (1908). "The Composition of a Field of Maize." Journal of Heredity **os-4**(1): 296-301.

Shull, G. H. (1914). "Duplicate genes for capsule-form inBursa bursa-pastoris." Zeitschrift für induktive Abstammungs- und Vererbungslehre **12**(1): 97-149.

Simpson, J. T., K. Wong, S. D. Jackman, J. E. Schein, S. J. Jones and I. Birol (2009). "ABySS: a parallel assembler for short read sequence data." Genome research: gr. 089532.089108.

Sipiczki, M. (2008). "Interspecies hybridization and recombination in Saccharomyces wine yeasts." FEMS Yeast Research **8**(7): 996-1007.

Slon, V., F. Mafessoni, B. Vernot, C. de Filippo, S. Grote, B. Viola, M. Hajdinjak, S. Peyrégne, S. Nagel, S. Brown, K. Douka, T. Higham, M. B. Kozlikin, M. V. Shunkov, A. P. Derevianko, J. Kelso, M. Meyer, K. Prüfer and S. Pääbo (2018). "The genome of the offspring of a Neanderthal mother and a Denisovan father." Nature **561**(7721): 113-116.

Smukowski Heil, C. S., C. G. DeSevo, D. A. Pai, C. M. Tucker, M. L. Hoang and M. J. Dunham (2017). "Loss of Heterozygosity Drives Adaptation in Hybrid Yeast." Molecular biology and evolution **34**(7): 1596-1612.

Sniegowski, P. D., P. G. Dombrowski and E. Fingerman (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**(4): 299-306.

Spellman, P. T., G. Sherlock, M. Q. Zhang, V. R. Iyer, K. Anders, M. B. Eisen, P. O. Brown, D. Botstein and B. Futcher (1998). "Comprehensive identification of cell cycle–regulated genes of the yeast Saccharomyces cerevisiae by microarray hybridization." Molecular biology of the cell **9**(12): 3273-3297.

Stanchi, F., E. Bertocco, S. Toppo, R. Dioguardi, B. Simionati, N. Cannata, R. Zimbello, G. Lanfranchi and G. Valle (2001). "Characterization of 16 novel human genes showing high similarity to yeast sequences." Yeast **18**(1): 69-80.

Stebbins, G. L. (1971). Processes of organic evolution, Prentice-Hall.

Stelkens, R. B., M. A. Brockhurst, G. D. D. Hurst and D. Greig (2014). "Hybridization facilitates evolutionary rescue." Evolutionary Applications **7**(10): 1209-1217.

Stern, C. (1936). "Somatic crossing over and segregation in Drosophila melanogaster." Genetics **21**(6): 625.

Stukenbrock, E. H. (2016). "The Role of Hybridization in the Evolution and Emergence of New Fungal Plant Pathogens." Phytopathology **106**(2): 104-112.

Stukenbrock, E. H. (2016). "The Role of Hybridization in the Evolution and Emergence of New Fungal Plant Pathogens." Phytopathology **106**(2): 104-112.

Suarez-Gonzalez, A., C. Lexer and Q. C. B. Cronk (2018). "Adaptive introgression: a plant perspective." Biology Letters **14**(3): 20170688.

Suzuki, R. and H. Shimodaira (2006). "Pvclust: an R package for assessing the uncertainty in hierarchical clustering." Bioinformatics **22**(12): 1540-1542.

Sweeney, J. Y., H. A. Kuehne and P. D. Sniegowski (2004). "Sympatric natural Saccharomyces cerevisiae and S. paradoxus populations have different thermal growth profiles." FEMS Yeast Research **4**(4-5): 521-525.

Sylvester, K., Q.-M. Wang, B. James, R. Mendez, A. B. Hulfachor and C. T. Hittinger (2015). "Temperature and host preferences drive the diversification of Saccharomyces and other yeasts: a survey and the discovery of eight new yeast species." FEMS Yeast Research **15**(3): fov002-fov002.

Tajima, F. (1989). "Statistical method for testing the neutral mutation hypothesis by DNA polymorphism." Genetics **123**(3): 585-595.

Taylor, E., J. Boughman, M. Groenenboom, M. Sniatynski, D. Schluter and J. Gow (2006). "Speciation in reverse: morphological and genetic evidence of the collapse of a three-spined stickleback (Gasterosteus aculeatus) species pair." Molecular Ecology **15**(2): 343-355.

Templeton, A. R. (1981). "Mechanisms of Speciation - A Population Genetic Approach." Annual Review of Ecology and Systematics **12**(1): 23-48.

The Heliconius Genome Consortium, K. K. Dasmahapatra, J. R. Walters, A. D. Briscoe, J. W. Davey, A. Whibley, N. J. Nadeau, A. V. Zimin, D. S. T. Hughes, L. C. Ferguson, S. H. Martin, C. Salazar, J. J. Lewis, S. Adler, S.-J. Ahn, D. A. Baker, S. W. Baxter, N. L. Chamberlain, R. Chauhan, B. A. Counterman, T. Dalmay, L. E. Gilbert, K. Gordon, D. G. Heckel, H. M. Hines, K. J. Hoff, P. W. H. Holland, E. Jacquin-Joly, F. M. Jiggins, R. T. Jones, D. D. Kapan, P. Kersey, G. Lamas, D. Lawson, D. Mapleson, L. S. Maroja, A. Martin, S. Moxon, W. J. Palmer, R. Papa, A. Papanicolaou, Y. Pauchet, D. A. Ray, N. Rosser, S. L. Salzberg, M. A. Supple, A. Surridge, A. Tenger-Trolander, H. Vogel, P. A. Wilkinson, D. Wilson, J. A. Yorke, F. Yuan, A. L. Balmuth, C. Eland, K. Gharbi, M. Thomson, R. A. Gibbs, Y. Han, J. C. Jayaseelan, C. Kovar, T. Mathew, D. M. Muzny, F. Ongeri, L.-L. Pu, J. Qu, R. L. Thornton, K. C. Worley, Y.-Q. Wu, M. Linares, M. L. Blaxter, R. H. ffrench-Constant, M. Joron, M. R. Kronforst, S. P. Mullen, R. D. Reed, S. E. Scherer, S. Richards, J. Mallet, W. Owen McMillan and C. D. Jiggins (2012). "Butterfly genome reveals promiscuous exchange of mimicry adaptations among species." Nature **487**: 94.

Therneau, T. M. (2015). "A Package for Survival Analysis in S."

Thomas, D. Y. and D. Wilkie (1968). "Recombination of mitochondrial drug-resistance factors in Saccharomyces, cerevisiae." Biochemical and Biophysical Research Communications **30**(4): 368-372.

Tigano, A. and V. L. Friesen (2016). "Genomics of local adaptation with gene flow." Molecular Ecology **25**(10): 2144-2164.

Trujillo, J. M., C. Stenius, L. C. Christian and S. Ohno (1962). "Cromosomes of the horse, the donkey, and the mule." Chromosoma **13**(3): 243-248.

Tsai, I. J., D. Bensasson, A. Burt and V. Koufopanou (2008). "Population genomics of the wild yeast Saccharomyces paradoxus: Quantifying the life cycle." Proc Natl Acad Sci U S A **105**(12): 4957-4962.

Tsai, I. J., D. Bensasson, A. Burt and V. Koufopanou (2008). "Population genomics of the wild yeast *Saccharomyces paradoxus*: Quantifying the life cycle." Proceedings of the National Academy of Sciences **105**(12): 4957-4962.

Turelli, M. and L. C. Moyle (2007). "Asymmetric Postmating Isolation: Darwin's Corollary to Haldane's Rule." Genetics **176**(2): 1059-1088.

Van der Jeugd, H. and R. McCleery (2002). "Effects of spatial autocorrelation, natal philopatry and phenotypic plasticity on the heritability of laying date." Journal of Evolutionary Biology **15**(3): 380-387.

Vernot, B. and J. M. Akey (2014). "Resurrecting Surviving Neandertal Lineages from Modern Human Genomes." Science.

Via, S. (1991). "The genetic structure of host plant adaptation in a spatial patchwork: demographic variability among reciprocally transplanted pea aphid clones." Evolution **45**(4): 827-852.

Via, S. (1999). "Reproductive isolation between sympatric races of pea aphids. I. gene flow restriction and habitat choice." Evolution **53**(5): 1446-1457.

Villanueva, R. (2016). "Cryptic speciation in the stony octocoral Heliopora coerulea: temporal reproductive isolation between two growth forms." Marine Biodiversity **46**(2): 503-507.

Vision, T. J., D. G. Brown and S. D. Tanksley (2000). "The origins of genomic duplications in Arabidopsis." Science **290**(5499): 2114-2117.

Wagih, O. and L. Parts (2014). "gitter: a robust and accurate method for quantification of colony sizes from plate images." G3 (Bethesda) **4**(3): 547-552.

Wake, D. B. (1997). "Incipient species formation in salamanders of the *Ensatina* complex." Proceedings of the National Academy of Sciences **94**(15): 7761.

Walsh, J. B. (1982). "Rate of accumulation of reproductive isolation by chromosome rearrangements." The American Naturalist **120**(4): 510-532.

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

Wertheim, B., L. Beukeboom and L. Van de Zande (2013). "Polyploidy in animals: effects of gene expression on sex determination, evolution and ecology." Cytogenetic and genome research **140**(2-4): 256-269.

West-Eberhard, M. J. (1983). "Sexual selection, social competition, and speciation." The Quarterly Review of Biology **58**(2): 155-183.

Whiteley, A. R., S. W. Fitzpatrick, W. C. Funk and D. A. Tallmon (2015). "Genetic rescue to the rescue." Trends in Ecology & Evolution **30**(1): 42-49.

Whitman, W. B., D. C. Coleman and W. J. Wiebe (1998). "Prokaryotes: the unseen majority." Proceedings of the National Academy of Sciences of the United States of America **95**(12): 6578-6583.

Whitney, K. D., R. A. Randell and L. H. Rieseberg (2006). "Adaptive introgression of herbivore resistance traits in the weedy sunflower Helianthus annuus." The American Naturalist **167**(6): 794-807.

Willis, K. J., K. D. Bennett, D. Walker and G. M. Hewitt (2004). "Genetic consequences of climatic oscillations in the Quaternary." Philosophical Transactions of the Royal Society of London. Series B: Biological Sciences **359**(1442): 183-195.

Woese, C. (1998). "The universal ancestor." Proceedings of the National Academy of Sciences **95**(12): 6854-6859.

Wolfe, K. H. (2001). "Yesterday's polyploids and the mystery of diploidization." Nature Reviews Genetics **2**(5): 333.

Won, Y.-J., A. Sivasundar, Y. Wang and J. Hey (2005). "On the origin of Lake Malawi cichlid species: a population genetic analysis of divergence." Proceedings of the National Academy of Sciences **102**(suppl 1): 6581-6586.

Xia, W., L. Nielly-Thibault, G. Charron, C. R. Landry, D. Kasimer, J. B. Anderson and L. M. Kohn (2017). "Population genomics reveals structure at the individual, host-tree scale and persistence of genotypic variants of

the undomesticated yeast Saccharomyces paradoxus in a natural woodland." Molecular Ecology **26**(4): 995-1007.

Yandell, B. S. (1997). Practical Data Analysis for Designed Experiments, Taylor & Francis.

Yang, F., B. Fu, P. C. M. O'Brien, W. Nie, O. A. Ryder and M. A. Ferguson-Smith (2004). "Refined genome-wide comparative map of the domestic horse, donkey and human based on cross-species chromosome painting: insight into the occasional fertility of mules." Chromosome Research **12**(1): 65-76.

Yona, A. H., Y. S. Manor, R. H. Herbst, G. H. Romano, A. Mitchell, M. Kupiec, Y. Pilpel and O. Dahan (2012). "Chromosomal duplication is a transient evolutionary solution to stress." Proc Natl Acad Sci U S A **109**(51): 21010-21015.

Yue, J.-X., J. Li, L. Aigrain, J. Hallin, K. Persson, K. Oliver, A. Bergström, P. Coupland, J. Warringer, M. C. Lagomarsino, G. Fischer, R. Durbin and G. Liti (2017). "Contrasting evolutionary genome dynamics between domesticated and wild yeasts." Nature Genetics **49**: 913.

Zaremba-Niedzwiedzka, K., E. F. Caceres, J. H. Saw, D. Bäckström, L. Juzokaite, E. Vancaester, K. W. Seitz, K. Anantharaman, P. Starnawski, K. U. Kjeldsen, M. B. Stott, T. Nunoura, J. F. Banfield, A. Schramm, B. J. Baker, A. Spang and T. J. G. Ettema (2017). "Asgard archaea illuminate the origin of eukaryotic cellular complexity." Nature **541**: 353.

Zeyl, C., C. Curtin, K. Karnap and E. Beauchamp (2005). "Antagonism between sexual and natural selection in experimental populations of Saccharomyces cerevisiae." Evolution **59**(10): 2109-2115.

Zhang, H., A. Skelton, R. C. Gardner and M. R. Goddard (2010). "Saccharomyces paradoxus and Saccharomyces cerevisiae reside on oak trees in New Zealand: evidence for migration from Europe and interspecies hybrids." FEMS Yeast Research **10**(7): 941-947.

Zhu, Y. O., M. L. Siegal, D. W. Hall and D. A. Petrov (2014). "Precise estimates of mutation rate and spectrum in yeast." Proceedings of the National Academy of Sciences **111**(22): E2310-E2318.

Zhu, Y. O., M. L. Siegal, D. W. Hall and D. A. Petrov (2014). "Precise estimates of mutation rate and spectrum in yeast." Proc Natl Acad Sci U S A **111**(22): E2310-2318.

ANNEXE A: Matériels et méthodes supplémentaire pour le chapitre I

1. Strain collection and sampling

We assembled a collection of 153 *S. paradoxus* strains representative of the genetic diversity and geographical distribution of *S. paradoxus* found in Northeastern America (Naumov *et coll.* 1998, Kuehne *et coll.* 2007, Liti *et coll.* 2009, Charron *et coll.* 2014, Leducq *et coll.* 2014, Sylvester *et coll.* 2015) (Figure I.1a; Table I.S1). This collection includes 32 unpublished strains that were sampled in 2013 in the contact zone between *SpB* and *SpC* (Quebec, New-Brunswick, Maine, Massachusetts; Table I.S1). Strains were mostly isolated from the bark of deciduous trees (oak, maple) according to standard protocols (Sniegowski *et coll.* 2002). Strains were assigned to *S. paradoxus SpA* (14), *SpB* (89) and *SpC* (50) based upon three unlinked nuclear markers (Charron *et coll.* 2014, Leducq *et coll.* 2014). Whole-genome sequencing confirms that these markers are discriminant of the three groups.

2. Library preparation (Illumina TruSeq) for high coverage (HC) genome sequencing of 24 *S. paradoxus* strains.

A representative subset of 24 strains was used for high coverage and quality (HC; i.e. a high (~100×) and uniform coverage along the genome) whole genome sequencing to perform *de novo* assembly and SNP calling. For each strain, 4 µl of liquid culture from original glycerol stocks were streaked on solid YPD (Yeast Peptone Dextrose) medium. Petri dishes were incubated overnight at 25°C and one colony per strain was picked to inoculate 4 ml of liquid YPD medium. DNA was extracted using the DNeasy® Blood & Tissue Kit (QIAGEN) adapted to yeast. Genomic DNA was suspended in 100 µl of elution buffer and each sample was incubated for 10 minutes with 3µl of RNase A (Roche) at 37°C followed by purification using standard ethanol precipitation. Before library preparation, genomic DNA was quality-controlled and genotypes of strains were re-confirmed

using a RFLP method we previously developed for *S. paradoxus* American lineages identification (Leducq *et coll.* 2014).

For each strain, 1µg of genomic DNA was fragmented by nebulization following Roche standard nebulization protocol. Nebulization conditions (liquid nitrogen at 2.07 bar for 2 minutes) were optimized to obtain an average fragment size of 500bp. Library preparation (end repair, A-tail and ligation setups) was performed following three different protocols from New England Biolabs®, Lucigen® and TruSeq Illumina®, with eight randomly chosen samples per protocol. In all cases, the adaptors used in the ligation reactions came from the TruSeq Illumina® kit and the final library amplification followed the TruSeq Illumina® protocol. Library purification and sizing were performed according to standard protocols and optimized to remove fragments shorter than 300bp: 50µl of post-PCR library and 45µl of Agencourt® AMpure XP beads were mixed together. Library sequencing (Paired-end 100bp) was done on a HiSeq 2500 Illumina® sequencer at the Genome Québec sequencing platform (McGill University). Reads can be retrieved from NCBI under the BioProject number PRJNA277692 (BioSamples SAMN03389655-SAMN03389678).

3. Library preparation (Illumina Nextera) for low coverage (LC) genome sequencing of 131 *Saccharomyces* strains (137 libraries).

An additional set of 129 strains, representing almost the entire collection of *S. paradoxus* strains available for Northeastern America at the time, was sequenced at low coverage (LC i.e. a low (~20×) and variable depth is expected along the genome) for SNP calling. To empirically verify that the false positive rate of heterozygous sites from sequencing and mapping errors was low compared to population polymorphism, 43 strains were sporulated to obtain homozygous strains before sequencing. Cells from a single haploid spore can switch their mating type, allowing a mother and a daughter cell to mate such that the resulting diploid cell is expected to be 100% homozygous. Sporulation and separation of haploid

meiotic products (spores) were performed according to standard protocols (Charron *et coll.* 2014). For 4 out of the 43 strains, two meiotic products were randomly chosen to be sequenced (LL2012_002, LL2012_16, LL2012_18 and LL2012_28; Table S1). For a single strain (LL2012_012), one meiotic product was sequenced twice as a control (Table I.S1). Including replicates, we prepared a total of 48 homozygous strains for sequencing. No sporulation was done for the 89 remaining strains, resulting in a total of 137 libraries. Genomic DNA preparation was performed as described above. Library preparation followed the standard protocol from Illumina Nextera® with the following modifications: PCR amplification was based on 18 cycles and PCR cleanup was performed with 50µl of AMPure XP beads. Library sequencing (Paired-end 100) was performed as described above. Reads can be retrieved from NCBI under the BioProject number PRJNA277692 (BioSamples SAMN03389659 - SAMN03389817).

4. Read mapping on the *S. paradoxus* reference genome

We mapped sequencing reads on the *S. paradoxus* reference genome CBS432, which phylogenetically belongs to lineage *SpA* (Liti *et coll.* 2009). Sequences of the 16 chromosomes from the *S. paradoxus* reference genome were retrieved from Scannell *et al.* (Scannell *et coll.* 2011). We mapped reads using Bowtie2 (Langmead et Salzberg 2012) in paired-end mode with the default settings, except for the "local" mode, which was used instead of the default "end-to-end" mode, since local alignment enhances the detection of indels. Duplicated reads were marked using Picard (http://broadinstitute.github.io/picard/) with a pixel distance threshold of 100. For the 24 HC libraries (TruSeq illumina®), an average of 78.3±9.9% of reads were successfully mapped on the reference genome, resulting in an average depth of coverage of 83.0±12.3× and a coverage of 97.6±0.4% of the reference genome. For the 137 *S.paradoxus* LC libraries (Illumina Nextera®), 87.9±5.7% of reads on average were successfully mapped to the

reference genome, resulting in an average depth of coverage of 32.5±22.3× and a coverage of 94.5±4.9% of the reference genome. A significant fraction of the unmapped reads belong to the mitochondrial genomes.

5. Aneuploidy detection from mapping (HC+LC)

We investigated chromosomal copy number variation in the genome of natural *S. paradoxus* strains. Aneuploidy would result in total/partial duplications or losses of chromosomes and could be detected by variation in the coverage depth along the genome. For each strain, we measured the average depth of coverage (*D*) in discrete 5,000bp windows along the genome. For each HC libraries, we normalized D (D_N) by the median calculated over the genome and removed windows containing highly repeated regions such as the rRNA genes on chromosome XII ($D_N > 3$) and the telomeric regions. We observed no notable increase of depth of coverage, except for the right arm of chromosome III for strain UTM6, suggesting a partial duplication of this chromosome (Figure I.S7).

For each LC libraries (Nextera XT Illumina®) and each window, depth was first normalized by a randomly chosen LC library (LL2013_205; *D*=21×). This normalization allowed to control for the biases caused by the non-random enzymatic fragmentation (tagmentation) of genomic DNA. For each library, D_N was normalized again by the median calculated over the whole genome and highly repetitive regions were removed, as described for the HC libraries. Using this process, we substantially decreased the tagmentation-induced coverage variation among positions, which however remained high because of low global depth of coverage in some libraries (Figure I.S8). Most strains exhibited no aneuploidy, except two obvious cases of complete (LL2012_014: Chr. IX; UCD62−268: chr. I) and six cases of partial chromosome duplications (yHKS343: chrs. IV and X; yHKS344: chrs. II and IV; UCD62−186: chr. V; LL2013_010: chr. III; Figure I.S8). All analyses were performed in R (R Development Core Team 2013).

6. Preliminary SNP calling (HC)

We performed a preliminary SNP calling for each of the 24 HC libraries separately, using the BCFtools toolkit from SAMtools with default parameters (Li *et coll.* 2009) after HC read mapping on the CBS432 reference genome (Liti *et coll.* 2009). The consensus sequence of each chromosome was then reconstructed for each strain from the resulting variants (566,402 SNPs) using SAMtools, without filtering. These SNPs are referred to as preliminary SNPs throughout this supplementary material.

7. Variant calling and filtering in 161 libraries (HC+LC)

SNPs and indels were called in all HC and LC sequenced strains at once using FreeBayes (Garrison et Marth 2012) using the default parameters. We then filtered the resulting variants with VCFlib (Garrison et Marth 2012), removing sites with a polymorphism probability under 99%, genotypes with a depth of coverage of less than 4× and alleles with a balance probability under 1% in heterozygous genotypes. These variants are to as filtered variants throughout this supplementary material (479547 SNPs).

8. Heterozygozity estimate (HC+LC)

The number of heterozygous variants (N_{HV}) in a given strain was estimated using the following formula:

$$N_{HV} \approx \frac{N_H}{N_G} \times N_V$$

where N_H is the number of called heterozygous genotypes for a given strain, N_G is the total number of called genotypes for this strain and N_V is the total number of called variants among individuals.

N_{HV} is higher in diploid (2N) than in homozygous (1N) strains ($p < 0.001$; Kruskal-Wallis rank sum test; Figure I.S9a) as expected, suggesting low but significant heterozygosity in natural diploid strains ($N_{HV} \approx 2.5\times10^3$). Few individuals exhibit slightly high N_{HV} ($N_{HV} > 4.1\times10^3$) that is evenly

distributed along the genome (data not shown). N_{HV} is higher in HC than in LC libraries ($N_{HV} \approx 1.8 \times 10^3$; $p < 0.001$) suggesting that lower ($D<60\times$) and non-uniform depth of coverage along the genome (i.e. biased by tagmentation-induced coverage variation) in LC libraries may bias the estimation of N_{HV}. Five HC strains corresponding to *SpA* strains have almost null N_{HV} ($N_{HV} \approx 6 \times 10^2$), which reflects the isogenicity of *SpA* strains in North America (Kuehne *et coll.* 2007). This could also be influenced by the mapping on the *SpA*-like reference genome CBS432 that induces overestimation of N_{HV} in non-*SpA* strains, or underestimation of N_{HV} in *SpA* strains.

In order to verify that the SNP calling error rate is much smaller than the actual polymorphism, we compared per-site genotypes of two LC libraries independently obtained from one spore of the strain LL2012_012 (monosporic clones Table I.S1). Among the 264,263 polymorphic sites identified ("N" positions removed in this case), replicated libraries had only 0.31% (830 sites) of different genotypes (Figure I.S9b). We observed similar proportions for pairs of spores derived from a same strain (0.28-0.39%; LL2012_016, LL2012_28 and LL2012_002) and no pairs of different strains exhibited lower proportion. We thus assumed that our error in assigning genotypes, including sequencing and mapping error, was around 0.3%. In the two replicated libraries, most of different genotypes (86.1%) were assigned to a heterozygous genotype ("H"), while heterozygous genotypes only represented 0.03% of identical genotypes, indicating that heterozygous sites most often result from calling errors (Figure I.S9c). Differences in genotypes between replicated libraries were evenly distributed along the genome, indicating than these errors could not bias further analyses in particular genomic regions (Figure I.S9d).

9. Global phylogenies (HC+LC)

A maximum likelihood phylogenetic tree was produced from the filtered variants in complete deletion using PhyML (Guindon *et coll.* 2010) with the TN93 model and an approximate Likelihood-Ratio Test (aLRT) as branch support (Figures I.1b, I.3a and I.S14c).

10. Population structure analysis

We randomly sampled 10,000 variable positions (over 479547 SNPs) in the 161 *S. paradoxus* genomes to assess population structure and admixture using the program STRUCTURE (Pritchard *et coll.* 2000). Our set consisted in 6,881 SNPs after filtering of positions harboring more than 40% of strains with missing genotypes. We carried out analyses under the admixture model over 20,000 iterations of the MCMC chain with a 10,000 burn-in period, assuming diploid and homozygote individuals. After 10 independent runs for each *K* value tested in the range from 1 to 6 (homogeneous populations), we observed no convergence to a unique solution. Because we failed at identifying the most likely *K* using the classical methods (Evanno *et coll.* 2005), we selected the run with the highest likelihood (*LnP(D)*) for each tested *K* value and chose 80% as the minimum threshold for *a posteriori* assignment of a strain to a given populations. For the most likely run of *K*=3 (*LnP(D)*= 101669.2), populations corresponded to lineages *SpA* (*n*=11), *SpB* (*n*=87) and *SpC* (*n*=53), respectively (Figure I.1c). For the most likely run (*K*= 6, *LnP(D)* = -75996.1; Figure I.1c), all strains from lineage *SpA* were assigned to the same population (A), while *SpB* and *SpC* strains were assigned to 3 and 2 populations, respectively. Among *SpB* strains, 38% (*n*=33) showed admixture between at least two populations, while only three *SpC* strains (YPS667: PE; UCD62-186 and UCD62-268: SK) showed admixture. All other runs were combinations of the six aforementioned populations (data not showed). Populations in lineage *SpB* followed the global phylogeny (Figure I.1b-c) but did not correspond to monophyletic groups, while *SpC* populations corresponded to two monophyletic groups in the *SpC* clade,

namely *SpC** (without SK, located at the root of *SpC** clade) and *SpC* (without PE, located at the root of *SpC* clade).

11. Growth rate measurement at different temperatures

We measured the growth rate of 182 *S. paradoxus* strains along a gradient of temperature ranging from 10 to 35°C with 5°C intervals. We achieved high-throughput measurements by performing growth assays on solid medium. We randomly arrayed 12 replicate colonies from each of the 182 strains on two high-density omnitrays (86 x 128 mm petri dish) containing 35 ml of solid YPD medium. The two omnitrays were incubated for 24 hours at 25°C and colonies were transferred onto 2×6 new omnitrays (two per temperature) and incubated again for 24h at 25°C to homogenize colony growth. Each array was transferred again onto an omnitray on solid YPD and incubated 72 hours at different temperatures (10 to 35°C). Pictures were taken after colony transfers (time point = 0) and then every 12 hours, while making sure that omnitrays were not exposed longer than 15 minutes to room temperature. After 72 hours, colony growth reached a plateau (data not shown) and the 12 arrays were transferred again on a fresh YPD medium for a second round of selection that increases the assay sensitivity. The whole procedure was repeated during 72 hours and growth analysis assayed by colony size was performed from pictures taken at the last time point. All steps were performed using a robotic device (BMC-BC, S&P ROBOTICS Inc.) as described in Rochette *et coll.* (2015). Images of plates were analyzed using ImageJ 1.45 m (Schneider *et coll.* 2012). The fitness of each strain at each temperature was calculated as the median of the log of colony size (measured as the number of pixels on the image) of the 12 replicates after 2×72 hours of growth.

Among-strain variance in growth rate was mostly explained by differences among lineages at any temperature (ANOVA: $p < 0.01$; Table I.2), and latitude was also a significant factor explaining adaptation to high

temperature (35°C; p<0.0001), as previously observed (Leducq *et coll.* 2014). Strains from the *SpB* lineage grow better than *SpC* at all temperatures on rich media (p < 0.001; Tukey test; Table I.3; Figs I.1e). Interestingly, *SpC** strains show no significant difference with *SpB* at any temperature (p > 0.05), and like *SpB*, *SpC** has a strong fitness advantage over *SpC*, especially at high temperatures (30-35°C: + 19-42%; p<0.001). All analyses were performed in R (R Development Core Team 2013).

12. Fitness assay for growth rate measurement on different carbon and nitrogen sources

The ability of strains to use different nitrogen and carbon sources was measured using phenotypic microarrays PM1, PM2A and PM3 (BIOLOG, Inc., Hayward, CA). This platform enables to test for cellular respiration based on tetrazolium redox chemistry. A strain able to use the provided carbon or nitrogen source will have more metabolic activity, which will lead to a stronger violet color in the corresponding well of the plate. Each plate contains 95 different carbon (PM1 and PM2A) or nitrogen source (PM3) and a negative control well (no carbon or nitrogen source). A representative subset of 10 strains (5 *SpB*: LL2012_025, MSH-604, LL2012_002, LL2012_022, UTM6 and 5 *SpC*: LL2011_012, LL2011_004, MSH587-1, LL2012_027, YPS667) representative of our collection was selected for these experiments. Plates were inoculated with suspensions of cells prepared following the standard protocol established by BIOLOG and then incubated at 25°C. Color intensity (absorbance at 590 nm) in was measured using an Infinite F200 pro microplate reader (TECAN, Männedorf, Switzerland). Measurements were taken every 4 hours following inoculation (t=0) and to the 48 hours mark (skipping the 36h time point for every experiment). At each time point plates were removed from the incubator and the A_{590nm} was measured 3 times in intervals of two minutes with the reading chamber set at 25°C. Each reading took 8 minutes after which the plate was returned to the incubator. For each time

point, we only kept the third measurement because values fluctuated with temperature re-adjustment of the microplate reader. After 48 hours, the A_{590nm} reached a plateau for most nutrients and we measured the metabolic performance of each strain on each source as the sum of A_{590nm} measurements over all time points.

We identified seven carbon and eight nitrogen (amino acids) sources for which *SpB* and *SpC* have opposite metabolic performance (Figure I.S2). We measured growth-based fitness (log of colony size) in limiting conditions for the 15 aforementioned carbon and nitrogen sources in the complete *S. paradoxus* collection (181 strains; Figure I.S1d) as performed for the temperature experiment described. This allow to examine the growth of *SpC** strains in the conditions found to differ between *SpC* and *SpB* based on the Biolog results. We used a Principal Component Analysis (PCA) to describe fitness variation among strains for the 15 carbon and nitrogen sources. Most among-strain growth variation (PCA; axis 1: 43.1%) was explained by latitude (ANOVA: $p<0.001$; Table I.2) suggesting that, like for high temperature, standing genetic variation due to local adaptation or isolation by distance, may segregate within lineages. Differences between lineages explained the main part of the among-strain growth variation (PCA axes 2 and 3: 21.2 and 10.2% respectively; ANOVA: $p<0.001$; Table I.2, Figure I.1g). *SpB* was favored over other lineages on most carbon sources, especially maltose, while *SpC* was favored over *SpB* when supplemented with sources of nitrogen, especially glutamine, asparagine, lysine, histidine and glycine ($p < 0.001$; Tukey test; Table I.3, Figure III.1g). *SpC** strains have intermediate and non-overlapping global phenotypes between *SpB* and *SpC* ($p< 0.01$; Tukey test; Table I.3, Figure I.1g).

13. Fitness assay for survival to a freeze-thaw cycle

In a previous survey, we measured survival to a freeze-thaw cycle in 27 American *S. paradoxus* strains and two controls (*S. paradoxus* CBS432 and *S. cerevisiae* BY4743). Here, we repeated the measurement for seven

S. paradoxus strains and the two controls, and added 36 strains, as described in Leducq *et coll.* (2014). Survival measurement was highly repeatable, as revealed by strong correlation between the two assays (7 strains and two controls, Pearson's correlation; $r = 0.96$, p.value <0.001). For the following analyses, we thus combined data from the two assays. Latitude was the most significant factor explaining survival to a freeze-thaw cycle (ANOVA; $p<0.001$; Table I.2), as previously observed (Leducq *et coll.* 2014). Survival to a freeze-thaw cycle was also explained by differences between lineages (ANOVA; $p<0.01$) and was significantly lower in *SpC* and *SpC** (8.9% and 5.2% survival, respectively) than in *SpB* strains (74.4%; $p < 0.05$; Tukey test; Table I.3; Figure I.1f,). All analyses were performed in R (R Development Core Team 2013).

14. Controlled crosses (strain preparation and previous data)

To determine whether *SpC** strains are reproductively isolated from the two parental lineages, we performed controlled crosses among haploid heterothallic strains derived from natural diploid strains and measured reproductive success (RS) as progeny (spore) survival, as described in Charron *et al.* (Charron *et coll.* 2014). We performed 64 controlled crosses with heterothallic haploid strains that we constructed from 22 natural diploid strains (Table I.4). We combined these results with data available from 81 previous crosses in which the *SpC** strains had been identified to the *SpC* group (Charron *et coll.* 2014) to increase the representativeness of *SpB*, *SpC* and *SpC** diversity. All *SpC** types were represented, except for SK (UCD_62-186, UCD_62-268), which did not sporulate and could not be used for controlled crosses. Heterothallic strains were constructed as previously described (Charron *et coll.* 2014), with the following modifications: the forward primer used to integrate the resistance cassettes at the *HO* locus for lineage *SpB* strains was slightly modified to increase the rate of successful deletion by homologuous recombination (*par-HO-SpB*: ACATCCTTATAGGCAGCA

ATCAATTCCATCTAAACTTCAACCAGCTGAAGCTTCGTACGC). Mating type verification, crosses, yeast sporulation, tetrad dissection (24 per cross) and evaluation of spore survival were all performed as described in Charron *et al.* Charron *et coll.* (2014).

15. Analysis of spore survival within and among lineages

We analyzed reproductive success (RS) from 145 controlled crosses performed among strains using exact Tukey test (Yandell 1997) among categories of crosses. For crosses performed among strains from the same lineage, we found no significant difference between selfing (RS = 79.7-97.9%; n = 21) and outcrossing (RS= 55.2-92.7%; n = 44) for any lineage (Table S5, Figure I.2). RS for outcrossing performed within *SpA* (83.3%), *SpB* (55.2%), *SpC** (92.7%) and *SpC* (75.8%) were overall not significantly different from each other (but between *SpB* and *SpC**; $p<0.001$) and globally higher than RS for outcrossing performed among these categories (n=80; RS = 24.0-44.3%; $p < 0.001$). RS was systematically higher among *SpC** strains (89.6-94.8%) than with strains from other lineages (20.3-49%), suggesting that *SpC** strains are not reproductively isolated from each other, regardless of their type (composition in *SpB*-like regions) and thus form a distinct *S. paradoxus* lineage. All analyses were performed in R (R Development Core Team 2013).

16. Genomic islands of high diversity (He) and low differentiation (F_{ST}) in HC strains

We used preliminary SNPs to measure diversity within lineages (expected heterozygosity H_e at biallelic sites) and of differentiation between lineages (F_{ST}) in 500bp discrete windows along the genome of 24 strains from HC libraries. In each window, we calculated the H_e at monomorphic and biallelic sites within each genetic group with the following formula:

$$H_e = 2p(1-p)$$

where p is an estimated allelic frequency. We computed pairwise fixation index F_{ST} between genetic groups with the R package *HIERFSTAT* (Goudet 2005). *SpA* strains from North America show almost no diversity (H_e = 0.0004 ± 0.0013; Figure S3a), most likely caused by their recent introduction in North America (Kuehne *et coll.* 2007). American lineages *SpB* and *SpC* show higher diversity (H_e = 0.0019 ± 0.0026 and H_e = 0.0017±0.0037, respectively; Figure I.S10b-c) than the *SpA* introduced lineage, as previously observed (Kuehne *et coll.* 2007, Liti *et coll.* 2009, Leducq *et coll.* 2014). F_{ST} between lineages of *S. paradoxus* is close to one (F_{ST} = 0.888-0.963), showing that lineages share few polymorphisms and exchange little to no genes. Differentiation measured between *SpB* and *SpC* (F_{ST} = 0.888±0.132; Figure I.S10f) is lower than the differentiation between *SpA* and *SpB* (F_{ST} = 0.959±0.041; Figure I.S10d) or between *SpA* and *SpC* (F_{ST} = 0.963±0.053; Figure I.S10e), in agreement with a divergence between American lineages that is more recent than the divergence with the European lineage (Leducq *et coll.* 2014).

Diversity (H_e) and differentiation index (F_{ST}) show extensive variation along the genome and strongly asymmetrical distributions, with medians biased towards lower values for H_e (0.0000-0.0014; Figure IS10a-c) and toward higher values for F_{ST} (0.924-0.977; Figure I.S10d-f). Diversity measured within *SpC* shows a more asymmetrical distribution, with 2.6% of 500bp windows showing high level of diversity ($H_e \geq 0.01$), against 0.1% and 0.9% of 500bp windows in *SpA* and *SpB*, respectively. Similarly, differentiation between *SpB* and *SpC* dramatically decreases in 10.6% of 500bp windows ($F_{ST} \leq 0.8$), which was only the case in 0.6% and 1.9% of windows for the comparisons of *SpA* with *SpB* and *SpC*, respectively.

Peaks of diversity and gaps of differentiation are not randomly distributed along the genome (Figure I.S10a-f). For each statistic (H_e and F_{ST}), we performed all possible pairwise *t*-tests (Welch two-sample *t*-tests) between two contiguous (i.e. non-overlapping) 10kb windows in the genome to

assess the extent of heterogeneity. To do so, we grouped together values measured within 20 contiguous windows of 500bp and compared their distribution to the 20 following windows located on the same chromosome. We repeated this test by sliding 10kb windows 500bp along chromosomes. We defined larger windows within which no significant difference between two contiguous 10kb windows was observed at a highly conservative threshold (*p*-value ≤ 0.0001) to define genomic regions showing homogenous diversity or differentiation (Figure I.S10g). All analyses were performed in R (R Development Core Team 2013).

Our approach revealed large genomic islands of high diversity and low differentiation along the genome. Median diversity measured among homogeneous regions within *SpA* and *SpB* remains low (x_{He} = 0 and 0.0015, respectively), with most regions corresponding to entire chromosomes (>200kb) and only one 27kb region on chromosome 9 for lineage *SpB* (*He* = 0.0055) exhibiting high diversity. Inversely, the distribution of diversity within *SpC* is highly heterogeneous, with many chromosomes fragmented in regions exhibiting contrasted H_e values. Median diversity along the genome remains low (x_{He} = 0.0012) but 12 large windows (28-123 kb) with a total size of 0.61 Mb, exhibit high diversity (H_e > 0.005). Differentiation measured between *SpA* and *SpB* is high (x_{Fst} = 0.965) all along the genome, with most homogeneous regions corresponding to entire chromosomes (>200kb). The exception is the previously mentioned regions of 27kb on chromosome 9, which reflects less divergence (F_{ST} = 0.870). The median differentiation measured between *SpA* and *SpC* is high and on the same order of magnitude than between *SpA* and *SpB* (x_{FST} = 0.961), but dramatically decreases (F_{ST} < 0.880) in 14 large windows (21-118kb) with a total size of 0.615 Mb. This effect is more pronounced for differentiation between *SpB* and *SpC*, which remains globally high (x_{FST} = 0.872), but dramatically decreases in 23 regions (22-119kb) representing a total size of 0.98 Mb (F_{ST} < 0.75). Genomic islands of

high diversity within *SpC* and low differentiation between *SpA* and *SpC*, and *SpB* and *SpC* mostly overlap (Figure I.S10g) and most likely correspond to remnants of ancient genomic exchanges by hybridization (i.e. introgression; (Gosset et Bierne 2013)) between *SpB* and *SpC* (see below).

17. Identification of introgressions by a sliding phylogeny approach (HC)

We used preliminary SNPs in the HC genomes to identify introgressed genomic regions using a sliding phylogeny approach (The Heliconius Genome Consortium *et coll.* 2012). The analyses of consensus sequences using this approach were conducted in R (R Development Core Team 2013), according to the steps described below. For each of the 16 *Saccharomyces* chromosomes, we generated a single alignment of the 24 HC strains and the reference strain CBS432. All ambiguous positions and gaps were replaced by Ns. After this step, any position with missing information for more than two strains were removed from the analyses. We splitted each chromosome alignment in discrete windows of about 2,000bp and counted the number of varying positions and positions with missing information in each strain. Windows having no variable position or missing information for at least five strains (corresponding to the number of strains in *SpA*) were removed from the analyses. We established the phylogeny based upon pairwise evolutionary distance (E_D) among strains within each window, using the mutation probabilities (μ_P) estimated from *Saccharomyces cerevisiae* (Zhu *et coll.* 2014) (Figure I.S11a). We converted mutation probabilities from one state to another in approximate evolutionary distances ($E_D \approx 1/\mu_P$) between observed states (Figure I.S4a). Overall, pairwise E_D between two strains at a given position could be summarized in four different values: 0 (no change), 2 (T↔C; A↔G), 3 (T↔G; A↔C) and 5 (T↔A; G↔C) (Figure I.S11b). To obtain an E_D matrix with exactly these four possible values, we first converted nucleotide sequences in two distinct matrices M_A and M_B (Figure I.S11b). Matrix M_A is

weighted 3 For T/C and 0 for A/G positions, and matrix M_B is weighted 2 for A/C and 0 for T/G positions. Then the pairwise evolutionary distance between two individuals i (M_{Ai}, M_{Bi}) and j (M_{Aj}, M_{Bj}) at a given position is $E_{Dij}= |M_{Ai} - M_{Aj}| + |M_{Bi} - M_{Bj}|$. For instance, for two sequences i(TCG) and j(CCC), M_A and M_B matrices could be combined in

$$i\begin{pmatrix}3,0\\3,2\\0,0\end{pmatrix} \text{ and } j\begin{pmatrix}3,2\\3,2\\3,2\end{pmatrix}$$

and the pairwise evolutionary distance between i and j can be formulated as the sum of the absolute differences between individual matrices (Figure I.S11c):

$$E_{Dij} = \sum \left| i\begin{pmatrix}3,0\\3,2\\0,0\end{pmatrix} - j\begin{pmatrix}3,2\\3,2\\3,2\end{pmatrix} \right| = \sum \begin{pmatrix}0,2\\0,0\\3,2\end{pmatrix} = 7$$

To identify potential introgressions, we determined whether at least strain from a lineage significantly clustered with strains from another lineage. In each 2,000bp window, we computed the total pairwise E_D («manhattan» distances) and constructed the evolutionary tree between the 25 individuals using the function *pvclust()* (R package *pvclust* (Suzuki et Shimodaira 2006)). We performed 1,000 bootstrap resampling to identify significant clusters in the tree ($a \geq 0.95$) and removed clusters containing less than three strains for the following analyses (Figure I.S10a). Significant clusters were divided in two categories according to the number (n) of strains they contained: large clusters ($n > 14$) and small clusters ($n \leq 14$). Given the frequencies of *SpA*, *SpB* and *SpC* in our dataset (n= 6, 10 and 9, respectively), large clusters are likely to contain strains from two lineages and small clusters are likely to contain strains from one lineage. In each window, we only kept the largest non-overlapping clusters and assigned them to one lineage if its frequency in the cluster was at least half plus one (Figure I.S10b). We prioritized reassignment of each strain to one lineage, unless no small cluster was significant. In this case, the strain was assigned to two lineages according to the large cluster.

Using this approach, strains were correctly assigned to lineages *SpA*, *SpB* and *SpC* in 99.9±0.0, 97.1±0.2 and 97.4±1.7% of 2,000bp windows, respectively (Figure I.S10h). In 1.8±0.6% of windows (about 200kb), strains from *SpB* and *SpC* significantly clustered together but could not be distinguished from each other. Three strains from *SpC* (*SpC** strains: LL2011_005, LL2012_016 and LL2012_018) clustered with *SpB* in a substantial proportion of windows ("*SpB*-like" windows; 2.6-3.6%; 300-408 kb), whereas it was not the case for other strains from *SpC* (0.1±0.1%). We observed a similar proportion of windows for which strains from *SpB* clustered with *SpC* (0.2±0.1%). Most of windows for which three *SpC** strains clustered with *SpB* (*SpB*-like), correspond to large genomic islands of high diversity in *SpC* and low divergence between *SpB* and *SpC* described above (Figure I.S10h).

Most *SpB*-like regions in *SpC** can be referred back to introgression events during hybridization between *SpB* and *SpC*, and are shared between the three aforementioned strains. These regions are well delimited on chromosomes IV, VI, VIII, XIII, XV and XVI. However, large *SpB*-like regions in chromosome II and IX are found in *SpC** strains LL2012_016 and LL2012_018 (from Pointe-Platon, QC) but not in LL2011_005 (Île d'Orléans, QC). In contrast, a *SpB*-like region is found on chromosome XI of LL2011_005 but not in LL2012_016 and LL2012_018.

18. Identification of introgressed regions by a site-wise clustering approach (HC+LC)

Filtered variants were used to determine whether introgression from lineage *SpB* to *SpC* is common in natural populations of *Saccharomyces paradoxus* or if the three *SpC** strains (LL2011_005, LL2012_016 and _018) are an exception. We screened the genome of 151 *S. paradoxus* strains (LC and HC libraries combined), representative of the known diversity in northeast America (Figures I.1a and I.S1). Because of the larger number of strains, we could not use a sliding phylogeny approach

as previously, since this method requires bootstrapping within each window.

In order to visualize the variation of phylogeny along the genomes, we used CBS432 and the 16 presumably non-introgressed HC strains (10 *SpB* and 6 *SpC*) as references. Among the filtered variants, we selected the sites where these genomes show fixed alleles within *SpA, SpB* and *SpC*. We then extrapolated the observed allele-lineage associations to all HC and LC strains, thus redefining the three lineages at each selected site. This approach allowed us to classify genotypes in 10 categories: 'A', 'B' and 'C ' are homozygous for alleles associated with a single lineage, 'AB', 'AC' and 'BC' are homozygous for alleles associated with two lineages, 'R' is homozygous for a rare allele (no association with lineages), "H" is heterozygous, "N" is an unknown genotype and "P" is a genotype at a site where genetic references show polymorphism within lineages. We identified 280,724 sites that are variable between at least two strains.

We calculated the frequencies of each allele categories for each strain in 5,000bp discrete windows along chromosomes. We considered that a given window was an introgressed region if a majority of alleles that were associated with a lineage (the "donor" lineage") is found in at least one strain from a lineage not associated with this allele (the "receiver" lineage). We identified extensive *SpB*-like windows in genomes of eight additional strains from *SpC* (Figure I.3a). In these strains, *SpB*-like regions represent 2.2-5.8% of the genome (Table I.6). From this point, we will also refer to these strains as *SpC**.

As previously observed, *SpB*-like windows are not randomly distributed along the genome but cluster together in large and well-delimited regions, and could represent substantial portions of chromosomes (up to 195kb representing 69% of chromosome III in LL2011_006). All *SpC** strains share seven well-delimited *SpB*-like regions in chromosomes VI (55kb), XIII (35kb), XVI (30kb), VIII (15kb), V (2×10kb) and IV (15kb), suggesting that *SpC** strains likely share a common

ancestor that already harbored these *SpB*-like regions (Table I.6). Other *SpB*-like regions are either shared between at least two strains (670kb) or are unique to one strain (355kb), resulting in six types of *SpC** strains distinguishable by different sorting of *SpB*-like regions and referred as types SK, QC, IO1, IO2, WI and PP (Table I.6).

All but two *SpC** types (SK, WI) come from the narrow contact area between lineages *SpB* and *SpC* (South Quebec), often in locations where both parental lineages were isolated (Figure I.1a, Table I.6). The *SpC** strains isolated from this geographical area also harbor the highest proportion of *SpB*-like regions (2.8-5.8% of the genome), corresponding to up to three hybridization events (H0-H3), suggesting that the biogeographical context may have facilitated backcrosses between *SpC** and its parental lineages. The *SpC** strain yHKS414 (WI type) comes from Wisconsin, a region where only *SpB* was isolated (Sylvester *et coll.* 2015), and only harbors two types of *SpB*-like regions (H0 and H1b; 2.5% of the genome). Strains UCD62-186 and UCD62-268 (SK type) that show admixture between *SpC** and *SpC* (see STRUCTURE analysis in section 10) harbor 2.2% of *SpB*-like regions, mostly consisting in H0 regions. However, the origin of these strains is highly uncertain (likely isolated in Saskatchewan), and to our knowledge, there is no further confirmation about the presence of *S. paradoxus* in this region. The *SpC* strain YPS667 from Pennsylvania (Kuehne *et coll.* 2007) (PE type) also showed admixture between *SpC** and *SpC* in the STRUCTURE analysis (Section 10, Figure I.1c), but we detected no *SpB*-like region in its genome, suggesting that YPS667 is part of a diverging *SpC* population. However, more sampling is needed to confirm the presence of the *SpC* population in this region where *SpB* is dominant (Figures I.1a and I.S1). It is also important to note that we obtained the aforementioned strains types (SK, PE) because they were suspected or shown to be of the *SpC* lineage (Kuehne *et coll.* 2007, Leducq *et coll.* 2014) and are not thus representative of these regions.

19. S. paradoxus karyotype analysis by PFGE

To identify chromosomal rearrangements within and among lineages, we performed karyotypic analysis of 116 strains based on PFGE chromosome profiling performed in Charron *et coll.* (2014). We first converted raw pictures of PFGE gels (Figure I.S12) in pixel intensity matrices using ImageJ 1.45 m (Schneider *et coll.* 2012). All gels contained the same *S. cerevisiae* (strain S288C) standard that we used as control to normalize band sizes among images. To do so, we retrieved pixel intensity profiles for each standard using distance from the well as origin. Then we varied this distance proportionally for each gel image (i.e. by conserving the same relative spacing between bands) until we found the maximum match among different images, using Pearson correlation coefficient in pixel intensities per position. Pixel intensities of each image were converted accordingly. We removed bands corresponding to four large chromosomes (IV, XII, XIV, VII; >1.1 Mb) for which variation could not be detected with these PFGE settings. To decrease residual experimental noise, we calculated mean intensities in five band size classes (*A*-*E*) that we could clearly distinguish based on the combined profiles of the 116 *S. paradoxus* strains (Figure I.S13a,b). These categories correspond to *S. cerevisiae* chromosomes (in decreasing size) XIII, XVI (*A*); II, XIV, X, XI (*B*); V, VII (*C*); IX (*D*); III, VI and I (*E*). Using this approach, *S. paradoxus* lineages *SpB*, *SpC* and *SpC** could be clearly distinguished based upon their karyotype profiles. Strains from the same lineage clustered together in a similarity matrix based upon mean intensities in five size classes (Figure I.S13a) and form well separated groups in a principal component analysis (Figure I.S13d). Strains from lineage *SpC** are characterized by a significant decreased intensity in classes *A* and *E* ($p<0.001$ and $p<0.05$, respectively; Figure I.S13c) and a significant increased intensity in classes *B* and *C* ($p<0.001$; Tukey test) as compared to lineages *SpB* and *SpC*, corresponding to major changes of migration of chromosomes XIII and VI, (Figure I.S13b). Strains from lineage *SpC* have a significant increased intensity in class *A*

and a significant decreased in class *B* (p<0.001; Figure I.S13c), corresponding to major changes of migration of chromosomes II and X (Figure I.S13b). These apparent gains and losses reflect extensive variation in chromosome size, likely resulting from chromosomal translocations that are fixed within lineages *SpC* (IItXr) and *SpC** (VItXIIIr), the latter being further confirmed by *de novo* genome assembly (see below). All analyses were performed in R (R Development Core Team 2013).

20. *De novo* assembly of 24 HC strains using ABySS

We performed *de novo* genome assembly of 24 HC genomes using ABySS (Simpson *et coll.* 2009) to identify chromosomal rearrangements responsible for the difference in karyotypic profiles observed. We tested *k*-mer lengths (parameter *k* in ABySS) in a range between 25 to 64 (k=64 is the maximum allowed value by default in ABySS) in three strains from lineages *SpA* (LL2012_026), *SpB* (UWOPS79.140) and *SpC* (LL2012_027), respectively, using depth of coverage, *N50* (both statistics provided by ABySS), the number of scaffolds (*N*) and the total genome size (*S*) as descriptors for assembly quality (Figure I.S5). We calculated *N* and *S* after removing scaffold smaller than 200bp, corresponding to the minimum scaffold length expected with 100bp paired-end reads. Despite an overall decrease of depth for high k values (19-30×; Figure I.S5a), we observed the maximum N50 (80-121 kb; Figure I.S5b), the minimum number of scaffolds (N=295-342; Figure I.S5c) and the maximum genome size (S = 11.7-11.8 Mb; Figure I.S5d) for k=64. We therefore used this value for further genome assemblies. *De novo* genome assembly with ABySS yielded a large number (N=1749-10107) of scaffolds per strain for a total genome size in the expected range for *Saccharomyces* genomes (S=11.9-12.7 Mb) (Kellis *et coll.* 2003). After removing scaffolds shorter than 200 bp, the number of scaffolds per strain dramatically decreased (n=248-422) but the total genome size was almost unaffected (S=11.7-12.0 Mb).

21 Scaffold alignment and filtering using MAUVE

We first distinguished scaffolds from mitochondrial and nuclear genomes. We performed scaffolds ordering and alignment on the CBS432 mitochondrial reference genome (GenBank accession: JQ862335.1; (Prochazka *et coll.* 2012)) for each strain separately, using the "Move contigs" option implemented in MAUVE (Darling *et coll.* 2004) with default parameters (minimum LCB weight = 200bp). For one strain from lineage *SpA* (LL2012_026) we identified no mitochondrial scaffold and so we assumed it harbored no mitochondrion, at least at the DNA extraction step. We successfully identified mitochondrial scaffolds for all remaining strains. Then we removed the mitochondrial-like scaffolds from the dataset and aligned again the remaining scaffolds onto the S288C *S. cerevisiae* nuclear genome (Goffeau *et coll.* 1996), which is the most complete to date, using the "Move contigs" option implemented in MAUVE (Darling *et coll.* 2004). We increased LCB weight (2,000bp) to detect rearrangements supported by genomic blocks larger than 2kb. We identified scaffolds that did not align with any region of the reference as orphan scaffolds. The mitochondrial genome consisted in few scaffolds per strain (N = 1-18) for a total size (S = 64.0-87.8 kb) and a GC content (13.84-15.44%) typical for *S. paradoxus* mitochondrial genomes (14.06%) (Prochazka *et coll.* 2012). Nuclear genome sizes (S = 11.67-11.92 Mb) and GC contents (38.45-38.72%), including orphan scaffolds, were in the range observed for *S. paradoxus* CBS432 (S = 11.75 Mb; 38.65%) and *S. cerevisiae* S288c (S= 12.16 Mb; 38.30%) (Kellis *et coll.* 2003). Nuclear genome size is significantly higher in both lineages *SpC* (S=11.81±0.03 Mb) and *SpC** (S=11.86±0.06 Mb) than in lineages *SpA* (S=11.74±0.01 Mb; p<0.01; Tukey test) and *SpB* (S=11.75±0.03 Mb; p<0.05; Tukey test). Nuclear GC content is significantly higher in lineage *SpA* (38.71±0.02 %) than in any other lineages (38.50±0.02 %; p<0.001; Tukey test).

22. Identification of chromosomal rearrangements

To identify chromosomal variations (translocations and inversions) in *S. paradoxus* lineages, we used the *S. cerevisiae* S288C genome (Goffeau *et coll.* 1996) as reference. We retrieved MAUVE output files containing coordinates of homologous genomic regions between ordered scaffolds of each *S. paradoxus* strain and the reference (Backbone files). For each strain, we identified chimeric scaffolds, i.e. scaffolds that aligned with at least two non-contiguous regions of the reference genome. Chimeric scaffolds consist of several "blocks", each of which aligns with two distinct genomic regions, either distal (i.e. separated by another scaffold) on the same or different chromosomes (translocations), or matching with two contiguous genomic regions with different strands (inversion). From these chimeric scaffolds, we removed those with harboring block sizes smaller than 400bp. We considered a rearrangement to be the same between two different strains if the homologous fusion point coordinates were less than 1,000bp apart. Depending on the level of fragmentation of each genome, the size estimation of blocks could vary. We identified 122 different types of translocations, 95 involving small-scale telomeric exchanges among chromosomes (<20kb), and 20 different types of inversions, 18 of them being of small scale (<20kb). Only 26 translocations and most inversions were present in at least two strains, and most of these occurrences supported the *S. paradoxus* phylogeny (Figure I.S14a). One small-scale translocation is present in all *S. paradoxus* strains and is thus likely ancestral in this species or represent a change in the *S. cerevisiae* lineage. Three other small-scale translocations are fixed within American lineages *SpB*, *SpC** and *SpC*, and two are fixed within *SpA*. One and two additional small-scale translocations are fixed in lineage *SpB* and *SpC*, respectively.

All large-scale translocations but one are unique to one strain (Figure I.S14a). The exception is a translocation between the right arm of chromosome XIII and chromosome VI (VItXIIIr) that is only present in *SpC** strains (Figure I.3b, I.S14a). This translocation is supported by the absence of band for chromosomes VI (~300kb) and XIII (~900kb) at

expected sizes in all *SpC** karyotypes based on PFGE (Figure I.S13a). The increased intensities observed in PFGE class sizes *B* (~700kb) and *C* (~600kb) for all *SpC** strains likely correspond to the newly formed chromosomes (XIIIL and VItXIIIr, respectively). Two large-scale inversions are found on chromosomes VI (iVI; 40 kb) and VII (iVII; 30-60kb), respectively. Inversion iVII is present in all American lineages (Figure I.S14a), while iVI is only present in *SpB* and *SpC**. One translocation (VItXIIIr) and one inversion (iVI) thus distinguish *SpC** from its parental lineages *SpB* and *SpC*. The iVI inversion is flanked by two ~100bp inverted repeated (IR; Figure I.S6a) regions that may be a remnant of transposable element activities responsible for the inversion (Kim *et coll.* 1998). The iVI boundaries also coincide with the limits of a *SpB*-like region that is shared between all *SpC** strains (H0; Figures I.3 and I.5a).

23. Validation and distribution of the VItXIIIr translocation

From MAUVE alignments, we identified chimeric scaffolds corresponding to the putative VItXIIIr translocation in three *SpC** strains. We searched for them in the NCBI database using BLAST and in the *S. paradoxus* database from the SANGER institute (Liti *et coll.* 2009) (Figure I.S14b). The three scaffolds align together along a 150kb region, which is homologous with 20kb of the right arm of *S. cerevisiae* S288c chromosome VI (GenBank: BK006940.2, 86% of identity) and 115kb of the right arm of *S. cerevisiae* S288c chromosome XIII (GenBank: BK006946.2, 88% of identity). The two regions are separated by a 22kb fusion region, for which no close homology was identified. To validate and estimate the distribution of the VItXIIIr translocation in *S. paradoxus*, we mapped reads from 161 libraries (HC and LC combined) with Bowtie2 onto the chimeric VItXIIIr scaffold of the *SpC** strain LL2012-018 (172,029 bp). For each strain, we calculated the averaged depth of coverage in 1,000 bp sliding windows along the chimeric scaffold (Figure I.S14c). For all *SpA* and *SpC* and most (97%) of *SpB* strains, no read mapped in the fusion region. The fusion

region was only present in the eleven *SpC** and three *SpB* (referred to as *SpBf*) strains (LL2013_183, LL2012_014, UWOPS80-13), suggesting that these strains harbor the VItXIIIr translocation. The three *SpBf* strains come from highly disconnected places from Canada (New Brunswick, Québec, Ontario), in sampling locations where no *SpC* or *SpC** strain was isolated.

24. Validation, distribution and phylogeny of the iVI inversion

We assessed the presence of the iVI inversion in all *S. paradoxus* strains by PCR. We designed two reverse diagnostic primers, one at each end of iVI, in regions that are fully conserved across *S. paradoxus* lineages regardless of the strand of iVI, and two forward primers, one in the inverted-repeated (IR) regions flanking iVI for positive control PCR with reverse primers, and one upstream the 5' IR region for diagnostic PCR with reverse primers (Table S11, Figure I.S6a-b). For each PCR reaction, the mixture contained 2 µl of 10X Bioshop® Buffer, 20ng of genomic DNA, 1.2 µl of $MgCl_2$ 25mM, 1.6 µl of dNTP mix 2.5 mM, 1.6 µl of each primer at 2.5 µM (Table S11) and 0.12 µl of Taq Polymerase Bioshop® 5 U/µl, in a final volume of 20 µl. PCR reactions were carried out in a thermocycler MasterCycler ProS Eppendorf© with the following steps: 5' at 95°C; then 38 cycles of 20" at 98°C, 15" at 59°C and 1'30" at 72°C; and a final extension of 10' at 72°C. The presence or absence of PCR products was verified on agarose gel. Genomic DNA was the same as the one used for the genomic Illumina library preparation described above. We confirmed that the "ancestral" configuration (i.e. no inversion) observed in *S. cerevisiae* is fixed in lineages *SpA* and *SpC*, while the iVI inversion is fixed in *SpC**, found in all in all *SpB* strains but six, suggesting a secondary loss in these strains (Figure I.S6b-c, Table S1). The fact that the inversion is present in both *SpB* and *SpC** lineage and correspond to a *SpB*-like region that is fixed within *SpC** (H0) suggests that the inversion was transmitted from *SpB* to *SpC** during the original hybridization event. From a phylogeny of

the iVI region (1135 varying positions, complete deletion) using 134 strains (strains with >50% missing positions were removed; Figure I.S6c), we confirmed that *SpC** strains form a monophyletic group within the *SpB* clade, which confirms that they inherited the iVI region from *SpB*. Interestingly, the three *SpBf* strains form a monophyletic group with *SpC** within the *SpB* clade, indicating that *SpC** likely inherited iVI from *SpBf* strains (Figure I.S14d).

25. Bias in segregation of some chromosomal regions and introgressed regions correlate with reproductive isolation

We whole-genome sequenced pools of 325 haploid spores that we obtained from the sporulation of eight *SpC* × *SpC** diploid F1 hybrids (Table I.7). For each *SpC* × *SpC** cross, 24 tetrads were dissected and surviving spores (19-47 per cross) were incubated in one mL of liquid YPD overnight at 25°C with agitation. We pooled together spores from a same cross according to the number of surviving spores per tetrad (*n*= 1-4), after normalizing cell concentration (OD_{600nm}). For all crosses but one, no complete tetrad was obtained (*n*=1-3) so we used three pools of spores for each of them. For one cross, we also obtained one complete tetrad, so we obtained four pools for this cross (*n*=1-4). We thus obtained a total of 25 pools of spores (3-30 spores per pools; Table I.7) that we sequenced following the Illumina Nextera® protocol. Reads can be retrieved from NCBI under the BioProject number PRJNA277692 (BioSamples SAMN03389818-SAMN03389842).

We mapped reads onto the reference *S. paradoxus* genome CBS432 using Bowtie2 (Table I.7). In order to assess segregation biases at the genomic level, we called SNPs and indels in the 25 pools of spores at once using Freebayes (Garrison et Marth 2012). The parameters were chosen so that each library was modeled as a pool of DNA from a discrete number of individuals ("-J" option) expected to show only 2 alleles per variant site ("-p 2" option). We filtered the resulting variants using VCFlib by excluding

sites with a polymorphism probability under 99% and genotypes with a number of reads under 4. We combined the variant sites from the spores with those of the parents (as described in section 7) into one VCF file using VCFlib(Garrison et Marth 2012). From this file, we retrieved the origin and frequency of each allele at each variant site and in each library, as well as information about genetic divergence between the two groups of parent strains.

For each library, we calculated the frequency of genotypes inherited by the *SpC** parent in 5kb windows along the genome (Figure I.4d). Regions of high *SpC*-*SpC** divergence coincided with *SpB*-like regions, as expected. We calculated Pearson correlation coefficients between the number of surviving spores per tetrad and the *SpC** genotype frequency in 20kb windows. We calculated correlation coefficients by performing 1,000 independent randomization of frequency values among libraries from the same cross to determine whether *SpC** genotype frequencies and spore surviving are more correlated than expected by chance alone.

*SpC** genotypes in *SpB*-like regions of chromosomes II (IIL) and VI (VIr) are most likely transmitted to the progeny than expected by chance, and bias in the transmission increases with spore lethality, while *SpC** genotypes in the *SpB*-like region of chromosome XII (XIIr) show the reverse pattern (Figure I.5d). We observed other regions where *SpC** genotypes show bias in segregation, but that do not correspond to *SpB*-like regions.

For each pooled library, we estimated the ploidy level along the genome. To do so, we normalized the depth of coverage per position (D) by the median over the genome and filtered outlier positions (D > 1.5×3rd quartile) to remove highly repetitive regions. We then calculated the mean depth of coverage in discrete 20,000bp windows along the genome. We normalized again each library by a control library (from a complete tetrad) to reduce coverage variation along the genome caused by non-random tagmentation of genomic DNA. In each window, we calculate the Pearson's correlation coefficient between the normalize depth and the number of surviving

spores per tetrad and calculated the expected distribution of correlation coefficient by 1,000 randomization of coverage values, as described above for genotype frequencies.

Complete duplications of chromosomes XI and XII occurred in one and two crosses, respectively, regardless of the number of surviving spores per tetrad (Figure I.4e). Two partial aneuploidies occurred in the eight crosses, namely on the right arms of chromosomes II (IIr; 160kb) and XIII (XIIIr; 240kb). Increase of depth of coverage was systematic in these regions and was significantly and negatively correlated with the number of surviving spores per tetrad in both cases (Pearson's correlation test; $p<0.001$ and $p<0.05$, respectively; Figure I.5c). The breakpoint in ploidy on chromosome XIII coincides with the fusion point with chromosome VI in the VItXIIIr translocation in *SpC** strains (Figure I.4e). We mapped reads from spore libraries with Bowtie2 onto the chimeric VItXIIIr scaffold of the *SpC** strain LL2012-018 (172,029 bp) and calculated ploidy level as described above. We found that ploidy level was negatively correlated with the number of surviving spores per tetrad, as observed for the linked XIIIr region (Pearson's correlation test; $p<0.05$; Figure I.5c). For chromosome II, we identified one major chromosomal variation in the PFGE karyotype profile of lineage *SpC* involving chromosome II (see section19; Figures I.S12 and I.S13), which could result from a translocation from chromosome II to chromosome X that is typical of *SpC* strains..

26. Genotyping of iVI regions in spores from crosses between *SpC* and *SpC** strains

Using the diagnostic protocol for the iVI inversion (section 24), we assessed the configuration of the region of chromosome VI in 154 spores obtained from five experimental crosses between *SpC* and *SpC** strains (LL11_005×LL12_006, LL11_005×MSH587-1, LL12_016×LL11_002, LL12_016×LL11_004 and LL12_016×LL11_009). DNA extraction followed Lõoke *et al.* (Lõoke *et coll.* 2011) with the following modifications in order to

perform the extraction in 96-well plates: 1-Cells were suspended in 50µL of lysis buffer; 2-100µL of ethanol 95% were added for the DNA precipitation; 3-Plates were spun at 3,000 rpm for 48 minutes (5810R, Eppendorf); 4-Pellets were washed with 100µL 70% ethanol; 5-Pellets were dissolved in 100µL of water and 6- plates were spun at 3,000 rpm for 3 minutes at 15,000g. The genotype of each individual spore could be assigned by the presence of either band (*SpC* and *SpC** genotypes) or both bands (*SpC/SpC** genotypes) on an agarose gel.

We found that the iVI inversion (*SpC** configuration) is present in 66% of surviving spores, while only 17% have the *SpC* configuration. Interestingly, 18% of surviving spores have both configurations, suggesting frequent aneuploidy for the iVI region. Moreover, we found that the iVI is more often transmitted in tetrads with only 1 or 2 surviving spores than expected by chance ($p<0.001$ and $p<0.05$, respectively; 1,000 randomization as described in section 25; Figure I.5b).

27. Role of chromosomal rearrangements in the dynamics of introgressed regions and in reproductive isolation

We examined the role of chromosomal rearrangements distinguishing *SpC* and *SpC** in decreasing spore viability and bias in *SpB*-like regions transmission, namely the iVI inversion, the VItXIIIr translocation (both typical of *SpC**; Figure I.S14) and the XtIIr translocation (typical of *SpC*; Figure I.S13).

The translocated regions (XIIIr and IIr) show increased ploidy in *SpC* × *SpC** spores (Figures I.4e and I.5c), indicating that surviving spores inherit extra copies of these regions. Translocated regions IIr and XIIIr (240 and 180 kb, respectively) contain 28 and 13 genes, respectively, which are essential in *S. cerevisiae* (Giaever *et coll.* 2002) indicating that the absence of these regions ($N=0$) is likely lethal in *S. paradoxus* (Figure I.S12a). Assuming that a translocated region segregates like a bi-allelic locus, we expect that its frequency in meiotic products follows Mendelian

frequencies. In the case of the VItXIIIr translocation, 25% of spores are expected to carry either parental configuration (VI, XIII) or (VItXIIIr, XIIIL), each having one copy of the translocated region XIIIr (*N*=1), which is viable. Non-parental configurations (VItXIIIr, XIII) and (VI, XIIIL) exhibit respectively two (*N*=2) and no (*N*=0) copy of XIIIr, the latter being lethal. The same expectations could be done for XtII. Hence, for both the VItXIIIr and the XtIIr translocation, 75% of spores are expected to survive and to carry on average 1.33 copies of translocated regions, in agreement with what we measured in surviving spores (*N*=1.23±0.10 and 1.24±0.11, respectively). By combining effects of both translocations, we expect that 56.25% of spore carry at least one copy of each translocated region, indicating that translocations contribute to most (80%) of the observed spore mortality (43.75% in theory against 54.7% observed in actual crosses; Figure I.2).

The iVI inversion coincides with an introgressed region that is fixed within the *SpC** lineage (H0; Figures I.3 and I.5a, Table I.6). These regions are most likely transmitted in surviving *SpC*×*SpC** progeny (Figure I.5b,d), while other fixed introgressed regions, which do not coincide with identified chromosomal rearrangements, normally segregate in the progeny (Figure I.4d). Because the iVI inversion and the translocated region XIIIr are on the same arm of chromosome VItXIIIr (separated by 50kb; Figure I.S15a), they are likely co-segregating during meiosis. Similarly, regions located on same arms than translocated regions (LT regions), show strong bias in transmission of *SpC** genotypes (Figure I.S15a), suggesting that these regions co-segregate with translocated regions.

We evaluated the extent of co-segregation in LT regions, which is expected to decrease from translocation breakpoints due to recombination (Figure I.S15b). To do so, we determined whether *SpC** genotype frequencies in 25 spore libraries were associated among distant regions of the genome. Genotype frequencies were retrieved for each library and each discrete 5kb window of the genome (see section 25) and association was

calculated as the correlation coefficient (Pearson test, Bonferroni correction) among *SpC** genotype frequencies in 25 pools for all possible window pair located within 400kb in the same chromosomes (80 classes in the range 5-400 kb, n = 1,067 – 2,095 pairwise comparisons per class; Figure I.S15b). The extent of co-segregation in LT region decreases with distance, as expected, but is larger in the iVI inversion (90kb) than in other LT regions (<10kb), indicating that the inversion causes reduced recombination (Figure I.S15b). Hence, the over-transmission of the iVI inversion and of the *SpB*-like regions likely results from combined effect of their linkage with the XIIIr region and reduced recombination in the iVI region.

Given that the iVI inversion co-segregates with XIIIr, we expect that it is present in two thirds (67%) of surviving spores (Figure I.5a), which is in agreement with what we observed (66%; see section 26 and Figure I.5b). We also observed that 18% of viable spores inherited both configurations of the iVI inversion, which is unexpected given the overall normal ploidy we observed for this region in pool libraries (N~1). One explanation is that casual recombination occurs in the iVI region between configurations of *SpC* (normal chromosome VI) and *SpC** (VItXIIIr), which is expected to produce two new chromosomes: an acentric chromosome, which would be lethal for the inheriting spore, and a dicentric chromosome, which would be viable. Considering that our PCR diagnostic is based on amplification of the iVI end located next to the centromere (Figure I.S6a), it is possible that the double amplification we observed for some spores results from the presence of both configurations of the diagnostic region in the dicentric chromosome. The fact that 18% of surviving spores harbor dicentric chromosomes would indicate that the same proportion of dead spores inherited acentric chromosomes. This suggests that recombination in the iVI inversion also plays a role in decreasing overall spore viability, independently from the translocations.

We provide other examples of non-fixed introgressed regions that do not normally segregate in the progeny (Figure I.5d), but with no link with chromosomal rearrangements. For instance, a non-fixed *SpB*-like region in left arm of chromosome II (IIL) typical of *SpC** strains from type PP (Table S6) is most likely transmitted in surviving progeny, indicating that it could become rapidly fixed in outcrossing populations with mixed *SpC* and *SpC**. Another non-fixed *SpB*-like region in right arm of chromosome XII (XIIR) typical of *SpC** types IO1 and QC (Table I.6) shows the opposite pattern. Because we cannot directly link these observations with chromosomal changes, they suggest that, like chromosomal rearrangements, introgressed regions are dynamic. This also indicates that new *SpB*-like regions can still arise in *SpC**. Relative time divergence between different *SpB*-like regions and their homologous regions in *SpB* lineage support this hypothesis (See section 28, Figure I.S16, Table I.9). For instance, the *SpB*-like region of chromosome II is 4 times younger than that of chromosome VI, suggesting that it was secondary acquired by backcrosses between *SpB* and *SpC**. Accordingly, we showed that reproductive barriers among lineages are not complete, and the acquisition or loss of *SpB*-like regions by backcrossing with *SpB* and *SpC*, respectively, is still feasible in sympatric populations.

28. Timing of introgression and biogeographic scenario

We attempted to date the original hybridization between *SpB* and *SpC* that gave rise *SpC** and to determine whether the departing *SpB*-like regions we observed among different *SpC** strains results from successive and/or independent backcrosses with parental lineages *SpB* and *SpC*. We estimated the divergence time between *S. paradoxus* lineages in different genomic regions. First, we splitted the 24 HC consensus genomes according to *SpB*- and *SpC*-like regions we identified in the eleven *SpC** strains by a site-wise clustering approach (see section 18), using as units

500bp sliding windows. From the consensus alignment of chromosomes I, II, IV, VI, IX, XIII, XV and XVI, we generated new alignments by grouping together 500pb windows with similar phylogenetic signals. Non-introgressed regions represented the majority of each chromosome and were thus split into ~50kb regions, while *SpB*-like regions (15-77kb) were not sliced. For instance, chromosome II was sliced into eleven alignments consisting in ten non-introgressed windows (10×51.4kb) and one *SpB*-like window that is typical to *SpC** strains from type PP (77kb; Table S6), respectively.

We used a Bayesian approach to estimate divergence time between evolutionary categories using the software BEAST v1.8.1 for each alignment (Drummond *et coll.* 2012). Evolutionary categories represented in HC strains are CBS432 (European), *SpA* (5 strains; European isolated in North America), *SpB* (10 strains), *SpC* (5 strains), PE (1 strain), PP (2 strains) and IO2 (1 strain). We considered PE strains separately from other *SpC* strains (strain YPS667; see section 18 for justification). *SpC** strains from types PP and IO2 differ with respect to introgressions on chromosomes II/IX, and XI, respectively (Table I.6). We used the divergence between *SpB* and *SpC* as calibration time, assuming that it started at the beginning of the last glaciation 110,000 years ago (±10,000 years; normal distribution) (Dyke 2004, Charron *et coll.* 2014) and $1.67x10^{-10}$ as initial substitution rate previously estimated in *S. cerevisiae* (Zhu *et coll.* 2014), with default distribution probability (mean=0, standard deviation=1). We compared two substitution rate models (strict *vs.* Lognormal relaxed clock) using AICM comparison (1,000 bootstraps) implemented in Tracer v1.6 (Drummond *et coll.* 2012).

This approach allows the substitution rate to vary among evolutionary categories (Drummond *et coll.* 2006). We used three different regions of chromosome VI as controls: the H0 region (*SpB*-like in all *SpC** types; 36kb) and the main remaining part of the chromosome with no introgression, sliced in two equal regions (55 kb each) and performed

$10x10^6$ MCMC iterations with 10% of burn-in and statistics sampling every 10,000 iteration. The Lognormal relaxed clock model has systematically higher AICM than the strict clock model, based on likelihood, while introducing 1.5 to 2.8 times more variance in estimation of divergence time (data not shown). We also observed a greater variance among different regions in estimating divergence time with the relaxed clock model, which may reflect the contrasted evolutionary histories among genomic regions due to different recombination rates, selective pressures and more complex introgression scenarios.

Using the relaxed clock model, we first estimated the divergence times between evolutionary categories in 53 ancestral regions of 50kb (2.56 Mb) for which the MCMC stabilized for all estimated statistics after burn-in (effective sampling size ≥100). The 16 regions of 50kb did not stabilize and were removed. We combined time estimations from these 53 independent simulations using Tracer v1.6 (Drummond *et coll.* 2012) and retrieved the mean, median and standard deviation for each divergence time. From this, we estimated that *SpA* strains from North America diverged from each other 286 ± 174 years ago, in agreement with a radiation after a trans-oceanic migration event (Kuehne *et coll.* 2007) during the colonization of North America by Europeans in the last 500 years, as proposed for the origin of these strains out of Europe (for instance in New Zealand (Zhang *et coll.* 2010)). In addition, based on our assumption, we dated the divergence between European and American populations to 176,200 ± 36,600 years ago, in the same order of magnitude as previous estimations (124,268–994,147 years ago (Liti *et coll.* 2006)). Our estimates based on the timing of *SpB-SpC* divergence are therefore in line with biogeographical events that are independent from the one we are studying (Figure I.S16, Table I.9). Radiation within lineages *SpB* and *SpC* (PE included) occurred 15,400 ± 4,000 and 8,200 ± 2,300 years ago, respectively, and coincide with the ice sheet retreat 18,000 to 5,000 years ago (Dyke 2004) , suggesting that these radiations result from

their establishment at the end of the glacial period in this region. We dated the divergence between *SpC* and the *SpC* parental contributor to *SpC** 10500 ± 4300 years ago, hence contemporaneously to these radiations and the end of the glacial period.

We used radiation times estimated from non-introgressed regions as priors for the estimation of time within *SpB*-like regions. We performed analyses in BEAST as described above with the following modifications. For six ancestral *SpB*-like regions analyzed separately (H0 and H1; Table I.6), we assumed that the divergence time between CBS432 and *SpA*, and the radiation times within lineages *SpA*, *SpB*, *SpC** (IO2 and PP combined) and *SpC* (PE removed), followed a normal distribution, with mean and standard deviation estimated from 53 non-introgressed regions as priors, and median as initial value. For three unshared *SpB*-like regions (H2), we removed radiation within *SpC** as prior, and considered the divergence between *SpB* and PP (H2c; chromosomes II and IX) and IO2 (H2b; chromosome XI) in separate analyses. From the H0 and H1 regions we estimated that the common *SpB* contributor to all *SpC** diverged from *SpB* from 10,000 to 28,500 years ago, hence not earlier than the divergence between *SpC* and the *SpC* parental contributor to *SpC** (10,500 ± 4,300). Unshared *SpB*-like regions diverged later from *SpB* (6,500 to 10,400 years ago), supporting the acquisition of these regions by secondary hybridizations (Figure I.S16, Table I.9).

In order to assess the consistency of patterns of genetic diversity with our demographic scenario, we used VCFtools (Danecek *et coll.* 2011) to compute Tajima's D statistics (Tajima 1989) in lineages *SpA*, *SpB*, *SpC* and *SpC** across chromosome 7 (10Kb windows), for which we did not detect evidence on introgression that could affect Tajima's D (Figure I.3a). Hence, we expected that only natural selection and population size variations would explain Tajima's D estimates. We corrected the effects of different sample sizes (14 in *SpA*, 92 in *SpB*, 40 in *SpC* and 15 in *SpC**) by

recalculating the statistics in 100 random subsets of 15 *SpB* and 15 *SpC* strains. We then compared the values across lineages (mean values from *SpB* and *SpC*) using a sign test. Global Tajima's D across chromosome 7 in *SpC* (-1.04±0.60) is significantly higher than in *SpA* (-2.46±0.42; sign test, $p<0.001$) but lower than in *SpB* (-0.60±0.28; $p<0.001$) or *SpC** (-0.57±0.64; $p<0.001$). There is no significant difference between *SpB* and *SpC** (bilateral sign test, $p=1$). Tajima's D for each lineage is significantly lower than zero (sign tests, $p<0.001$), which suggests recent population expansion. This effect is more pronounced for *SpA*, in agreement with its recent colonization in North America, and for *SpC,* which could reflect recolonization that followed the glacial refuge that could have been very limited in terms of geographical area.

Table I.S1: List of strains used in this study.

Name of the strain, whether sporulation was done before sequencing (c.s.: strain cannot sporulate: 4 out of 47 tested), sample location, nature of the substrate, reference (Ref.), lineage (Lin.; *Scer* for *Saccharomyces cerevisiae*), presence of the translocation between chromosomes VI and XIII (VItXIIIr; based on read mapping on a chimeric scaffold; Figure S7b), presence of the inversion on chromosome VI (iVI; based on PCR; Figure S14). Bars indicate missing information. U.S. states and Canadian provinces abbreviations: British Columbia (BC), Indiana (IN), Massachusetts (MA), Maine (ME), Michigan (MI), Missouri (MO), New-Brunswick (NB), New Hampshire (NH), New Jersey (NJ), Ontario (ON), Pennsylvania (PE), Quebec (QC), Saskatchewan (SK), Wisconsin (WI).

Strain	Sporulation	Sample location	Substrate	Reference	Lineage	VItXIIIr	iVI
yHKS172	NO	Indiana Dunes State Park, IN	Bark - Populus deltoides	2	*SpA*	NO	NO
yHKS320	NO	Sheboygan,WI	Bark - Acer saccharum	2	*SpA*	NO	NO
yHKS341	NO	Fayette, MI	Soil - Acer	2	*SpA*	NO	NO
yHKS402	NO	Fayette, MI	Soil - Acer	2	*SpA*	NO	NO
yHKS421	NO	Fayette, MI	Soil - Acer	2	*SpA*	NO	NO
yHRM71	NO	Madison, WI	Soil - Unknown	2	*SpA*	NO	NO
yHRM73	NO	Madison, WI	Soil - Unknown	2	*SpA*	NO	NO
YPS 644	YES	Buck Hill Falls, PE	Oak tree	3	*SpA*	NO	NO
YPS 695	YES	Tuscarora, PE	Oak tree	3	*SpA*	NO	NO
LL2012_001	NO	Hull, QC	Bark - Oak tree	8	*SpA*	NO	NO
LL2012_026	c. s.	Sherbrooke, QC	Bark - Mapple tree	8	*SpA*	NO	NO
X	NO	DNA from non-identified strain	-	-	*SpA*	NO	NO
YPS642	NO	Buck Hill Falls, PE	Bark - Oak tree	3	*SpA*	NO	NO
YPS744	NO	Tuscarora, PE	Bark - Oak tree	3	*SpA*	NO	NO
UWOPS-99-501.1	YES	Long Point, ON	Flux - Quercus rubra	4	*SpB*	NO	NO
yHKS175	NO	Madison, WI	Soil - Quercus alba	2	*SpB*	NO	NO
yHKS339	NO	Tahquamenon Falls, MI	Soil - Tsuga canadensis	2	*SpB*	NO	NO
yHKS69	NO	Kirksville, MO	Soil - Quercus alba	2	*SpB*	NO	NO

UWOPS_79-140	NO	Saint-Joseph Island, ON	Knot - Apple tree		4	*SpB*	NO	NO
UWOPS_91-202	NO	Long Point, ON	Flux - Oak tree		4	*SpB*	NO	NO
95-1-5A	YES	Saugatuck, MI	Oak tree		3.6	*SpB*	NO	YES
LL2012_002a	YES (spore 1/2)	Pointe Platon, QC	Bark - Oak tree		8	*SpB*	NO	YES
LL2012_002b	YES (spore 2/2)	Pointe Platon, QC	Bark - Oak tree		8	*SpB*	NO	YES
LL2012_009	YES	Québec, QC	Bark - Oak tree		8	*SpB*	NO	YES
LL2012_015	YES	Pointe Platon, QC	Bark - Oak tree		8	*SpB*	NO	YES
LL2012_021	YES	Pointe Platon, QC	Bark - Oak tree		8	*SpB*	NO	YES
LL2012_023	NO	Saint-Justin, QC	Bark - Oak tree		8	*SpB*	NO	YES
LL2012_024	YES	Saint Justin, QC	Bark - Oak tree		8	*SpB*	NO	YES
LL2012_028a	YES (spore 1/2)	Sherbrooke, QC	Bark - Mapple tree		8	*SpB*	NO	YES
LL2012_028b	YES (spore 2/2)	Sherbrooke, QC	Bark - Mapple tree		8	*SpB*	NO	YES
LL2012_029	YES	St Thècle, QC	Moss on tree		8	*SpB*	NO	YES
LL2012_030	YES	Vancouver, BC	Bark - Mapple tree		8	*SpB*	NO	YES
LL2013_025	NO	Rockport, MA	Bark - Oak tree	this study		*SpB*	NO	YES
LL2013_027	NO	Biddeford, ME	Bark - Oak tree	this study		*SpB*	NO	YES
LL2013_028	NO	Biddeford, ME	Bark - Oak tree	this study		*SpB*	NO	YES
LL2013_029	NO	Biddeford, ME	Bark - Oak tree	this study		*SpB*	NO	YES
LL2013_030	NO	Biddeford, ME	Bark - Oak tree	this study		*SpB*	NO	YES
LL2013_041	NO	Woburn, MA	Bark - Oak tree	this study		*SpB*	NO	YES
LL2013_043	NO	Woburn, MA	Bark - Oak tree	this study		*SpB*	NO	YES
LL2013_049	NO	Rockport, MA	Bark - Oak tree	this study		*SpB*	NO	YES
LL2013_050	NO	Woburn, MA	Bark - Oak tree	this study		*SpB*	NO	YES
LL2013_051	NO	Woburn, MA	Bark - Oak tree	this study		*SpB*	NO	YES
LL2013_055	NO	Rockport, MA	Bark - Oak tree	this study		*SpB*	NO	YES
LL2013_056	NO	Rockport, MA	Bark - Oak tree	this study		*SpB*	NO	YES
LL2013_143	NO	Sherbrooke, QC	Bark - Oak tree	this study		*SpB*	NO	YES
LL2013_212	NO	Orford sur le lac, QC	Bark - Cedar tree	this study		*SpB*	NO	YES
MSH-3004-2	YES	Mont St Hilaire, QC	Bark - Oak tree		4	*SpB*	NO	YES

MSH-475	YES	Mont St Hilaire, QC	Bark - Oak tree	4	*SpB*	NO	YES
MSH-483	YES	Mont St Hilaire, QC	Bark - Oak tree	4	*SpB*	NO	YES
MSH-573-4	YES	Mont St Hilaire, QC	Bark - Oak tree	4	*SpB*	NO	YES
MSH-578a-4-E	YES	Mont St Hilaire, QC	Bark - Oak tree	4	*SpB*	NO	YES
MSH-605	YES	Mont St Hilaire, QC	Bark - Oak tree	4	*SpB*	NO	YES
MSH3061-3	NO	Mont St Hilaire, QC	Bark - Oak tree	4	*SpB*	NO	YES
MSH498	NO	Mont St Hilaire, QC	Bark - Oak tree	4	*SpB*	NO	YES
MSH503	NO	Mont St Hilaire, QC	Bark - Oak tree	4	*SpB*	NO	YES
MSH544	NO	Mont St Hilaire, QC	Bark - Oak tree	4	*SpB*	NO	YES
MSH567-4	NO	Mont St Hilaire, QC	Bark - Oak tree	4	*SpB*	NO	YES
MSH584	NO	Mont St Hilaire, QC	Bark - Oak tree	4	*SpB*	NO	YES
MSH850-1	NO	Mont St Hilaire, QC	Bark - Oak tree	4	*SpB*	NO	YES
NRLL YB-2047	YES	Salem, MO	Frass for Hickory	4	*SpB*	NO	YES
UTM10	NO	Toronto, ON	Bark - Oak tree	4	*SpB*	NO	YES
UTM9	NO	Toronto, ON	Bark - Oak tree	4	*SpB*	NO	YES
UWOPS-03-346t1	YES	Long Point, ON	Flux - Quercus rubra	4	*SpB*	NO	YES
yHAB204	NO	Tahquamenon Falls, MI	Soil - Tsuga canadensis	2	*SpB*	NO	YES
yHBJ1	YES	Iron River, WI	Decayed wood - Oak tree	4	*SpB*	NO	YES
yHBJ11	NO	Iron River, WI	Soil - Quercus	2	*SpB*	NO	YES
yHBJ12	NO	Iron River, WI	Soil - Abies balsamea	2	*SpB*	NO	YES
yHBJ16	NO	Iron River, WI	Soil - Birch tree	4	*SpB*	NO	YES
yHBJ17	YES	Iron River, WI	Decayed wood - Birch tree	4	*SpB*	NO	YES
yHBJ24	YES	Iron River, WI	Decayed wood - Oak tree	4	*SpB*	NO	YES
yHBJ5	YES	Iron River, WI	Soil - Oak tree	4	*SpB*	NO	YES
yHBJ6	NO	Iron River, WI	Mushroom	4	*SpB*	NO	YES
yHBJ9	YES	Iron River, WI	Soil - Birch tree	4	*SpB*	NO	YES
yHKS10	YES	Sussex, WI	Leaves - Apple tree	4	*SpB*	NO	YES
yHKS145	NO	Madison, WI	Bark - Quercus alba	2	*SpB*	NO	YES

yHKS157	NO	Madison, WI	Bark - Quercus alba	2	*SpB*	NO	YES
yHKS182	NO	Indiana Dunes State Park, IN	Sand - Populus deltoides	2	*SpB*	NO	YES
yHKS183	NO	Madison, WI	Bark - Unknown	2	*SpB*	NO	YES
yHKS223	NO	Plainfield, NH	Soil - Fagus grandifolia	2	*SpB*	NO	YES
yHKS224	NO	Plainfield, NH	Soil - Fagus grandifolia	2	*SpB*	NO	YES
yHKS226	NO	Plainfield, NH	Soil - Betula populifolia	2	*SpB*	NO	YES
yHKS246	NO	Plainfield, NH	Soil - Fagus grandifolia	2	*SpB*	NO	YES
yHKS248	NO	Plainfield, NH	Soil - Fagus grandifolia	2	*SpB*	NO	YES
yHKS253	NO	Plainfield, NH	Soil - Betula populifolia	2	*SpB*	NO	YES
yHKS267	NO	Columbia, MO	Soil - Quercus alba	2	*SpB*	NO	YES
yHKS306	NO	Plainfield, NH	Soil - Fagus grandifolia	2	*SpB*	NO	YES
yHKS331	NO	Tahquamenon Falls, MI	Soil - Tsuga canadensis	2	*SpB*	NO	YES
yHKS343	NO	Plainfield, NH	Soil - Fagus grandifolia	2	*SpB*	NO	YES
yHKS344	NO	Plainfield, NH	Soil - Betula populifolia	2	*SpB*	NO	YES
yHKS5	NO	Slinger, WI	Bark - Unknown	2	*SpB*	NO	YES
yHKS58	NO	Columbia, MO	Soil - Quercus alba	2	*SpB*	NO	YES
yHRM48	NO	Salem, OR	Soil - Quercus	2	*SpB*	NO	YES
yHRM69	NO	Madison, WI	Soil - Unknown	2	*SpB*	NO	YES
YPS 618	YES	New Brunswick, NJ	Oak tree	3	*SpB*	NO	YES
YPS 631	YES	New Brunswick, NJ	Oak tree	3	*SpB*	NO	YES
95-7-1D	NO	East Lansing, MI	Bark - Oak tree	3.6	*SpB*	NO	YES
LL2012_002	NO	Pointe Platon, QC	Bark - Oak tree	8	*SpB*	NO	YES
LL2012_022	NO	Saint Justin, QC	Bark - Oak tree	8	*SpB*	NO	YES
LL2012_025	NO	Sherbrooke, QC	Bark - Oak tree	8	*SpB*	NO	YES
MSH-604	NO	Mont St Hilaire, QC	Bark - Oak tree	4	*SpB*	NO	YES
UTM6	NO	Toronto, ON	Bark - Oak tree	4	*SpB*	NO	YES
YPS484	NO	Grand Bend, ON	Bark - Oak tree	3	*SpB*	NO	YES
YPS616	NO	New Brunswick, NJ	Bark - Oak tree	3	*SpB*	NO	YES
LL2012_014	YES	St Michel de Bellechasse, QC	Bark - Mapple tree	8	*SpBf*	YES	YES

LL2013_183	NO	Oak Point Provincial Parc, NB	Bark - Quercus rubra	this study		*SpBf*	YES	YES
UWOPS-80-13	c. s.	London, ON	Black knot - Prunus virginiana		4	*SpBf*	YES	YES
LL2011_001	NO	Québec, QC	Bark - Oak tree		8	*SpC*	NO	NO
LL2011_002	YES	Cap Chat, QC	Bark - Mapple tree		8	*SpC*	NO	NO
LL2011_003	NO	Cap Chat, QC	Bark - Mapple tree		8	*SpC*	NO	NO
LL2011_007	NO	Île d'Orleans, QC	Bark - Mapple tree		8	*SpC*	NO	NO
LL2011_008	YES	Île d'Orleans, QC	Bark - Mapple tree		8	*SpC*	NO	NO
LL2011_009	YES	Squatec, QC	Soil - Birch tree		8	*SpC*	NO	NO
LL2011_010	NO	Squatec, QC	Soil - Fir tree		8	*SpC*	NO	NO
LL2012_004	YES	Cap Chat, QC	Bark - Mapple tree		8	*SpC*	NO	NO
LL2012_005	YES	Cap Chat, QC	Bark - Birch tree		8	*SpC*	NO	NO
LL2012_006	YES	Cap Chat, QC	Bark - Wild Cherry tree		8	*SpC*	NO	NO
LL2012_007	YES	St Jean Port Joli, QC	Bark - Mapple tree		8	*SpC*	NO	NO
LL2012_008	NO	Cap Chat, QC	Bark - Mapple tree		8	*SpC*	NO	NO
LL2012_010	YES	Parc du Bic, QC	Bark - Mapple tree		8	*SpC*	NO	NO
LL2012_011	NO	Parc du Bic, QC	Bark - Mapple tree		8	*SpC*	NO	NO
LL2012_012a	YES (replicate 1/2)	L'Anse au Griffon, QC	Bark - Mapple tree		8	*SpC*	NO	NO
LL2012_012b	YES (replicate 2/2)	L'Anse au Griffon, QC	Bark - Mapple tree		8	*SpC*	NO	NO
LL2012_019	NO	Pointe Platon, QC	Bark - Oak tree		8	*SpC*	NO	NO
LL2013_006	NO	Québec, QC	Bark - Oak tree	this study		*SpC*	NO	NO
LL2013_140	NO	Lac Témiscouata, QC	Bark - Cedar tree	this study		*SpC*	NO	NO
LL2013_152	NO	Mont Ste Anne, QC	Bark - Mapple tree	this study		*SpC*	NO	NO
LL2013_158	NO	Mont Ste Anne, QC	Bark - Birch tree	this study		*SpC*	NO	NO
LL2013_166	NO	Mont Ste Anne, QC	Bark - Mapple tree	this study		*SpC*	NO	NO
LL2013_200	NO	Orford sur le lac, QC	Bark - Mapple tree	this study		*SpC*	NO	NO
LL2013_201	NO	Orford sur le lac, QC	Bark - Mapple tree	this study		*SpC*	NO	NO
LL2013_202	NO	Orford sur le lac, QC	Bark - Mapple tree	this study		*SpC*	NO	NO
LL2013_204	NO	Orford sur le lac, QC	Bark - Mapple tree	this study		*SpC*	NO	NO

LL2013_205	NO	Orford sur le lac, QC	Bark - Mapple tree	this study		*SpC*	NO	NO
LL2013_206	NO	Orford sur le lac, QC	Bark - Mapple tree	this study		*SpC*	NO	NO
LL2013_207	NO	Orford sur le lac, QC	Bark - Mapple tree	this study		*SpC*	NO	NO
LL2013_210	NO	Rimouski, QC	Bark - Mapple tree	this study		*SpC*	NO	NO
MSHD1S11	YES	Mont St Hilaire, QC	Bark - Oak tree		4	*SpC*	NO	NO
MSHD2B12	NO	Mont St Hilaire, QC	Bark - Oak tree		40	*SpC*	NO	NO
yHKS225	NO	Plainfield, NH	Soil - Betula papyrifera		2	*SpC*	NO	NO
yHKS254	NO	Plainfield, NH	Soil - Betula papyrifera		2	*SpC*	NO	NO
LL2011_004	NO	Cap Chat, QC	Slug - Fir tree		8	*SpC*	NO	NO
LL2011_011	NO	Squatec, QC	Bomblebee		8	*SpC*	NO	NO
LL2011_012	NO	Station Dushesnay, QC	Bark - Mapple tree		8	*SpC*	NO	NO
LL2012_027	NO	Sherbrooke, QC	Bark - Mapple tree		8	*SpC*	NO	NO
MSH-587-1	NO	Mont St Hilaire, QC	Bark - Oak tree		4	*SpC*	NO	NO
YPS667	NO	Buck Hill Falls, PE	Bark - Oak tree		3	*SpC*	NO	NO
LL2011_006	YES	Île d'Orleans, QC	Rotten apple		8	*SpC**	YES	YES
LL2012_016a	YES (spore 1/2)	Pointe Platon, QC	Bark - Oak tree		8	*SpC**	YES	YES
LL2012_016b	YES (spore 2/2)	Pointe Platon, QC	Bark - Oak tree		8	*SpC**	YES	YES
LL2012_017	NO	Pointe Platon, QC	Bark - Oak tree		8	*SpC**	YES	YES
LL2012_018a	YES (spore 1/2)	Pointe Platon, QC	Bark - Oak tree		8	*SpC**	YES	YES
LL2012_018b	YES (spore 2/2)	Pointe Platon, QC	Bark - Oak tree		8	*SpC**	YES	YES
LL2012_020	YES	Pointe Platon, QC	Bark - Oak tree		8	*SpC**	YES	YES
LL2013_010	NO	Québec, QC	Bark - Oak tree	this study		*SpC**	YES	YES
LL2013_012	NO	Québec, QC	Soil - Oak tree	this study		*SpC**	YES	YES
UCD 62-186	c. s.	Saskatoon, SK	Wasp nest		4	*SpC**	YES	YES
UCD 62-268	c. s.	unknown	Flower		4	*SpC**	YES	YES
yHKS414	NO	Sheboygan,WI	Bark - Acer saccharum		2	*SpC**	YES	YES
LL2011_005	NO	Île d'Orleans, QC	Bark - Mapple tree		8	*SpC**	YES	YES
LL2012_016	NO	Pointe Platon, QC	Bark - Oak tree		8	*SpC**	YES	YES
LL2012_018	NO	Pointe Platon, QC	Bark - Oak tree		8	*SpC**	YES	YES

LL2013_040	NO	Rockport, MA	Bark - Oak tree	this study	*Scer*	-	-
LL2012_054	NO	Rockport, MA	Bark - Oak tree	this study	*Scer*	-	-

ANNEXE B: Matériel et méthode Suppémentaires pour le chapitre II

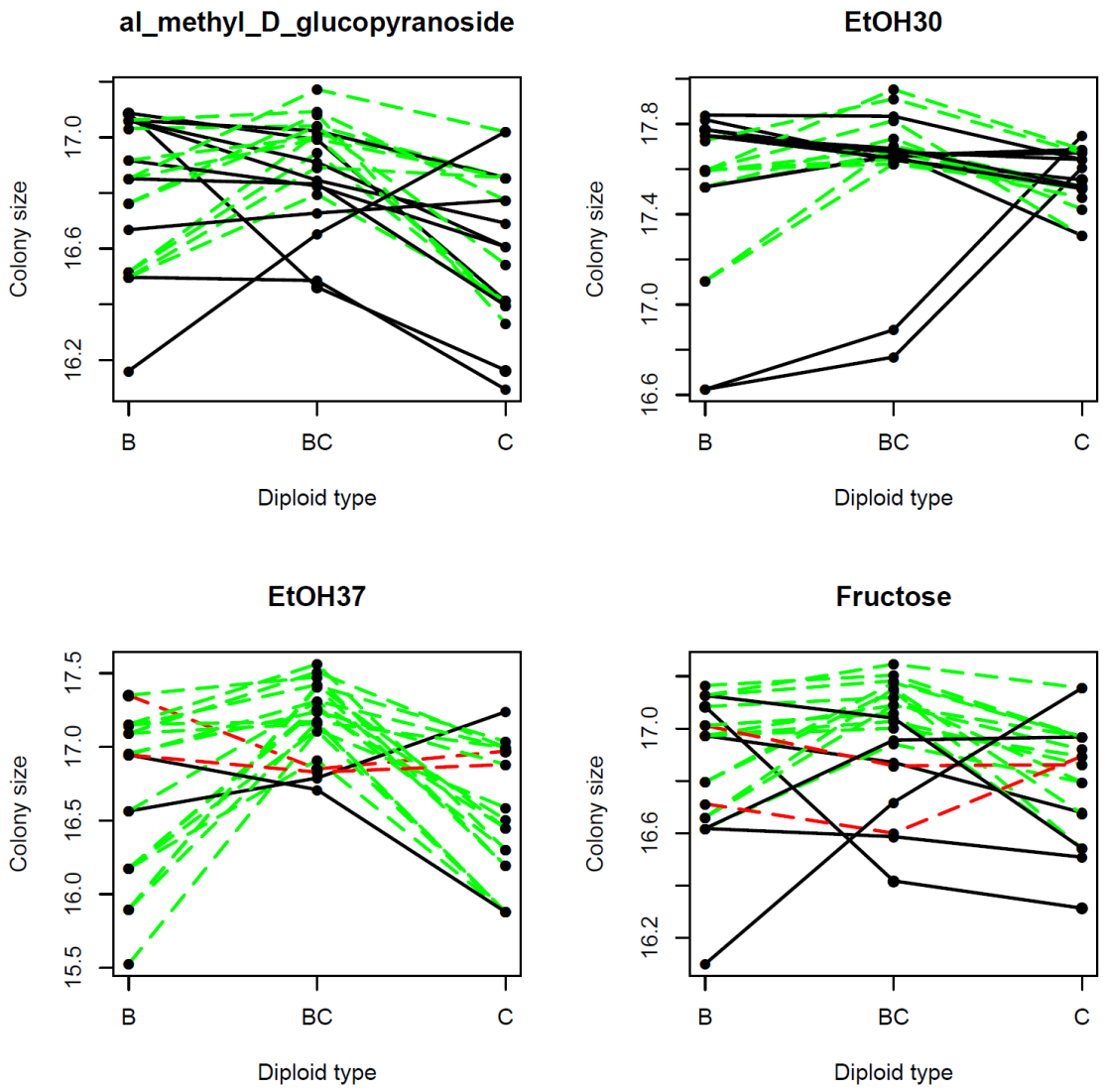
Methods Supplement

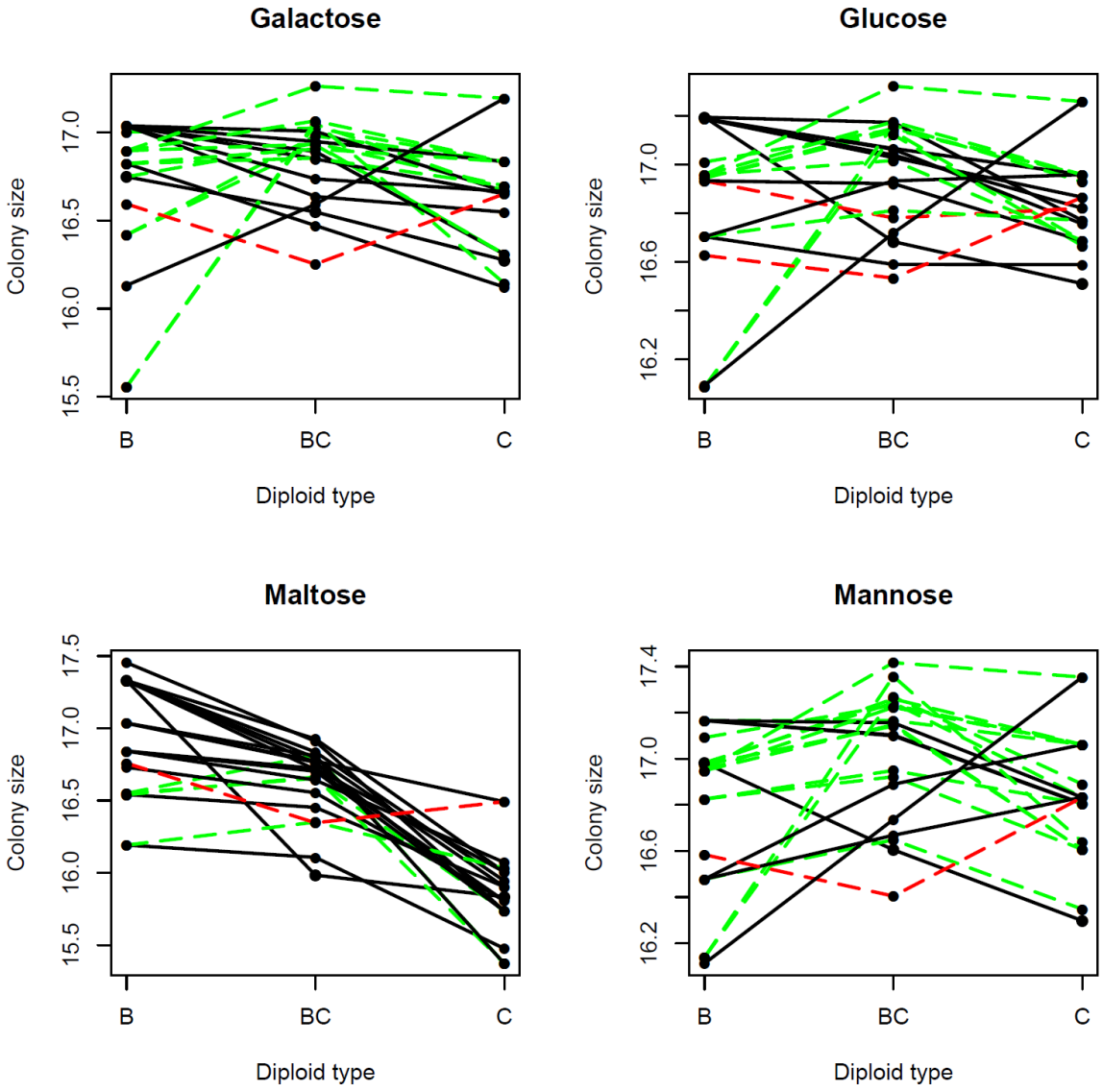
R packages used were flowCore v1.26.3 and mixtools (http://www.jstatsoft.org/v32/i06/) for freeze-thaw survival data. The map was generated using maps (https://CRAN.R-project.org/package=maps), mapdata (https://CRAN.R-project.org/package=mapdata) and maptools (https://CRAN.R-project.org/package=maptools).

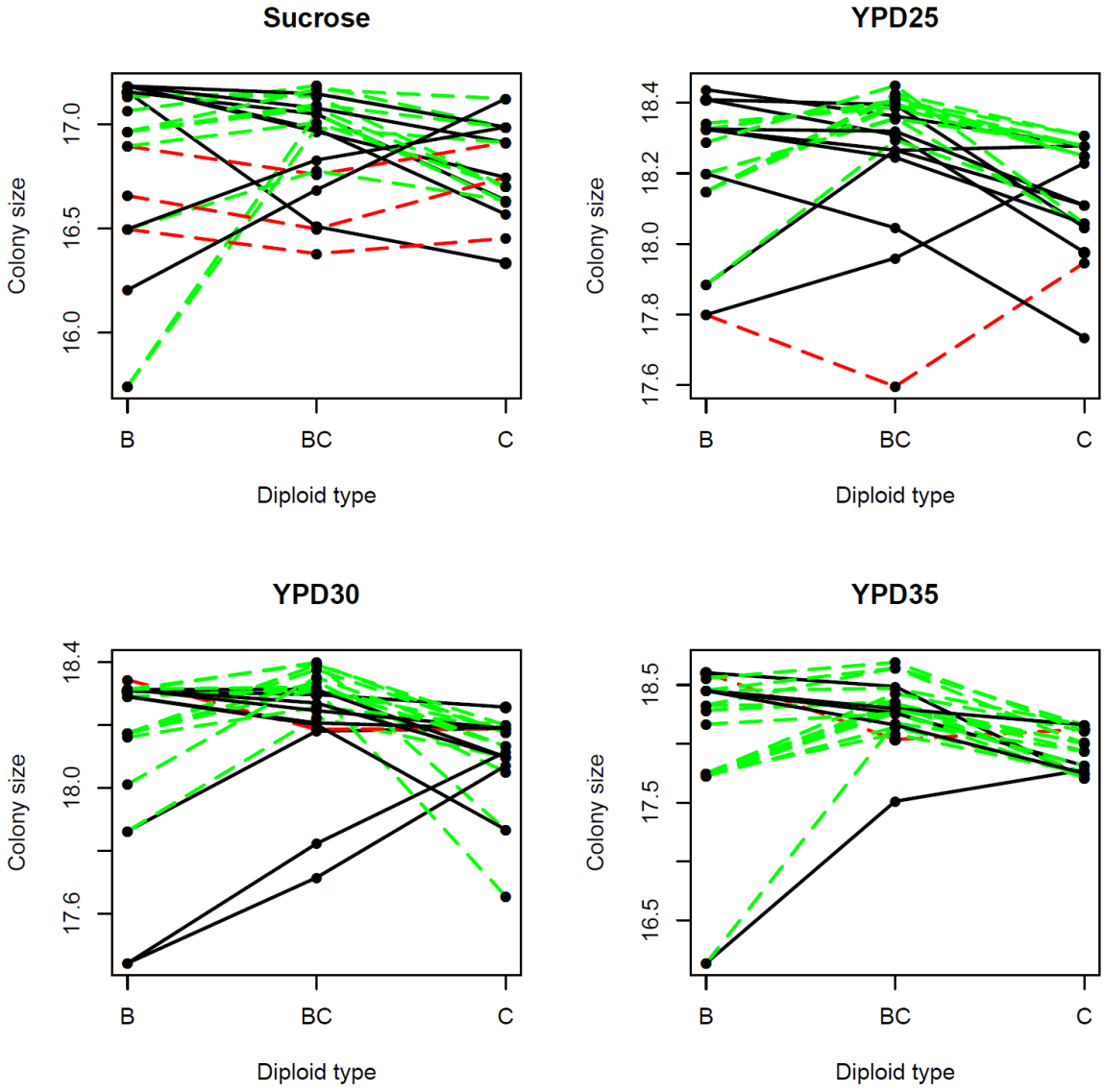
Table II.S1: List of strains used in this study

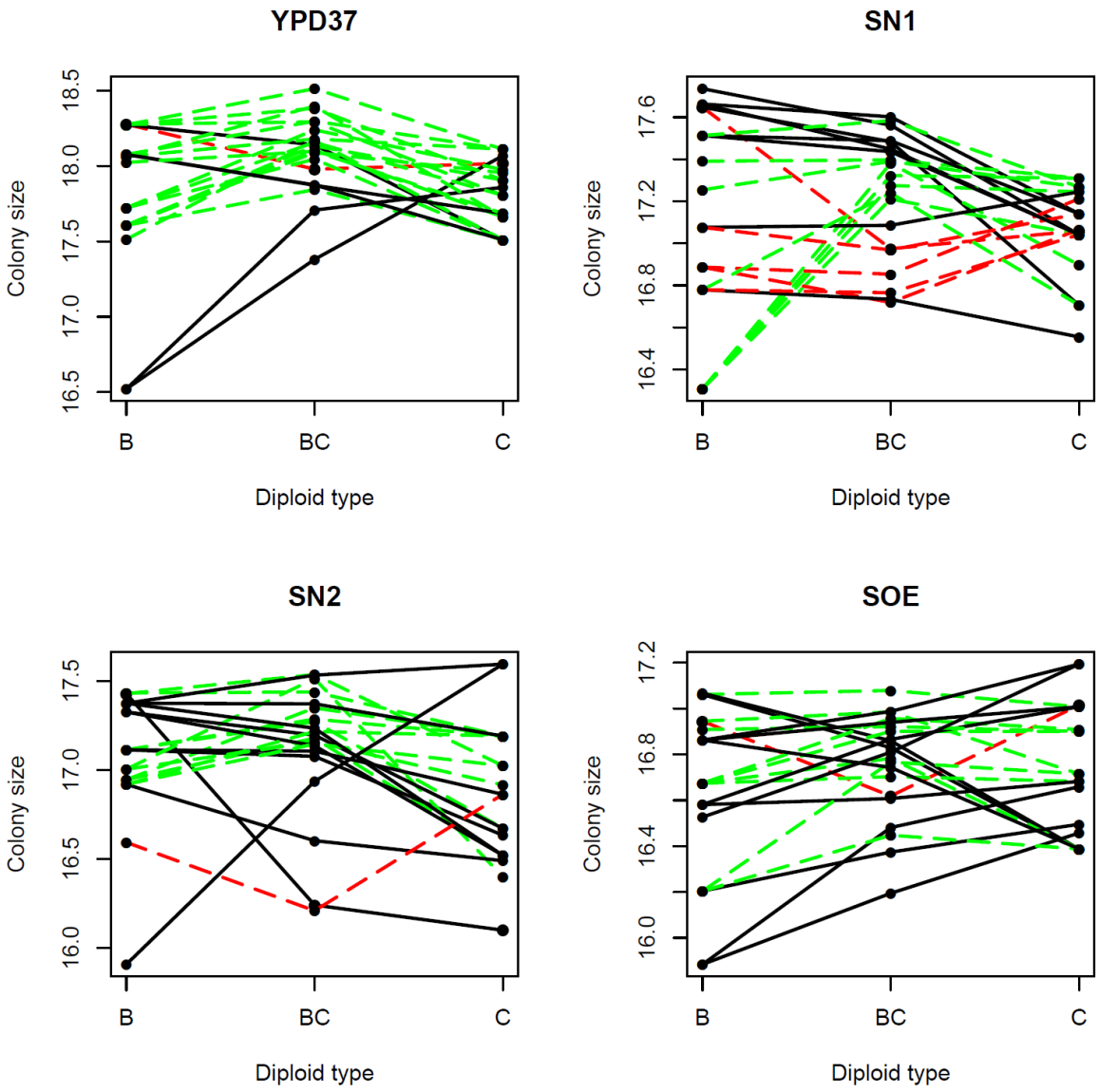
Strain	ID_Screen	Genotype	Reference
LL12_014	LL17_69_W	B	Charron *et al*. 2014a
LL12_021	PPC6_W	B	Charron *et al*. 2014a
LL12_028	VC2_W	B	Charron *et al*. 2014a
LL13_183	LL13_183_W	B	Leducq *et al*. 2016
MSH 604	GB23_W	B	Leducq *et al*. 2014
UWOPS 79-140	79_140_W	B	Leducq *et al*. 2014
UWOPS 91-202	91_202_W	B	Leducq *et al*. 2014
yHBJ001	yHBJ1_W	B	Sylvester *et al*. 2015
yHKS226	yHKS226_W	B	Sylvester *et al*. 2015
YPS484	YPS484_W	B	Kuehne *et al*. 2007
LL11_001	L9_W	C	Charron *et al*. 2014a
LL11_002	G63a_W	C	Charron *et al*. 2014a
LL11_004	G56_W	C	Charron *et al*. 2014a
LL11_007	IO168_W	C	Charron *et al*. 2014a
LL11_009	SQ107_W	C	Charron *et al*. 2014a
LL11_011	SQ130_W	C	Charron *et al*. 2014a
LL12_019	PPC3_W	C	Charron *et al*. 2014a
LL12_027	SH23_W	C	Charron *et al*. 2014a
MSH 587-1	GB20_W	C	Leducq *et al*. 2014
yHKS225	yHKS225_W	C	Sylvester *et al*. 2015
YPS667	YPS667_W	C	Kuehne *et al*. 2007
YPS484_x_LL2011_011	YPS484_x_SQ130	BC	Charron *et al*. 2014b
UWOPS_91-202_x_LL2011_011	91_202_x_SQ130	BC	Charron *et al*. 2014b
UWOPS_91-202_x_MSH-587-1	91_202_x_GB20	BC	Charron *et al*. 2014b
LL2011_002_x_UWOPS-79-140	G63a_x_79_140	BC	Charron *et al*. 2014b
YPS667_x_yHBJ1	YPS667_x_YHBJ1	BC	Charron *et al*. 2014b
LL2012_028_x_LL2011_004	VC2_x_G56	BC	Charron *et al*. 2014b
LL2012_028_x_LL2011_002	VC2_x_G63a	BC	Charron *et al*. 2014b
LL2012_028_x_MSH-587-1	VC2_x_GB20	BC	Charron *et al*. 2014b
LL2012_021_x_LL2011_011	PPC6_x_SQ130	BC	Charron *et al*. 2014b
YPS484_x_YPS667	YPS484_x_YPS667	BC	Charron *et al*. 2014b
YPS484_x_MSH-587-1	YPS484_x_GB20	BC	Charron *et al*. 2014b
UWOPS_79-140_x_LL2011_009	79_140_x_SQ107	BC	Charron *et al*. 2014b
LL13_183_x_MSH587-1	LL13_183_x_GB20	BC	Leducq *et al*. 2016
LL12_014_x_LL11_007	LL17_69_x_IO168	BC	Leducq *et al*. 2016
LL13_183_x_LL12_027	LL13_183_x_SH23	BC	Leducq *et al*. 2016
yHKS225_x_yHKS226	yHKS225_x_yHKS226	BC	Leducq *et al*. 2016
MSH587-1_x_MSH604	GB20_x_GB23	BC	Leducq *et al*. 2016
LL13_183_x_yHKS225	LL13_183_x_yHKS225	BC	Leducq *et al*. 2016
LL12_019_x_LL12_021	PPC3_x_PPC6	BC	Leducq *et al*. 2016
LL13_183_x_LL12_019	LL13_183_x_PPC3	BC	Leducq *et al*. 2016
LL12_014_x_LL11_001	LL17_69_x_L9	BC	Leducq *et al*. 2016
LL2011_009_x_LL2011_011	SQ107_x_SQ130	CC	Charron *et al*. 2014b
LL2011_004_x_LL2011_011	G56_x_SQ130	CC	Charron *et al*. 2014b
LL2011_001_x_MSH-587-1	L9_x_GB20	CC	Charron *et al*. 2014b

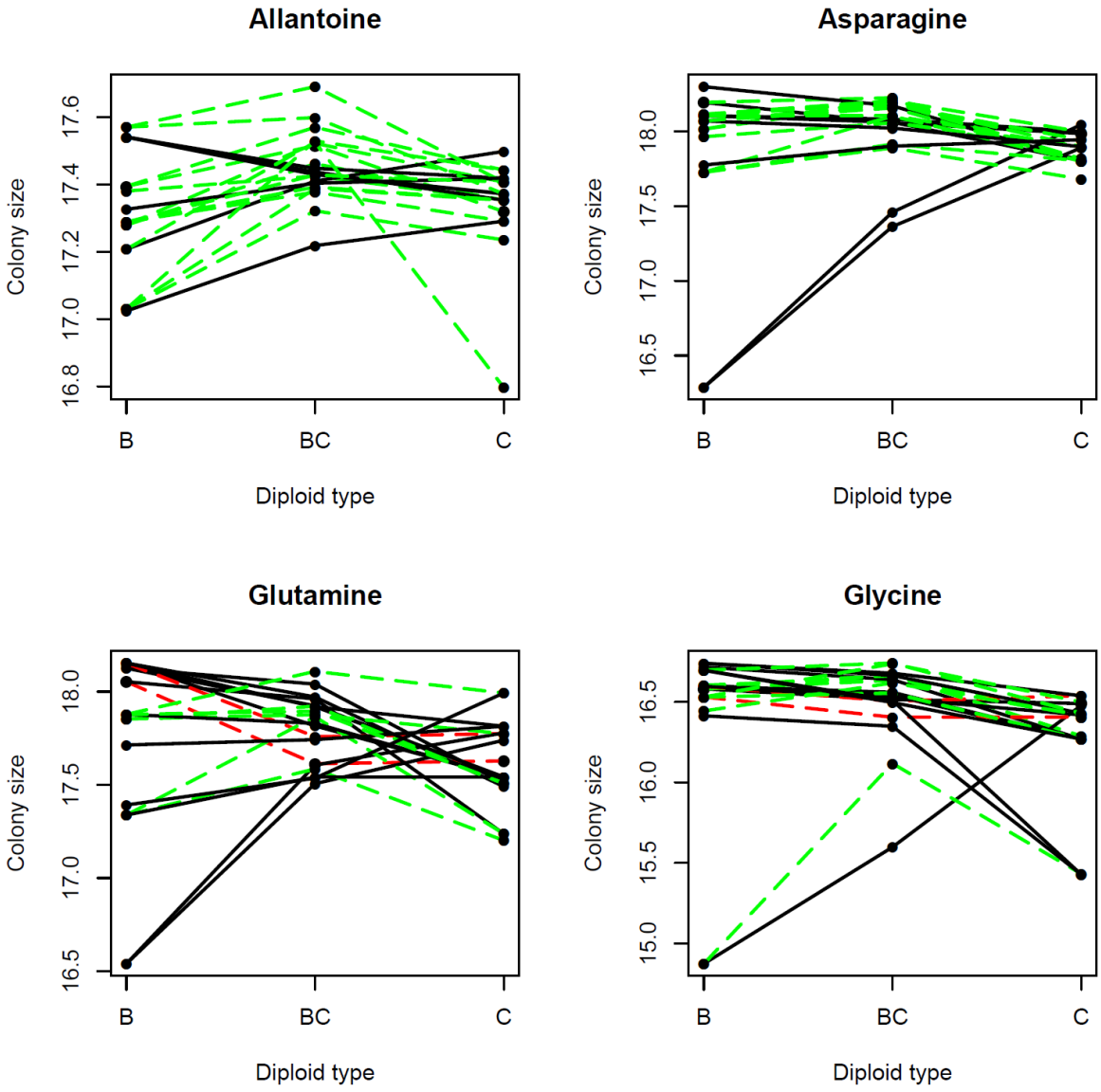
LL2011_001_x_LL2011_002	L9_x_G63a	CC	Charron *et al*. 2014b
LL2011_004_x_LL2011_001	G56_x_L9	CC	Charron *et al*. 2014b
YPS667_x_LL2011_009	YPS667_x_SQ107	CC	Charron *et al*. 2014b
MSH-587-1_x_YPS667	GB20_x_YPS667	CC	Charron *et al*. 2014b
LL12_014_x_LL12_028	LL17_69_x_VC2	BB	Leducq *et al*. 2016
LL12_014_x_yHBJ001	LL17_69_x_yHBJ1	BB	Leducq *et al*. 2016
LL12_014_x_LL13_183	LL17_69_x_LL13_183	BB	Leducq *et al*. 2016
LL12_014_x_MSH604	LL17_69_x_GB23	BB	Leducq *et al*. 2016
LL13_183_x_LL12_021	LL13_183_x_PPC6	BB	Leducq *et al*. 2016
LL12_014_x_yHKS226	LL17_69_x_yHKS226	BB	Leducq *et al*. 2016
YPS484_x_UWOPS_79-140	YPS484_x_79_140	BB	Charron *et al*. 2014b
UWOPS_91-202_x_yHBJ1	91_202_x_yHBJ1	BB	Charron *et al*. 2014b
UWOPS_91-202_x_LL2012_021	91_202_x_PPC6	BB	Charron *et al*. 2014b
UWOPS_79-140_x_LL2012_021	79_140_x_PPC6	BB	Charron *et al*. 2014b
UWOPS-91-202_x_LL2012_028	91_202_x_VC2	BB	Charron *et al*. 2014b
yHBJ1_x_LL2012_028	yHBJ1_x_VC2	BB	Charron *et al*. 2014b
YPS484_x_LL2012_021	YPS484_x_PPC6	BB	Charron *et al*. 2014b

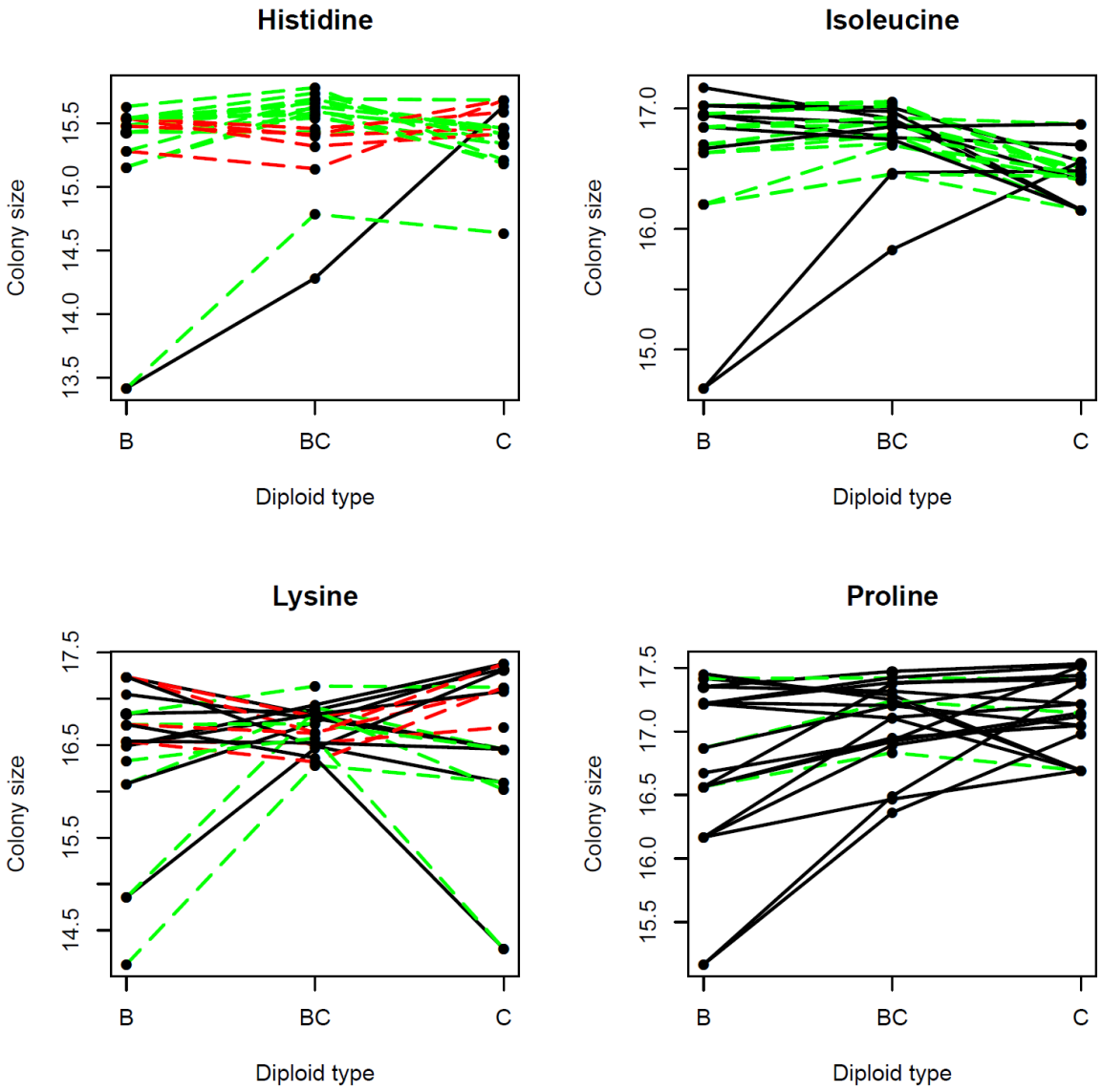

al_methyl_D_glucopyranoside
Colony size
16.2
16.6
17.0
B
BC
C
Diploid type
EtOH30
Colony size
16.6
17.0
17.4
17.8
B
BC
C
Diploid type
EtOH37
Colony size
15.5
16.0
16.5
17.0
17.5
B
BC
C
Diploid type
Fructose
Colony size
16.2
16.6
17.0
B
BC
C
Diploid type

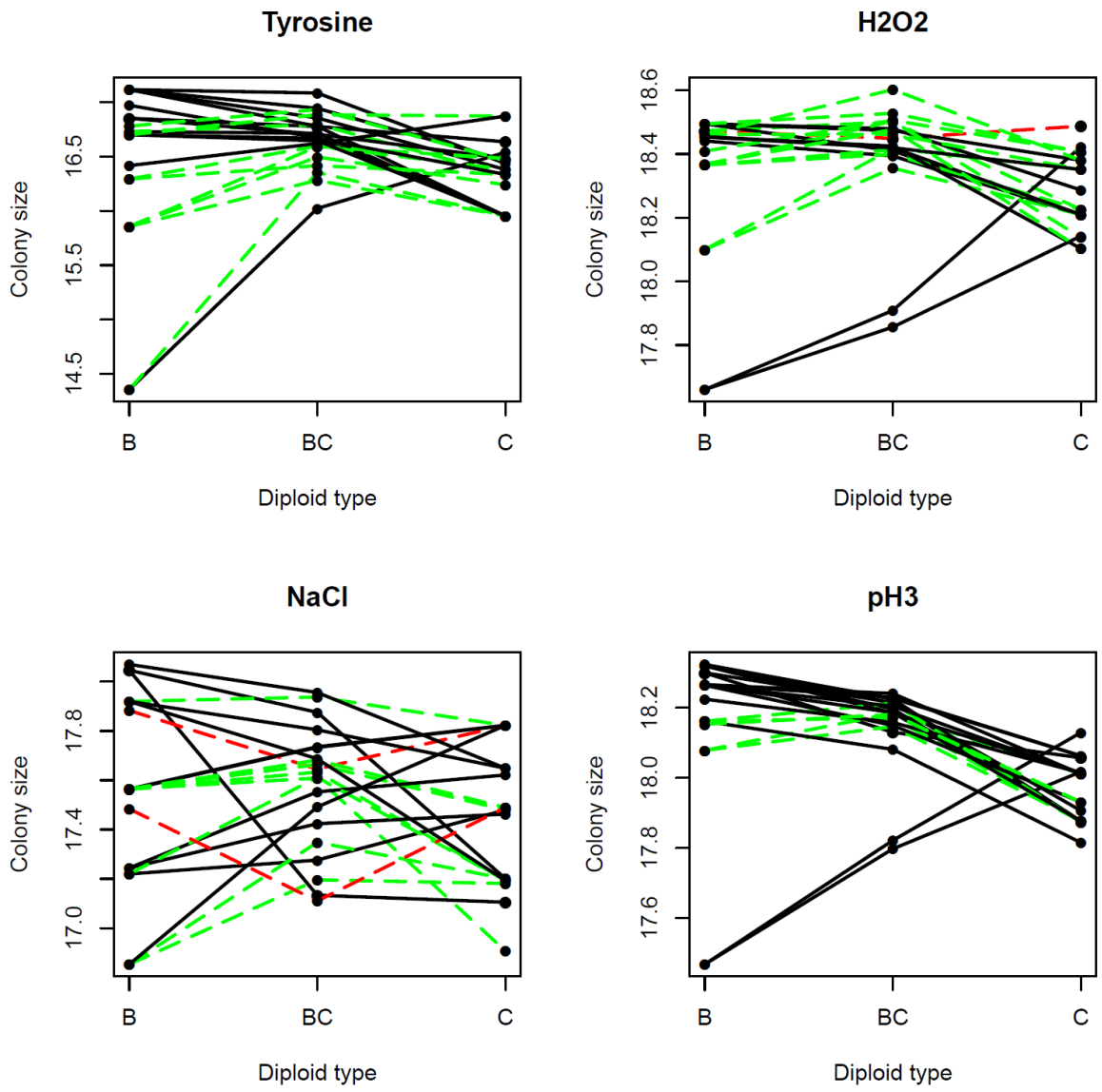

Galactose
Colony size
15.5
16.0
16.5
17.0
B
BC
C
Diploid type
Glucose
Colony size
16.2
16.6
17.0
B
BC
C
Diploid type
Maltose
Colony size
15.5
16.0
16.5
17.0
17.5
B
BC
C
Diploid type
Mannose
Colony size
16.2
16.6
17.0
17.4
B
BC
C
Diploid type


Sucrose
Colony size
16.0
16.5
17.0
B
BC
C
Diploid type
YPD25
Colony size
17.6
17.8
18.0
18.2
18.4
B
BC
C
Diploid type
YPD30
Colony size
17.6
18.0
18.4
B
BC
C
Diploid type
YPD35
Colony size
16.5
17.5
18.5
B
BC
C
Diploid type


YPD37
SN1
SN2
SOE
Colony size
Diploid type
B
BC
C
16.5
17.0
17.5
18.0
18.5
16.4
16.8
17.2
17.6
16.0


Allantoine
Colony size
16.8
17.0
17.2
17.4
17.6
B
BC
C
Diploid type
Asparagine
Colony size
16.5
17.0
17.5
18.0
B
BC
C
Diploid type
Glutamine
Colony size
16.5
17.0
17.5
18.0
B
BC
C
Diploid type
Glycine
Colony size
15.0
15.5
16.0
16.5
B
BC
C
Diploid type


Histidine
Colony size
13.5
14.0
14.5
15.0
15.5
B
BC
C
Diploid type
Isoleucine
Colony size
15.0
16.0
17.0
B
BC
C
Diploid type
Lysine
Colony size
14.5
15.5
16.5
17.5
B
BC
C
Diploid type
Proline
Colony size
15.5
16.0
16.5
17.0
17.5
B
BC
C
Diploid type


Tyrosine
H2O2
NaCl
pH3
Colony size
Diploid type
B
BC
C
14.5
15.5
16.5
17.8
18.0
18.2
18.4
18.6
17.0
17.4
17.6

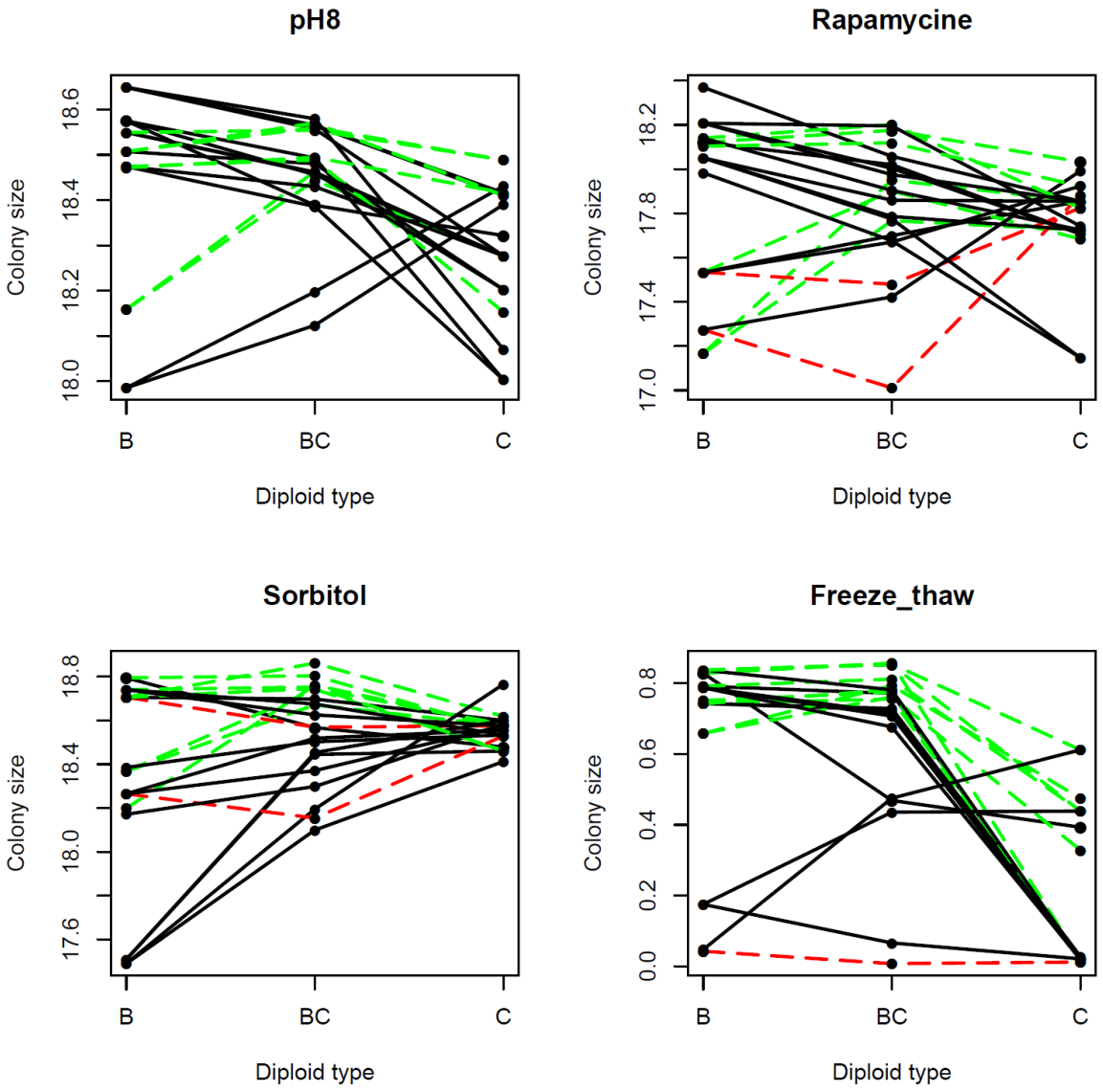


Figure II.S1: Representation of the mode of inheritance for each individual strain in each condition.

Each line represent a cross. A solid black line represents partial dominance of either the fitter or the less fit parent and colored dashed lines stands for either over-dominance (green) or under-dominance (red).

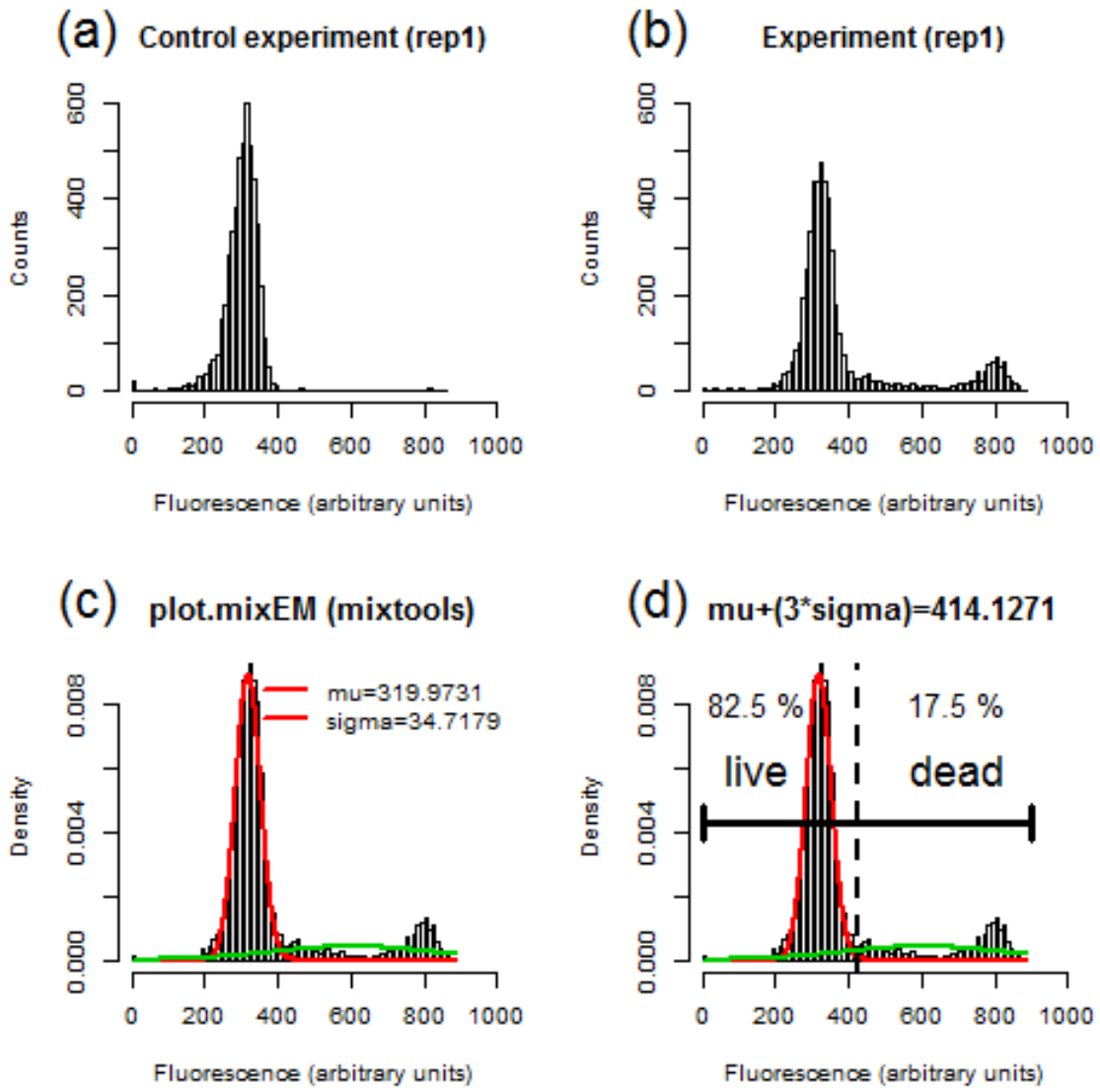


Figure II.S2: Evaluation of freeze-thaw stress resistance by flow cytometry.

(a) Distribution of fluorescence in non-treated cells establishes the fluorescence distribution for viable cells (b) Freeze-thaw experiment shows bimodal distribution. (c) mixtools is used to get the *mu* (mean) and *sigma* (standard deviation) parameters of the left distribution (viable cells). (d) Threshold for live/dead cells is established as *mu* + (3**sigma*) and the empirical cumulative distribution is calculated for the interval comprised between 0 and the threshold.

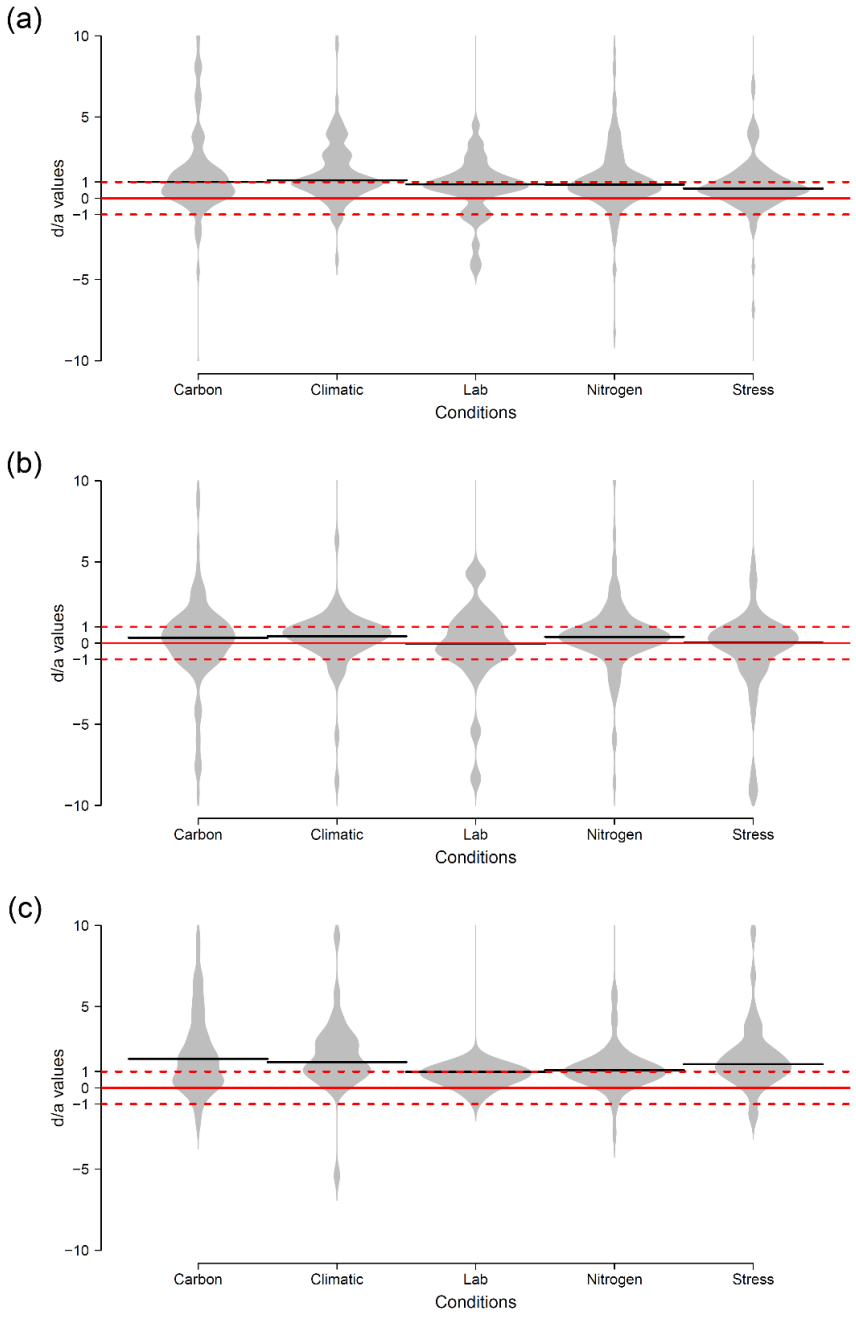


Figure II.S3: Distributions of d/a per condition type.
The d/a distribution between -10 and 10 for (a) *SpB*/*SpC*, (b) *SpB*/*SpB* and (c) *SpC*/*SpC* crosses. Dotted red lines indicate d/a values of -1 and 1, full red lines indicate d/a = 0 and the full black lines on each distribution indicate the median d/a value for this condition type.

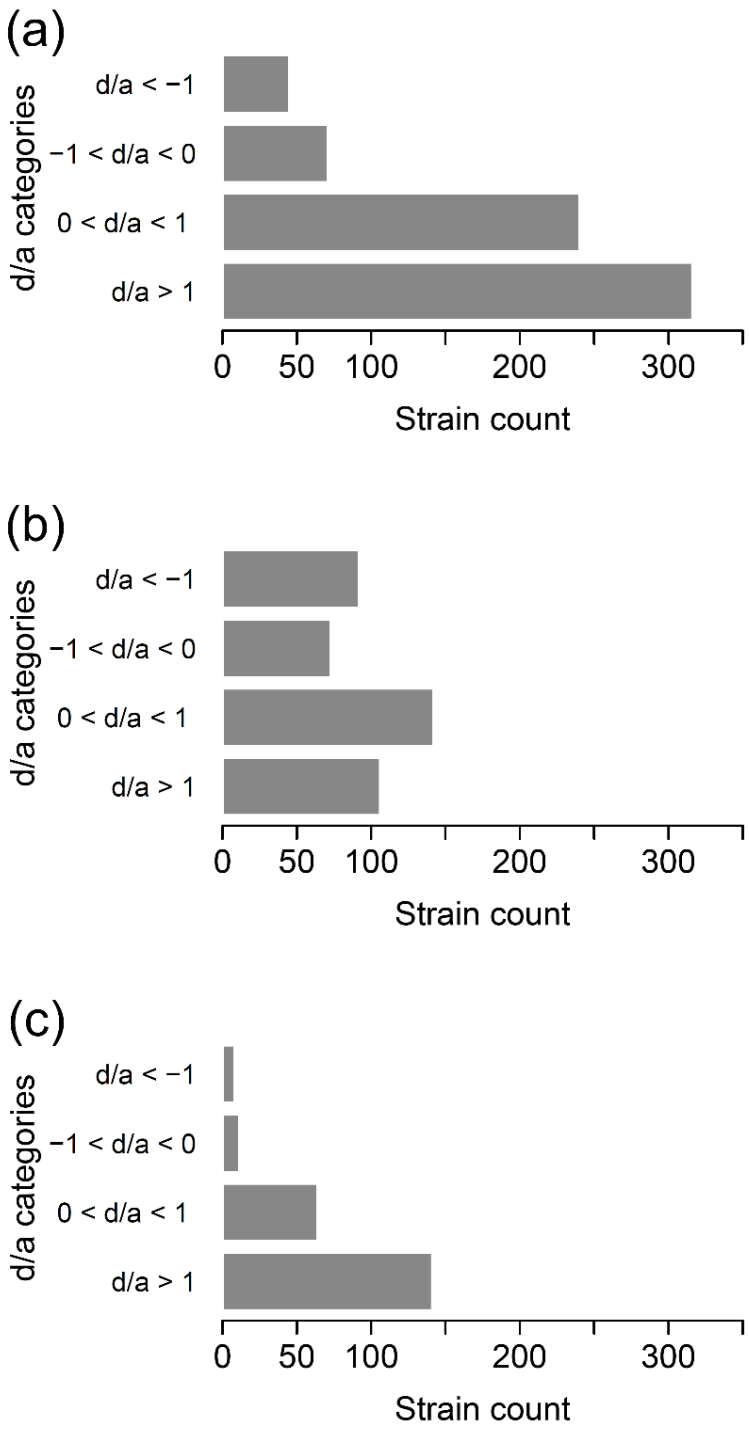


Figure II.S4: Counts of strains per mode of inheritance.
For (a) *SpB*/*SpC*, (b) *SpB*/*SpB* and (c) *SpC*/*SpC* crosses**.** One strain corresponds to one strain × condition combination.

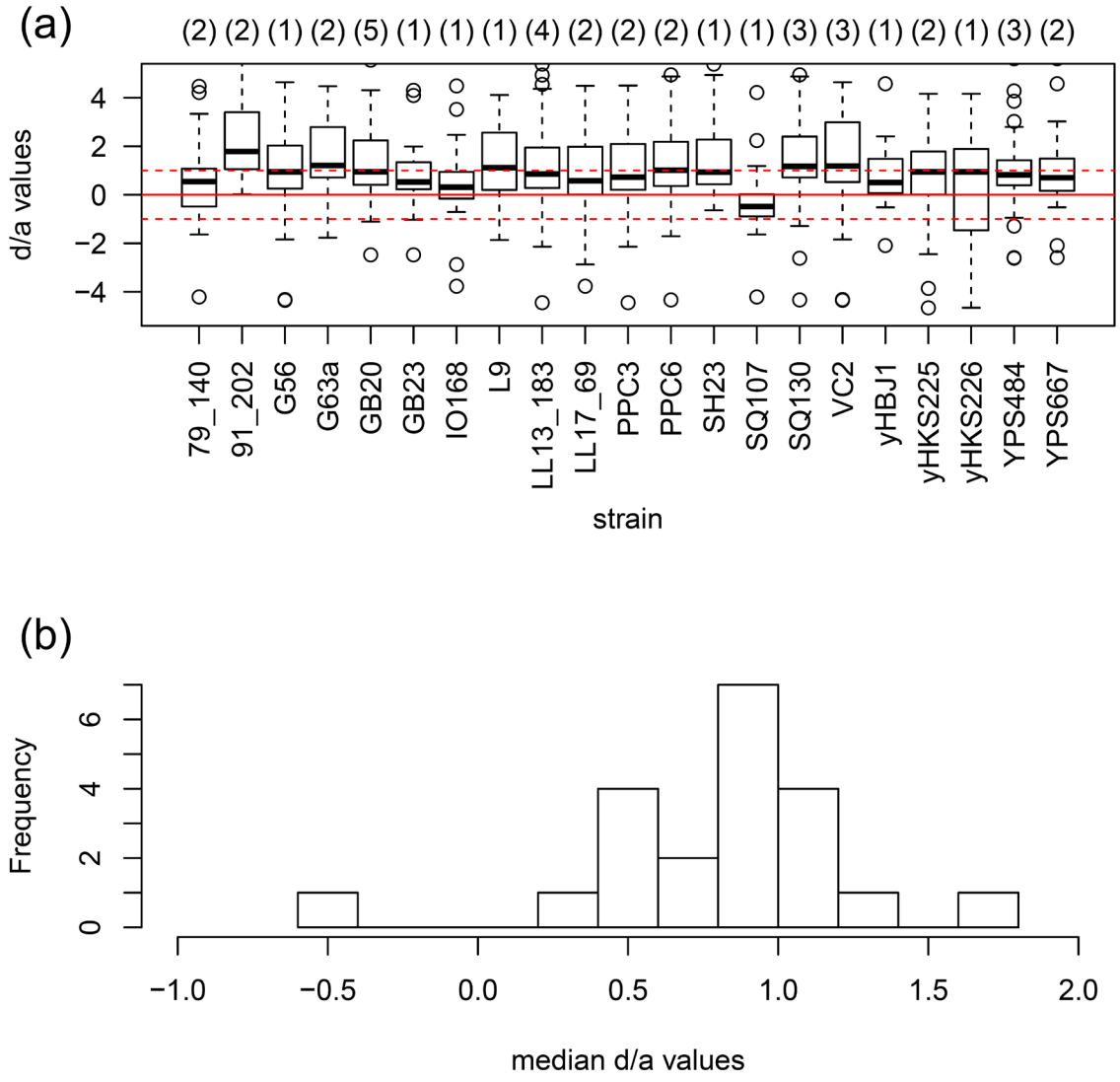


Figure II.S5: Strain effects on d/a values for *SpB*/*SpC* crosses.
(a) Boxplots of d/a values associated with the parental strains. Indicated in parenthesis above each plot is the number of crosses implicating each strain. (b) distribution of d/a median values for the parental strains.

ANNEXE C: Matériel et méthode supplémantaire pour le chapitre III

Supplementary Text

Mechanisms of whole genome duplication

To investigate the potential mechanisms of whole genome doubling that led to tetraploid hybrids, by endoreduplication or mating between two diploid hybrids, we sequenced the genome of all tetraploids at T_{ini} and T_{end}. The main way by which autotetraploidization could occur by mating in our study is to have two hybrids with damage to the opposite MAT loci in proximity to mate. Such tetraploid hybrids should thus have two MAT loci of opposite mating type with damaged function. By combining read coverage and allele frequency analysis, we investigated the presence of total or partial loss of chromosome III containing the MAT locus (Figures III.S19 and III.S20) or LOH events around the MAT locus (Figure III.S21). We identified two tetraploids that show aneuploidy on chromosome III at T_{end}. One of the L1 tetraploids has one more copy of chromosome III from the *SpC* parent and the H2 tetraploid has lost one copy from the *S. cerevisiae* parent. In both cases, aneuploidy affects only one copy of the mating type. Considering the fact that total or partial chromosome III loss should affect the two mating type loci in the tetraploid hybrid, these results show that the loss of chromosome III is not the molecular mechanism leading to whole genome doubling. Allele frequency analysis confirms these aneuploidies and show that there is no LOH events around the MAT locus in all tetraploids.

We also examined the copy number of MATa and MATα sequences from each parent (Figure III.S22). For an accurate analysis, we had to take in account the silent mating type copies. Indeed, in addition to the MAT locus, *Saccharomyces* yeasts carry two unexpressed, but complete, copies of mating-type genes at the silent loci, *HML* and *HMR* localised also on chromosome III (Figure III.S21)(Klar *et coll.*

1982). The *HML* locus carries MATα sequence while *HMR* carries MATa sequence. Thus, knowing that the *SpB* parent is MATa and the *SpC*, *SpA* and *S. cerevisiae* parents are MATα, each hybrid carries both mating type copies from each parent with a different copy number of MATa or MATα corresponding to each parental origin.

Copy number and allele frequency variations observed in MATa/MATα copies from each parent (Figure III.S22) are due either to the aneuploidies mentioned above (L1_51) (Figure III.S19) or to an LOH in the *HML* locus (L2_36) (Figure III.S21) or a combined effect of aneuploidy and LOH in the *HML* locus (H2_38) (Supplementary Figures III.S19, III.S21 and III.S22) which could not have led to mating.

Our results show no evidence of damages in the MAT locus in tetraploid hybrids that could have caused mating between diploid hybrids. The cases of aneuploidies or LOH observed affect only one parental copy of the MAT locus or the silent *HML* locus. Also, there is no evidence for damaging mutations occurring in the MAT loci. Thus, according to these results, autodiploidization is the most likely mechanism of whole genome doubling

Return to diploidy after tetraploidy

These analyses allowed us to examine the genome of the M1 tetraploid (M1_40) whose ploidy and fertility increased at T_{mid} and then decreased at T_{end}. This result could be explained by the presence of heterogenous colonies during our evolution experiment. The isolated colony at T_{mid} might segregating diploid and tetraploid cells. One type may have fixed in the glycerol stock and be lost in the next round of propagation, explaining why we observe only tetraploids at T_{mid}. However, the colony isolated during the subsequent passage was a diploid one explaining why we observe only diploids in subsequent passages.

Besides, GBS data are consistent with ploidy and fertility data for M1_40 line, however genome sequencing show a triploid state at T_{end} instead of diploid. The detection of chromosome I loss of the SpA parent copy at T_{ini} and T_{end} (Figure III.S19) confirms that both sequenced isolated colonies correspond to the same M1_40 line at two different times. The same variations are observed for the M2_36 line at T_{ini} that show a different ploidy state between Ploidy and GBS data (2n) and sequencing data (4n, with many aneuploidies) while the same LOH is observed in chromosome VIII at T_{ini} and T_{end} (Figure III.S15). These results suggest that there is still heterogeneity among isolated colonies from the glycerol stock probably due to genomic instability of hybrids.

Table III.S1: List of strains used in this study

Strain	Sampling site	Lineage	Mating type	Genotype	Reference Wild	Reference hoΔ	Reference ade2Δ
YPS644	Buck Hill Falls, PE	*SpA*	α	hoΔ::KAN, ade2Δ::HPH	Kuehne *et al*. 2007	Charron *et al*. 2014	This study
YPS644	Buck Hill Falls, PE	*SpA*	a	hoΔ::KAN, ade2Δ::HPH	Kuehne *et al*. 2007	Charron *et al*. 2014	This study
YPS744	Tuscarora Forest, PE	*SpA*	α	hoΔ::NAT, ade2Δ::HPH	Kuehne *et al*. 2007	Charron *et al*. 2014	This study
YPS744	Tuscarora Forest, PE	*SpA*	a	hoΔ::NAT, ade2Δ::HPH	Kuehne *et al*. 2007	Charron *et al*. 2014	This study
UWOPS_91_202	Long Point, ON	*SpB*	a	hoΔ::KAN, ade2Δ::HPH	Kuehne *et al*. 2007	Charron *et al*. 2014	This study
UWOPS_91_202	Long Point, ON	*SpB*	a	hoΔ::NAT	Kuehne *et al*. 2007	Charron *et al*. 2014	-
MSH604	Mont-St-Hilaire, QC	*SpB*	a	hoΔ::NAT, ade2Δ::HPH	Leducq *et al*. 2014	Leducq *et al*. 2016	This study
LL12_021	Pointe Platon, QC	*SpB*	α	hoΔ::NAT, ade2Δ::HPH	Leducq *et al*. 2014	Leducq *et al*. 2016	This study
LL12_028	Sherbrooke, QC	*SpB*	α	hoΔ::KAN, ade2Δ::HPH	Leducq *et al*. 2014	Leducq *et al*. 2016	This study
YPS484	Grand Bend, ON	*SpB*	a	hoΔ::HPH	Kuehne et al. 2007	Charron *et al*. 2014	-
yHKS226	Plainfield, NH	*SpB*	α	hoΔ::HPH	Sylvester *et al*. 2015	Leducq *et al*. 2016	-
LL11_004	Cap Chat, QC	*SpC*	α	hoΔ::KAN, ade2Δ::HPH	Leducq *et al*. 2014	Charron *et al*. 2014	This study
LL11_009	St-Michel-du-Squatec, QC	*SpC*	α	hoΔ::NAT, ade2Δ::HPH	Leducq *et al*. 2014	Charron *et al*. 2014	This study
LL11_009	St-Michel-du-Squatec, QC	*SpC*	a	hoΔ::NAT, ade2Δ::HPH	Leducq *et al*. 2014	Charron *et al*. 2014	This study
MSH_587-1	Mont-St-Hilaire, QC	*SpC*	α	hoΔ::HPH	Leducq et al. 2014	Charron et al. 2014	-
YPS667	Buck Hill Falls, PE	*SpC*	α	hoΔ::HPH	Kuehne et al. 2007	Charron et al. 2014	-
yHKS225	Plainfield, NH	*SpC*	a	hoΔ::NAT	Sylvester et al. 2015	Leducq et al. 2016	-
LL12_019	Pointe Platon, QC	*SpC*	a	hoΔ::HPH	Leducq et al. 2014	Leducq et al. 2016	-
LL13_040	Rockport, MA	*Scer*	α	hoΔ::KAN, ade2Δ::HPH	This study	This study	This study
LL13_040	Rockport, MA	*Scer*	a	hoΔ::KAN, ade2Δ::HPH	This study	This study	This study
LL13_054	Rockport, MA	*Scer*	α	hoΔ::NAT, ade2Δ::HPH	This study	This study	This study
LL13_054	Rockport, MA	*Scer*	a	hoΔ::NAT, ade2Δ::HPH	This study	This study	This study

Table III.S2: Genetic divergence among the different S. paradoxus lineages and among different strains of the same lineage (Eberlein et al., 2019).

Reference	Compared to	% Mean nucleotide divergence
SpB	*SpB*	0.424
SpB	*SpC*	2.237
SpB	*SpA*	3.748
SpC	*SpC*	0.288
SpC	*SpA*	3.432
SpA	*SpA*	0.101

Table III.S3: List of crosses made for this study.

Cross	a strain	a strain resistance	α strain	α strain resistance
$VL_{div}1$	MSH604	NAT	LL12_028	G418
$L_{div}1$	MSH604	NAT	LL11_004	G418
$M_{div}1$	MSH604	NAT	YPS644	G418
$H_{div}1$	MSH604	NAT	LL13_040	G418
$VL_{div}2$	UWOPS_91_202	G418	LL12_021	NAT
$L_{div}2$	UWOPS_91_202	G418	LL11_009	NAT
$M_{div}2$	UWOPS_91_202	G418	YPS744	NAT
$H_{div}2$	UWOPS_91_202	G418	LL13_054	NAT
$L_{div}3$	UWOPS_91_202	NAT	MSH587	HYG
$L_{div}4$	YPS484	NAT	YPS667	HYG
$L_{div}5$	yHKS226	HYG	yHKS225	NAT
$L_{div}6$	LL12_021	NAT	LL12_019	HYG
SpC×*SpA*-1	YPS744	NAT	LL11_004	G418
SpC×*SpA*-2	YPS644	G418	LL11_009	NAT
SpC×*SpC*-1	LL11_009	NAT	LL11_004	G418
SpC×*Scer*-1	LL13_054	NAT	LL11_004	G418
SpC×*Scer*-2	LL13_040	G418	LL11_009	NAT

Table III.S4: Oligonucleotides used in this study.

Primer name	Sequence	Description
CLOP97-F1	acaatcaaggaatcaagaaaccgtggtaaaaaattcaagtCAGCTGAAGCTTCGTACGC	Forward primer to construct the deletion cassette for the ADE2 gene in *SpA*
CLOP97-F2	gtaattgttcgctggccaagtatatcaatacatttatataGCATAGGCCACTAGTGGATC	Reverse primer to construct the deletion cassette for the ADE2 gene in *SpA*
CLOP97-F3	acaatcaaggaatcaagaaaccgtgataaaaaattcaagtCAGCTGAAGCTTCGTACGC	Forward primer to construct the deletion cassette for the ADE2 gene in *SpB*
CLOP97-F4	tgtgattgttgctggccaagtacatcaatacatttatacaGCATAGGCCACTAGTGGATC	Reverse primer to construct the deletion cassette for the ADE2 gene in *SpB*
CLOP97-F5	acaattaaggaatcaagaaaccgtgataaaaaattcaagtCAGCTGAAGCTTCGTACGC	Forward primer to construct the deletion cassette for the ADE2 gene in *SpC*
CLOP97-F6	gtaattgttcgctggccaagtatattaatacatttatataGCATAGGCCACTAGTGGATC	Reverse primer to construct the deletion cassette for the ADE2 gene in *SpC*
CLOP97-F7	caatcaagaaaaacaagaaaatcggacaaaacaatcaagtCAGCTGAAGCTTCGTACGC	Forward primer to construct the deletion cassette for the ADE2 gene in *Scer*
CLOP97-F8	ataattatttgctgtgcaagtatatcaataaacttatataGCATAGGCCACTAGTGGATC	Reverse primer to construct the deletion cassette for the ADE2 gene in *Scer*
CLOP97-F9	GGACCAGATGGACGATATTA	Forward primer for the verification of the ADE2 gene deletion in *S. paradoxus*
CLOP97-F10	CGGTAATGCATTGAGCAGAT	Reverse primer for the verification of the ADE2 gene deletion in *S. paradoxus*
CLOP97-F11	GACTCTTGTTGCATGGCTAC	Forward primer for the verification of the ADE2 gene deletion in *S. cerevisiae*
CLOP97-F12	CGCAGACTTAAGCAGGTAAT	Reverse primer for the verification of the ADE2 gene deletion in *S. cerevisiae*
verifMATa_F	AGTCACATCAAGATCGTTTATGG	Forward primer for mating type diagnostic (a specific)
verifMATalpha_F	ACTCCACTTCAAGTAAGAGTTTG	Forward primer for mating type diagnostic (alpha specific)
verifMATa/alpha_R	GCACGGAATATGGGACTACTTCG	Reverse primer for mating type diagnostic
CLOP80-A3	TGAGGTCCCGCATGAATGAC	Forward primer for rnl diagnostic
CLOP80-B3	ACGTACTTGTTTCACTCGTTTGT	Reverse primer for rnl diagnostic
CLOP80-C3	GGTTCAAGATGATTAATTTCACAAG	Forward primer for atp6 diagnostic
CLOP80-D3	ACCAGCAGGTACGAATAATGA	Reverse primer for atp6 diagnostic

Table III.S5: Number of generations equivalence for the different timepoints sampled.

Last Glycerol	T_{ini} (generation number)	T_{mid} (generation number)	T_{end} (generation number)
P4	P0	NA	P4 (88)
P7	P0	P4 (88)	P7 (176)
P10	P0	P4 (88)	P10 (242)
P13	P0	P7 (176)	P13 (308)
P16	P0	P7 (176)	P16 (352)
P19	P0	P10 (242)	P19 (440)
P22	P0	P10 (242)	P22 (506)
P25	P0	P13 (308)	P25 (572)
P28	P0	P13 (308)	P28 (638)
P31	P0	P16 (352)	P31 (704)
P35	P0	P16 (352)	P35 (770)

Table III.S6: Logrank test P-values for all survival curves

Cross type

	M	VL	L
VL	0.010114	NA	NA
L	0.000305	2.49E-06	NA
H	0.000786	3.45E-06	0.667319

Blocks

	set1(1-64)	set1(65-96)	set2(1-64)	set2(65-96)	VL1
set1(65-96)	0.075385	NA	NA	NA	NA
set2(1-64)	0.018042	0.907645	NA	NA	NA
set2(65-96)	0.147398	0.768516	0.689276	NA	NA
VL1	0.003756	0.042042	0.042042	0.025266	NA
VL2	0.00077	0.009335	0.009335	0.006103	0.388572

Groups

	LH
VLM	5.22E-09

Table III.S7: P-values of the Fisher's exact test on the proportions of viable spores for the 23 lines that were significantly different between T_{ini} and T_{end} (correction with FDR).

Lines	T_{ini}-T_{mid}	T_{ini}-T_{mid} corrected	T_{mid}-T_{end}	T_{mid}-T_{end} corrected	T_{ini}-T_{end}	T_{ini}-T_{end} corrected
A30	0.0002	0.0033	0.4702	0.8408	0.0047	0.0360
A31	0.0371	0.1720	0.3865	0.7330	0.0020	0.0189
A41	0.1155	0.3965	0.0736	0.2872	0.0004	0.0055
A51	1.28E-11	4.40E-10	0.1975	0.5164	5.68E-16	2.92E-14
A63	0.1275	0.4184	0.2223	0.5612	0.0038	0.0304
A87	0.0699	0.2789	7.25E-17	4.67E-15	6.99E-11	2.25E-09
D10	0.0032	0.0266	1	1	0.0032	0.0266
D22	0.0107	0.0693	0.3914	0.7330	0.0003	0.0044
D36	1.71E-23	1.76E-21	0.4973	0.8655	9.48E-27	1.22E-24
D42	0.0195	0.1128	0.3480	0.6920	0.0006	0.0074
D45	9.82E-11	2.97E-09	0.4975	0.8655	1.54E-08	4.17E-07
D89	0.0006	0.0071	1	1	2.60E-04	0.004
B40	1.49E-22	1.28E-20	5.91E-28	1.01E-25	0.2214	0.5612
B43	0.0371	0.1720	0.0087	0.0584	1.48E-06	3.30E-05
B49	3.34E-13	1.32E-11	0.0008	0.0086	7.66E-05	0.0013
B75	2.55E-07	5.97E-06	1	1	2.55E-07	5.97E-06
E86	0.5711	0.9278	9.90E-04	0.0102	5.83E-05	0.0010
E95	0.0561	0.2387	1.46E-03	0.0142	2.13E-07	5.49E-06
F38	7.46E-48	3.84E-45	2.27E-05	0.0004	1.19E-28	3.06E-26
H6	0.0973	0.3631	1.41E-14	6.07E-13	1.01E-20	7.40E-19
H25	1	1	5.60E-16	2.92E-14	1.85E-15	8.67E-14
I1	0.8852	1	0.0126	0.0773	0.0054	0.0411
I43	0.0031	0.0266	0.3206	0.6527	4.39E-05	7.80E-04
I45	4.21E-06	9.04E-05	0.0304	0.1468	1.19E-11	4.37E-10

Table III.S8: Ploidy distribution of 94 independent hybrids in three replicates of the $L_{div}1$ and $L_{div}2$ crosses.

	Ploidy		
Cross	2n (%)	3n (%)	4n (%)
$L_{div}1$-1*	48.0	52.0	0.0
$L_{div}1$-2	98.0	2.0	0.0
$L_{div}1$-3	67.0	32.0	1.0
$L_{div}2$-1*	42.5	57.5	0.0
$L_{div}2$-2	91.5	8.5	0.0
$L_{div}2$-3	99.0	1.0	0.0