UNIVERSITÉ LAVAL

Saccharomyces paradoxus en tant que système d'étude de la spéciation dans les populations naturelles

Mémoire

Guillaume Charron

Maîtrise en biologie
Maître ès sciences (M.Sc.)

Québec, Canada

RÉSUMÉ

La spéciation implique l'établissement de barrières à la reproduction entre les espèces en formation. Bien que leurs bases moléculaires soient identifiées, l'émergence de ces barrières dans le temps et la manière dont elles se propagent à l'intérieur des populations reste peu connue. Les objectifs de cette étude sont i) de déterminer la présence de *Saccharomyces paradoxus* au Québec ii) de déterminer le niveau d'isolement reproducteur post-zygotique entre les lignées américaines de *S. paradoxus* et iii) de le corréler avec la divergence génétique et les réarrangements chromosomiques. Nous avons échantillonné divers substrats au Québec afin d'isoler *S. paradoxus*. L'identification des levures au niveau de l'espèce s'est fait à l'aide des séquences ITS. En laboratoire nous avons vérifié le caryotype des souches de *S. paradoxus* par PFGE, procédé à des croisements entre les lignées et estimé l'isolement reproducteur à l'aide de la survie des spores hybrides. Notre étude confirme la présence des lignées américaines de *S. paradoxus* au Québec. Un isolement reproducteur est présent entre et à l'intérieur de ces lignées. La variation au niveau des profils chromosomiques à l'intérieur des lignées est corrélée avec l'isolement reproducteur. Nos résultats suggèrent que la variation au niveau des réarrangements chromosomiques ségrégerait au sein des populations et jouerait un rôle dans l'initiation de l'isolement reproducteur.

ABSTRACT

Speciation involves the emergence of reproductive barriers between the incipient species. Although the molecular bases of these barriers have been identified, we lack knowledge about the timing of their emergence and their propagation through populations. The objectives of this project are i) to look for the presence of *S. paradoxus* in the province of Québec, ii) to determine the level of post-zygotic reproductive isolation between the American lineages of *S. paradoxus* and iii) to correlate this reproductive isolation with genetic divergence and chromosomal rearrangements. We sampled different substrates through the Province of Québec to isolate *S. paradoxus*. Yeasts were identified to the species level using ITS sequences. We used a PFGE method to obtain the karyotypes from the S. *paradoxus* yeast, made experimental crosses and estimated reproductive isolation by counting surviving hybrid spores. Our study confirms the presence of the two *S. paradoxus* lineages in the province of Québec. We also find that reproductive isolation is not only present between strains of the two lineages but also between strains from the same lineage. A great intra-lineage variation in the chromosomal profiles is correlated with reproductive isolation. Our results suggest that variation of chromosomal rearrangements could segregate between populations and could be involved in the initiation of reproductive isolation.

TABLE DES MATIÈRES

RÉSUMÉ III
ABSTRACT V
TABLE DES MATIÈRES VII
LISTE DES FIGURES IX
LISTE DES ANNEXES XI
REMERCIEMENTS XIII
AVANT-PROPOS XV
INTRODUCTION GÉNÉRALE 1
SPÉCIATION 1
ISOLEMENT REPRODUCTEUR 2
ISOLEMENT REPRODUCTEUR POST-ZYGOTIQUE 4
ISOLEMENT REPRODUCTEUR POST-ZYGOTIQUE INTRINSÈQUE 4
RÉARRANGEMENTS CHROMOSOMIQUES 5
CHANGEMENTS AU NIVEAU DE LA PLOÏDIE 5
INCOMPATIBILITÉS GÉNÉTIQUES 6
INCOMPATIBILITÉ ENTRE GÉNOME NUCLÉAIRE ET ENDOSYMBIOTE 7
LES LEVURES DU GENRE SACCHAROMYCES 8
SACCHAROMYCES PARADOXUS 9
DÉMOGRAPHIE DES POPULATIONS DE S. PARADOXUS 10
DIVERGENCE GÉNÉTIQUE ET PHÉNOTYPIQUE CHEZ S. PARADOXUS 11
OBJECTIF 11
CHAPITRE 1: EXPLORING THE NORTHERN LIMIT OF THE DISTRIBUTION OF *SACCHAROMYCES CEREVISIAE* AND *SACCHAROMYCES PARADOXUS* IN NORTH AMERICA 13
CHAPITRE 2: CHROMOSOMAL VARIATION SEGREGATES WITHIN INCIPIENT SPECIES AND CORRELATES WITH REPRODUCTIVE ISOLATION 29
CONCLUSION GÉNÉRALE 49
DISTRIBUTION DES LEVURES SACCHAROMYCES AU QUÉBEC 50
ANALYSE DE L'ISOLEMENT REPRODUCTEUR CHEZ S. PARADOXUS 50
VARIATION DANS LA VIABILITÉ INTRAGROUPE 51
PERSPECTIVES 52
BIBLIOGRAPHIE 55
ANNEXES 69

LISTE DES FIGURES

Figure 1: Map of the 29 sites sampled in this study.....26
Figure 2: Effect of temperature on yeast isolation success. 27
Figure 3: Season-dependent isolation success. 28
Figure 4: Two diverging lineages of *S. paradoxus* are partially sympatric in North America.....44
Figure 5: Reproductive isolation and karyotyping in *S. paradoxus* North-American strains.....45
Figure 6: Spore survival (*S*) among 25 *S. paradoxus* North-American strains correlates with variation among and within lineages in genetic divergence (*Dg*) and chromosome profiles (*Dc*). 46
Figure 7: Proposed scenario for the divergence and secondary contact of the North American lineages of *Saccharomyces paradoxus*.....47

LISTE DES ANNEXES

Table S1: List of sampling sites. 69
Table S2: ANOVA analysis of yeast isolation success 70
Table S3: List of strains isolated in this study....71
Table S4: Uncharacterized yeasts isolated in this study. 79
Table S5: List of yeast strains used in this study. 81
Table S6: List of oligonucleotides and plasmids used for *HO* locus deletion. 83
Table S7: List of crosses realized in this study....84
Table S8: Results from linear model (*lm*) of spore survival (*S*) variation in function of genetic divergence (*Dg*) and karyotypic divergence (*Dc*) and their interaction (*Dg*Dc*) as additional factors (*lm Formula = S ~ dg+dc+ dg*dc*). 87

Figure S1: Principle of permutation tests used in this study to control for the effects of non-independence of pairwise comparisons and of bias in strain usage in pairwise comparisons. 88
Figure S2: Violin plots showing the expected distribution (grey lines) of *t*-values from Welch t-tests comparisons of *S* values among categories of crosses (see Figure 3a). ...90
Figure S3: Spore survival (*S*) is negatively correlated with genetic divergence (*Dg*) and chromosomal divergence (*Dc*)....91
Figure S4: Violin plots showing the expected distribution (grey lines) of estimated coefficients from linear models (see Table S4). 93

Appendix 1: scripts used to generate figures from data....94
Appendix 2: Dataset used to generate figures. 116

REMERCIEMENTS

Ça y est! Nous y sommes! C'est ici que s'achève mon cheminement à la maitrise. Je suis moi-même le premier surpris d'y être rendu, il me semble que je n'ai pas vu le temps passer. Si le temps a passé si vite ce n'est surtout pas parce que je n'ai rien fait, mais bien parce que j'ai eu du plaisir à travailler sur mon projet avec les gens qui m'ont entourés.

Tout d'abord, je voudrais remercier mon directeur Christian Landry, qui a tenté par tous les moyens possibles de faire de moi un meilleur étudiant ainsi qu'une meilleure personne. Je ne saurais dire si ça a fonctionné, mais en deux ans il m'a donné plusieurs opportunités de participer à des publications scientifiques ainsi que de présenter mes résultats dans de prestigieux congrès. Toutes ces expériences m'ont beaucoup appris sur moi-même, autant mes forces que mes faiblesses, et je voulais prendre le temps de remercier Christian pour tout.

Tout aussi important à mes yeux, j'aimerais remercier Jean-Baptiste Leducq, que j'appelle souvent « mon post-doc français » pour gens ne le connaissant pas. Je ne pouvais mieux tomber que sur JB pour m'initier à plusieurs secrets importants de la science (utilisation de la suite office, de R et la création d'un poster). Il m'a aussi aidé à rester dans le droit chemin avec ses maintenant célèbres « tartes aux doigts » derrière la tête!

Il me faut aussi remercier les gens que j'ai côtoyés régulièrement depuis plus de trois ans. Tous les membres du laboratoire que j'ai taquinés comme on taquine la truite, j'espère que vous ne m'en tiendrez pas rigueur. L'expression « qui aime bien châtie bien » n'aura jamais fait plus de sens pour moi. Un remerciement spécial à Alexandre Dubé (Toutoune), que je considère presque comme un frère.

Merci à Emilie Bernatchez ma conjointe, avec qui je trouve toujours une raison pour argumenter. Je voulais te remercier pour ta patience et ta bonne humeur, celle-ci m'a beaucoup aidé dans mon quotidien quelques fois stressant. Merci de m'avoir écouté raconter mes problèmes, même lorsque tu n'avais aucune idée de ce dont je parlais.

Finalement, j’aimerais remercier mes parents qui m’ont soutenu dans toutes mes prises de décisions depuis que je suis de retour aux études. Vous avez toujours été là pour me faire réfléchir sur des questions importantes et si je suis rendu où je suis présentement, une grande partie de cette réussite vous reviens. Merci aussi à Amélie et Olivier, avec qui je partage une partie de mon patrimoine génétique. Vous me comprenez et n’avez jamais hésité à m’aider s’il vous était possible de le faire.

Le soutien financier du projet a été assuré par le regroupement québécois de recherche sur la fonction, la structure et l'ingénierie des protéines (PROTEO) et le fond de recherche du Québec-nature et technologies (FRQNT).

AVANT-PROPOS

Ce mémoire présente deux chapitres portant sur des sujets différents, mais qui sont quand même relié par le fait que le système utilisé dans le deuxième chapitre a été « découvert » en partie par les résultats découlant du premier. Ce premier chapitre, qui a été publié dans la revue *FEMS Yeast Research* sous le titre «Exploring the northern limit of the distribution of *Saccharomyces cerevisiae* and *Saccharomyces paradoxus* in North America. », est une description de la diversité des levures que l'on retrouve au Québec ainsi qu'une exploration de l'aire de répartition des levures *Saccharomyces cerevisiae* et *S. paradoxus*. J'ai activement participé à l'échantillonnage, isolé et identifié par PCR une grande partie des souches avec Jean-Baptiste Leducq, chercheur post-doctoral. Dubé. J'ai aussi contribué à la rédaction du manuscrit en collaboration avec Jean-Baptiste Leducq et Christian R. Landry, mon directeur de recherche. Nous avons aussi reçu un soutien de la part de Chloé Bertin (stagiaire master 1) et d'Alexandre K. Dubé (professionnel de recherche)

Le deuxième chapitre lui consiste en un article intitulé « Chromosomal variation segregates within incipient species and correlates with reproductive isolation », publié dans la revue *Molecular Ecology*. Il rapporte la présence d'un isolement entre les lignées américaines sauvages et divergentes de *Saccharomyces paradoxus*. Il fait ensuite état de la corrélation entre cet isolement reproducteur et des divergences au niveau moléculaire entre les lignées. Cet encore une fois a été écrit en collaboration avec Christian Landry et Jean-Baptiste Leducq. Ils ont contribué grandement au processus d'élaboration, de création, de correction et de révision du manuscrit. En collaboration avec les co-auteurs j'ai participé à la préparation du design expérimental. J'ai réalisé seul les différentes expériences nécessaires à la vérification des hypothèses. J'ai réalisé l'analyse des données avec l'appui de Jean-Baptiste Leducq et finalement j'ai activement participé à la rédaction des premières versions du texte. La majorité des corrections demandées par le comité de révisions ont été effectuées par les co-auteurs.

INTRODUCTION GÉNÉRALE

SPÉCIATION

Comme une grande partie des sujets en biologie évolutive, l'étude de la spéciation commence réellement avec Darwin. Avant même la première parution de *l'Origine des espèces*, sont publiés conjointement un chapitre de son essai de 1844 ainsi qu'un article original d'Alfred Russel Wallace intitulé : *On the tendency of species to form varieties; and on the perpetuation of varieties and species by natural means of selection* où ils décrivent l'émergence des espèces comme une succession de petites variations qui par la suite éloigne les « variétés » de plus en plus loin du type originel (Darwin & Wallace 1858). À une époque où l'on parlait d'immutabilité des espèces, on ne pouvait probablement pas concevoir que les différents organismes que l'on pouvait observer descendaient d'un ancêtre commun chez qui certaines populations se sont différenciées. Darwin en rajouta avec la parution de *l'Origine des espèces* en évoquant la possibilité que le mécanisme d'évolution par sélection naturelle puisse, sur de longues périodes de temps, mener à la « multiplication » des espèces (Darwin 1859). Ces travaux ont posé les premières bases de l'étude de la spéciation. Les définitions d'espèce sont multiples (Coyne & Orr 2004), mais lorsqu'utilisée dans le contexte de spéciation on parle d'un groupe d'organismes se reproduisant entre eux tout en étant isolés reproductivement d'autres groupes (« biological species concept »)(Mayr 1942). Cette définition n'est en aucun cas parfaite et il existe des groupes auxquels on ne peut l'appliquer (Wilson 1999; Wheeler & Meier 2000) et certains groupes comme les bactéries n'ont pas de mode de reproduction sexuée.

Avant l'avènement de la biologie moléculaire, la systématique était basée majoritairement sur la base de critères morphologiques et la définition d'espèce biologique convenait bien puisque les membres d'une même espèce se ressemblent tous et diffèrent des membres des autres espèces. Même avec la génétique classique, cette définition est tout à fait pertinente puisque si les différences entre les espèces sont régies par les gènes, la présence de barrières à l'échange des gènes est nécessaire afin qu'elles puissent maintenir l'intégrité de leur différence. Il est aussi possible de quantifier l'isolement reproducteur et donc, selon le concept biologique de l'espèce, de définir si deux populations sont bel et bien des espèces. Pour en apprendre plus sur l'émergence des nouvelles espèces il faut donc comprendre les

bases moléculaires des incompatibilités reproductives, comment elles émergent ainsi que les rôles des différentes forces évolutives dans leur émergence (Coyne & Orr 2004).

Une des principales difficultés de l'étude de la spéciation est que ce peut autant être un processus graduel et continu qu'un processus instantané. On a souvent tenté de le découper en stades plus ou moins subjectifs qui permettent une meilleure compréhension de ce processus dans son entièreté (recensé dans Lowry 2012) et de se situer le long du continuum temporel. Alfred Russell Wallace, inspiré par les papillons d'Indonésie, séparait en différents niveaux de variation morphologique allant de simple variabilité à vraie espèce (Wallace 1864). Ces observations venaient corroborer l'hypothèse selon laquelle les espèces émergeaient par modifications successives. Plus récemment, on a utilisé la différenciation des séquences d'ADN pour séparer les stades en catégories basées sur le nombre de loci impliqués dans la divergence et l'isolement reproducteur entre deux espèces, allant de quelques loci à tout le génome (Wu 2001). Cette vision génique de la spéciation souligne l'importance des connaissances de l'écologie, de la biologie reproductive et de la génétique afin de pouvoir bien définir les espèces. Wu (2001) nous prévient aussi de la nécessité de réviser le concept biologique d'espèce et que l'on devrait rediriger l'attention donnée à l'isolement reproducteur sur les adaptations différentielles dans les séquences des génomes.

ISOLEMENT REPRODUCTEUR

Comme mentionné précédemment, la spéciation implique le développement d'un isolement reproducteur entre les deux espèces en devenir. Avec le temps, les populations accumulent alors des différences génétiques qui se traduisent en différences phénotypiques et comportementales. Ces différences ont le potentiel de se traduire en barrières directes ou indirectes au flux de gènes entre populations et elles ont été compilées pour la première fois par Dobzhansky (Dobzhansky 1937a, 1951). L'isolement reproducteur peut être qualifié d'isolement reproducteur prézygotique ou postzygotique, dépendamment de s'il agit comme barrière à la formation d'un hybride ou encore s'il affecte l'aptitude (fitness) de cet hybride par rapport aux espèces parentales (Coyne & Orr 2004). L'isolement reproducteur prézygotique peut être géographique, impliquant une barrière physique entre les

populations comme c'est le cas pour les pinsons des îles Galapagos décrit par Darwin (Darwin 1859). Il peut aussi être temporel, par exemple lorsque les moments de reproductions des populations ne sont pas synchrones (Blair 1941). Le comportement des populations peut aussi poser une barrière si la reproduction est accompagnée de comportements spécifiques comme l'appel chez les amphibiens (Blair 1964). L'isolement reproducteur prézygotique peut aussi découler d'incompatibilités entre les organes sexuels où la copulation ne sera pas efficace comme c'est le cas entre différentes espèces de demoiselles (Paulson 1974). Des incompatibilités entre les gamètes peuvent aussi être en cause (Gregory & Howard 1993). L'isolement reproducteur post-zygotique est observable chez les hybrides entre deux espèces. La plus évidente barrière est la non-viabilité de ces hybrides. Bien que certains soient viables, ils peuvent être stériles ou encore ils peuvent être incapables de fonctionner normalement ou de se reproduire dans les environnements parentaux. Ces barrières post-zygotiques seront détaillées un peu plus loin dans cette introduction.

Bien que les différentes barrières à la reproduction soient bien connues, établir leurs impacts réels dans le processus de spéciation demeure difficile. D'une part, le problème est lié aux organismes utilisés pour l'étudier. Ces modèles sont souvent des paires d'espèces sœurs ou très proches, mais pour lesquelles le processus est déjà complété et qui constitue une approche rétrospective de la spéciation (Via 2009). D'autre part, les barrières reproductrices sont multiples puisqu'elles peuvent émerger les unes après les autres dans les populations, il devient complexe de déterminer laquelle a joué un rôle prépondérant dans l'initiation de la spéciation (Presgraves 2003; Ramsey *et al.* 2003). Une fois l'isolement reproducteur complété, d'autres barrières peuvent continuer à émerger ou se renforcer, mais les étudier n'est pas pertinent dans le contexte de la spéciation (Coyne & Orr 1998). Donc un bon système d'étude des évènements précoces de la spéciation serait facilité par l'étude de populations pour lesquelles l'isolement reproducteur ne serait pas encore complet.

ISOLEMENT REPRODUCTEUR POST-ZYGOTIQUE

L'isolement reproducteur post-zygotique agit chez un hybride entre deux espèces. Cet hybride sera soit capable de fonctionner, en exhibant des phénotypes ou des comportements qui diminueront ses chances de survie ou de reproduction, ou il ne sera pas en mesure de se développer ou se reproduire (Coyne & Orr 2004). On le divisera en deux catégories selon qu'il dépend d'une interaction de l'hybride avec son environnement (extrinsèque) ou encore d'une interaction entre les gènes ou les chromosomes des parents (intrinsèque). L'isolement reproducteur extrinsèque dépend de l'environnement dans lequel l'hybride va évoluer. L'hybride présente alors certains phénotypes inexistants chez les espèces parentales, ce qui entraîne une diminution de son aptitude à survivre dans l'environnement des deux parents. Ce genre d'isolement a été observé dans plusieurs systèmes différents (chez l'épinoche dans Hatfield & Schluter 1999; chez les papillons Heliconius dans Jiggins *et al.* 2001; chez Baccharis halimifolia: Melo *et al.* 2014). Il se peut aussi que l'hybride fonctionne bien dans l'environnement, mais affiche des comportements qui l'empêchent de se trouver un partenaire de reproduction chez l'une ou l'autre des espèces parentales (Hobel & Gerhardt 2003). Des travaux récents suggèrent que ce type d'isolement reproducteur est plus important que l'isolement reproducteur intrinsèque dans les premiers stades de divergence (Schluter & Conte 2009; Johannesson *et al.* 2010). Ce genre d'isolement reproducteur n'est cependant pas permanent puisqu'il est possible que des changements environnementaux puissent mener à l'émergence d'une niche ou d'un environnement favorable aux hybrides (Grant & Grant 1993).

ISOLEMENT REPRODUCTEUR POST-ZYGOTIQUE INTRINSÈQUE

L'isolement reproducteur intrinsèque résulte quant à lui de défauts du développement de l'hybride qui mène soit à la stérilité ou à la non-viabilité des hybrides (Coyne & Orr 2004). Ces phénotypes extrêmes facilement quantifiables ainsi que l'aisance relative avec laquelle on peut en évaluer les bases génétiques ont fait que l'étude de cet isolement reproducteur est la base sur laquelle s'est fondée la génétique de la spéciation (Wolf *et al.* 2010). Les recherches ont jusqu'ici identifié quatre origines génétiques pour expliquer l'isolement intrinsèque chez différents organismes : les réarrangements chromosomiques, les changements du niveau de ploïdie, les incompatibilités génétiques et les incompatibilités

entre génome et endosymbiotes (Rieseberg 2001; Coyne & Orr 2004; Hoffmann & Rieseberg 2008).

Réarrangements chromosomiques

Les réarrangements entre où à l'intérieur des chromosomes comme des translocations non réciproques, des inversions ou encore de grosses insertions ou délétions peuvent avoir des impacts sur l'isolement reproducteur entre deux populations. En 1922, Harry Federley, en étudiant la ségrégation d'un trait pour la couleur des larves hybrides entre deux espèces de *Pygaera*, a mis en évidence que ces hybrides présentaient souvent un nombre irrégulier de chromosomes. Ce nombre anormal de chromosomes faisait en sorte que le processus de la méiose n'était pas capable de se dérouler normalement lors de la formation des gamètes des individus hybrides, entraînant leur stérilité (Federley 1922). Ces observations ont été corroborées par d'autres travaux (Wodsedalek 1916; Sax 1923; Jeffrey & Hicks 1925; Sturtevant & Beadle 1936) et d'autres expérimentations sur la restauration de la fertilité des hybrides par tétraploïdisation (Karpechenko 1927; Clausen 1928). En effet, la recombinaison entre des génomes ayant des réarrangements comme des translocations ou des fusions chromosomiques différentes, ou encore un nombre de chromosomes différents entre les deux espèces peut engendrer des aneuploïdies dans les produits de méiose et ainsi une diminution de la fertilité des hybrides (Capanna 1982; Turelli *et al.* 2001). Un autre effet des réarrangements chromosomiques peut être la suppression de la recombinaison pour les gènes situés dans les endroits réarrangés (Rieseberg 2001). Cette suppression favoriserait l'accumulation de la divergence génétique dans les régions concernées, ce qui faciliterait l'émergence d'incompatibilités entre les gènes (Coluzzi 1982; Noor *et al.* 2001).

Changements au niveau de la ploïdie

Les changements au niveau de la ploïdie diffèrent d'autres origines de l'isolement reproducteur post-zygotiques puisqu'ils peuvent mener à un isolement complet et instantané, même en sympatrie (Coyne & Orr 2004). Ces changements, rarement observés chez les animaux (Muller 1925; Orr 1990; Mable 2004), auraient joué un grand rôle dans l'évolution et la diversification des plantes vu la grande variabilité des tailles de leurs génomes, du nombre de leurs chromosomes et de leur ploïdie (Kellogg & Bennetzen 2004).

Ces altérations du nombre de chromosomes ont comme origine différentes erreurs lors des divisions méiotiques et mitotiques ou encore lors de l'hybridation entre deux espèces aux chromosomes divergents (Clausen & Goodspeed 1925; Veilleux 1985; Ramsey & Schemske 1998). Les nouveaux individus formés par ces différents mécanismes se retrouvent avec un génome dupliqué (tétraploïdes) ou des jeux de chromosomes supplémentaires (triploïdes et polyploïdes) (Stebbins 1950). Ces chromosomes supplémentaires entraînent le mésappariement des chromosomes durant la méiose, générant ainsi des gamètes aneuploïdes non viables (Coyne & Orr 2004).

INCOMPATIBILITÉS GÉNÉTIQUES

Les observations de Theodosius Dobzhansky en 1936 à la suite de rétrocroisements expérimentaux entre deux « races » de drosophiles donnant des hybrides stériles, mais dont les produits méiotiques semblaient normaux, lui ont permis d'émettre des hypothèses sur les bases génétiques de cette stérilité. Sa conclusion était que plusieurs chromosomes portaient des gènes impliqués dans la stérilité avec le chromosome X ayant un effet plus important que les autosomes. Les individus stériles (avec les plus petits testicules) possédaient un chromosome X d'une race et certains autosomes correspondants à l'autre et que plus il y avait d'autosomes correspondant au « génotype » du chromosome X, plus les individus étaient fertiles (Dobzhansky 1936). Les incompatibilités entre les gènes résultent donc d'interaction épistatique entre deux ou plusieurs loci (Coyne & Orr 2004). Ces interactions sont définies comme ayant un effet sur le phénotype qui n'est pas explicable par la notion de dominance ou d'additivité (Cordell 2002). Plusieurs expérimentations ont révélé l'importance des incompatibilités génétiques dans l'isolement reproducteur intrinsèque, et ce chez plusieurs espèces différentes (Xiphophorus maculatus: Wittbrodt *et al.* 1989; Drosophile: Johnson *et al.* 1992; riz: Matsubara *et al.* 2003). Comment expliquer que les allèles de deux populations issues d'un ancêtre commun puissent devenir incompatibles? Avant Dobzhansky, William Bateson avait déjà formulé un modèle pour expliquer comment la stérilité pouvait évoluer entre deux variétés. Il y évoque la nécessité d'émergence de deux facteurs nouveaux acquis séparément par deux variétés sans causer d'effet chez leur possesseur avant leur réunion dans un hybride (Bateson 1909). Pourquoi deux changements sont-ils nécessaires? Parce que selon ce modèle, si une seule

modification est nécessaire pour causer des incompatibilités avec le génome de l'autre population issue du même ancêtre, l'individu portant cette modification sera automatiquement stérile (Bateson 1909). Selon le modèle de Bateson-Dobzhansky-Muller, lorsque le flux de gènes est interrompu entre deux populations, leurs génomes évoluent indépendamment et accumulent alors des différences dans leur séquence. Sur une longue période de temps, ces différences se retrouvent alors fixées dans les populations (Dobzhansky 1951). Tant qu'elles restent dans le contexte génétique dans lequel elles ont évolué, les différences entre les populations n'ont aucun effet sur les individus, les incompatibilités peuvent donc s'accumuler sans opposition de la sélection naturelle (Dobzhansky 1934). Cependant, lorsque deux ensembles de gènes sont mis en interactions et qu'ils contiennent des allèles n'ayant jamais été « testés » ensemble, les incompatibilités sont alors révélées chez l'hybride. Selon ce modèle, les allèles incompatibles doivent émerger et être fixés après la séparation des deux populations.

Incompatibilité entre génome nucléaire et endosymbiote

Ce type d'incompatibilité a été découvert vers les années 1970, mais il n'en est pas moins important dans le cadre de l'isolement reproducteur. Ce sont chez les bactéries du genre *Wolbachia*, parasites intracellulaires des invertébrés retrouvés principalement dans les tissus des organes reproducteurs et transmis verticalement par le cytoplasme, que l'on a d'abord découvert les incompatibilités cytoplasmiques (Yen & Barr 1971). Elles sont connues pour causer des incompatibilités menant à la perte des chromosomes du spermatozoïde au moment de la fécondation (Werren 1997). En étudiant l'isolement reproducteur entre deux espèces sœurs de guêpes du genre *Nasonia* toutes deux infectées par *Wolbachia*, Bordenstein *et al.* (2001) ont découvert que l'isolement reproducteur post-zygotique présent entre ces deux espèces pouvait être éliminé en traitant les guêpes avec des antibiotiques. Les hybrides F_2 étant totalement viables et fertiles, l'isolement reproducteur causé par *Wolbachia* a donc évolué avant tous les autres types d'isolement et peut mener à la spéciation (Bordenstein *et al.* 2001). Ces résultats ont ouvert la voie à l'étude des incompatibilités causées par les mitochondries. Des croisements expérimentaux faits avec des populations de copépodes ont mis en évidence des incompatibilités mitonucléaires (Ellison & Burton 2008). Les hybrides F_2 et F_3 entre ces populations

souffraient d'une diminution prononcée de leur fertilité et de leur viabilité que seul un rétrocroisement maternel pouvait améliorer (Ellison & Burton 2008). La seule différence génétique entre les rétrocroisements paternels et maternels étant des éléments cytoplasmiques, cela suggère fortement un rôle de l'ADN mitochondrial. La production d'énergie nécessite une coordination entre génome mitochondrial et nucléaire (Blier *et al.* 2001) et l'on peut observer des performances métaboliques diminuées chez les hybrides entre les populations de ces copépodes (Ellison & Burton 2006) ce qui renforce l'hypothèse des incompatibilités cytonucléaires.

La variation génétique et chromosomique joue un grand rôle dans l'isolement reproducteur. Si l'émergence de nouveaux allèles incompatibles est responsables de l'isolement reproducteur post-zygotique intrinsèque, qu'en est-il de la variation présente à l'intérieur de l'espèce ancestrale ou des nouvelles lignées sous forme d'allèles polymorphiques ou de réarrangements chromosomiques? Nous disposons déjà de plusieurs exemples où cette variation peut causer de l'isolement reproducteur, et ce chez plusieurs espèces (Mammifère: Good *et al.* 2008; Plante: Bikard *et al.* 2009; Nématode: Kozlowska *et al.* 2012). Cutter (2012) a déjà proposé un modèle incluant le polymorphisme ancestral ou intrapopulation qui est compatible avec le modèle original de Bateson-Dobzhansky-Muller, mais il reste encore à étudier l'apport de ces polymorphismes dans l'établissement des barrières à la reproduction entre les populations. C'est cette question que j'ai abordée au cours de mes travaux de maitrise. Pour ce faire, il faut des systèmes dans lesquels on peut étudier le début d'un évènement de spéciation. Le système que nous avons choisi est la levure *Saccharomyces paradoxus*, dont les caractéristiques des populations nord-américaines semblent indiquer qu'elle constitue un excellent modèle pour l'étude de l'émergence des barrières reproductives.

Les levures du genre Saccharomyces

Les levures sont des rouages importants de plusieurs écosystèmes et sont présentes dans tous les biomes existants. Elles y participent à la transformation de la matière organique et interagissent avec une multitude d'autres organismes autant comme compétiteur, parasite, coopérateur ou encore pathogène (Starmer & Lachance 2011). Le terme levure est utilisé

pour décrire les champignons unicellulaires qui se reproduisent principalement par fission ou bourgeonnement et dont les formes de reproduction sexuée ne sont pas compartimentées dans des organes de fructification (Kurtzman et al. 2011). Le mot levure est cependant souvent le synonyme de *Saccharomyces cerevisiae*, la levure du boulanger. Premier eucaryote pour lequel le génome a été séquencé (Goffeau et al. 1996), ce champignon unicellulaire joue un rôle important dans plusieurs sphères de l'activité humaine, dont l'alimentation et les biotechnologies. Cet organisme est aussi un modèle de choix dans plusieurs champs d'études telles la biologie cellulaire et la génétique (Botstein & Fink 2011). Le genre *Saccharomyces* fait partie de l'ordre des Saccharomycetales avec les autres levures bourgeonnantes et de la famille des Saccharomycetaceae avec d'autres genres comme Kazachstania, Lachancea et Kluyveromyces qui prospèrent dans les environnements où des glucides sont disponibles (Vaughan-Martini & Martini 2011). Il compte sept espèces qui sont morphologiquement semblables, mais qui diffèrent pour certains phénotypes comme la tolérance à différentes températures (Goncalves et al. 2011). Au cours des dernières années, beaucoup de groupes de recherche se sont intéressés à l'écologie et à la génomique du genre *Saccharomyces* sauvages, ce qui l'a propulsé en tant que « genre modèle » pour la génomique comparative et évolutive (Landry *et al.* 2006; Replansky *et al.* 2008).

SACCHAROMYCES PARADOXUS

Beaucoup moins connue que *S. cerevisiae*, S. *paradoxus* est quant à elle beaucoup plus utilisée pour la recherche en écologie et en évolution (Johnson et al. 2004; Koufopanou et al. 2006; Tsai et al. 2008). L'absence de lien étroit avec l'homme facilite l'étude de ses populations naturelles dont la structure génétique semble très influencée par la géographie (Liti et al. 2009). On retrouve cette espèce principalement associée avec des arbres de la famille des Fagaceae, vivant bien souvent en sympatrie avec d'autres espèces de Saccharomyces (Sniegowski et al. 2002; Sampaio & Goncalves 2008; Zhang et al. 2010). Bien que ce soit les deux espèces les plus proches dans l'arbre des Saccharomyces, il y a une divergence nucléotidique d'environ 15 % entre *S. paradoxus* et *S. cerevisiae* (Cliften et al. 2001), mais très peu de réarrangements chromosomiques (Fischer et al. 2000). Un isolement reproducteur prézygotique est présent sous la forme d'une différence dans le

temps de germination des spores (Maclean & Greig 2008; Murphy & Zeyl 2012). Un isolement reproducteur post-zygotique lié à la divergence génétique empêchant le bon fonctionnement de la recombinaison menant à une faible survie des spores résultant de la méiose a aussi été montré (Chambers et al. 1996). Bien que différentes génétiquement, ces deux espèces sont très similaires biologiquement et physiologiquement. Il est donc possible de tirer profit des méthodes et des outils moléculaires développés pour *S. cerevisiae*, qui nécessitent souvent certaines modifications pour qu'elles soient efficaces chez son espèce sœur.

DÉMOGRAPHIE DES POPULATIONS DE S. PARADOXUS

À l'échelle mondiale, *S. paradoxus* est subdivisée en groupes européen, asiatique, américain sur la base de milliers de marqueurs polymorphes (Liti et al. 2009). En se basant sur la séquence de neuf gènes, Kuehne et al. 2009 ont montré que la population américaine de *S. paradoxus* est à son tour subdivisée en trois lignées génétiquement divergentes. Une première lignée endémique de toute l'Amérique du Nord (lignée B) et compose la majorité de la population. Une deuxième lignée d'origine européenne (lignée A), probablement emmenée sur le continent américain à la suite d'une migration transocéanique de nature anthropique (Kuehne et al. 2007). Ces deux premières lignées sont partiellement isolées reproductivement, ce qui suggèrerait un évènement de contact secondaire. La troisième lignée considérée comme rare, puisque retrouvée sous forme d'un isolat unique (lignée C), semble aussi se retrouver seulement en Amérique du Nord. Une analyse de l'échantillonnage (chapitre. 1) du nord-est de l'Amérique du Nord a cependant montré que cette lignée occupait seule la limite nordique de la distribution de *S. paradoxus*, soit de l'île d'Orléans jusqu'à la pointe gaspésienne (Charron *et al.* 2014; Leducq *et al.* 2014). Les lignées américaines de *S. paradoxus* ont une distribution partiellement chevauchante avec la lignée B qui occuperait majoritairement les régions du sud et de l'ouest jusqu'à la hauteur de Montréal où commencerait une zone de cooccurrence entre les deux lignées qui finirait aux alentours de Québec (Leducq et al. 2014).

Divergence génétique et phénotypique chez S. paradoxus

Génétiquement, les trois lignées présentes en Amérique du Nord ont une divergence nucléotidique variant entre 2,06 % et 3,01 %. Un certain niveau de polymorphisme nucléotidique est aussi présent à l'intérieur des lignées B (0,33 %) et C (0,04 %) (Leducq et al. 2014). En plus d'être divergentes génétiquement, les souches des lignées B et C montrent de la divergence pour certains phénotypes ayant de probables bases écologiques. Des tests en laboratoire sur la capacité des souches à croître en différentes conditions ont montré que les souches de la lignée B ont une bonne croissance à 37 °C alors que la souche type A (d'origine européenne) et les souches de la lignée C ont une croissance très faible dans cette condition (Leducq et al. 2014). Ces résultats sont concordants avec le fait que l'adaptation à différentes températures semble avoir eu un rôle dans l'évolution et l'écologie des espèces de Saccharomyces (Goncalves et al. 2011; Salvado et al. 2011). Il est possible d'observer les mêmes tendances pour ce qui est de la survie à un cycle gel-dégel pour laquelle la lignée B survit mieux que la lignée C. D'autant plus intéressants, ces phénotypes sont corrélés avec des facteurs tels la température moyenne et le nombre de jours avec au moins un cycle gel-dégel du lieu d'échantillonnage pour les deux groupes (Leducq et al. 2014). Ces observations suggèrent que les lignées B et C sont localement adaptées aux conditions climatiques (Leducq et al. 2014). De telles adaptations pourraient être des facteurs de spéciation écologique entre ces lignées, qui représenteraient ainsi un évènement de spéciation récent.

Objectif

L'isolement reproducteur post-zygotique intrinsèque est principalement causé par des incompatibilités génétiques ou des réarrangements chromosomiques. Il est possible de localiser ces barrières à la reproduction dans le génome, mais identifier le moment de leur émergence ou encore la manière dont elles se propagent dans les populations est un peu moins facile à aborder dans un système où les espèces sont déjà bien séparées. L'objectif principal de ce projet de maitrise est de vérifier la présence de ségrégation de facteurs connus pour causer de l'isolement reproducteur intrinsèque au sein des populations naturelles avant l'interruption du flux génique. Nos objectifs spécifiques sont i) de caractériser la diversité des populations de *Saccharomyces paradoxus* présente dans le nord

est de l'Amérique du Nord. ii) de tester la présence d'un isolement reproducteur post-zygotique intrinsèque entre deux lignées divergentes nord-américaines de *S. paradoxus* et iii) d'identifier la contribution de la divergence génétique et des réarrangements chromosomiques à cet isolement reproducteur. La majorité des souches utilisées dans ce projet de recherche ont été obtenues à l'aide d'un effort d'échantillonnage et d'isolement de levures qui constitue le premier chapitre de ce mémoire. En se basant sur les connaissances disponibles sur ces lignées, notre hypothèse de travail est que les lignées B et C sont en cours de spéciation et présenteront un certain niveau d'isolement reproducteur qui sera corrélé avec leur divergence génétique. Sur cette base on s'attend à ce que l'isolement reproducteur soit plus faible entre ces lignées que lorsque croisées avec la lignée A plus divergente génétiquement. Cette étude permettra l'établissement de *S. paradoxus* en tant que rare modèle eucaryote unicellulaire pour l'étude moléculaire de la spéciation et de l'isolement reproducteur. Ce projet de maitrise devrait contribuer à l'avancement des connaissances sur les rôles du polymorphisme dans le processus de spéciation et, dans un contexte plus général, sur les prémisses moléculaires de la spéciation.

CHAPITRE 1: EXPLORING THE NORTHERN LIMIT OF THE DISTRIBUTION OF *SACCHAROMYCES CEREVISIAE* AND *SACCHAROMYCES PARADOXUS* IN NORTH AMERICA.

RÉSUMÉ

La présente étude s'est penchée sur la distribution nordique de *Saccharomyces cerevisiae* et *Saccharomyces paradoxus* en Amérique du Nord. En tout 876 échantillons de divers substrats répartis en 29 sites d'échantillonnages ont été prélevés de l'environnement et ramenés au laboratoire pour procéder à un enrichissement et un isolement des levures mésophiles. Une vaste diversité de levure a été révélée, certaines d'entre-elles associées avec des substrats particuliers. Le séquençage du locus ITS1-5,8S-ITS2 a permis l'identification au niveau de l'espèce de 226 souches qui incluent 41 souches de *S. paradoxus*. Notre effort d'échantillonnage suggère cependant que *S. cerevisiae* est rare aux latitudes nordiques. Selon les taux de succès d'isolement, les communautés de levures connaissent un pic d'activité durant l'été pour ensuite se raréfier autour du début de l'automne. Les données obtenues corroborent les connaissances disponibles sur les niches écologiques des familles de levures comme les *Pichiaceae*, *Saccharomycodaceae*, *Debaryomycetaceae* et les *Phaffomycetaceae*. De plus nous n'avons pas été en mesure d'identifier 24 des isolats récoltés, isolats qui pourraient être importants pour la taxonomie, les biotechnologies ou la recherche biomédicale. Notre étude constitue de nouvelles données sur la diversité taxonomique des levures ainsi que des nouvelles ressources pour l'étude de l'évolution et de l'écologie du genre *Saccharomyces*.

ABSTRACT

We examined the northern limit of *Saccharomyces cerevisiae* and *Saccharomyces paradoxus* in northeast America. We collected 876 natural samples at 29 sites and applied enrichment methods for the isolation of mesophilic yeasts. We uncovered a large diversity of yeasts, in some cases associated with specific substrates. Sequencing of the *ITS1, 5.8S* and *ITS2* loci allowed to assign 226 yeast strains at the species level, including 41 *S. paradoxus* strains. Our intensive sampling suggests that if present, *S. cerevisiae* is rare at these northern latitudes. Our sampling efforts spread across several months of the year revealed that successful sampling increases throughout the summer and diminishes significantly at the beginning of the fall. The data obtained on the ecological context of yeasts corroborate what was previously reported on *Pichiaceae*, *Saccharomycodaceae*, *Debaryomycetaceae* and *Phaffomycetaceae* yeast families. We identified 24 yeast isolates that could not be assigned to any known species and that may be of taxonomic, medical, or biotechnological importance. Our study reports new data on the taxonomic diversity of yeasts and new resources for studying the evolution and ecology of *Saccharomyces paradoxus*.

Budding yeasts of the genus *Saccharomyces*, and particularly the model species *S. cerevisiae*, are the most studied group of fungi in the laboratory. *S. cerevisiae*, also known as brewer's yeast, has been used for several decades as a model in genetics, cell biology and genomics. Additionally, *S. cerevisiae* and its closely related species have also served as models in ecological genetics and genomics (Reviewed in Landry *et al.* 2006; Replansky & Bell 2009; Hittinger 2013). Extensive genetic variation has been documented for several traits, including the ability to grow on different media, at different temperatures and under different stress conditions (Gasch & Werner-Washburne 2002; Liti *et al.* 2009; Will *et al.* 2010), gene expression (Landry *et al.* 2007; Brown *et al.* 2008) and colony morphology variation in response to different carbon sources (Granek & Magwene 2010). Recently, the brewer's yeast has also been genetically engineered to produce an array of biofuels, which is of major interest in biotechnology (reviewed in Buijs *et al.* 2013). Because *S. cerevisiae* populations that are sampled in the wild have been affected by human activities, the distribution of genetic variation in this model species may not reflect natural population dynamics and biogeography (Liti *et al.* 2009; Hyma & Fay 2013), which makes it a limited model for the study of population genetics and adaptation. As a consequence, there is a growing interest for the use of other species of the genus in the study of natural populations and natural variation in relation with abiotic and biotic factors, particularly in the use of its sister species *S. paradoxus* (Dunham & Louis 2011).

Natural isolates from the genus *Saccharomyces* have been successfully isolated from a variety of substrates such as deciduous tree bark, exudates or associated soils, vineyard grapes, wild fruits and insects such as *Drosophila* (Phaff *et al.* 1956; Naumov *et al.* 1996; Naumov *et al.* 1998; Hyma & Fay 2013). However, the most common substrate for the isolation of natural *Saccharomyces paradoxus* is the bark from trees of the Fagaceae family, especially from the genus *Quercus*. *Saccharomyces* yeasts usually share these habitats with other Saccharomycetaceae such as *Torulaspora* ssp. and *Lachancea* ssp. (Sampaio & Goncalves 2008). *S. paradoxus* has been shown so far to be distributed where the deciduous trees with which it is associated are prosperous, namely in North America, Europe, the far East Asia and New-Zealand (Sniegowski *et al.* 2002; Naumov *et al.* 2003;

Liti *et al.* 2009; Zhang *et al.* 2010). In all cases, both *S. paradoxus* and *S. cerevisiae* have been isolated in the same regions, suggesting that their distribution areas largely overlap. At the northern limit of deciduous tree distribution in the north eastern America (Mont St-Hilaire; 45°33′N 73°09′W), a single site has been intensively sampled and this sampling suggests that *S. cerevisiae* is rare or absent at this latitude because only *S. paradoxus* and other distantly related species have been isolated (Replansky & Bell 2009).

Given the importance of *S. cerevisiae* and *S. paradoxus* in several fields of the life-sciences, including evolution and ecology as well as in biotechnology, it is of prime importance to better understand the ecological factors that determine the distribution of these species. Many regions, particularly regions where we expect to find the margin of their distribution, have not been or have been poorly sampled. The aim of this study was to examine whether the distributions of *S. paradoxus* and *S. cerevisiae* extend further north than what has been previously described and to test if they could be found on the same substrates as reported before in other regions as well as on other substrates. More than 800 samples were collected in 26 areas of the low St Lawrence valley and throughout the Gaspé peninsula. Our samples cover an area of about 150,000 km^2. We sampled a large diversity of substrates and we used an enrichment method commonly applied for the isolation of *S. cerevisiae* and *S. paradoxus* (Sniegowski *et al.* 2002).

METHOD

Sample collection

We collected 876 samples between April and October 2011 and 2012 on various substrates on sites indicated in Fig. 1 following methods described in Sniegowski *et al.* (2002) (see also Table S1). Samples mostly included exudate material, fragments of bark and soil associated with trees, insects, slugs, flowers and decaying fruits. Samples were collected using a 14 mL sterile plastic tube (BD Falcon) or 1.5 mL sterile microtubes (Axygen). Tubes were maintained for a maximum of 1 week at 4°C before enrichment.

Enrichment culture

All enrichments were performed as described in Sniegowski *et al.* (2002). Ten mL (14 mL tubes) or one mL (1.5 mL tubes) of medium was added to the sample before incubation and tubes were airtight sealed. In 2011, tubes from samples collected in paired replicates (two samples from the same substrate) were incubated at 18°C and 30°C, respectively, to ensure that the differences in optimal temperature growth between *S. cerevisiae* and *S. paradoxus* did not bias the isolation success (Sweeney *et al.* 2004). Tubes corresponding to unique samples were only incubated at 18°C. Because Saccharomycetaceae isolation efficiency was higher at 30°C (Table S2; Fig. 2), we incubated all samples at this temperature in 2012. All samples were incubated for a period of 10 days. After incubation, tubes were agitated and 10 µL of each culture was deposited on solid synthetic medium for colony isolation. The medium was as described in Sniegowski *et al.* (2002), with the following modification: Methyl α-D-glucopyranoside was replaced with glucose since some strains could be unable to use Methyl α-D-glucopyranoside as carbon source (Sniegowski *et al.* 2002). Plates were incubated for 3 to 4 days at the same temperatures used for enrichment cultures and examined for colony growth. All colonies were considered unless they were mouldy. Cells from all colonies were observed under an optical microscope using phase contrast at 400X to identify clones corresponding to round and budding cells. These colonies were streaked on yeast extract peptone dextrose (YPD) agar plates in order to obtain isolated clones, after which a single isolated clone for each strain was inoculated into liquid YPD medium, grown to saturation and archived in 25% glycerol at -80°C.

PCR identification of yeasts

We identified putative yeast isolates using a 850 bp of *ITS1-5.8S-ITS2* sequence following Montrocher *et al.* (1998). Genomic DNA was extracted as described in Kaiser *et al.* (1994) from a 5 mL overnight culture for all yeast candidates obtained from the enrichment cultures. PCR reactions 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 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: 3' at 95°C; then 35 cycles of 30" at 95°C, 30" at 50°C and 2' at 72°C; and a final

extension of 10' at 72°C. PCR products were then sequenced by Sanger sequencing. Sequences obtained were manually curated according to the original chromatogram and classified using NCBI BLAST against the nucleotide collection (nr/nt). An isolate was identified as a known taxon when it's *ITS1-5.8S-ITS2* sequence identity with the reference was higher than 97% with a longer than 90% coverage (Ciardo *et al.* 2006). We carefully discarded NCBI BLAST reference strains that matched with our sequences but for which we suspected identification errors (*i.e.* strains assigned to a taxon but matching with uncharacterized strains rather than with the type strains of the taxon). Additionally, we validated our identification by using type strains clearly identified in Kurtzman *et al.* (2011). Finally, we identified strains that did not match any clearly named taxon as uncharacterized (identity < 97%). We validated the status of uncharacterized strains after they had been frozen, by repeating *ITS1-5.8S-ITS2* amplification and sequencing.

Statistical analyses of isolation success

We performed statistical analyses on isolation success to test if our survey was biased with respect to sampling size and to detect the effects of substrate, incubation temperature during enrichment and time of the year on isolation yield. We considered three metrics of efficiency: the proportion of samples giving at least one yeast isolate after enrichment (including Saccharomycetaceae yeasts), the proportion of samples that yielded at least one isolate belonging to Saccharomycetaceae (including from genus *Saccharomyces*) and the proportion of samples giving at least one *Saccharomyces* isolate. We considered samples from years 2011 and 2012 separately, given the differences in locations, isolation and sampling strategies. For each category and year, we first performed an ANOVA on isolation success rate, estimated as the ratio between the number of strains from the given category and the total number of samples. We considered the nature of the substrate (i.e. decayed fruit *vs.* insect, tree or associated soil; 2011 only), the temperature of isolation (2011 only), the number of samples and the month of sampling as additive factors in a linear model. To investigate more deeply the effect of time of the year on isolation success in each category, we performed a chi-squared test on isolation success at different months. We considered separately years 2011 and 2012 and the nature of substrate. All analyses were performed in R (Chambers & Hastie 1992).

RESULTS AND DISCUSSION

Yeast isolation efficiency depends on the temperature of isolation, time of the year and substrate

We sampled 876 substrates in 26 locations in the St Lawrence Valley, Gaspé peninsula and three other places in Canada (Ontario, Manitoba, British Columbia). Five locations yielded no isolate, despite intensive sampling in two of them (Table S1; Fig. 1). Interestingly, most of these locations were either located at high altitude (Parc de Gaspésie, Grands-Jardins: 600-800 m) or at the northern limit of our sampling area (Saguenay: 40°20' N, 70°18' W; Manitoba: 52°08' N, 106°40' W), suggesting that cold weather and the scarcity of deciduous trees in these regions could have an effect on yeast diversity. From all the samples obtained, 28% (243 positives samples on 876) yielded putative yeast strains as identified by NCBI BLAST of the *ITS1-5.8S-ITS2* sequence (Table S3). In some rare cases (less than 5%) unidentified bacteria were obtained and in some rarer cases (1%) the filamentous fungus *Syncephalastrum* was isolated (Table S3). For some of the samples, isolation on solid medium gave rise to up to three different yeast taxa. The entire procedure allowed the isolation of 250 yeast strains covering at least 69 taxa (see Table S3).

We carried out our sampling over two years (2011-2012) from April to October. In 2011, we systematically isolated our strains at two different temperatures (18 and 30°C) or only at 18°C when only one sample was available. Greater isolation efficiency was obtained at 18°C (45%) with a majority of non-Saccharomycetaceae yeast (40%) and few Saccharomycetaceae (5%; Fig. 2). At 30°C, isolation efficiency substantially decreased (16%) but provided Saccharomycetaceae (5 and 7% respectively) and *Saccharomyces* (3% in both cases) in the same proportions as at 18°C, suggesting that 30°C is a better temperature to efficiently reduce the isolation of non-Saccharomycetaceae species, without affecting the yield of Saccharomycetaceae and mesophilic *Saccharomyces* ssp. (Fig. 2). In 2011, we also sampled a wide variety of substrates, including bark and exudate of trees, soil associated to these trees, insects and various decaying fruits. Interestingly, decayed fruits gave the highest yeast recovery (51% on 148 samples; Fig. 3a), which was more than

twice the recovery rate observed for other substrates (21% on 250 samples; Fig. 3b). However, almost no Saccharomycetaceae were recovered from decaying fruits (Fig. 3a), whereas trees, soil and insects provided 8% of Saccharomycetaceae and 4% of *Saccharomyces* (Fig. 3b). Overall, the nature of the substrates appeared to be the more critical factor for Saccharomycetaceae and *Saccharomyces* yield efficiency (ANOVA $p < 0.05$, Table S2). Thus in 2012 we sampled only bark of trees, soil associated to these trees and insects, and we carried isolation at 30°C to enhance the yield of *Saccharomyces*. Isolation success was slightly higher than in 2011 for comparable samples isolated in similar conditions (24% on 470 samples) and provided a lower proportion of non-Saccharomycetaceae species (6%) and a majority of Saccharomycetaceae (18.5%), including *Saccharomyces* (6%; Fig. 2).

We found that isolation success continually and significantly increased from April to August-September and substantially decreased at the end of summer (Fig. 3; Table S2). This observation suggests that yeast abundance could progressively increase during the year, with a peak at the end of summer, and suddenly collapses in the fall. Our survey along the year was however done on different sites and this result has thus to be validated by a survey of the same sites along the year to discard any potential among-site variation. Differences among years could be another source of variation because of difference in isolation temperature and of the higher variety of substrates we sampled in 2011. We controlled for this source of variation and found that the evolution of yeast abundance during the year could be observed independently in 2011 (Figs. 3a-b) and 2012 (Fig. 3c), regardless the number of samples and the nature of the substratum. This increase was stronger in our sampling of 2012 and for *Saccharomyces* isolation efficiency (ANOVA: $p< 0.1$, Table S2; chi-squared test: $p << 0.0001$, Fig. 3).

Non-Saccharomycetaceae and new potential taxa

The isolation method we used was not specific to *S. cerevisiae* and *S. paradoxus* and allowed the isolation of yeasts that were not specifically targeted by our sampling. Decaying fruits provided a high proportion (60%) of non-Saccharomycetaceae yeasts (Fig. 3a). These included two main families, Pichiaceae and Saccharomycodaceae, most often

represented by the genera *Pichia* and *Hanseniaspora* (Table. S3), which supports previous studies providing evidence that yeast diversity on fruits was dominated by these taxa in Europe (Davenport 1976; Vadkertiova *et al.* 2012). Additionally to Saccharomycetaceae, trees and their associated soils yielded Debaryomycetaceae (Table. S3). For instance, we found *Debaryomyces hansenii*, which has been previously isolated from bark of tree in India (Bhadra *et al.* 2008). Many Phaffomycetaceae (formerly described as Wickerhamomycetaceae; Kurtzman & Robnett (2013); Table. S3), mostly represented by *Wickerhamomyces anomalus*, were found to be associated with fruits and trees. This species was previously known to be mainly associated with food products and clinical environment but some isolates had already been found in the wild on fruits and various plant surfaces (Daniel *et al.* 2011). It is also known to produce killer agents and is used in the industry as a biocontrol agent against fungal crops pathogens (Lassois *et al.* 2008) and some bacterial crop pathogens (Wang *et al.* 2009).

Out of these non-Saccharomycetaceae yeasts, about 18% (24 out of 137) could not be formally identified to the species level. These isolates could be of interest in taxonomic studies as they belong to multiple families in the yeasts current taxonomy such as the Pichiaceae, Saccharomycodaceae and Phaffomycetaceae (Table S4). Two of these are potentially new yeasts species, found on bark of trees, and were closely related to yeasts from the genus *Williopsis* (KF057726-7; 88% sequence homology), which are of biotechnological and medical interest (reviewed in Schmitt & Breinig 2002). Moreover, one of these isolates (KF057493), also found on trees, may be a completely new species, since the closest in homology (85%) sequence belonged to an uncultured compost fungus (Hultman *et al.* 2008). Our identification of yeast strains was however based on only one marker (*ITS1-5.8S-ITS2*), which is likely unavailable for some known species. Hence, the status of potentially new taxa we isolated should be validated with other genetic markers also frequently used in the identification of fungi such as *LSU* rRNA, *SSU* rRNA, EF-1α or Cytochrome oxidase II (Kurtzman *et al.* 2011).

Saccharomycetaceae and Saccharomyces *diversity*

We obtained 113 strains of Saccharomycetaceae. The vast majority of them came from bark samples from deciduous trees (92%; Figs. 3c) and were present in most of the sites sampled, excepted northern or elevated sites as discussed above (Saguenay, Grands-Jardins, Parc de Gaspésie; Fig. 1). These Saccharomycetaceae strains were represented by the genus *Kazachstania* (6), *Kluyveromyces* (8), *Torulaspora* (21), *Lachancea* (36) and *Saccharomyces* (41; Table. S3), as previously found in different regions of the world for instance in Ontario (Maganti *et al.* 2012) and British Columbia in Canada, in Germany and in Portugal (Sampaio & Goncalves 2008).

The budding yeast *Saccharomyces cerevisiae* was not found in any of the samples while 41 strains of *Saccharomyces paradoxus* were recovered. This is opposite to what has been observed in other regions of North America, for instance in Pennsylvania (Sniegowski *et al.* 2002), Oregon and Missouri (Hyma & Fay 2013). However, as it is the case in Québec, *S. paradoxus* but not *S. cerevisiae* was recovered in Ontario (Maganti *et al.* 2012). The scarcity or absence of *S. cerevisiae* in these regions, could tentatively be explained by the difference in optimal growth temperature for these species (Sweeney *et al.* 2004), as the climate in the provinces of Québec and Ontario is significantly colder than the other areas where isolation of *S. cerevisiae* and *S. paradoxus* on the same substrates was reported (Sniegowski *et al.* 2002; Sampaio & Goncalves 2008; Liti *et al.* 2009; Zhang *et al.* 2010). Our survey also suggests that mesophilic *Saccharomyces* ssp. are absent or rare in cooler regions for instance in the mountains and the north of the St Lawrence River (Fig. 1). However, further sampling and isolation at lower temperature should be used in future investigations of *S. paradoxus* in these regions (Sampaio & Goncalves 2008). For instance, *S. uvarum* was reported in southern British Columbia and was only isolated at 10°C (Sampaio & Goncalves 2008).

Regarding the effect of substrates on the taxa isolated, trees belonging to the Fagaceae family (mostly oak species) yielded a three fold higher percentage of *S. paradoxus* isolates (12%) than other trees, including maple (4%). These rates remain lower than *Saccharomyces* isolation success in British Columbia, Germany, Portugal (33% on

Fagaceae and 9% on other trees; Sampaio & Goncalves 2008) and in Pennsylvania (23% on Fagaceae; Sniegowski *et al.* 2002) but are equivalent to what was observed in U.K. (8% on Fagaceae; Johnson *et al.* 2004) and Ontario (10% on Fagaceae and 4% on other trees; Maganti *et al.* 2012). The low recovery rates were observed in cooler regions supports the hypothesis that the northern distribution of mesophilic *Saccharomyces* ssp. could be limited by low temperatures. Insects, mostly from the genus *Musca* and *Bombus*, yielded some isolates in the same proportion as trees not belonging to the Fagaceae family (4%). This low recovery rate suggests that insects represent a poor substrate for the isolation of *S. paradoxus*. This is in contrast with several studies that successfully isolated *Saccharomyces cerevisiae* from *Drosophila* and suggested that insects could be a major vector of yeast dispersal (Phaff *et al.* 1956; Naumov *et al.* 1996; Coluccio *et al.* 2008; Stefanini *et al.* 2012). Our result could either suggest that, unlike *Drosophila*, *Musca* and *Bombus* are only incidental vectors of *Saccharomyces* or that contrary to *S. cerevisiae*, *S. paradoxus* is not or rarely associated with insects. However, *S. paradoxus* was widespread across our sampling area and its distribution was continuous among sites, suggesting that there is no limit to its dispersal at the scale of our study area (Fig. 1).

Conclusion

Our sampling of the North East of North America unveiled a portion of the cultivable yeast diversity that resides along the St Lawrence valley and within the Gaspé peninsula in the province of Québec. Our results suggest that this diversity comprises many species that form dynamic communities that prosper and decline with the cycling of the seasons. Some of the yeasts isolated could not be assigned to known species based on ITS sequences, which suggests that there may be some indigenous yeasts in the province of Québec that have not been reported elsewhere in the world. The biotechnology industry uses non-*Saccharomyces* ascomycetous yeasts in diverse applications like environmental biotechnology, biomedical research, food industry and many other areas (reviewed in Johnson 2013). The more yeasts species or strains are uncovered in studies like this one, the more likely we are to find sources of enzymatic activities or species for biological applications that fulfil our needs. Our study therefore represents a useful resource for this community. Finally, our study support previous reports suggesting that climatic variables

are major factors in shaping the ecological distribution of *Saccharomyces* yeasts (Sweeney *et al.* 2004; Sampaio & Goncalves 2008; Liti *et al.* 2009; Goncalves *et al.* 2011).

Saccharomyces cerevisiae isolates could not be obtained with methods that were successful in isolating it in other regions of the world, which indicates that the overlap of its distribution with *Saccharomyces paradoxus* is limited at northern latitudes, at least in North America. Finally, several *S. paradoxus* isolates were obtained and could be used in the context of ecological genomics studies that will provide more information on the ecology, phylogeography and biology of this species.

ACKNOWLEDGEMENTS

This project was funded by a NSERC discovery grant to CRL. CRL is a CIHR New Investigator. Thanks to A. Aubin-Landry, N. Aubin-Horth, J. Landry, L. Landry, G. Diss, B. Vauthier, A. Gerstein, S. Charette,V. Bourret, S. Renaut, N. Brodeur, D. L. Charron, A. Charron, F. Carrier, E. Bernatchez, U. Cruz-Martinez, M. Bélanger and D. Charbonneau for sampling and two anonymous reviewers for helpful comments on the manuscript. GC was supported by a PROTEO graduate student scholarship. JBL was supported by a fellowship from the Fonds de Recherche en Santé du Québec (FRSQ) and a HFSP Grant to CRL. The authors have no conflict of interest to declare.

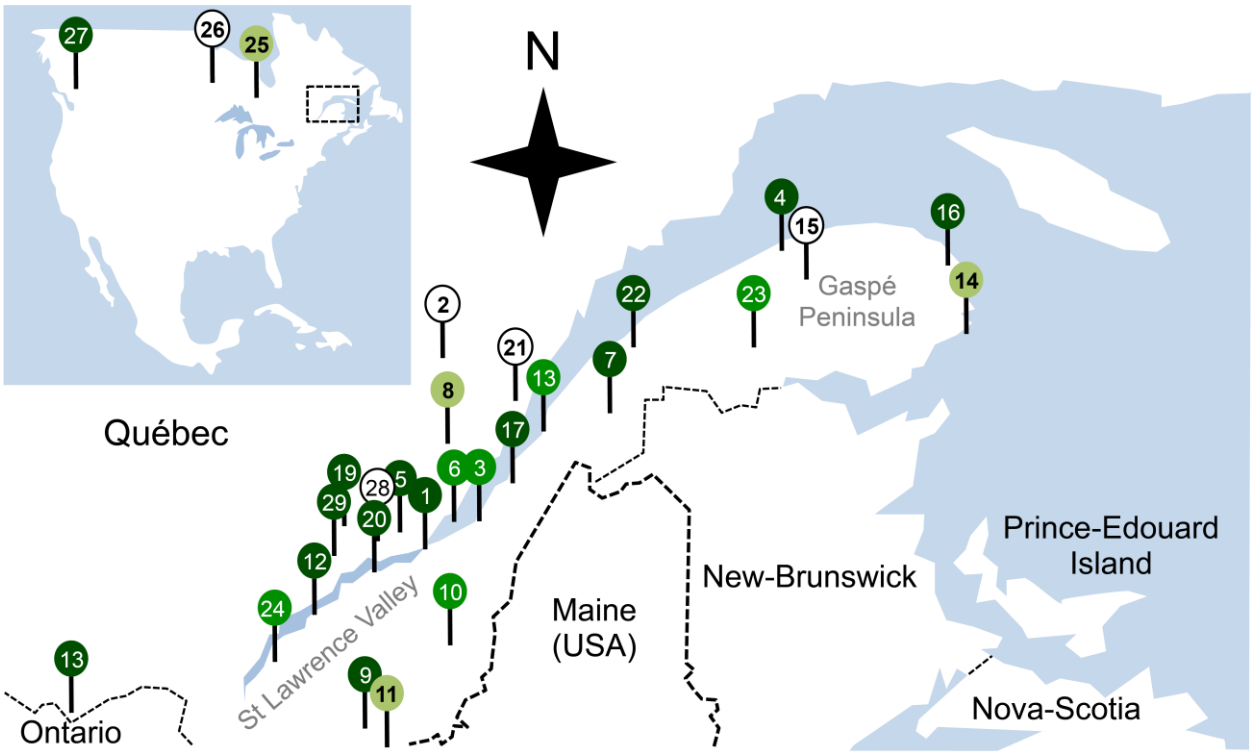

Figure 1: Map of the 29 sites sampled in this study.

The majority of sites (1-24; 28-29) were located in the province of Québec (St Lawrence Valley and Gaspé Peninsula) and three elsewhere in Canada (Map on top left; the dotted frame indicate our study area). Sites are represented by circles (see Table S1 for details of locations) and are coloured in dark green when *Saccharomyces* ssp. were found, in green when Saccharomycetaceae yeasts but no *Saccharomyces* ssp. were found, in light green when only non-Saccharomycetaceae yeasts were found and white when no yeast was isolated.

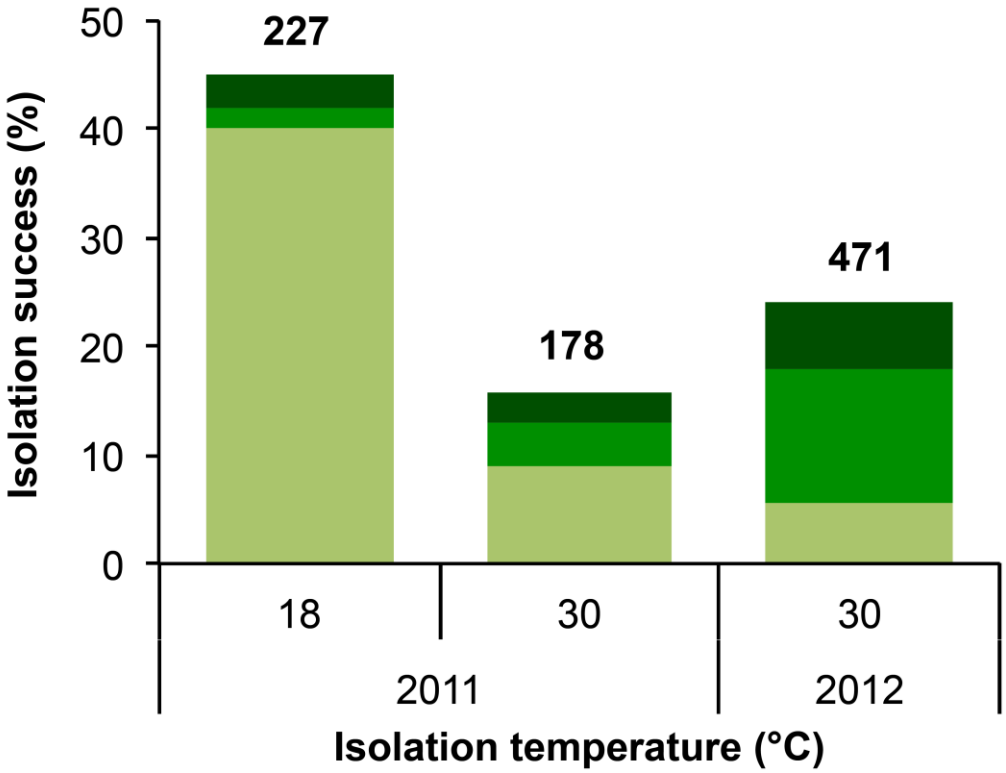

Figure 2: Effect of temperature on yeast isolation success.

Isolation was performed following the methods described in Sniegowski *et al.* 2002 and tested at two different temperatures for the same samples in 2011 (18°C and 30°C) and at 30°C in 2012. Bars represent proportion of samples that gave at least one strain. Bars are filled in light green when only non-Saccharomycetaceae yeasts were found, in green when Saccharomycetaceae yeasts (excluding *Saccharomyces* ssp.) were found or in dark green when *Saccharomyces* ssp. were found. The number of samples in each temperature category is shown above the bars.

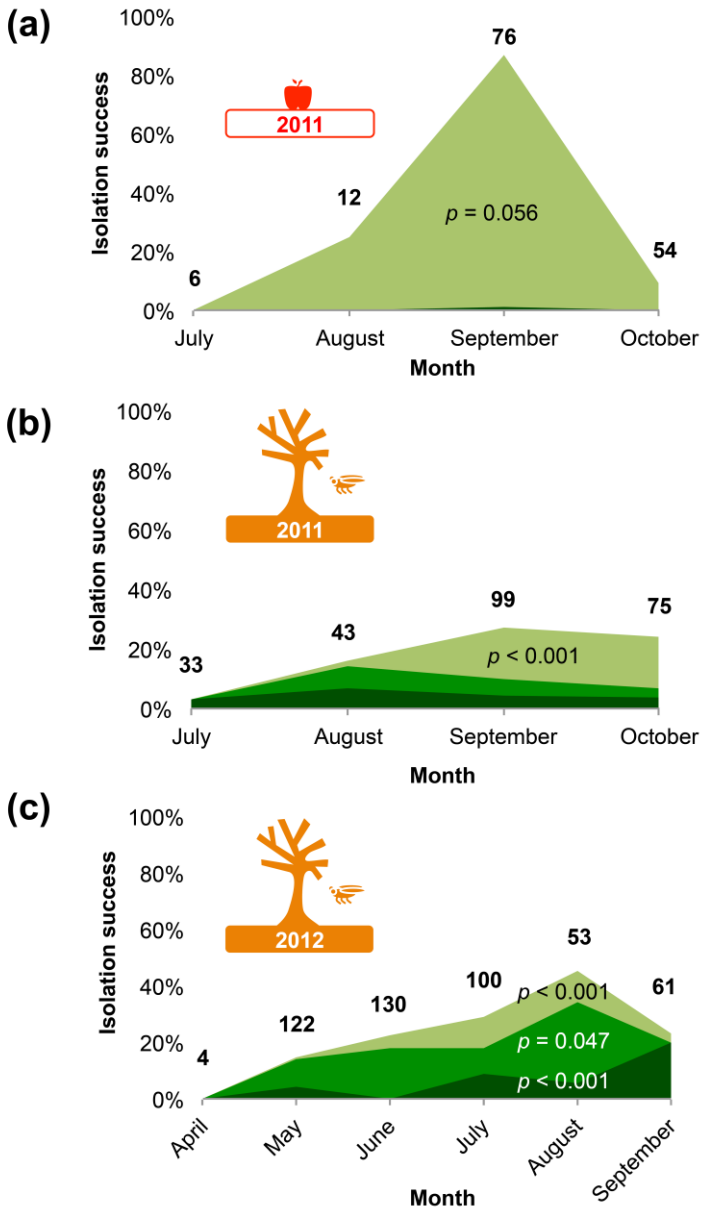

Figure 3: Season-dependent isolation success.

A) Isolation success from decayed fruits in 2011. B) Isolation success from trees, associated soils and insects in 2011 and C) in 2012. In each case, areas indicate the proportion of samples that gave at least one yeast strain (light green) for each month, one Saccharomycetaceae strain (green) or one *Saccharomyces* ssp. strain (dark green). The number of samples in each month is represented above distributions. *P*-values indicate whether isolation success significantly vary among months in each category according to a chi-square test (only significant or marginally significant differences are shown).

CHAPITRE 2: CHROMOSOMAL VARIATION SEGREGATES WITHIN INCIPIENT SPECIES AND CORRELATES WITH REPRODUCTIVE ISOLATION.

RÉSUMÉ

L'isolement reproducteur est une étape critique du processus de spéciation. Parmi les facteurs engendrant cet isolement reproducteur, les incompatibilités génétiques sont parmi les plus étudiés. Ces incompatibilités sont-elles présentes avant l'interruption du flux de gènes entre les espèces naissantes? Obtenir une réponse à cette question pose un défi particulier puisque cela nécessite de trouver des populations en cours de spéciation avant que l'isolement reproducteur soit complet entre celle-ci. Nous avons mesuré en laboratoire l'étendue de l'isolement reproducteur post-zygotique intrinsèque à l'intérieur de et entre deux lignées divergentes de la levure *Saccharomyces paradoxus* dont l'aire de répartition est partiellement chevauchante. Telle qu'attendu la viabilité des hybrides entre les lignées diminue avec la divergence génétique. Cependant, une large proportion de l'isolement post-zygotique à l'intérieur des lignées est corrélé avec des réarrangements chromosomiques. Ceci suggère qu'avant leur fixation, les différences caryotypiques pourraient contribuer aux stades précoces de l'isolement reproducteur intra-lignée. Nos observations montrent que la variation existante dans les populations pour certains des facteurs de l'isolement reproducteur intrinsèque pourrait ségréger à l'intérieur des espèces avant de contribuer à interrompre le flux génique et souligne le rôle des réarrangements chromosomiques dans la spéciation. Nous proposons différentes hypothèse utilisant les adaptations, la biogéographie ainsi que l'évolution des traits d'histoire de vie pour tenter d'expliquer nos observations.

ABSTRACT

Reproductive isolation is a critical step in the process of speciation. Among the most important factors driving reproductive isolation are genetic incompatibilities. Whether these incompatibilities are already present before extrinsic factors prevent gene flow between incipient species remains largely unresolved in natural systems. This question is particularly challenging because it requires that we catch speciating populations in the act before they reach the full-fledged species status. We measured the extent of intrinsic post-zygotic isolation within and between phenotypically and genetically divergent lineages of the wild yeast *Saccharomyces paradoxus* that have partially overlapping geographic distributions. We find that hybrid viability between lineages progressively decreases with genetic divergence. A large proportion of postzygotic inviability within lineages is associated with chromosomal rearrangements, suggesting that chromosomal differences substantially contribute to the early steps of reproductive isolation within lineages before reaching fixation. Our observations show that polymorphic intrinsic factors may segregate within incipient species before they contribute to their full reproductive isolation and highlight the role of chromosomal rearrangements in speciation. We propose different hypotheses based on adaptation, biogeographic events and life-history evolution that could explain these observations.

Many fundamental questions regarding the molecular mechanisms of speciation and reproductive isolation (RI) remain unanswered (Wolf *et al.* 2010; Butlin *et al.* 2012). Years of research on speciation have revealed that chromosomal changes (CCs) and genetic incompatibilities [negative epistatic genetic interactions, also referred as Dobzhansky-Muller incompatibilities (DMIs)], are important intrinsic mechanisms that cause reproductive barriers (reviewed in Johnson 2010; Presgraves 2010; Wolf *et al.* 2010). Another major question that recently emerged is whether these intrinsic incompatibilities are already present before geographic or other extrinsic factors prevent gene flow between emerging species (reviewed in Cutter 2012). This question is particularly challenging because it requires that we catch speciating populations in the act, that is, identifying species at early or intermediate stages of divergence before they reach the full-fledged species status (Via 2009). Observations in plants and animals (reviewed in Cutter 2012) supporting the presence of polymorphic DMIs within species, challenge the assumption that reproductive incompatibilities requires the emergence of new incompatible alleles after isolation (Seward 1910; Dobzhansky 1937b; Muller 1942). For instance, Corbett-Detig *et al.* (2013) showed that such DMIs are polymorphic within populations of the model species *Drosophila melanogaster*.

There is strong evidence for the role of DMIs and CCs in creating post-zygotic incompatibilities. How many and what type of genes underlie DMIs are major questions that have been addressed in model species (e.g. mice in Mihola *et al.* 2009; *Drosophila* in Phadnis & Orr 2009). The role of CCs in initiating and maintaining reproductive barriers has been noticed early in the study of speciation cytogenetics (White 1978) and they play important roles in particular taxonomic groups such as plants (reviewed in Hoffmann & Rieseberg 2008; Widmer *et al.* 2009). CCs may contribute to species incompatibilities in several ways (Rieseberg 2001; Hoffmann & Rieseberg 2008). First, heterokaryotypes from crosses between individuals with chromosomal differences are often infertile due to the accumulation of chromosomal abnormalities during meiosis, which leads to hybrid-sterility (Sturtevant & Beadle 1936; Davisson & Akeson 1993). Another role of CCs is that they enhance the emergence and maintenance of some genetic elements by suppressing

recombination (Navarro & Barton 2003; Feder & Nosil 2009), by reducing gene flow between ecologically similar populations (Navarro & Barton 2003; Noor & Feder 2006; Widmer *et al.* 2009), or by enhancing adaptation to new environments (Kirkpatrick & Barton 2006; Hoffmann & Rieseberg 2008; Widmer *et al.* 2009; Lowry & Willis 2010). Finally, large-scale changes such as changes in ploidy may directly contribute to speciation by creating individuals that become reproductively isolated from their parental species within a single generation (Otto & Whitton 2000). This mechanism appears to be important in plants where a large proportion of speciation events are accompanied by ploidy increase (Wood *et al.* 2009).

The role of CCs in speciation may also be important in fungi where there are many reports of extensive CCs among closely related species as well as extensive variation in RI within and among species (Giraud & Gourbiere 2012). Interestingly, many of these CCs may be adaptive, as shown in natural and experimental yeast populations (Gordon *et al.* 2009; Dunn *et al.* 2013). These observations suggest that natural selection could accelerate the fixation of CCs in natural populations and thus contribute to the onset of RI and speciation at the same time. There is strong experimental evidence for a direct causal link between CCs and RI in the model fungal species of the genus *Saccharomyces* and *Schizosaccharomyces*. For instance, Delneri *et al.* (2003) showed that a single chromosomal inversion was sufficient to affect progeny survival in a cross between *Saccharomyces cerevisiae* and *S. mikatae*. In other studies of species of the same genus, both CCs and DMI have been suggested to contribute to RI (Liti *et al.* 2006). More recently, Hou *et al.* (2014) showed that reciprocal translocations play a role in creating RI among strains within *S. cerevisiae*. Finally, Zanders *et al.* (2014) showed that chromosomal rearrangements were playing a role in RI between *Schizosaccharomyces pombe* and *S. kambucha*, which can mate to form F1 hybrids but produce very few viable F2s. These observations stress the potentially fundamental role of polymorphic CCs in initiating RI among incipient species and show that fungi are powerful and tractable model systems for investigating these effects. However, due to the lack of knowledge on the ecology and biogeography of these species, it is difficult to link the mechanisms of reproductive isolation with the ecological and demographic factors that initiated the process in these

organisms (Landry *et al.* 2006; Replansky *et al.* 2008). Moreover, Zeyl (2014) recently pointed out that our current knowledge on the genetic bases of early speciation is mostly limited to few laboratory and mostly inbreeding organisms, stressing the need to investigate these processes in a completely natural context.

We recently identified a new system within natural populations of *Saccharomyces paradoxus*, the sister species of the laboratory model *S. cerevisiae*, that offers an unprecedented case study in ecological and speciation genomics (Figure 4a). North American *S. paradoxus* populations are composed of three principal lineages: two American lineages (B and C) and a European lineage (A) that was introduced recently in a transoceanic migratory event from a European source population (Kuehne *et al.* 2007; Leducq *et al.* 2014; Figure 4b). We recently showed that the two American lineages occupy different but overlapping regions (Figure 4c), are genetically distinct and are phenotypically divergent in a manner that suggests the influence of historical and ecological factors in driving their divergence (Leducq *et al.* 2014). Based on these observations, we propose that the two American lineages of *S. paradoxus* B and C are in an early speciation process. Here, we tested this hypothesis by measuring postzygotic RI between 25 *S. paradoxus* strains representative of the American lineages. Intrinsic post-zygotic RI was measured as the fraction of viable progeny within a cross. Genetic divergence globally explained variation in RI measured among strains. In a second step, we performed chromosomal profiling analysis of the parental strains to examine the potential role of CCs in RI. Surprisingly, we found more variation in karyotypic profiles within than among lineages, and this variation correlated with RI measured within lineages. We discuss the potential role of CCs in initiating RI within incipient species and the putative ecological and biogeographical origins of these rearrangements.

MATERIAL AND METHODS

Strain collection

We used 25 natural strains of *S. paradoxus* representative of the geographical distribution and of the three genetic lineages found in North-East America (Leducq *et al.* 2014): A (4),

B (7) and C (14) (Figure 4c, Table S5). Most strains were isolated on the bark of deciduous trees, and all were isolated according to the same standard protocol (Kuehne *et al.* 2007; Leducq *et al.* 2014).

Heterothallic strain construction and spore viability

We deleted the *HO* gene in wild diploid strains (Table S5) to prevent mating-type switching after sporulation. We amplified *HO* deletion cassettes comprising an antibiotic marker by PCR using oligonucleotides specific to each lineage. We designed lineage A- and B-specific oligonucleotides from available genomic data (Cubillos *et al.* 2009) and lineage C-specific oligonucleotides from the *HO* locus sequence that we obtained from five representative strains (LL2012-006, LL2011-005, LL2011-001, LL2012-016 and YPS667; Leducq *et al.* 2014) after genomic DNA extraction (Amberg *et al.* 2005), standard PCR amplification and Sanger sequencing (NCBI accession numbers KJ410184-88). Cassette amplification, transformation by homologous recombination and cassette integration verification followed Leducq *et al.* (2012), with heat shock performed at 37°C instead of 42°C for lineage A and C (oligonucleotides and plasmids are listed in Table S6).

We obtained heterothallic haploid strains by sporulation (Ramaswamy *et al.* 1998) and dissection of the parental diploid strains constructed above. Sporulated cells were dissected (six tetrads per strain) on a YPD (yeast extract, peptone, dextrose) agar plate (2%) using a Sporeplay dissection microscope (Singer, UK) to obtain one spore of each mating type. Mating types were verified by crosses with tester strains of both mating types with antibiotic resistance markers other than the one of the tested strain. Mating types were inferred from growth of the strains on YPD agar plate containing both antibiotics.

We performed 87 crosses by optimizing possible pairwise comparisons in terms of geographical distribution and genetic divergence (Leducq *et al.* 2014) among 25 haploid *S. paradoxus* American strains (Table S7). As controls, we included six crosses of strains from each of the three lineages with two other members of the *Saccharomyces* genus, *S. cerevisiae* strain BY4741 (Baker Brachmann *et al.* 1998) and *S. uvarum* strain JRY8147 (Gallagher *et al.* 2009; Table S5). We inoculated 5 mL of fresh YPD medium with 5 µL of

each haploid strain to be crossed and incubated them at 30°C for 24h with agitation. We selected diploids on YPD-agar plates with the two appropriate antibiotics (G418: 200µg/mL; CloNAT: 100µg/mL; Hygromycine: 250µg/mL) that were incubated at 25°C for 48h and isolated one diploid clone for each cross. We sporulated the clones and dissected 24 tetrads per cross (sporulation and dissection as described above), for a total of 96 possible spores. Dissected spores were incubated at 25°C for 72h and we measured spore survival (*S*) for each cross as the proportion of spores forming a colony visible to the unassisted eye (Figure 5a).

Karyotype analysis

We prepared yeast chromosomes following Maringele & Lydall (2006) with the following modifications: incubations steps in EDTA-Tris-β-mercaptoethanol and proteinase solutions were conducted overnight and voltage gradient used for the migration was set to 5.9V/cm. All migrations were performed in a CHEF DR® III variable angle system (Biorad, USA) and the buffer was systematically changed every two migrations. Each gel contained 9 samples of *S. paradoxus* (unmodified parental strains) and a *S. cerevisiae* CHEF DNA size standard (Biorad, USA). Gels were stained with ethidium bromide and imaged using a Biorad GEL DOC XR+ molecular imager with the image lab software (Biorad, USA; Figure 5b). Images were analyzed using Bionumerics version 7.1 by Applied Math NV (available from http://www.applied-maths.com). Dice's similarity coefficient was used to establish the similarity distance of band patterns based on their position after gel migration. The migration distance of a chromosome is inversely proportional to its length and a change in chromosomes length by large insertions and deletions translates into a shift in banding pattern. Parameters were determined as to obtain an optimal clustering of the *S. cerevisiae* standard markers. Standard settings were used with the following modifications: Optimization 0.5%, Tolerance 1.2% and uncertain bands included.

Statistical analyses

Our aim was to test whether we could explain variation in spore survival (*S*) using two potential intrinsic factors of post zygotic reproductive isolation: genetic divergence (*Dg*) and karyotypic divergence based on banding patterns (*Dc*) between parental strains. We estimated *Dg* from two unlinked markers (1460 bp; Leducq *et al.* 2014) using pairwise

nucleotide divergence among strains (Maximum Composite Likelihood model; Tamura *et al.* 2011). We used a linear model (*lm*) to explain variation in *S* by considering *Dg* and *Dc* as additive and interacting factors (model M1: *S~Dg+Dc+Dg*Dc*) and estimated the significance of the correlation (Pearson's test) between *S* and each intrinsic factor. We repeated this test by discarding the possible interaction between intrinsic factors, i.e. by calculating the correlations between *Dg* and *S* corrected by *Dc* (residuals from *lm*: *S~Dc*) and between *Dc* and *S* corrected by *Dg* (residuals from *lm*: *S~Dg*). In order to distinguish small- from large -scale effects of intrinsic factors, we repeated the analysis within lineages only (i.e. by excluding interlineage and interspecies crosses; model M2) and among lineages or species only (i.e. by excluding intralineage crosses; model M3). Details of the models are shown in Table S8. We used Welch *t*-tests to compare averaged *S* among categories of crosses according to genetic lineages. For each test (Pearson's tests, linear models and Welch *t*-tests), we re-estimated the statistics after 100,000 randomizations of *S* values (intrinsic factor fixed to correct for the fact that not all strains were involved in the same number of crosses). We reported *P*, the proportion of randomly re-estimated statistics that were equal or lower than the observed value. Figure S1 summarizes the method of randomization used, considering that pairwise comparisons between strains were not independent (i.e. strains were involved in different crosses) and unbalanced (i.e. not all combinations of crosses were tested and some strains were over-represented). We then considered that the significance of the test was not due to the structure of our data when the observed statistic was significantly higher ($P \geq 0.95$) or lower ($P \leq 0.05$) than the re-estimated statistics. All analyses were conducted in R 3.0.0 (R Development Core Team 2010). All datasets and scripts used are available as Appendices 1 and 2, respectively.

RESULTS

Intra- and interlineage variation in reproductive isolation

We tested whether the North American lineages of *S. paradoxus*, A, B and C (Figure 4), were postzygotically reproductively isolated. We performed pairwise crosses between strains of lineages A (European lineage but isolated in North America), B and C. Crosses with *S. cerevisiae* and *S. uvarum* yielded low hybrid viability (0-2% of progeny survival;

Figure 6a; Table S7) as reported before (Naumov 1987; Greig 2009). Within *S. paradoxus*, we observed the lowest average spore survival rate (*S*) in crosses between lineages, with 42.1% (A×B), 31.9% (A×C) and 39.7% (B×C) spore viability (Figure 6a), which is in the range previously observed among lineages (Sniegowski *et al.* 2002; Liti *et al.* 2006; Kuehne *et al.* 2007). B×C crosses showed a significant decrease in average *S* as compared to intralineage crosses B×B (58.6%; *p*-value = 0.014, Welch t-test) and C×C (61.6%; *p*-value<0.001, Welch t-test; Figure 6a) and these reductions are significantly stronger than expected over 100,000 random permutations (*P*=0.997 and *P*=0.999, respectively; Figure S2). Surprisingly, we found extensive intralineage reproductive isolation in both lineages B and C, with minimum *S* values as low as 35.4% (B×B) and 19.8% (C×C). This is particularly extreme in lineage C, with some C×C spores having less viability than A×B, A×C or B×C hybrids. This effect was mainly explained by four strains from lineage C from two distinct locations (LL2011_005, LL2011_006, LL2012_016 and LL2012_18; Table S7), which were involved in 16 crosses averaging 42.5% spore viability with other lineage C strains, while they yielded significantly increased spore viability when crossed with each other (86.2%, four crosses; *p*-value<0.001, Welch *t*.test; Table S7). Removing crosses involving these four strains from the analyses (20 crosses) increases the C×C progeny viability to 73.9%, which remains significantly higher than B×C (*p*-value<0.001, Welch *t*.test; *P*=0.999).

Extensive within-lineage chromosomal variation and interlineage nucleotide divergence explain variation in reproductive isolation

We examined whether the extent of reproductive isolation among strains (*S*; Figure 6a) was correlated with the extent of genetic divergence (*Dg*; Figure 6b) and differences in their chromosome profiles (*Dc*; Figure 6c). The extent of variation in *Dc* based on chromosomal banding patterns observed among lineages (0.263-0.387) is similar to what was measured between *S. cerevisiae* and *S. paradoxus* (0.196-0.404; Figure 6c). Surprisingly, we observed similar amount of variation within lineages B and C (0.044-0.337; Figures 5b and 6c). This is mostly because chromosomal profiles are more variable within lineage C (*Dc* = 0.337 ± 0.207) than within lineages A and B (*Dc*=0.044±0.011 and 0.271±0.089, respectively). This extensive variation in lineage C chromosome profiles was explained by

some pairs of strains from lineage C being less similar than some B-C pairs. For instance, the four aforementioned strains (LL2011_005, LL2011_006, LL2012_016 and LL2012_18) have on average more karyotypic similarity with each other (*Dc*=0.080±0.062) than with other lineage C strains (*Dc*=0.423±0.062; *p*-value<0.001, Welch *t*.test). We found that globally, *S* was significantly negatively correlated with both nucleotide divergence *Dg* (*r*=-0.64, *p*-value<0.001, Pearson; Figure S3a, f) and chromosomal divergence *Dc* (*r*=-0.34, *p*-value<0.001, Pearson; Figure S3b, f) among strains. These correlations remain stronger than expected over 100,000 random permutations (*P*<0.001) and were intensified when correcting *S* by karyotypic divergence (*S~Dg*: *r*=-0.66, *p*-value<0.001, Pearson; *P*<0.001; Figure 6d) or by genetic divergence (*S~Dc*: *r*=-0.40, *p*-value<0.001, Pearson; *P*<0.001; Figure 6e). Our linear model (M1) showed that both *Dc* and *Dg* explain a significant part of variation in *S* among crosses (*p*-value<0.001 in both cases) but that these two parameters are not correlated with each other (*p*-value=0.130; Figure S3c, Table S8). The part of variation in *S* explained by *Dg* and *Dc* remain significantly higher than expected over 100,000 random permutations (*P*=0.011 and *P*=0.004, respectively; Figure S4). Analyzing separately within-lineage crosses (model M2) from among-lineage/species crosses (model M3) revealed that the intra-lineage variation in *S* is only explained by chromosomal differences (M2; *Dg*: *p.value*=0.279, *P*=0.291; *Dc*: *p.value*<0.001, *P*=0.005; Table S8), which was confirmed by a significant correlation between *S* and *Dc* within lineages (*r*=-0.46, *p*-value<0.001, Pearson; *P*=0.001; Figure S3d,f), but not among lineages (*r*=-0.061, *p*-value=0,707, Pearson; Figure S3e). The interlineage variation in *S* is only marginally explained by genetic divergence (M3; *Dg*: *p.value*=0.084; *Dc*: *p.value*=0.559; Table S8) and the effect is rather due to the structure of data (*P*=0.490 and *P*=0.513, respectively; Figure S4).

DISCUSSION

Intrinsic postzygotic reproductive isolation results from various genetic mechanisms such as differences in ploidy levels, gene-gene incompatibilities, interaction between endosymbionts and the nuclear genome, and chromosomal rearrangements (Orr & Turelli 2001; Rieseberg 2001; Coyne & Orr 2004; Noor & Feder 2006; Hoffmann & Rieseberg

2008). From these, gene-gene incompatibilities are a very important mechanism, as they were reported to play a role in hybrid inviability and sterility in both animals and plants (Orr & Turelli 2001; Coyne & Orr 2004). Earlier work on the *Saccharomyces* genus found that nucleotide divergence together with chromosomal changes may play an important role in reproductive isolation via the mismatch repair system (Greig *et al.* 2003). Recent data suggests that cytonuclear interactions may also generate some incompatibilities, resulting in reproductive isolation between more distant species (Chou *et al.* 2010).

Our results show a negative correlation between spore viability and genetic divergence among lineages (Figure 6d), which supports the accumulation of genetic incompatibilities as a factor in the reproductive isolation among natural lineages of *S. paradoxus* found in North American. This result is consistent with the hypothesis that genetic incompatibilities between lineages accumulate with time and may be correlated with or be caused by global nucleotidic divergence. This result is also in line with the continuum of genetic differentiation and incompatibility that exists between and within species of the *Saccharomyces sensu stricto* group, which may result from the decay in sequence homology along this continuum (Liti *et al.* 2006). The partial intrinsic post-zygotic reproductive isolation between lineages B and C may represent an advanced case of this continuum of reproductive isolation. Unexpectedly, we found that a fraction of this reproductive isolation may have been present in the ancestral population of these two lineages and be segregating within lineages or to have accumulated recently, as we observed within lineage reproductive isolation in both lineages B and C. We observed strains within lineage C that, when crossed with other strains of the lineage, have spore viability in the range of the B×C hybrids. However, crosses among the same strains produced spores with the viability we would expect within a lineage. These strains were isolated in two different locations distant from more than 50 km (Pointe Platon and Île d'Orléans; Figure 4c, Table S5) and thus represent a part of diversity segregating within lineage C. This suggests that alleles segregating within lineage C contribute to reproductive isolation and that the sorting of these elements is incomplete. Our results support the hypothesis proposed by Corbett-Detig *et al*. (2013) that genetic incompatibilities are common within species.

Chromosome analyses revealed a correlation between reproductive isolation and the divergence of chromosome profiles. Lineage C has more variation in chromosome profiles than the other two lineages, suggesting that it has accumulated more genomic insertions or deletions or rearrangements than other strains. For instance, the four aforementioned strains present about 40% of chromosome divergence with other strains from lineage C, in the same range observed between different lineages or species (Figure 6c). This correlates with a higher variation in reproductive isolation within this lineage. The evolutionary forces that have contributed to the accumulation of CCs are currently unknown but several scenarios involving adaptive and non-adaptive forces can be put forward. Genomic rearrangements and CCs in yeasts have been observed as having adaptive roles under stressful environmental conditions. Experimental evolution studies reported that long time starvation leads to large-scale genomic rearrangements in *S. cerevisiae,* which improved fitness in new starvation episodes (Dunham *et al.* 2002; Coyle & Kroll 2008; Hong & Gresham 2014). In addition, wild *S. cerevisiae* strains isolated from the Evolution Canyon harbored genomic rearrangements that were correlated with their resistance to toxic concentration of copper in their environment (Chang *et al.* 2013). Thus, one possibility is that sub-populations of the C lineage may have experienced stressful local environmental conditions during the years since they colonized this environment, and that these rearrangements were a mean of rapid adaptation (Figure 7a) and contribute to reproductive isolation among sub-populations as a side effect. Interestingly, some chromosomal rearrangements associated with adaptation to starvation resistance in *S. cerevisiae* were shown to influence the viability of hybrids in experimental evolution settings (Kroll *et al.* 2013), thus coupling the mechanisms of adaptation and reproductive isolation.

Neutral changes may also play a role in the accumulation of CCs. Evolutionary models implying CCs in speciation have been criticized based on the fact that strong genetic drift would be necessary to fix CCs because heterozygous individuals for these CCs are expected to have lower fitness early in the process (Walsh 1982; Lande 1985; Rieseberg 2001). Models developed since then suggested that CCs primarily acted by reducing recombination and that they may have little effects on fitness in heterozygotes (Rieseberg

2001; Navarro & Barton 2003; Noor & Feder 2006). However, the mode of reproduction of budding yeasts may allow for the neutral accumulation of these CCs by genetic drift. The life cycle of the *Saccharomyces* yeasts is comprised of mostly asexual reproduction cycles punctuated by rare sexual reproduction events that, most of the time, end up in the mating of spores from the same tetrad (inbreeding; Tsai *et al.* 2008). The colonization of new environments (Figure 7a) could have changed these life history traits (Nosil 2012) and thus played a role in the accumulation of CCs. Strains of lineage C could mate or outbreed at lower frequencies. Clonal reproduction or sexual reproduction mostly limited to inbreeding would render reproductive incompatibilities among genotypes effectively neutral because heterozygous individuals are rarely produced, leading to their accumulation (Figure 7b). Two lines of evidence support this model. Lineage C appears to have a more limited geographical distribution than lineage B and a lower density, as suggested by the lower sampling success for *S. paradoxus* strains in the north of its geographic distribution (Charron *et al.* 2014). In addition, average nucleotide diversity in lineage C (0.04+0.01%) is much lower than in lineage B (0.33+0.09%; Leducq *et al.* 2014), consistent with a reduced effective population size. Another possibility is that these changes are generally deleterious and accumulate in lineage C through random genetic drift, without the need for changes in life history.

Both the adaptive (natural selection favoring the accumulation of CCs) and non-adaptive mechanisms (drift allowing the accumulation of CCs) of our models are tightly linked with the recent changes in the habitats occupied by *S. paradoxus* in North America (Figure 7b). The repeated glaciation events during the Pleistocene have been shown to be a factor that has promoted allopatric speciation via the Mississippian and Atlantic glacial refugia in several taxonomic groups, for instance in freshwater fishes species (April *et al.* 2013) and in *Drosophila* (Ford & Aquadro 1996). This may have also been the case for yeasts, as the break in the B-C distribution corresponds to the biogeographic pattern observed in North-East American taxonomic groups above (Figure 4c). Freshwater fishes and some *Drosophila* species are separated in west and east lineages with a secondary contact zone situated along Lake Ontario and the St. Lawrence River, exactly where a geographic co-occurrence zone between B and C was found (Leducq *et al.* 2014). The northern

distribution and the putatively low population size of lineage C suggest a colonization event from a glacial refugium, which is also congruent with the evidence of such refugium in the North-America ice shelf for deciduous trees on which yeasts reside, for instance *Quercus* (Jackson *et al.* 2000). Historical biogeographic barriers may thus have contributed not only to shape the plant and animal biodiversity of this region but also the microbial one.

Overall, our work suggests that the two North American endemic lineages of *S. paradoxus* are at the initial stage of speciation and thus offer a new model for the study of this process at the molecular level in a genetically tractable and naturally evolving species. The large remaining, yet unexplained, variance in reproductive isolation we observed within lineages also suggests complex intrinsic mechanisms such as polymorphic incompatibilities segregating within lineages. In addition, the mechanism by which RI could be caused by CCs need further investigation and will require additional analyses that allow finer resolution than the method used here, which can only detect changes that affect chromosome sizes and could miss inversions, which are known to contribute to reproductive isolation in other taxonomic groups (Noor *et al.* 2001; Kirkpatrick & Barton 2006; Hoffmann & Rieseberg 2008; Feder & Nosil 2009; Lowry & Willis 2010). Combined with the genetics and genomics resources of the budding yeasts, this model offers an unprecedented opportunity for the study of the role of evolutionary forces such as natural selection and genetic drift on genes and genome architecture at the early stage of speciation.

ACKNOWLEDGEMENTS

We thank Emilie Bernatchez, Marie Filteau, Anne-Marie Dion-Côté, Nadia Aubin-Horth, Guillaume Diss and Alexandre K. Dubé for comments on the manuscript. We thank Roger C. Levesque and Iréna Kukavica-Ibrulj for the use of their CHEF-PFGE apparatus. We thank Raghav Nuwal for help in performing crosses during the review process. This work was funded by a NSERC discovery grant to CRL and partially by a HFSP grant to CRL. JBL was supported by a FRQS fellowship and GC by a PROTEO fellowship. CRL is a CIHR New investigator.

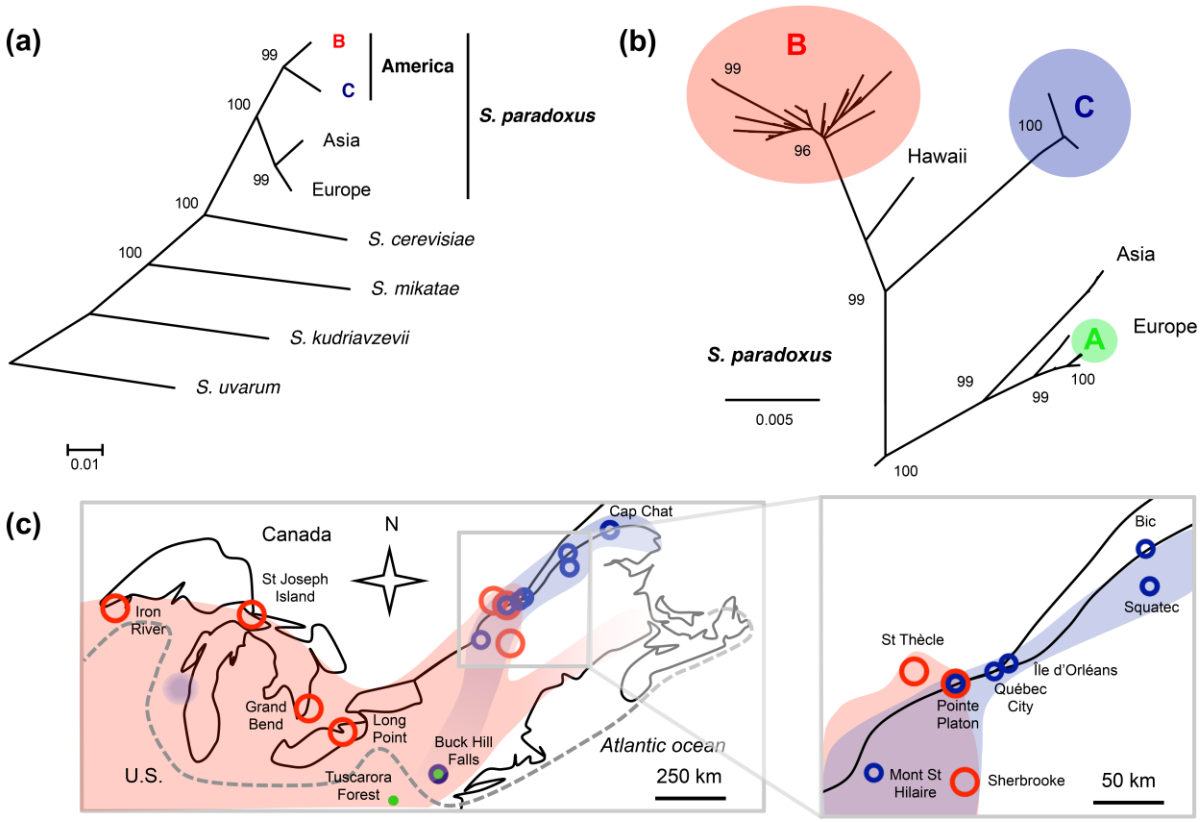

Figure 4: Two diverging lineages of *S. paradoxus* are partially sympatric in North America.

a) Phylogenetic position of main *S. paradoxus* lineages in *Saccharomyces sensus stricto*. b) Details of the *S. paradoxus* phylogeny. American lineages are denoted as A (recent European introduction; green), B (red) and C (dark blue). The phylogenetic trees are adapted from Leducq et al. (2014) (pairwise nucleotide divergence). c) The geographic distribution of lineages B (red) and C (dark blue) in North-East America could reflect colonization and secondary contact after the last glaciation approximately 110,000 to 12,000 years ago (southern limit of the ice shelf indicated by a dotted line; Andersen & Borns 1997). Lineage distributions were updated based on additional sampling efforts in 2013 (personal observations; CT Hittinger, personal communication). Gradients indicate that the eastern and southern limits of lineages B and C are defined by rare isolates. The sampling locations of 25 strains used in this study are indicated by colored circles.

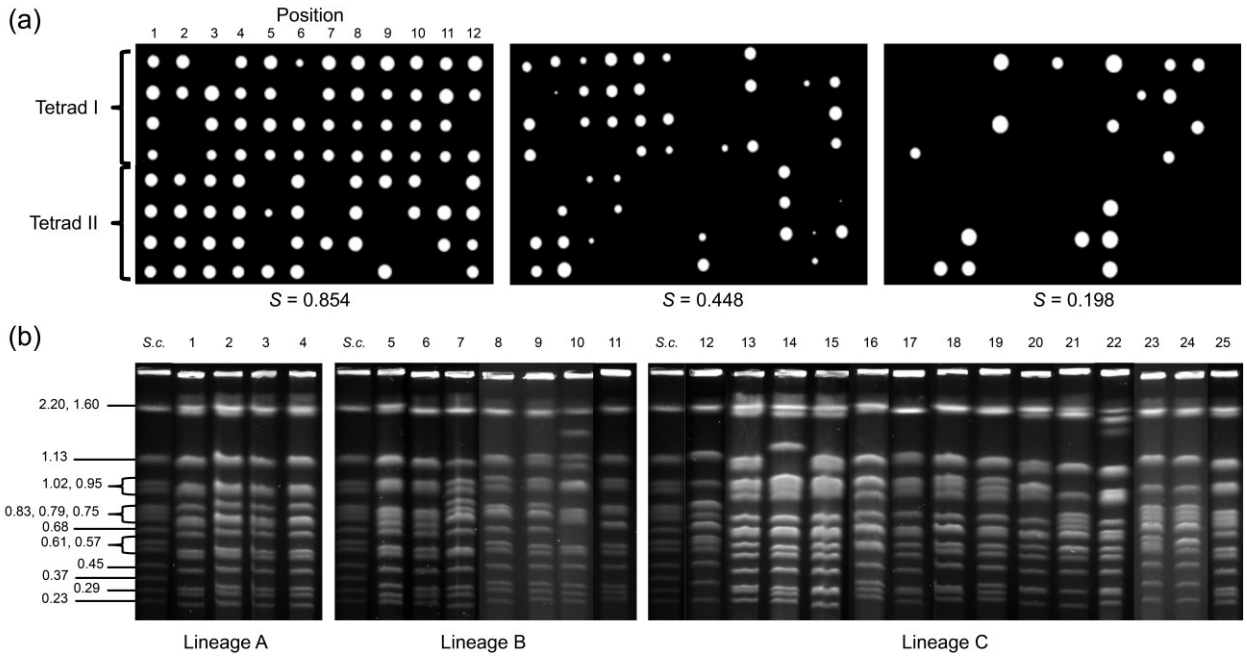

Figure 5: Reproductive isolation and karyotyping in *S. paradoxus* North-American strains.

a) Three examples of spore viability (*S* value indicated below) measured as colony growth after tetrad dissection. Each position represents two dissected tetrads (I and II), with four spores for each tetrad (aligned vertically). b) Karyotypes of strains used in this study as measured by CHEF-PFGE. Each band represents at least one chromosome. Numbers above the karyotypes refer to strains, grouped by lineages (see Table S5). *S.c.* represents *S. cerevisiae* chromosomes standard (chromosome lengths are given in Mbp on the left).

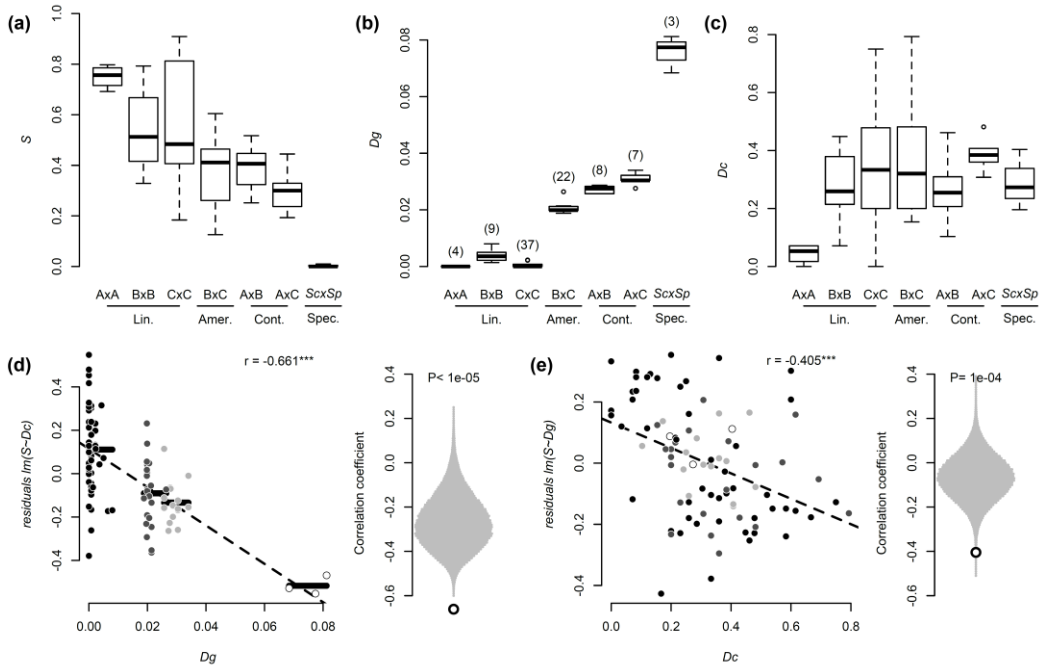

Figure 6: Spore survival (*S*) among 25 *S. paradoxus* North-American strains correlates with variation among and within lineages in genetic divergence (*Dg*) and chromosome profiles (*Dc*).

(a) Boxplot showing the distribution of *S* among strains within each evolutionary category: within-lineage crosses (Lin.); among American lineages, B×C crosses (Am.); between European and American lineages, A×B and A×C crosses (Cont.) and among species, *S. paradoxus* × *S. cerevisiae* crosses (Spec.). (b) Boxplot illustrating the distribution of *Dg* among strains within each evolutionary category. (c) Boxplot illustrating the distribution of *Dc* among strains within each evolutionary category. (d) *S* significantly decreases with *Dg*. *S* values were corrected by *Dc* (residues from linear model *S*~*Dg*). (e) *S* significantly decreases with *Dc*. *S* values were corrected by *Dg* (residues from linear model *S*~*Dg*). In the left panels of (d) and (e), different colors symbolize different evolutionary categories: within-lineage (black); crosses between American lineages B×C (dark grey); crosses among European and American lineages, A×B and A×C (grey) and crosses between species *S. paradoxus* × *S. cerevisiae* (white). Horizontal black bars represent mean *S* value in each category. The dotted line shows the linear regression of *S* as a function of *Dc* (d) or *Dg* (e). Correlation coefficients *r* (Pearson's test) are highly significant ("***": *p*-value<0.001). Violin plots on the panels indicate that the observed correlation coefficients (black circles) are stronger than expected (grey distribution) over 100,000 random permutations (*P*<0.001; see Figure S1).

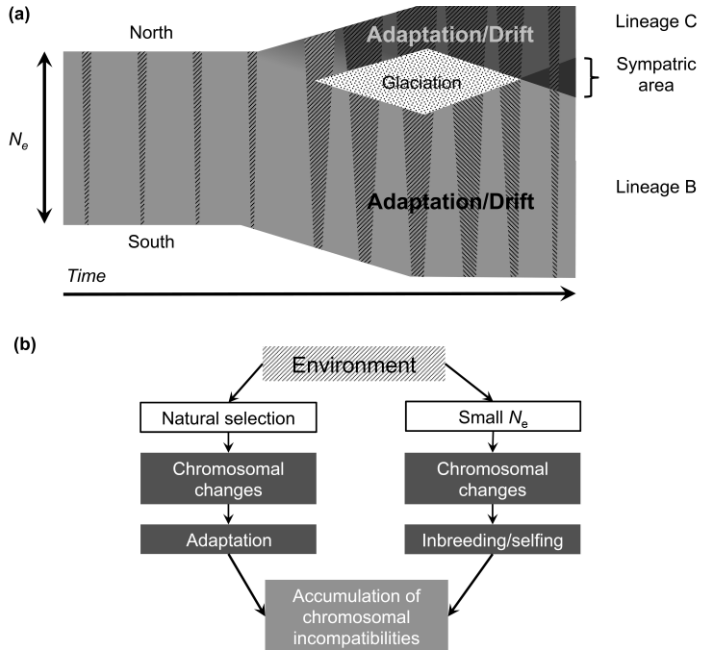

Figure 7: Proposed scenario for the divergence and secondary contact of the North American lineages of *Saccharomyces paradoxus*.

a) The horizontal timeline represents the evolution of a putative ancestral American *S. paradoxus* population over time (horizontal arrow) into Lineages B (grey) and C (dark grey). Lineages may have diverged after allopatric isolation of lineage C ancestor in a glacial refugium (dotted area). A secondary contact (sympatric area, black) took place after the last glaciation event in the region (110,000 to 12,000 years ago). Given its northern distribution, lineage C was likely exposed to more frequent and long episodes of extreme conditions (hatched areas) than lineage B. The effective population sizes of the two lineages may have been influenced by these historical and biogeographical events. DMI and chromosomal changes may have accumulated through natural selection or genetic drift in each lineage. Selection could have been divergent selection or similar selective pressures that would have fixed alternative alleles. b) Neutral and selective processes may have led to the accumulation of chromosomal incompatibilities within and between *S. paradoxus* American lineages. A stressful environment at the northern limit of the two lineages may have favored evolution of adaptive chromosomal changes in local populations. For instance, chromosomal changes could be adaptive in themselves through changes in gene dosage (Chang *et al.* 2013) or by the capture of locally adapted alleles (Kirkpatrick & Barton 2006). Alternatively, these environments may have favored the maintenance of small populations, which favored inbreeding and made chromosomal changes effectively neutral due to the absence of outcrossing and thus of the formation of heterozygous individuals.

CONCLUSION GÉNÉRALE

L'étude de l'émergence des barrières à la reproduction est nécessaire à notre compréhension du processus qu'est la spéciation. Au cours des dernières décennies, un grand intérêt a été porté sur le rôle de l'écologie dans la divergence des populations et dans l'émergence de nouvelles espèces (Hatfield & Schluter 1999; Nosil *et al.* 2009; Schluter & Conte 2009; Via 2009). Plus récemment, une question a été soulevée quant au rôle du polymorphisme au tout début de la spéciation (Cutter 2012). Étudier ces questions serait fortement facilité par le développement d'un système de populations naturelles sur lesquelles des données génomiques, écologiques et fonctionnelles pourraient être obtenues ou inférées en laboratoire. Depuis le début des années 2000, le genre Saccharomyces s'est développé en tant que genre modèle pour l'écologie, la génomique et l'évolution (Fischer *et al.* 2000; Landry *et al.* 2006; Liti *et al.* 2006; Replansky *et al.* 2008). Parmi les levures du genre Saccharomyces, les populations nord-américaines de S. paradoxus représentent une occasion unique de répondre aux questions mentionnées précédemment. Elles offrent une histoire évolutive fortement influencée par l'écologie (Leducq *et al.* 2014) ainsi que l'opportunité d'étudier leur génome, leur métabolisme et leur isolement reproducteur avec des méthodes moléculaires puissantes.

Nos objectifs étaient i) de caractériser la diversité des populations de *Saccharomyces paradoxus* présente dans le nord-est de l'Amérique du Nord, ii) de tester la présence d'un isolement reproducteur post-zygotique intrinsèque entre deux lignées divergentes nord-américaines de *S. paradoxus* et iii) d'identifier la contribution de la divergence génétique et des réarrangements chromosomiques à cet isolement reproducteur. Différentes souches de levures ont été utilisées pour répondre à ces questions. Ces souches provenaient d'un échantillonnage de levures du nord-est de l'Amérique du Nord et sur une variété de substrat. L'analyse subséquente des souches de *S. paradoxus* obtenus dans cet échantillonnage a révélé la présence de trois lignées génétiquement divergentes de *S. paradoxus*. Deux de ces lignées (B et C) sont retrouvées en plusieurs endroits, mais semblent occuper des aires de répartition distinctes et partiellement chevauchantes (Leducq *et al.* 2014). Ces lignées présentent aussi une certaine divergence phénotypique, ce qui nous porte à penser qu'elles pourraient faire partie d'un évènement de spéciation en cours

(Leducq *et al.* 2014). Suite à des croisements entre individus de ces lignées en laboratoire, on a pu déterminer la présence d'un isolement reproducteur post-zygotique intrinsèque partiel entre les souches des deux lignées principales. Parallèlement, avec une approche de caryotypage par PFGE, on a pu observer une variation importante des profils chromosomiques à l'intérieur des groupes. La survie des spores, utilisée pour estimer l'isolement reproducteur post-zygotique intrinsèque, est corrélée négativement avec la divergence génétique entre les lignées et avec la divergence caryotypique à l'intérieur des lignées.

Distribution des levures Saccharomyces au Québec

Notre échantillonnage d'une grande partie du sud de la province de Québec a révélé une grande diversité de levures mésophiles présentes sur différents substrats (chp. 1). Beaucoup de levures de la famille des Saccharomycetaceae ont pu être isolées de l'écorce des arbres feuillus comme les chênes et les érables. Parmi celles-ci, plusieurs souches de *S. paradoxus* ont été isolées, mais nous n'avons pas été en mesure d'établir si elles vivent en sympatrie avec d'autres espèces de *Saccharomyces* comme cela a souvent été rapporté à d'autres endroits dans le monde (Naumov *et al.* 1998; Sniegowski *et al.* 2002; Sampaio & Goncalves 2008; Zhang *et al.* 2010). La capacité de croissance de *S. cerevisiae* étant plus faible à basse température pourrait jouer en faveur de *S. paradoxus* si elles sont en compétition pour la même niche écologique, ce qui pourrait expliquer l'exclusion de *S. cerevisiae* de notre aire d'échantillonnage. Pour ce qui est des espèces psychrophiles comme *S. uvarum* et *S. kudriavzevii*, il est possible que l'incubation à 30 °C soit trop spécifique aux levures mésophiles, entraînant la perte des levures moins adaptées aux températures élevées dans les cultures d'enrichissement. Un échantillonnage à plus grande échelle et une modification des paramètres d'isolement des levures pourraient nous donner une meilleure vue d'ensemble de la diversité des levures Saccharomyces présentes au nord-est de l'Amérique du Nord.

Analyse de l'isolement reproducteur chez S. paradoxus

L'isolement reproducteur post-zygotique partiel nous indique la présence de certaines incompatibilités entre les souches des lignées américaines de *S. paradoxus*. Cependant, en

se basant l'identification des souches par PCR (méthode chp. 1) et par une méthode RFLP (Leducq *et al.* 2014), aucun hybride n'a pu être isolé de l'environnement. Il est possible qu'un autre type d'isolement reproducteur plus complet soit présent entre les lignées B et C et que l'inviabilité partielle des hybrides soit observable en laboratoire, mais ne soit pas la barrière la plus importante au flux de gènes. Maclean and Greig (2008) ont observé la présence d'un isolement reproducteur prézygotique entre les espèces de levures *S. cerevisiae* et *S. paradoxus* par une préférence de conjugaison entre individus de la même espèce. Étant capable de discriminer les spores de leurs lignées respectives et de se conjuguer préférentiellement avec ces derniers diminuerait de beaucoup les probabilités d'observer un hybride en nature. Un second type d'isolement reproducteur susceptible d'être trouvé pourrait être de l'isolement reproducteur post-zygotique extrinsèque. Selon les phénotypes étudiés par Leducq *et al.* (2014), les lignées B et C pourraient être adaptée à des environnements différents. Dans l'éventualité où les hybrides possèderaient des phénotypes intermédiaires ou inexistants chez l'une ou l'autre des lignées parentales, il se pourrait qu'ils ne perdurent pas assez longtemps dans l'environnement et que les probabilités d'en isoler soient très faibles ou encore qu'ils ne soient pas en mesure de concurrencer avec les lignées parentales lors du processus d'isolement en laboratoire (méthode chp. 1) et soient perdues par la suite. D'autres pistes comme des incompatibilités entre gènes mitochondriaux et nucléaires ou encore des aneuploïdies causées par les réarrangements chromosomiques sont aussi envisageables. L'étude comparative de tels isolements reproducteurs pourrait nous indiquer l'importance relative de chaque type d'isolement reproducteur dans un contexte de spéciation écologique en populations naturelles. Il est possible qu'avec ces nouvelles études, nous constations que le flux de gènes ait cessé entre les lignées B et C. Cependant, il sera toujours possible d'étudier l'isolement reproducteur à l'intérieur des populations où le flux de gènes est moins susceptible d'être déjà interrompu.

Variation dans la viabilité intragroupe

La haute variabilité dans l'isolement reproducteur à l'intérieur de la lignée C est surprenante, mais pourrait indiquer que du polymorphisme présent dans la population ségrège entre deux groupes de cette lignée (Cutter 2012; Corbett-Detig *et al.* 2013). Ce polymorphisme pourrait être d'origine génétique puisqu'un certain degré de variation

génétique au (0,04 ± 0,01 %) est présent dans cette lignée (Leducq *et al.* 2014). Nous observons aussi une grande variabilité dans les profils chromosomiques au sein de cette lignée (résultats chp. 2). Les réarrangements de chromosomes jouent aussi un rôle dans l'isolement reproducteur (introduction chp. 2) et ils pourraient être en mesure de ségréger dans la lignée C pour ainsi causer un isolement reproducteur variable. Il y a donc des possibilités pour que ce type de polymorphisme des réarrangements chromosomiques joue un rôle dans l'initiation de la spéciation dans les populations naturelles avant la divergence génétique lors de l'interruption du flux de gènes.

Perspectives

Nos résultats suggèrent que les lignées B et C pourraient faire partie d'un évènement de spéciation en cours à une échelle intracontinentale. En effet, la distribution géographique des lignées suggère qu'elles se sont probablement retrouvées en allopatrie lors de la dernière période de glaciation au pléistocène (discussion chp. 2) et que la zone de cooccurrence actuelle représenterait alors un évènement de contact secondaire au même titre que les migrations de nature anthropique qui ont emmené la lignée A en Amérique

En ce moment, l'intérêt pour la spéciation est dirigé vers l'influence de l'écologie sur l'émergence de la divergence entre les populations et l'isolement reproducteur incomplet à l'intérieur des espèces (Schluter 2001; Rundle & Nosil 2005; Egan & Funk 2009; Smadja & Butlin 2011). La spéciation écologique propose comme hypothèse principale que la sélection divergente entre deux niches écologiques mènerait à l'émergence d'incompatibilités reproductives (Schluter 2000). Selon ce modèle, deux populations à l'écologie divergente auront un isolement reproducteur plus fort que deux populations écologiquement similaires (Nosil *et al.* 2009). Sous cette prémisse, les gènes qui accumuleront le plus de divergence, et qui seraient susceptibles de mener à un isolement reproducteur sont ceux qui sont impliqués dans la survie et l'utilisation des ressources disponibles dans l'habitat respectif des populations divergentes. Cette divergence adaptative pourrait être impliquée dans l'isolement reproducteur entre les groupes B et C. Comme ces deux lignées occupent des aires de répartition différentes et présentent des phénotypes divergents pour des traits qui semblent écologiquement pertinents (température

de croissance et résistance aux cycles gel-dégel), les levures *S. paradoxus* pourraient faire partie d'un évènement de spéciation écologique. Sur cette base, on s'attendra à ce que les régions génomiques impliquées dans cette divergence adaptative soient sous sélection et soient plus différenciées entre les populations que les régions qui ne le sont pas (Beaumont 2005). Le séquençage du génome de plusieurs individus des deux lignées suivi de l'évaluation de la différenciation des séquences le long du génome nous indiquera la présence de telles régions. Par la suite, nous pourrons explorer le contenu en gènes de ces régions et d'identifier les gènes responsables de cette divergence. Pour confirmer le lien entre ces gènes divergents et les phénotypes, il serait possible de remplacer les gènes identifiés dans une lignée par ceux de l'autre et d'évaluer leurs effets sur le fitness ou encore l'isolement reproducteur.

Ce système eucaryote unicellulaire unique pourrait aussi s'avérer un outil utile dans l'étude moléculaire de la spéciation. Les outils moléculaires développés chez la levure telle l'insertion de gènes par transformation pourraient nous aider à disséquer l'impact de la divergence génétique dans l'isolement reproducteur. L'avènement des nouvelles technologies de séquençage rend aussi accessible des études de génomique des populations ou encore de génomique comparative. Couplées avec la possibilité d'obtenir une impressionnante quantité de ségrégants dans un temps relativement court, ces analyses utilisées avec le système décrit dans la présente étude ouvrent la voie à l'étude des facteurs génétiques menant à la divergence des populations et à l'établissement des barrières reproductrices dans un contexte de spéciation écologique dans des populations naturelles.

BIBLIOGRAPHIE

Amberg DC, Burke DJ, Strathern JN (2005) Methods in Yeast Genetics: A Cold Spring Harbor Laboratory Course Manual, 2005 Edition (Cold Spring). 119-120.

Andersen BG, Borns HW (1997) *The Ice Age World: An Introduction to Quaternary History and Research with Emphasis on North America and Northern Europe During the Last 2.5 Million Years* Scandinavian University Press North America.

April J, Hanner RH, Dion-Cote AM, Bernatchez L (2013) Glacial cycles as an allopatric speciation pump in north-eastern American freshwater fishes. *Mol Ecol* **22**, 409-422.

Baker Brachmann C, Davies A, Cost GJ*, et al.* (1998) Designer deletion strains derived from *Saccharomyces cerevisiae* S288C: A useful set of strains and plasmids for PCR-mediated gene disruption and other applications. *Yeast* **14**, 115-132.

Bateson W (1909) Heredity and Variation in Modern Lights. In: *Darwin and Modern Science* (ed. Seward AC), pp. 85-101. Cambridge University Press, Cambridge.

Beaumont MA (2005) Adaptation and speciation: what can Fst tell us? *Trends in Ecology & Evolution* **20**, 435-440.

Bhadra B, Rao RS, Singh PK, Sarkar PK, Shivaji S (2008) Yeasts and yeast-like fungi associated with tree bark: diversity and identification of yeasts producing extracellular endoxylanases. *Current microbiology* **56**, 489-494.

Bikard D, Patel D, Le Metté C*, et al.* (2009) Divergent Evolution of Duplicate Genes Leads to Genetic Incompatibilities Within *A. thaliana*. *Science* **323**, 623-626.

Blair AP (1941) Variation, Isolating Mechanisms, and Hybridization in Certain Toads. *Genetics* **26**, 398-417.

Blair WF (1964) Isolating Mechanisms and Interspecies Interactions in Anuran Amphibians. *The Quarterly Review of Biology* **39**, 334-344.

Blier PU, Dufresne F, Burton RS (2001) Natural selection and the evolution of mtDNA-encoded peptides: evidence for intergenomic co-adaptation. *Trends in Genetics* **17**, 400-406.

Bordenstein SR, O'Hara FP, Werren JH (2001) Wolbachia-induced incompatibility precedes other hybrid incompatibilities in Nasonia. *Nature* **409**, 707-710.

Brown KM, Landry CR, Hartl DL, Cavalieri D (2008) Cascading transcriptional effects of a naturally occurring frameshift mutation in *Saccharomyces cerevisiae*. *Mol Ecol* **17**, 2985-2997.

Buijs NA, Siewers V, Nielsen J (2013) Advanced biofuel production by the yeast *Saccharomyces cerevisiae*. *Current opinion in chemical biology*.

Butlin R, Debelle A, Kerth C*, et al.* (2012) What do we need to know about speciation? *Trends Ecol Evol* **27**, 27-39.

Capanna E (1982) Robertsonian numerical variation in animal speciation: *Mus musculus*, an emblematic model. In: *Mechanism of Speciation* (ed. Barigozzi C), pp. 155-177. Alan R. Liss, New York.

Chambers JM, Hastie T (1992) *Statistical models in S* Chapman & Hall London.

Chang SL, Lai HY, Tung SY, Leu JY (2013) Dynamic large-scale chromosomal rearrangements fuel rapid adaptation in yeast populations. *PLoS Genet* **9**, e1003232.

Charron G, Leducq JB, Bertin C, Dube AK, Landry CR (2014) Exploring the northern limit of the distribution of *Saccharomyces cerevisiae* and *Saccharomyces paradoxus* in North America. *FEMS Yeast Research* **14**, 281-288.

Chou JY, Hung YS, Lin KH, Lee HY, Leu JY (2010) Multiple molecular mechanisms cause reproductive isolation between three yeast species. *PLoS Biol* **8**, e1000432.

Ciardo DE, Schar G, Bottger EC, Altwegg M, Bosshard PP (2006) Internal transcribed spacer sequencing versus biochemical profiling for identification of medically important yeasts. *Journal of clinical microbiology* **44**, 77-84.

Clausen R (1928) Interspecific Hybridization in *Nicotiana*. VII. The cytology of Hybrids of the Synthetic species, Digluta, with its Parents, *glutinosa* and *tabacum*. *Botany* **11**, 177-211.

Clausen R, Goodspeed T (1925) Interspecific hybridization in *Nicotiana*. II. A tetraploid *glutinosa-tabacum* hybrid, an experimental verification of Winge's hypothesis. *Genetics* **10**, 278.

Coluccio AE, Rodriguez RK, Kernan MJ, Neiman AM (2008) The yeast spore wall enables spores to survive passage through the digestive tract of *Drosophila*. *PLoS One* **3**, e2873.

Coluzzi M (1982) Spatial distribution of chromosomal inversions and speciation in Anopheline mosquitoes. *Prog Clin Biol Res* **96**, 143-153.

Corbett-Detig RB, Zhou J, Clark AG, Hartl DL, Ayroles JF (2013) Genetic incompatibilities are widespread within species. *Nature* **504**, 135-137.

Cordell HJ (2002) Epistasis: what it means, what it doesn't mean, and statistical methods to detect it in humans. *Human Molecular Genetics* **11**, 2463-2468.

Coyle S, Kroll E (2008) Starvation induces genomic rearrangements and starvation-resilient phenotypes in yeast. *Mol Biol Evol* **25**, 310-318.

Coyne JA, Orr HA (1998) The evolutionary genetics of speciation. *Philos Trans R Soc Lond B Biol Sci* **353**, 287-305.

Coyne JA, Orr HA (2004) *Speciation* Sinauer Associates, Incorporated Publishers.

Cubillos FA, Louis EJ, Liti G (2009) Generation of a large set of genetically tractable haploid and diploid *Saccharomyces* strains. *FEMS Yeast Res* **9**, 1217-1225.

Cutter AD (2012) The polymorphic prelude to Bateson-Dobzhansky-Muller incompatibilities. *Trends Ecol Evol* **27**, 209-218.

Daniel HM, Moons MC, Huret S, Vrancken G, De Vuyst L (2011) *Wickerhamomyces anomalus* in the sourdough microbial ecosystem. *Antonie Van Leeuwenhoek* **99**, 63-73.

Darwin CR (1859) *On the origin of species by means of natural selection.* John Murray, London, UK.

Darwin CR, Wallace AR (1858) On the tendency of species to form varieties; and on the perpetuation of varieties and species by natural means of selection. *Zoology* **3**, 45-50.

Davenport RR (1976) *Distribution of yeasts and yeast-like organisms from aerial surfaces of developing apples and grapes.* Academic Press, London.

Davisson MT, Akeson EC (1993) Recombination suppression by heterozygous Robertsonian chromosomes in the mouse. *Genetics* **133**, 649-667.

Delneri D, Colson I, Grammenoudi S*, et al.* (2003) Engineering evolution to study speciation in yeasts. *Nature* **422**, 68-72.

Dobzhansky T (1934) Studies on hybrid sterility. I. Spermatogenesis in pure and hybrid *Drosophila pseudoobscura* hybrids. *Z. Zellforsch Mikrosk. Anat.* **21**, 169-221.

Dobzhansky T (1936) Studies on Hybrid Sterility. II. Localization of Sterility Factors in *Drosophila Pseudoobscura* Hybrids. *Genetics* **21**, 113-135.

Dobzhansky T (1937a) Genetic nature of species differences. *American Naturalist* **71**, 404-420.

Dobzhansky T (1951) *Genetics and the Origins of Species* Columbia University Press, New York, NY.

Dobzhansky TG (1937b) *Genetics and the Origin of Species.* Columbia University Press.

Dunham MJ, Badrane H, Ferea T*, et al.* (2002) Characteristic genome rearrangements in experimental evolution of *Saccharomyces cerevisiae*. *Proc Natl Acad Sci U S A* **99**, 16144-16149.

Dunham MJ, Louis EJ (2011) Yeast evolution and ecology meet genomics. *Embo Reports* **12**, 8-10.

Dunn B, Paulish T, Stanbery A*, et al.* (2013) Recurrent rearrangement during adaptive evolution in an interspecific yeast hybrid suggests a model for rapid introgression. *PLoS Genet* **9**, e1003366.

Egan SP, Funk DJ (2009) Ecologically dependent postmating isolation between sympatric host forms of *Neochlamisus bebbianae* leaf beetles. *Proc Natl Acad Sci U S A* **106**, 19426-19431.

Ellison CK, Burton RS (2006) Disruption of mitochondrial function in interpopulation hybrids of *Tigriopus californicus*. *Evolution* **60**, 1382-1391.

Ellison CK, Burton RS (2008) Interpopulation hybrid breakdown maps to the mitochondrial genome. *Evolution* **62**, 631-638.

Feder JL, Nosil P (2009) Chromosomal inversions and species differences: when are genes affecting adaptive divergence and reproductive isolation expected to reside within inversions? *Evolution* **63**, 3061-3075.

Federley H (1922) Über einen fall von criss-cross-vererbung bei einer artkreuzung. *Hereditas* **3**, 125-146.

Fischer G, James SA, Roberts IN, Oliver SG, Louis EJ (2000) Chromosomal evolution in *Saccharomyces*. *Nature* **405**, 451-454.

Ford MJ, Aquadro CF (1996) Selection on X-linked genes during speciation in the *Drosophila athabasca* complex. *Genetics* **144**, 689-703.

Gallagher JE, Babiarz JE, Teytelman L, Wolfe KH, Rine J (2009) Elaboration, diversification and regulation of the Sir1 family of silencing proteins in *Saccharomyces*. *Genetics* **181**, 1477-1491.

Gasch AP, Werner-Washburne M (2002) The genomics of yeast responses to environmental stress and starvation. *Functional & integrative genomics* **2**, 181-192.

Giraud T, Gourbiere S (2012) The tempo and modes of evolution of reproductive isolation in fungi. *Heredity (Edinb)* **109**, 204-214.

Goldstein AL, McCusker JH (1999) Three new dominant drug resistance cassettes for gene disruption in *Saccharomyces cerevisiae*. *Yeast* **15**, 1541-1553.

Goncalves P, Valerio E, Correia C, de Almeida JM, Sampaio JP (2011) Evidence for divergent evolution of growth temperature preference in sympatric *Saccharomyces species*. *PLoS One* **6**, e20739.

Good JM, Handel MA, Nachman MW (2008) Asymmetry and polymorphism of hybrid male sterility during the early stages of speciation in house mice. *Evolution* **62**, 50-65.

Gordon JL, Byrne KP, Wolfe KH (2009) Additions, losses, and rearrangements on the evolutionary route from a reconstructed ancestor to the modern *Saccharomyces cerevisiae* genome. *PLoS Genet* **5**, e1000485.

Granek JA, Magwene PM (2010) Environmental and genetic determinants of colony morphology in yeast. *PLoS Genet* **6**, e1000823.

Grant BR, Grant PR (1993) Eolution of Darwin's finches caused by a rare climatic event. *Proceedings of the Royal Society B-Biological Sciences* **251**, 111-117.

Gregory PG, Howard DJ (1993) Laboratory Hybridization Studies of *Allonemobius fasciatus* and *A. socius* (Orthoptera: Gryllidae). *Annals of the Entomological Society of America* **86**, 694-701.

Greig D (2009) Reproductive isolation in Saccharomyces. *Heredity (Edinb)* **102**, 39-44.

Greig D, Travisano M, Louis EJ, Borts RH (2003) A role for the mismatch repair system during incipient speciation in *Saccharomyces*. *J Evol Biol* **16**, 429-437.

Guldener U, Heck S, Fielder T, Beinhauer J, Hegemann JH (1996) A new efficient gene disruption cassette for repeated use in budding yeast. *Nucleic Acids Res* **24**, 2519-2524.

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

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

Hobel G, Gerhardt HC (2003) Reproductive character displacement in the acoustic communication system of green tree frogs (Hyla cinerea). *Evolution* **57**, 894-904.

Hoffmann AA, Rieseberg LH (2008) Revisiting the Impact of Inversions in Evolution: From Population Genetic Markers to Drivers of Adaptive Shifts and Speciation? *Annu Rev Ecol Evol Syst* **39**, 21-42.

Hong J, Gresham D (2014) Molecular specificity, convergence and constraint shape adaptive evolution in nutrient-poor environments. *PLoS Genet* **10**, e1004041.

Hou J, Friedrich A, de Montigny J, Schacherer J (2014) Chromosomal Rearrangements as a Major Mechanism in the Onset of Reproductive Isolation in *Saccharomyces cerevisiae*. *Current Biology* **24**, 1153-1159.

Hultman J, Ritari J, Romantschuk M, Paulin L, Auvinen P (2008) Universal ligation-detection-reaction microarray applied for compost microbes. *Bmc Microbiology* **8**.

Hyma KE, Fay JC (2013) Mixing of vineyard and oak-tree ecotypes of Saccharomyces cerevisiae in North American vineyards. *Molecular Ecology* **doi: 10.1111/mec.12155.**

Jackson ST, Webb RS, Anderson KH*, et al.* (2000) Vegetation and environment in Eastern North America during the Last Glacial Maximum. *Quaternary Science Reviews* **19**, 489-508.

Jeffrey EC, Hicks GC (1925) The Reduction Division in Relation to Mutation in Plants and Animals. *The American naturalist* **59**, 410-426.

Jiggins CD, Linares M, Naisbit RE*, et al.* (2001) Sex-linked hybrid sterility in a butterfly. *Evolution* **55**, 1631-1638.

Johannesson K, Panova M, Kemppainen P*, et al.* (2010) Repeated evolution of reproductive isolation in a marine snail: unveiling mechanisms of speciation. *Philos Trans R Soc Lond B Biol Sci* **365**, 1735-1747.

Johnson EA (2013) Biotechnology of non-*Saccharomyces* yeasts--the ascomycetes. *Applied microbiology and biotechnology* **97**, 503-517.

Johnson LJ, Koufopanou V, Goddard MR*, et al.* (2004) Population genetics of the wild yeast *Saccharomyces paradoxus*. *Genetics* **166**, 43-52.

Johnson NA (2010) Hybrid incompatibility genes: remnants of a genomic battlefield? *Trends in Genetics* **26**, 317-325.

Johnson NA, Perez DE, Cabot EL, Hollocher H, Wu CI (1992) A test of reciprocal X-Y interactions as a cause of hybrid sterility in *Drosophila*. *Nature* **358**, 751-753.

Kaiser C, Michaelis S, Mitchell A, Cold Spring Harbor Laboratory. (1994) *Methods in yeast genetics : a Cold Spring Harbor Laboratory course manual*, 1994 edn. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY.

Karpechenko GD (1927) The Production of Polyploid Gametes in Hybrids. *Hereditas* **9**, 349-368.

Kellogg EA, Bennetzen JL (2004) The evolution of nuclear genome structure in seed plants. *Am J Bot* **91**, 1709-1725.

Kirkpatrick M, Barton N (2006) Chromosome inversions, local adaptation and speciation. *Genetics* **173**, 419-434.

Kozlowska JL, Ahmad AR, Jahesh E, Cutter AD (2012) Genetic variation for postzygotic reproductive isolation between *Caenorhabditis briggsae* and *Caenorhabditis* sp. 9. *Evolution* **66**, 1180-1195.

Kroll E, Coyle S, Dunn B*, et al.* (2013) Starvation-associated genome restructuring can lead to reproductive isolation in yeast. *PLoS One* **8**, e66414.

Kuehne HA, Murphy HA, Francis CA, Sniegowski PD (2007) Allopatric divergence, secondary contact, and genetic isolation in wild yeast populations. *Curr Biol* **17**, 407-411.

Kurtzman C, Fell JW, Boekhout T (2011) *The yeasts: a taxonomic study* Access Online via Elsevier.

Kurtzman CP, Robnett CJ (2013) Relationships among genera of the Saccharomycotina (Ascomycota) from multigene phylogenetic analysis of type species. *FEMS Yeast Research* **13**, 23-33.

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

Landry CR, Lemos B, Rifkin SA, Dickinson WJ, Hartl DL (2007) Genetic properties influencing the evolvability of gene expression. *Science* **317**, 118-121.

Landry CR, Townsend JP, Hartl DL, Cavalieri D (2006) Ecological and evolutionary genomics of *Saccharomyces cerevisiae*. *Mol Ecol* **15**, 575-591.

Lassois L, de Bellaire LD, Jijakli MH (2008) Biological control of crown rot of bananas with *Pichia anomala* strain K and *Candida oleophila* strain O. *Biological Control* **45**, 410-418.

Leducq JB, Charron G, Diss G*, et al.* (2012) Evidence for the robustness of protein complexes to inter-species hybridization. *PLoS Genet* **8**, e1003161.

Leducq JB, Charron G, Samani P*, et al.* (2014) Local climatic adaptation in a widespread microorganism. *Proc Biol Sci* **281**, 20132472.

Liti G, Barton DB, Louis EJ (2006) Sequence diversity, reproductive isolation and species concepts in *Saccharomyces*. *Genetics* **174**, 839-850.

Liti G, Carter DM, Moses AM*, et al.* (2009) Population genomics of domestic and wild yeasts. *Nature* **458**, 337-341.

Lowry DB (2012) Ecotypes and the controversy over stages in the formation of new species. *Biological Journal of the Linnean Society* **106**, 241-257.

Lowry DB, Willis JH (2010) A widespread chromosomal inversion polymorphism contributes to a major life-history transition, local adaptation, and reproductive isolation. *PLoS Biol* **8**, e1000500.

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

Maclean CJ, Greig D (2008) Prezygotic reproductive isolation between *Saccharomyces cerevisiae* and *Saccharomyces paradoxus*. *BMC Evol Biol* **8**, 1.

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

Maringele L, Lydall D (2006) Pulsed-field gel electrophoresis of budding yeast chromosomes. *Methods Mol Biol* **313**, 65-73.

Matsubara K, Khin T, Sano Y (2003) A gene block causing cross-incompatibility hidden in wild and cultivated rice. *Genetics* **165**, 343-352.

Mayr E (1942) *Systematics and the origins of species* Columbia University Press, New-York, NY.

Melo MC, Grealy A, Brittain B, Walter GM, Ortiz-Barrientos D (2014) Strong extrinsic reproductive isolation between parapatric populations of an Australian groundsel. *New Phytol* **203**, 323-334.

Mihola O, Trachtulec Z, Vlcek C, Schimenti JC, Forejt J (2009) A mouse speciation gene encodes a meiotic histone H3 methyltransferase. *Science* **323**, 373-375.

Montrocher R, Verner MC, Briolay J, Gautier C, Marmeisse R (1998) Phylogenetic analysis of the *Saccharomyces cerevisiae* group based on polymorphisms of rDNA spacer sequences. *International journal of systematic bacteriology* **48 Pt 1**, 295-303.

Muller HJ (1925) Why Polyploidy is Rarer in Animals Than in Plants. *The American naturalist* **59**, 346-353.

Muller HJ (1942) Isolating mechanisms, evolution and temperature. In: *Temperature, Evolution, Development* (ed. Dobzhansky T), pp. 71-125. Jaques Cattell Press, Lancaster, PA.

Naumov GI (1987) Genetic basis for classification and identification of the ascomycetous yeast. *Studies in Mycology* **30**, 469-475.

Naumov GI, Gazdiev DO, Naumova ES (2003) [Identification of the yeast species *Saccharomyces bayanus* in Far East Asia]. *Mikrobiologiia* **72**, 834-839.

Naumov GI, Naumov ES, Sancho ED (1996) Genetic reidentification of *Saccharomyces* strains associated with black knot disease of trees in Ontario and *Drosophila* species in California. *Canadian Journal of Microbiology* **42**, 335-339.

Naumov GI, Naumova ES, Sniegowski PD (1998) *Saccharomyces paradoxus* and *Saccharomyces cerevisiae* are associated with exudates of North American oaks. *Canadian Journal of Microbiology* **44**, 1045-1050.

Navarro A, Barton NH (2003) Accumulating postzygotic isolation genes in parapatry: a new twist on chromosomal speciation. *Evolution* **57**, 447-459.

Noor MA, Feder JL (2006) Speciation genetics: evolving approaches. *Nat Rev Genet* **7**, 851-861.

Noor MA, Grams KL, Bertucci LA, Reiland J (2001) Chromosomal inversions and the reproductive isolation of species. *Proc Natl Acad Sci U S A* **98**, 12084-12088.

Nosil P (2012) *Ecological speciation* Oxford University Press, Oxford ; New York.

Nosil P, Harmon LJ, Seehausen O (2009) Ecological explanations for (incomplete) speciation. *Trends Ecol Evol* **24**, 145-156.

Orr HA (1990) "Why Polyploidy is Rarer in Animals Than in Plants" Revisited. *The American naturalist* **136**, 759-770.

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

Otto SP, Whitton J (2000) Polyploid incidence and evolution. *Annual Review of Genetics* **34**, 401-437.

Paulson DR (1974) Reproductive isolation in damselflies. *Systematic Biology* **23**, 40-49.

Phadnis N, Orr HA (2009) A single gene causes both male sterility and segregation distortion in *Drosophila* hybrids. *Science* **323**, 376-379.

Phaff HJ, Miller MW, Shifrine M (1956) The taxonomy of yeasts isolated from *Drosophila* in the Yosemite region of California. *Antonie Van Leeuwenhoek* **22**, 145-161.

Presgraves DC (2003) A fine-scale genetic analysis of hybrid incompatibilities in Drosophila. *Genetics* **163**, 955-972.

Presgraves DC (2010) The molecular evolutionary basis of species formation. *Nat Rev Genet* **11**, 175-180.

R Development Core Team (2010) R: A language and environment for statistical computing. R Foundation for Statistical Computing), Vienna, Austria.

Ramaswamy NT, Li L, Khalil M, Cannon JF (1998) Regulation of yeast glycogen metabolism and sporulation by Glc7p protein phosphatase. *Genetics* **149**, 57-72.

Ramsey J, Bradshaw HD, Jr., Schemske DW (2003) Components of reproductive isolation between the monkeyflowers *Mimulus lewisii* and *M. cardinalis* (Phrymaceae). *Evolution* **57**, 1520-1534.

Ramsey J, Schemske DW (1998) Pathways, Mechanisms, and Rates of Polyploid Formation in Flowering Plants. *Annual Review of Ecology and Systematics* **29**, 467-501.

Replansky T, Bell G (2009) The relationship between environmental complexity, species diversity and productivity in a natural reconstructed yeast community. *Oikos* **118**, 233-239.

Replansky T, Koufopanou V, Greig D, Bell G (2008) *Saccharomyces sensu stricto* as a model system for evolution and ecology. *Trends Ecol Evol* **23**, 494-501.

Rieseberg LH (2001) Chromosomal rearrangements and speciation. *Trends Ecol Evol* **16**, 351-358.

Rundle HD, Nosil P (2005) Ecological speciation. *Ecology letters* **8**, 336-352.

Sampaio JP, Goncalves P (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**, 2144-2152.

Sax K (1923) The Relation between Chromosome Number, Morphological Characters and Rust Resistance in Segregates of Partially Sterile Wheat Hybrids. *Genetics* **8**, 301-321.

Schluter D (2000) *The Ecology of Adaptative Radiation* Oxford University Press, Oxford.

Schluter D (2001) Ecology and the origin of species. *Trends Ecol Evol* **16**, 372-380.

Schluter D, Conte GL (2009) Genetics and ecological speciation. *Proc Natl Acad Sci U S A* **106 Suppl 1**, 9955-9962.

Schmitt MJ, Breinig F (2002) The viral killer system in yeast: from molecular biology to application. *Fems Microbiology Reviews* **26**, 257-276.

Seward AC (1910) Darwin and modern science. *Molecular and General Genetics MGG* **3**, 109-110.

Smadja CM, Butlin RK (2011) A framework for comparing processes of speciation in the presence of gene flow. *Mol Ecol* **20**, 5123-5140.

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

Stebbins C (1950) Variation and evolution in plants. *Variation and evolution in plants.*

Stefanini I, Dapporto L, Legras JL*, et al.* (2012) Role of social wasps in Saccharomyces cerevisiae ecology and evolution. *Proceedings of the National Academy of Sciences of the United States of America* **109**, 13398-13403.

Sturtevant A, Beadle G (1936) The relations of inversions in the X chromosome of Drosophila melanogaster to crossing over and disjunction. *Genetics* **21**, 554.

Sweeney JY, Kuehne HA, Sniegowski PD (2004) Sympatric natural *Saccharomyces cerevisiae* and *S. paradoxus* populations have different thermal growth profiles. *FEMS Yeast Research* **4**, 521-525.

Tamura K, Peterson D, Peterson N*, et al.* (2011) MEGA5: molecular evolutionary genetics analysis using maximum likelihood, evolutionary distance, and maximum parsimony methods. *Mol Biol Evol* **28**, 2731-2739.

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

Turelli M, Barton NH, Coyne JA (2001) Theory and speciation. *Trends Ecol Evol* **16**, 330-343.

Vadkertiova R, Molnarova J, Vranova D, Slavikova E (2012) Yeasts and yeast-like organisms associated with fruits and blossoms of different fruit trees. *Canadian Journal of Microbiology* **58**, 1344-1352.

Veilleux R (1985) Diploid and Polyploid Gametes in Crop Plants: Mechanisms of Formation and Utilization in Plant Breeding. In: *Plant Breeding Reviews*, pp. 253-288. John Wiley & Sons, Inc.

Via S (2009) Natural selection in action during speciation. *Proc Natl Acad Sci U S A* **106 Suppl 1**, 9939-9946.

Wallace AR (1864) On the phenomenon of variation and geographical distribution as illustrated by the Papilionidae of the malayan region. *Transaction of the Linnean society* **25**.

Walsh JB (1982) Rate of accumulation of reproductive isolation by chromosome rearrangements. *American Naturalist* **140**, 510-532.

Wang XD, Li GQ, Jiang DH, Huang HC (2009) Screening of plant epiphytic yeasts for biocontrol of bacterial fruit blotch (*Acidovorax avenae* subsp *citrulli*) of hami melon. *Biological Control* **50**, 164-171.

Werren JH (1997) Biology of *Wolbachia*. *Annual review of entomology* **42**, 587-609.

Wheeler QD, Meier RD (2000) *Species concept and Phylogenetic Theory: A debate* Columbia University Press, New York, NY.

White M (1978) Chain processes in chromosomal speciation. *Systematic Biology* **27**, 285-298.

Widmer A, Lexer C, Cozzolino S (2009) Evolution of reproductive isolation in plants. *Heredity (Edinb)* **102**, 31-38.

Will JL, Kim HS, Clarke J*, et al.* (2010) Incipient balancing selection through adaptive loss of aquaporins in natural *Saccharomyces cerevisiae* populations. *PLoS Genet* **6**, e1000893.

Wilson RA (1999) *Species: New Interdisciplinary Essays* MIT Press, Cambridge, MA.

Wittbrodt J, Adam D, Malitschek B*, et al.* (1989) Novel putative receptor tyrosine kinase encoded by the melanoma-inducing Tu locus in Xiphophorus. *Nature* **341**, 415-421.

Wodsedalek JE (1916) Causes of sterility in the mule. *The Biological Bulletin* **30**, 1-[56]-51.

Wolf JB, Lindell J, Backstrom N (2010) Speciation genetics: current status and evolving approaches. *Philos Trans R Soc Lond B Biol Sci* **365**, 1717-1733.

Wood TE, Takebayashi N, Barker MS*, et al.* (2009) The frequency of polyploid speciation in vascular plants. *Proceedings of the national Academy of sciences* **106**, 13875-13879.

Wu C-I (2001) The genic view of the process of speciation. *J Evol Biol* **14**, 851-865.

Yen JH, Barr AR (1971) New Hypothesis of the Cause of Cytoplasmic Incompatibility in Culex pipiens L. *Nature* **232**, 657-658.

Zanders SE, Eickbush MT, Jonathan SY*, et al.* (2014) Genome rearrangements and pervasive meiotic drive cause hybrid infertility in fission yeast. *eLife* **3**, e02630.

Zeyl C (2014) Evolution: A Collection of Misfits. *Current Biology* **24**, R394-R396.

Zhang H, Skelton A, Gardner RC, Goddard MR (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**, 941-947.

ANNEXES

Table S1: List of sampling sites.

The following information is shown: number of the sampling area (references for Fig. 1), names of sampling sites in the sampling area and number of samples.

Sampling area	Sampling sites	Samples
1	Québec City, Île d'Orléans, Saint-Michel-de-Bellechasse, Saint-Nicolas	280
2	Saguenay	18
3	Cap Saint-Ignace, Saint-Vallier, Berthier-sur-Mer, Montmagny, Saint-Fabien-de-Panet	23
4	Cap-Chat	50
5	Station Duchesnay	13
6	Grosse-Île	8
7	Squatec	99
8	Parc des Grands-Jardins	32
9	Sherbrooke	11
10	Saint-Georges, Saint-Honoré	49
11	Stanstead	15
12	Île Saint-Quentin	18
13	Gatineau, Hull	15
14	Île Bonaventure	14
15	Parc de la Gaspésie	7
16	Anse-au-Griffon, Parc Forillon	15
17	Saint-Jean-Port-Joli, Îslet-sur-Mer, Kamouraska	15
18	Notre-Dame-du-Portage	10
19	Saint-Thècle	11
20	Pointe Platon	26
21	Lac Louis, Sacré-Cœur, Baie Sainte-Catherine	25
22	Parc National du Bic, Trois-Pistoles	10
23	Routhierville	6
24	Lavaltrie, Saint-Guillaume Lanoraie, Saint-Jean-Baptiste	41
25	Ontario	2
26	Manitoba	3
27	British-Columbia	47
28	Lac Simon, Saint-Léonard-de-Portneuf	2
29	Saint-Justin	11

Table S2: ANOVA analysis of yeast isolation success

Year	Yeast group	Total variance	Factor (part and significance of the variance)			
			Substrate	Temperature	Sample size	Month
2011	all yeasts	1.1587	0.07	0.00	0.05	0.03
	Saccharomycetaceae	0.0509	**0.46***	0.00	0.00	0.00
	S. paradoxus	0.0079	**0.57***	0.02	0.06	0.00
2012	all yeasts	0.1133	-	-	0.05	0.49
	Saccharomycetaceae	0.0594	-	-	0.05	0.60
	S. paradoxus	0.0271	-	-	0.01	**0.66°**

Factors were considered as additive and are indicated in bold when explaining a significant part of isolation success variance in a given yeast group ("°": $p < 0.1$; "*": $p < 0.05$)

Table S3: List of strains isolated in this study.

The following information is shown: name of the isolate, temperature of incubation used for isolation, geographical location (see Fig. 1 and Table S1), substrate, accession number of the *ITS1, 5.8S, ITS2* sequence in NCBI, name of the closest taxon found by BLAST in NCBI, percentage of sequence coverage with the closest taxon, percentage of sequence identity with the closest taxon, reference for the closest taxon and accession number of the of the *ITS1, 5.8S, ITS2* sequence of the closest taxon in NCBI (unpub.: no article was associated to the accession number).

Isolated strain					Reference sequence			
Name	Isolation T°	Location	Substrate	Accession #	Closest known species	Coverage (%)	Identity (%)	Accession #
LL11_001	30°C	1	Oak tree bark	KF057635	*Saccharomyces paradoxus*	100	99	D89891
LL11_002	30°C	4	Maple tree bark	KF057516	*Saccharomyces paradoxus*	100	99	D89891
LL11_003	30°C	4	Maple tree bark	KF057517	*Saccharomyces paradoxus*	100	99	D89891
LL11_004	30°C	4	Slug	KF057514	*Saccharomyces paradoxus*	100	99	D89891
LL11_005	18°C	1	Maple tree bark	KF057527	*Saccharomyces paradoxus*	100	99	D89891
LL11_006	18°C	1	Fallen Apple	KF057551	*Saccharomyces paradoxus*	100	99	D89891
LL11_007	18°C	1	Maple tree bark	KF057607	*Saccharomyces paradoxus*	100	99	D89891
LL11_008	18°C	1	Maple tree bark	KF057609	*Saccharomyces paradoxus*	100	99	D89891
LL11_009	18°C	7	Poplar tree associated soil	KF057736	*Saccharomyces paradoxus*	100	99	D89891
LL11_010	18°C	7	Fir tree associated soil	KF057737	*Saccharomyces paradoxus*	100	99	D89891
LL11_011	18°C	7	Insect	KF057740	*Saccharomyces paradoxus*	100	99	D89891
LL11_012	30°C	5	Maple tree bark	KF057505	*Saccharomyces paradoxus*	100	99	D89891
LL11_014	30°C	3	Poplar bark	KF057515	*Kazachstania aerobia*	93	99	JF916514
LL11_015	30°C	3	Oak tree associated soil	KF057509	*Torulaspora delbrueckii*	100	99	HE616749
LL11_016	30°C	3	Oak tree associated soil	KF057510	*Kazachstania aerobia*	93	99	JF916514
LL11_017	30°C	4	Raspberry	KF057511	*Starmerella bombicola*	99	100	HQ111054
LL11_019	30°C	4	Bumblebee	KF057513	*Kluyveromyces dobzhanskii*	95	99	FR717867
LL11_020	30°C	4	Maple tree bark	KF057518	*Starmerella bombicola*	99	99	HQ111054
LL11_021	30°C	4	Maple tree bark	KF057519	*Tumularia aquatica*	98	98	FJ000399
LL11_022	30°C	5	Maple tree bark	KF057506	*Lachancea thermotolerans*	100	99	CU928180
LL11_023	30°C	5	Maple tree bark	KF057507	*Hanseniaspora osmophila*	98	98	AJ512431
LL11_024	30°C	6	Maple tree bark	KF057521	*Starmerella bombicola*	99	99	HQ111054

LL11_025	30°C	6	Soil	KF057522	*Kazachstania martiniae*	99	99	JF916547
LL11_026	30°C	6	Soil	KF057523	*Cryptococcus podzolicus*	94	100	AB035576
LL11_027	30°C	6	Acorn	KF057524	*Kluyveromyces dobzhanskii*	100	99	FR717867
LL11_028	30°C	6	*Achillea millefollium* flower	KF057525	*Starmerella bombicola*	99	99	HQ111056
LL11_029	30°C	6	Oak tree bark	KF057526	*Kluyveromyces dobzhanskii*	100	99	FR717867
LL11_030	30°C	1	Apple tree asociated soil	KF057703	*Pichia fluxuum*	98	99	DQ104720
LL11_031	30°C	1	Apple tree asociated soil	KF057704	*Pichia fluxuum*	99	99	DQ104720
LL11_032	30°C	1	Apple tree bark	KF057705	*Hanseniaspora osmophila*	97	98	AJ512431
LL11_033	18°C	1	Fallen apple	KF057706	*Issatchenkia hanoiensis*	87	99	FJ153177
LL11_034	30°C	1	Fallen apple	KF057707	*Pichia* sp.	99	97	FM177680
LL11_035	30°C	22	Maple tree bark	KF057708	*Wickerhamomyces anomalus*	95	99	FJ713067
LL11_037	30°C	22	*Malva moschata* flower	KF057710	*Starmerella bombicola*	99	100	HQ111054
LL11_038	18°C	1	Fallen apple	KF057528	*Candida norvegica*	100	98	KC542800
LL11_040	18°C	1	Maple tree bark	KF057530	*Guehomyces pullulans*	100	99	AF444418
LL11_041	18°C	1	Maple tree bark	KF057531	*Pichia membranifaciens*	99	99	DQ198954
LL11_042	18°C	1	Maple tree bark	KF057532	*Debaryomyces hansenii*	98	99	JN942909
LL11_043	18°C	1	Maple tree bark	KF057533	*Yamadazyma phyllophila*	98	97	AB734051
LL11_045	18°C	1	Fallen grapes	KF057535	*Rhodotorula nothofagi*	99	99	AY383749
LL11_046	18°C	1	Fallen grapes	KF057536	*Hanseniaspora uvarum*	99	99	GU237050
LL11_047	18°C	1	Fallen grapes	KF057537	*Hanseniaspora uvarum*	100	99	GU237050
LL11_048	18°C	1	Fallen grapes	KF057538	*Wickerhamomyces anomalus*	93	99	FJ713067
LL11_049	18°C	1	Fallen grapes	KF057539	*Pichia fermentans*	88	99	DQ665310
LL11_050	18°C	1	Fallen grapes	KF057540	*Rhodotorula* sp.	100	100	AM901704
LL11_051	18°C	1	Fallen grapes	KF057541	*Candida* sp.	99	99	DQ104728
LL11_052	18°C	1	Fallen apple	KF057542	*Candida* sp.	98	99	DQ104728
LL11_053	18°C	1	Fallen apple	KF057543	*Pichia fermentans*	88	99	DQ665310
LL11_054	18°C	1	Maple tree bark	KF057544	*Candida* sp.	98	99	DQ104728
LL11_055	18°C	1	Fallen apple	KF057545	*Candida* sp.	98	99	DQ104728
LL11_056	18°C	1	Fallen apple	KF057546	*Pichia membranifaciens*	100	97	DQ198954
LL11_057	18°C	1	Fallen apple	KF057547	*Pichia kudriavzevii*	99	99	JF715197
LL11_058	18°C	1	Fallen apple	KF057548	*Pichia kluyveri*	99	98	FM864201
LL11_059	18°C	1	Fallen apple	KF057549	*Candida* sp	99	99	DQ104728

LL11_060	18°C	1	Fallen apple	KF057550	*Hanseniaspora uvarum*	99	99	GU237050
LL11_061	18°C	1	Fallen apple	KF057552	*Hanseniaspora uvarum*	99	99	GU237050
LL11_062	18°C	1	Fallen apple	KF057553	*Candida* sp.	90	99	FJ210529
LL11_066	18°C	1	Maple tree bark	KF057557	*Arthroascus schoenii*	97	100	DQ361051
LL11_068	18°C	1	Maple tree bark	KF057559	*Candida norvegica*	100	98	KC542800
LL11_069	18°C	1	Fallen strawberry	KF057560	*Pichia kluyveri*	99	97	FM864201
LL11_070	18°C	1	Fallen strawberry	KF057561	*Pichia kluyveri*	99	99	FM864201
LL11_071	18°C	1	Fallen strawberry	KF057562	*Pichia kudriavzevii*	100	100	JF715197
LL11_072	18°C	1	Fallen strawberry	KF057563	*Pichia kudriavzevii*	100	99	JF715197
LL11_073	18°C	1	Fallen strawberry	KF057564	*Hanseniaspora uvarum*	100	99	GU237050
LL11_074	18°C	1	Fallen strawberry	KF057565	*Pichia kluyveri*	100	99	FM864201
LL11_075	18°C	1	Fallen strawberry	KF057566	*Hanseniaspora uvarum*	99	99	GU237050
LL11_076	18°C	1	Fallen strawberry	KF057567	*Hanseniaspora uvarum*	99	99	GU237050
LL11_077	18°C	1	Fallen strawberry	KF057568	*Hanseniaspora uvarum*	99	99	GU237050
LL11_078	18°C	1	Fallen strawberry	KF057569	*Pichia kudriavzevii*	100	99	JF715197
LL11_079	18°C	1	Fallen strawberry	KF057570	*Pichia kudriavzevii*	100	100	JF715197
LL11_080	18°C	1	Fallen strawberry	KF057571	*Hanseniaspora uvarum*	100	99	GU237050
LL11_081	18°C	1	Fallen strawberry	KF057572	*Issatchenkia terricola*	100	98	JQ993370
LL11_082	18°C	1	Fallen strawberry	KF057573	*Hanseniaspora uvarum*	100	99	GU237050
LL11_084	18°C	1	Fallen strawberry	KF057575	*Hanseniaspora uvarum*	100	99	GU237050
LL11_085	18°C	1	Fallen strawberry	KF057576	*Pichia kluyveri*	100	98	FM864201
LL11_086	18°C	1	Fallen strawberry	KF057577	*Hanseniaspora uvarum*	100	99	GU237050
LL11_087	18°C	1	Fallen strawberry	KF057578	*Pichia kudriavzevii*	100	99	JF715197
LL11_088	18°C	1	Fallen strawberry	KF057579	*Hanseniaspora uvarum*	100	99	GU237050
LL11_090	18°C	1	Fallen strawberry	KF057581	*Hanseniaspora uvarum*	100	99	GU237050
LL11_091	18°C	1	Maple tree bark	KF057582	*Wickerhamomyces anomalus*	94	99	FJ713067
LL11_092	18°C	1	Maple tree bark	KF057584	*Candida norvegica*	100	99	GU246258
LL11_095	18°C	1	Fallen apple	KF057587	*Pichia membranifaciens*	100	99	DQ198954
LL11_100	18°C	1	Fallen apple	KF057592	*Hanseniaspora uvarum*	99	99	GU237050
LL11_101	18°C	1	Fallen apple	KF057593	*Candida boidinii*	99	99	EF197945
LL11_102	18°C	1	Fallen apple	KF057594	*Pichia kudriavzevii*	100	100	JF715197
LL11_103	18°C	1	Fallen apple	KF057595	*Metschnikowia pulcherrima*	99	97	GU237059

LL11_104	18°C	1	Fallen apple	KF057596	*Wickerhamomyces anomalus*	93	98	FJ713067
LL11_106	18°C	1	Fallen apple	KF057598	*Hanseniaspora uvarum*	99	99	GU237050
LL11_107	18°C	1	Fallen apple	KF057599	*Pichia membranifaciens*	100	99	DQ198954
LL11_109	18°C	1	Fallen apple	KF057601	*Pichia kluyveri*	98	97	FM864201
LL11_113	18°C	1	Fallen apple	KF057605	*Candida zemplinina*	99	100	GU237057
LL11_114	18°C	1	Maple tree bark	KF057606	*Kluyveromyces dobzhanskii*	100	99	FR717867
LL11_115	18°C	1	Maple tree bark	KF057608	*Torulaspora delbrueckii*	99	99	HE616749
LL11_118	18°C	1	Fallen apple	KF057612	*Pichia* sp.	98	99	JX171189
LL11_119	18°C	1	Fallen apple	KF057613	*Hanseniaspora uvarum*	99	99	GU237050
LL11_120	18°C	1	Fallen apple	KF057614	*Pichia membranifaciens*	98	99	DQ198954
LL11_121	18°C	1	Fallen apple	KF057615	*Candida boidinii*	98	99	EF197945
LL11_122	18°C	1	Fallen apple	KF057616	*Pichia membranifaciens*	99	99	DQ198954
LL11_123	18°C	1	Fallen apple	KF057617	*Pichia membranifaciens*	99	99	DQ198954
LL11_127	18°C	1	Fallen apple	KF057621	*Pichia membranifaciens*	98	99	DQ198954
LL11_128	18°C	1	Fallen apple	KF057622	*Sporobolomyces roseus*	99	99	HQ913900
LL11_129	18°C	1	Fallen apple	KF057623	*Hanseniaspora uvarum*	99	99	GU237050
LL11_130	18°C	1	Fallen apple	KF057624	*Candida* sp.	99	99	DQ104728
LL11_131	18°C	1	Fallen apple	KF057625	*Pichia membranifaciens*	100	99	DQ198954
LL11_132	18°C	1	Fallen apple	KF057626	*Wickerhamomyces anomalus*	94	98	FJ713067
LL11_134	18°C	1	Fallen apple	KF057628	*Pichia membranifaciens*	99	99	DQ198954
LL11_135	18°C	1	Fallen apple	KF057629	*Candida* sp.	98	99	FJ210529
LL11_136	18°C	1	Fallen apple	KF057630	*Pichia membranifaciens*	100	99	DQ198954
LL11_137	18°C	1	Fallen apple	KF057631	*Pichia fermentans*	88	99	DQ665310
LL11_138	18°C	7	Maple tree bark	KF057719	*Candida* sp.	97	99	JF682351
LL11_139	18°C	7	Wild fruit	KF057720	*Rhodotorula laryngis*	98	98	EU149811
LL11_140	18°C	7	Poplar bud	KF057721	*Starmerella bombicola*	100	99	HQ111056
LL11_142	18°C	7	Poplar tree associated soil	KF057723	*Candida boidinii*	98	99	FJ914929
LL11_143	18°C	7	*Viburnum* sp. fruit	KF057724	*Nakazawaea holstii*	98	99	AB449811
LL11_144	18°C	7	Birch tree bark	KF057725	*Nakazawaea holstii*	100	99	AB449811
LL11_149	18°C	7	Maple tree bark	KF057730	*Yamadazyma phyllophila*	98	97	AB734050
LL11_150	18°C	7	Maple tree bark	KF057731	*Torulaspora delbrueckii*	99	99	HE616749
LL11_153	18°C	7	Fallen apple	KF057734	*Candida* sp.	90	99	FJ210529

LL11_154	18°C	7	Fallen apple	KF057735	*Candida stellimalicola*	100	97	FM199968
LL11_155	18°C	7	Spruce tree associated soil	KF057738	*Torulaspora delbrueckii*	99	99	HE616749
LL12_001	30°C	13	Oak tree bark	KF153060	*Saccharomyces paradoxus*	100	99	D89891
LL12_002	30°C	20	Oak tree bark	KF153067	*Saccharomyces paradoxus*	100	99	D89891
LL12_003	30°C	20	Oak tree bark	KF153068	*Saccharomyces paradoxus*	100	99	D89891
LL12_004	30°C	4	Maple tree bark	KF153056	*Saccharomyces paradoxus*	100	99	D89891
LL12_005	30°C	4	Other tree Bark	KF153057	*Saccharomyces paradoxus*	100	99	D89891
LL12_006	30°C	4	Other tree Bark	KF153075	*Saccharomyces paradoxus*	100	99	D89891
LL12_007	30°C	17	Maple tree bark	KF153058	*Saccharomyces paradoxus*	100	99	D89891
LL12_008	30°C	4	Maple tree bark	KF153059	*Saccharomyces paradoxus*	100	99	D89891
LL12_009	30°C	1	Oak tree bark	KF153055	*Saccharomyces paradoxus*	100	99	D89891
LL12_010	30°C	22	Maple tree bark	KF153063	*Saccharomyces paradoxus*	100	99	D89891
LL12_011	30°C	22	Maple tree bark	KF153064	*Saccharomyces paradoxus*	100	99	D89891
LL12_012	30°C	16	Maple tree bark	KF153061	*Saccharomyces paradoxus*	100	99	D89891
LL12_014	30°C	1	Maple tree bark	KF153065	*Saccharomyces paradoxus*	100	99	D89891
LL12_015	30°C	20	Oak tree bark	KF153066	*Saccharomyces paradoxus*	100	99	D89891
LL12_016	30°C	20	Oak tree bark	KF153069	*Saccharomyces paradoxus*	100	99	D89891
LL12_017	30°C	20	Oak tree bark	KF153070	*Saccharomyces paradoxus*	100	99	D89891
LL12_018	30°C	20	Oak tree bark	KF153071	*Saccharomyces paradoxus*	100	99	D89891
LL12_019	30°C	20	Oak tree bark	KF153072	*Saccharomyces paradoxus*	100	99	D89891
LL12_020	30°C	20	Oak tree bark	KF153073	*Saccharomyces paradoxus*	100	99	D89891
LL12_021	30°C	20	Oak tree bark	KF153074	*Saccharomyces paradoxus*	100	99	D89891
LL12_022	30°C	12	Oak tree bark	KF057715	*Saccharomyces paradoxus*	100	99	D89891
LL12_023	30°C	12	Oak tree bark	KF057716	*Saccharomyces paradoxus*	100	99	D89891
LL12_024	30°C	12	Oak tree bark	KF057717	*Saccharomyces paradoxus*	100	99	D89891
LL12_025	30°C	9	Oak tree bark	KF057503	*Saccharomyces paradoxus*	100	99	D89891
LL12_026	30°C	9	Maple tree bark	KF057713	*Saccharomyces paradoxus*	100	99	D89891
LL12_027	30°C	9	Maple tree bark	KF057712	*Saccharomyces paradoxus*	100	99	D89891
LL12_028	30°C	9	Maple tree bark	KF057751	*Saccharomyces paradoxus*	100	99	D89891
LL12_029	30°C	19	Oak tree associated soil	KF153062	*Saccharomyces paradoxus*	100	99	D89891
LL12_030	30°C	27	Maple tree associated soil	KF057642	*Saccharomyces paradoxus*	100	99	D89891
LL12_031	30°C	1	Oak tree bark	KF057648	*Lachancea thermotolerans*	100	99	CU928180

LL12_032	30°C	1	Oak tree bark	KF057649	*Lachancea thermotolerans*	100	99	GQ340441
LL12_033	30°C	1	Oak tree bark	KF057645	*Lachancea thermotolerans*	100	99	CU928180
LL12_034	30°C	1	Oak tree bark	KF057646	*Lachancea thermotolerans*	100	99	CU928180
LL12_035	30°C	1	Oak tree bark	KF057647	*Lachancea thermotolerans*	100	99	CU928180
LL12_036	30°C	9	Maple tree bark	KF057711	*Lachancea thermotolerans*	100	99	GQ340441
LL12_037	30°C	9	Maple tree bark	KF057520	*Wickerhamomyces anomalus*	100	99	FJ713067
LL12_038	30°C	27	Oak tree bark	KF057640	*Torulaspora delbrueckii*	99	99	HE616749
LL12_039	30°C	27	Oak tree bark	KF057641	*Torulaspora delbrueckii*	99	99	HE616749
LL12_040	30°C	27	Maple tree bark	KF057643	*Lachancea thermotolerans*	100	99	GQ340441
LL12_041	30°C	27	Oak tree bark	KF057644	*Lachancea thermotolerans*	100	99	GQ340441
LL12_042	30°C	1	Oak tree bark	KF057636	*Lachancea thermotolerans*	100	99	CU928180
LL12_043	30°C	1	Oak tree bark	KF057637	*Lachancea thermotolerans*	100	99	CU928180
LL12_044	30°C	11	Oak tree bark	KF057650	*Wickerhamomyces anomalus*	100	99	FJ713067
LL12_045	30°C	11	Maple tree bark	KF057651	*Wickerhamomyces anomalus*	100	99	FJ713067
LL12_046	30°C	24	Maple tree bark	KF057699	*Torulaspora delbrueckii*	99	99	HE616749
LL12_047	30°C	24	Maple tree asociated soil	KF057685	*Pichia fermentans*	98	98	DQ665310
LL12_049	30°C	24	Maple tree bark	KF057687	*Torulaspora delbrueckii*	99	99	HE616749
LL12_050	30°C	24	Maple tree bark	KF057688	*Torulaspora delbrueckii*	99	99	HE616749
LL12_051	30°C	24	Ant on a maple tree	KF057689	*Torulaspora delbrueckii*	99	99	HE616749
LL12_052	30°C	24	Maple tree bark	KF057690	*Torulaspora delbrueckii*	99	99	HE616749
LL12_053	30°C	24	Planted maple tree bark	KF057691	*Torulaspora delbrueckii*	100	99	HE616749
LL12_054	30°C	24	Planted maple tree bark	KF057692	*Torulaspora delbrueckii*	99	99	HE616749
LL12_055	30°C	24	Planted oak tree bark	KF057693	*Lachancea kluyveri*	94	100	AB037405
LL12_056	30°C	24	Planted oak tree bark	KF057694	*Lachancea thermotolerans*	100	99	GQ340441
LL12_057	30°C	24	Planted oak tree bark	KF057695	*Lachancea thermotolerans*	100	99	GQ340441
LL12_058	30°C	12	Oak tree bark	KF057696	*Pichia fermentans*	98	99	DQ665310
LL12_059	30°C	12	Maple tree bark	KF057697	*Lachancea kluyveri*	94	100	AB037405
LL12_061	30°C	1	Maple tree bark	KF057700	*Kazachstania aerobia*	95	99	JF916514
LL12_062	30°C	1	Oak tree bark	KF057701	*Lachancea kluyveri*	94	100	AB037405
LL12_063	30°C	1	Horsefly	KF057702	*Kluyveromyces waltii*	99	97	GQ340441
LL12_064	30°C	13	Oak tree bark	KF057652	*Kluyveromyces lactis*	100	99	KC609424
LL12_065	30°C	13	Oak tree bark	KF057653	*Torulaspora delbrueckii*	99	99	HE616749

LL12_066	30°C	13	Oak tree bark	KF057654	*Kluyveromyces lactis*	99	99	KC609424
LL12_067	30°C	13	Oak tree bark	KF057655	*Torulaspora delbrueckii*	99	99	HE616749
LL12_068	30°C	13	Oak tree bark	KF057659	*Lachancea thermotolerans*	99	99	GQ340441
LL12_069	30°C	13	Oak tree bark	KF057656	*Lachancea kluyveri*	94	100	AB037405
LL12_070	30°C	13	Oak tree bark	KF057657	*Lachancea thermotolerans*	99	99	GQ340441
LL12_071	30°C	13	Oak tree bark	KF057658	*Kazachstania servazzii*	95	99	AY046153
LL12_072	30°C	17	Maple tree bark	KF057504	*Kazachstania aerobia*	100	99	JF916514
LL12_073	30°C	22	Maple tree bark	KF057669	*Lachancea thermotolerans*	99	99	GQ340441
LL12_074	30°C	22	Maple tree bark	KF057671	*Torulaspora delbrueckii*	99	99	HE616749
LL12_075	30°C	22	Maple tree bark	KF057675	*Wickerhamomyces anomalus*	100	99	FJ713067
LL12_076	30°C	22	Maple tree bark	KF057684	*Wickerhamomyces anomalus*	100	99	FJ713067
LL12_077	30°C	16	Maple tree bark	KF057667	*Pichia membranifaciens*	100	99	DQ409165
LL12_078	30°C	16	Maple tree bark	KF057668	*Lachancea thermotolerans*	100	99	GQ340441
LL12_081	30°C	23	Maple tree bark	KF057673	*Lachancea kluyveri*	94	100	AB037405
LL12_082	30°C	23	Maple tree bark	KF057674	*Yamadazyma scolyti*	100	99	HE612109
LL12_083	30°C	23	Maple tree bark	KF057676	*Lachancea thermotolerans*	100	99	GQ340441
LL12_084	30°C	18	Hoverfly	KF057677	*Starmerella bombicola*	100	99	HQ111052
LL12_085	30°C	18	Hoverfly	KF057678	*Starmerella bombicola*	100	99	HQ111052
LL12_088	30°C	18	Bumblebee	KF057681	*Zygosaccharomyces rouxii*	99	99	AB302825
LL12_089	30°C	17	Bumblebee	KF057682	*Torulaspora delbrueckii*	99	99	HE616749
LL12_090	30°C	1	Maple tree bark	KF057683	*Torulaspora delbrueckii*	99	99	HE616749
LL12_092	30°C	25	Pine tree Bark	KF057639	*Ogataea zoltii*	99	98	AB440285
LL12_093	30°C	1	Oak tree bark	KF057500	*Lachancea thermotolerans*	100	99	GQ340441
LL12_094	30°C	1	Oak tree bark	KF057501	*Lachancea thermotolerans*	99	99	GQ340441
LL12_095	30°C	1	Maple tree bark	KF057502	*Lachancea thermotolerans*	100	99	GQ340441
LL12_098	30°C	1	Oak tree bark	KF057495	*Kluyveromyces lactis*	99	99	KC609424
LL12_101	30°C	1	Oak tree bark	KF057498	*Lachancea thermotolerans*	100	99	GQ340441
LL12_102	30°C	1	Oak tree bark	KF057499	*Ogataea parapolymorpha*	100	99	FJ914922
LL12_103	30°C	1	Oak tree bark	KF057743	*Pichia fermentans*	97	99	DQ665310
LL12_104	30°C	1	Oak tree bark	KF057744	*Lachancea thermotolerans*	99	99	GQ340441
LL12_105	30°C	1	Oak tree bark	KF057745	*Lachancea thermotolerans*	100	99	GQ340441
LL12_106	30°C	1	Oak tree bark	KF057746	*Ogataea parapolymorpha*	99	99	FJ914922

LL12_107	30°C	1	Maple tree bark	KF057747	*Lachancea thermotolerans*	100	99	GQ340441
LL12_108	30°C	1	Maple tree bark	KF057748	*Lachancea thermotolerans*	100	99	GQ340441
LL12_109	30°C	1	Maple tree bark	KF057749	*Lachancea thermotolerans*	100	99	GQ340441
LL12_110	30°C	1	Maple tree bark	KF057750	*Lachancea thermotolerans*	99	99	GQ340441
LL12_111	30°C	19	Unknown tree stump	KF057666	*Meyerozyma guiliermondii*	100	99	EF568004
LL12_113	30°C	12	Oak tree bark	KF057718	*Debaryomyces hansenii*	100	99	JN942909
LL12_114	30°C	10	Maple tree associated soil	KF057660	*Torulaspora delbrueckii*	99	99	HE616749
LL12_115	30°C	10	Maple tree bark	KF057661	*Lachancea thermotolerans*	100	99	GQ340441
LL12_116	30°C	10	Maple tree bark	KF057662	*Torulaspora delbrueckii*	99	99	HE616749
LL12_117	30°C	10	Maple tree bark	KF057663	*Lachancea thermotolerans*	100	99	GQ340441
LL12_118	30°C	10	Maple tree bark	KF057664	*Lachancea thermotolerans*	100	99	GQ340441
LL12_119	30°C	10	Maple tree bark	KF057665	*Torulaspora delbrueckii*	99	99	HE616749

Table S4: Uncharacterized yeasts isolated in this study.

The following information is shown: name of the yeast isolate, temperature of incubation used for isolation, geographical location (see Fig. 1 and Table S1), substrate, accession number of the *ITS1, 5.8S, ITS2* sequence in NCBI, name of the closest taxon found by BLAST in NCBI, percentage of sequence coverage with the closest taxon, percentage of sequence identity with the closest taxon, reference for the closest taxon and accession number of the of the *ITS1, 5.8S, ITS2* sequence of the closest taxon in NCBI.

Isolated strain					Reference sequence			
Name	Isolation T°	Location	Substrate	Accession #	Closest known species	Coverage (%)	Identity (%)	Accession #
LL11_063	18°C	1	Fallen apple	KF057554	*Pichia kluyveri*	100	90	FM864201
LL11_065	18°C	1	Fallen apple	KF057556	*Pichia kluyveri*	84	90	FM864201
LL11_067	18°C	1	Maple tree bark	KF057558	*Hanseniaspora vinea*	99	95	AY046201
LL11_083	18°C	1	Fallen strawberry	KF057574	Uncultured compost fungus	100	93	FJ176549
LL11_111	18°C	1	Fallen apple	KF057603	*Candida stellimalicola*	93	93	FM199968
LL11_112	18°C	1	Fallen apple	KF057604	*Pichia kluyveri*	99	96	FM864201
LL11_116	18°C	1	Fallen apple	KF057610	*Hanseniaspora uvarum*	98	94	GU237050
LL11_124	18°C	1	Fallen apple	KF057618	*Metschnikowia chrysoperlae*	99	95	DQ367881
LL11_133	18°C	1	Fallen apple	KF057627	*Candida* sp.**	100	93	EU570109
LL11_145	30°C	7	Maple tree bark	KF057726	*Williopsis saturnus*	85	88	EU307976
LL11_146	30°C	7	Birch tree bark	KF057727	*Williopsis saturnus*	85	88	EU307976
LL11_148	30°C	7	Maple tree bark	KF057729	*Komagataella pastoris*	98	90	FR839631
LL11_151	30°C	7	Maple tree bark	KF057732	*Komagataella pastoris*	93	91	FR839631
LL11_152	30°C	7	Maple tree bark	KF057733	*Komagataella pastoris*	97	90	FR839631
LL11_156	30°C	7	*Prunus cistena* fruit	KF057739	*Starmerella* sp.	99	91	HQ111058
LL12_048	30°C	24	Maple tree bark	KF057686	*Candida* sp.	99	95	FJ873586
LL12_060	30°C	12	Maple tree bark	KF057698	*Ogataea chonburiensis*	99	94	FJ914924

LL12_079	30°C	14	Fly	KF057670	*Starmerella* sp.	100	92	JX112044
LL12_086	30°C	18	Bumblebee	KF057679	*Candida apicola*	100	93	EU926486
LL12_087	30°C	18	Bumblebee	KF057680	*Candida apicola*	100	93	EU926486
LL12_091	18°C	25	Aspen tree bark	KF057638	*Candida* sp.	98	94	FJ873586
LL12_096	30°C	1	Oak tree bark	KF057493	Uncultured compost fungus	100	85	FM177656
LL12_099	30°C	1	Oak tree bark	KF057496	*Pichia* sp.	97	87	FM177656
LL12_112	30°C	12	Oak tree bark	KF057714	*Pichia membranifaciens*	87	89	FJ231463

*= Only one sequence obtained as blast result

Table S5: List of yeast strains used in this study.

HO deletion cassette, mating type and lineage are indicated.

Strain	Sampling site	Reference	Construction used			Lineage/ Species	Ref Figure 2
			ho::KAN	*ho::NAT*	*ho::HPH*		
YPS744	Tuscarora Forest, PE	(Kuehne *et al.* 2007)	a	–	–	A	1
YPS642	Buck Hill Falls, PE	(Kuehne *et al.* 2007)	–	a	–	A	2
YPS695	Tuscarora Forest, PE	(Kuehne *et al.* 2007)	α	–	a,α	A	3
YPS644	Buck Hill Falls, PE	(Kuehne *et al.* 2007)	–	a, α	α	A	4
UWOPS-91-202.1	Long Point, On	(Leducq *et al.* 2014)	–	a, α	α	B	5
YPS484	Grand Bend, On	(Kuehne *et al.* 2007)	–	a, α	–	B	6
LL2012_028	Sherbrooke, QC	(Leducq *et al.* 2014)	α	–	a,α	B	7
LL2012_021	Pointe Platon, QC	(Leducq *et al.* 2014)	–	a, α	α	B	8
yHBJ1	Iron River, WI	(Leducq *et al.* 2014)	α	–	a, α	B	9
LL2012_029	St-Thecle, QC	(Leducq *et al.* 2014)	α	–	a, α	B	10
UWOPS-79-140	Saint-Joseph Island, ON	(Leducq *et al.* 2014)	–	–	a, α	B	11
LL2012_016	Pointe Platon, QC	(Leducq *et al.* 2014)	α	–	a	C	12
LL2011_004	Cap Chat, QC	(Leducq *et al.* 2014)	a, α	–	a, α	C	13
LL2012_004	Cap Chat, QC	(Leducq *et al.* 2014)	a, α	–	a, α	C	14
LL2012_011	Parc du Bic, QC	(Leducq *et al.* 2014)	–	a, α	a	C	15
MSH-587-1	Mont St Hilaire, QC	(Leducq *et al.* 2014)	–	a, α	α	C	16
LL2011_002	Cap Chat, QC	(Leducq *et al.* 2014)	–	–	a	C	17
LL2011_001	Québec City, QC	(Leducq *et al.* 2014)	–	α	α	C	18
LL2011_011	Squatec, QC	(Leducq *et al.* 2014)	–	–	a, α	C	19
LL2011_009	Squatec, QC	(Leducq *et al.* 2014)	–	a, α	a	C	20
LL2011_006	Ile d'Orléans, QC	(Leducq *et al.* 2014)	a, α	–	–	C	21
LL2012_006	Cap Chat, QC	(Leducq *et al.* 2014)	–	a, α	–	C	22
LL2012_018	Pointe Platon, QC	(Leducq *et al.* 2014)	–	a, α	–	C	23
LL2011_005	Ile d'Orléans, QC	(Leducq *et al.* 2014)	–	a	–	C	24

YPS667	Buck Hill Falls, PE	(Kuehne *et al.* 2007)	–	a	α	C	25
BY4741	–	(Baker Brachmann *et al.* 1998)	–	a, α	–	*S. cerevisiae*	N/A
JRY 8147	–	(Gallagher *et al.* 2009)	–	a	α	*S. uvarum*	N/A

Table S6: List of oligonucleotides and plasmids used for *HO* locus deletion.

Primer	Sequence
C lineage *HO* locus sequencing	
Forward Primer	
OP40_E6	TTTTGTTTTCTCTTAACTTTGTATCCT
Reverse Primer	
OP40_E7	CGAAGAGATCCACGAAAAGC
Cassette construction	
Forward Primers	
PAR_HO_5 (A)	CACACCCTTATAAGCAGCAATCAATTTCATCTAACTT CAACCAGCTGAAGCTTCGTACGC
OP40_E9 (B & C)	ACATCCTTATAGGCAGCAATCAATTCCATCTAAACTT TAACCAGCTGAAGCTTCGTACGC
Reverse Primers	
PAR_HO_3 (A)	TTATCACATACAACTTTTTTTTAACTAATGTACACATT GCATAGGCCACTAGTGGATCTG
OP40_E8 (B & C)	TTAATTACATAACAATTTTTTTTTATAATATACATATTG CATAGGCCACTAGTGGATCTG
5’ verification	
Forward Primers	
OP48_C10 (A)	ACAGAAGCTTGTTGAAGTGC
OP48_C11 (B & C)	ACAGAAGCTTGTTGAAGCGC
Reverse Primers	
KAN_B	CTGCAGCGAGGAGCCGTAAT
HPH_B	GTCGCGGTGAGTTCAGGCTT
NAT_B	CGGTAAGCCGTGTCGTCAAG
3’ verification	
Forward Primer	
TEF_TERM	CCAGATGCGAAGTTAAGTGC
Reverse Primers	
CLO3-2 (A & C)	CATTCAAGTAAAGAGATCACC
OP48c12 (B)	CATACAAGTAAAGAGACCACC
Plasmids (Cassette construction)	
pAG25	(Goldstein & McCusker 1999)
pAG32	(Goldstein & McCusker 1999)
pUG6	(Guldener *et al.* 1996)

Table S7: List of crosses realized in this study.

Cross type, spore viability, Karyotypic and genotypic divergence are indicated.

Strain1	Strain2	Cross type	Spores viability	Karyotypic divergence	Genetic divergence
YPS744	YPS744	AA	0.7970*	0.0000	0.0000
YPS695	YPS744	AA	0.7450*	0.0345	0.0000
YPS642	YPS744	AA	0.8590*	0.0714	0.0000
YPS644	YPS744	AA	0.8330*	0.0714	0.0000
YPS744	YPS484	AB	0.4480*	0.1034	0.0257
YPS744	YHBJ1	AB	0.5100*	0.1724	0.0278
LL2012_029	YPS744	AB	0.3440	0.2413	0.0287
YPS642	UWOPS-79-140	AB	0.4270	0.2500	0.0257
YPS644	LL2012_028	AB	0.2710*	0.2592	0.0272
YPS744	LL2012_021	AB	0.4530*	0.2857	0.0279
LL2012_029	YPS642	AB	0.3540	0.3333	0.0287
YPS744	UWOPS-79-140	AB	0.5570*	0.4615	0.0257
YPS744	MSH-587-1	AC	0.3440*	0.3077	0.0304
YPS642	YPS667	AC	0.4790*	0.3600	0.0340
YPS644	LL2011_009	AC	0.3650*	0.3600	0.0304
YPS695	YPS667	AC	0.3230	0.3846	0.0340
YPS642	LL2012_004	AC	0.2080*	0.4074	0.0304
YPS744	LL2012_016	AC	0.2400*	0.4074	0.0276
YPS695	LL2011_011	AC	0.2710*	0.4815	0.0304
YHBJ1	LL2012_028	BB	0.4479	0.0714	0.0065
YPS484	UWOPS-79-140	BB	0.7190	0.1200	0.0022
UWOPS-91-202.1	LL2012_021	BB	0.6880	0.2143	0.0022
UWOPS-79-140	LL2012_021	BB	0.8540	0.2500	0.0043
YPS484	LL2012_021	BB	0.7660*	0.2592	0.0022
LL2012_029	YHBJ1	BB	0.3540	0.2857	0.0080
UWOPS-91-202.1	YHBJ1	BB	0.5520	0.3793	0.0050
LL2012_029	LL2012_021	BB	0.5100	0.4074	0.0036
UWOPS91.202.1	LL2012_028	BB	0.3854	0.4483	0.0014
YPS667	YHBJ1	BC	0.5100	0.1538	0.0264
LL2011_002	UWOPS-79-140	BC	0.4900	0.1818	0.0199
YPS484	MSH-587-1	BC	0.4900*	0.1999	0.0199
LL2012_016	LL2012_021	BC	0.4480	0.1999	0.0188
LL2012_018	LL2012_021	BC	0.4580	0.1999	0.0206
LL2012_029	LL2011_006	BC	0.5000	0.2143	0.0212
LL2011_005	LL2012_028	BC	0.3020	0.2307	0.0214
YPS484	YPS667	BC	0.6510*	0.3077	0.0199
UWOPS-79-140	LL2011_009	BC	0.5160*	0.3913	0.0199
YPS484	LL2012_004	BC	0.2660*	0.4285	0.0199
YPS484	LL2011_011	BC	0.4380	0.4815	0.0199
UWOPS-91-202.1	MSH-587-1	BC	0.6040	0.6153	0.0198
LL2012_021	LL2011_011	BC	0.3854	0.6923	0.0206
LL2012_028	MSH-587-1	BC	0.1979	0.1999	0.0214

LL2012_028	LL2011_002	BC	0.1354	0.3600	0.0214
LL2012_028	LL2011_004	BC	0.3438	0.3846	0.0214
LL2012_028	LL2012_016	BC	0.2813	0.3077	0.0196
LL2012_021	LL2012_004	BC	0.2292	0.4815	0.0206
UWOPS-91-202.1	LL2012_004	BC	0.2813	0.7931	0.0198
UWOPS-91-202.1	LL2011_001	BC	0.2083	0.3333	0.0198
UWOPS-91-202.1	LL2012_018	BC	0.5521	0.2592	0.0198
UWOPS-91-202.1	LL2011_011	BC	0.4479	0.5714	0.0198
YPS667	YPS667	CC	0.9580	0.0000	0.0000
LL2011_005	LL2012_016	CC	0.8540	0.0833	0.0008
LL2012_016	LL2012_018	CC	0.9170	0.0833	0.0008
LL2011_005	LL2011_001	CC	0.1980	0.1667	0.0000
MSH-587-1	YPS667	CC	0.6820*	0.2174	0.0022
LL2011_006	LL2011_011	CC	0.4380	0.2592	0.0008
LL2011_006	LL2012_011	CC	0.4900	0.2592	0.0008
LL2012_016	YPS667	CC	0.6150*	0.3333	0.0023
LL2012_004	LL2011_009	CC	0.5260*	0.3846	0.0000
YPS667	LL2011_009	CC	0.6610*	0.4167	0.0022
LL2012_016	LL2011_002	CC	0.4380	0.4783	0.0008
LL2011_005	MSH-587-1	CC	0.3960	0.4783	0.0000
LL2011_002	LL2012_004	CC	0.5210	0.5200	0.0000
LL2012_016	LL2011_009	CC	0.4690	0.5833	0.0008
LL2011_009	LL2011_011	CC	0.8330	0.6000	0.0000
LL2011_005	LL2012_006	CC	0.4690	0.6153	0.0000
LL2012_016	LL2011_004	CC	0.4900	0.7500	0.0008
LL2011_006	LL2012_018	CC	0.8958	0.1538	0.0008
LL2011_006	MSH-587-1	CC	0.4271	0.3600	0.0008
LL2011_006	LL2011_002	CC	0.3958	0.1999	0.0008
LL2011_006	LL2011_004	CC	0.3646	0.4615	0.0008
LL2012_011	LL2012_018	CC	0.5104	0.3600	0.0000
LL2012_011	MSH-587-1	CC	0.9167	0.1304	0.0000
LL2012_011	LL2011_002	CC	0.9063	0.0833	0.0000
LL2012_011	LL2011_004	CC	0.9271	0.6000	0.0000
LL2011_004	LL2011_011	CC	0.9792	0.1999	0.0000
LL2011_004	LL2012_004	CC	0.3958	0.2307	0.0000
LL2011_004	LL2012_018	CC	0.3854	0.5833	0.0000
LL2011_004	LL2011_001	CC	0.4479	0.6667	0.0000
LL2012_006	LL2012_016	CC	0.4688	0.5385	0.0008
LL2012_006	MSH-587-1	CC	0.9688	0.3600	0.0000
LL2012_006	LL2011_002	CC	0.9063	0.1200	0.0000
LL2012_006	LL2011_004	CC	0.8750	0.2307	0.0000
LL2011_001	LL2012_016	CC	0.2396	0.3333	0.0008
LL2011_001	MSH-587-1	CC	0.5208	0.3333	0.0000
LL2011_001	LL2011_002	CC	0.5417	0.3043	0.0000
LL2012_018	LL2011_005	CC	0.7813	0.0000	0.0000
YPS744	BY4741	A-*Scer*	0	0.2730	0.0684
LL2012_021	BY4741	B-*Scer*	0.0104	0.1959	0.0774

MSH-587-1	BY4741	C-*Scer*	0	0.4037	0.0812
YPS744	JRY8147	A-*Suva*	0	NA	0.1484
LL2012_028	JRY8147	B-*Suva*	0	NA	0.1566
LL2012_018	JRY8147	C-*Suva*	0.0208	NA	0.1534

*48 tetrads were dissected for those crosses instead of 24

Table S8: Results from linear model (*lm*) of spore survival (*S*) variation in function of genetic divergence (*Dg*) and karyotypic divergence (*Dc*) and their interaction (*Dg*Dc*) as additional factors (*lm Formula = S ~ dg+dc+ dg*dc*).

The model was performed by considering three cases: all cross considered (M1), only crosses within lineages considered (M2) or only crosses among lineages and species considered (M3). Comparisons including *S. uvarum* were removed from the analysis. For each model and each factor, the following information is shown: the estimated coefficient (E.C.) of the factor (the sign of the estimate indicates the direction of the correlation), its standard error (S.E.) and its *t.value*. The significance of each estimated coefficient (*t*-test) is represented by following symbols: "***" (*p.value* <0.001), "**" (*p.value* <0.01), "*" (*p.value* <0.05) and "°"(*p.value* <0.1). The proportion of permutations given a lower E.C. than observed (*P.*) is indicates in bold when lower than 0.05 (distribution of E.C. estimated other 100,000 independent random permutations, by taking into account the non-independence of pairwise comparisons and the bias in strain usage; see Figures S1 and S4).

Model	Factor	(E.C.)	S.E.	*t-value*	*P*
M1	Intercept	0.78	0.04	18.28***	0.9999
	Dg	-12.85	2.82	-4.56***	**0.0107**
	Dc	-0.52	0.12	-4.32***	**0.0039**
	*Dg*Dc*	12.85	8.40	1.53	0.8118
M2	Intercept	0.82	0.06	14.55***	0.9999
	Dg	-39.03	35.64	-1.10	0.2915
	Dc	-0.56	0.16	-3.38***	**0.0053**
	*Dg*Dc*	42.69	126.81	0.34	0.5715
M3	Intercept	0.62	0.12	5.26***	0.4277
	Dg	-7.90	4.44	-1.78°	0.4901
	Dc	-0.20	0.35	-0.59	0.5125
	*Dg*Dc*	2.27	13.82	0.16	0.4414

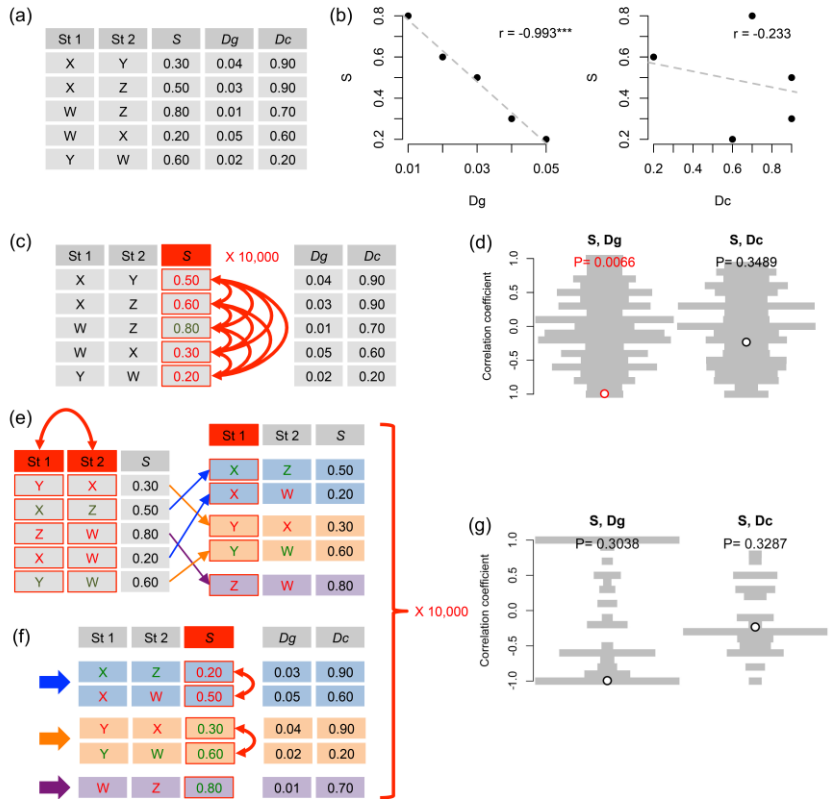

Figure S1: Principle of permutation tests used in this study to control for the effects of non-independence of pairwise comparisons and of bias in strain usage in pairwise comparisons.

(a) Hybrid spore survival (*S*), genetic divergence (*Dg*) and karyotypic divergence (*Dc*) were measured between two strains (St1 and St2) in five hypothetical crosses involving four hypothetical strains (W, X, Y and Z). Each strain was used in several crosses, inducing non-independence of pairwise comparisons among strains. Cross between Z and Y was not performed, introducing a bias in strain usage. (b) *S* and *Dg* are significantly (left; "***": $p<0.001$) and negatively correlated. *S* and *Dc* are not correlated (right). (c) Values of *S* were randomly permuted among comparisons (in red); *Dg* and *Dc* values were fixed. By chance, some *S* values were not permuted (in green). Random permutations were repeated 10,000 times. (d) Violin diagrams showing the distribution of correlation coefficients (in grey) between *S* and *Dg* (left) or *Dc* (right) obtained after 10,000 permutations. The low proportion ($P=0.0081$) of permutations given a lower coefficient than observed indicates that the *S*:*Dg* correlation (red circle) is stronger than expected, but not the *S*:*Dc* comparison (black circle; $P=0.3465$). (e) To control for non-independence of pairwise comparisons and

bias in strain usage, we first randomly permutated St1 and St2 names in each cross (left). By chance, some names were not permuted (in green). We divided the dataset according to the strains in column St1, generating three categories corresponding to strains X (blue), Y (orange) and W (purple), respectively (right). (f) In each category, values of *S* were randomly permuted among comparisons (in red); *Dg* and *Dc* values were fixed. By chance, some *S* values were not permuted (in green). Steps (e) and (f) were repeated 10,000 independent times. (g) The distribution of correlation coefficients between *S* and *Dg* (left) is biased for extreme values due to the structure of data, then the observed coefficient is not stronger than expected (black circle; P=0.3021).

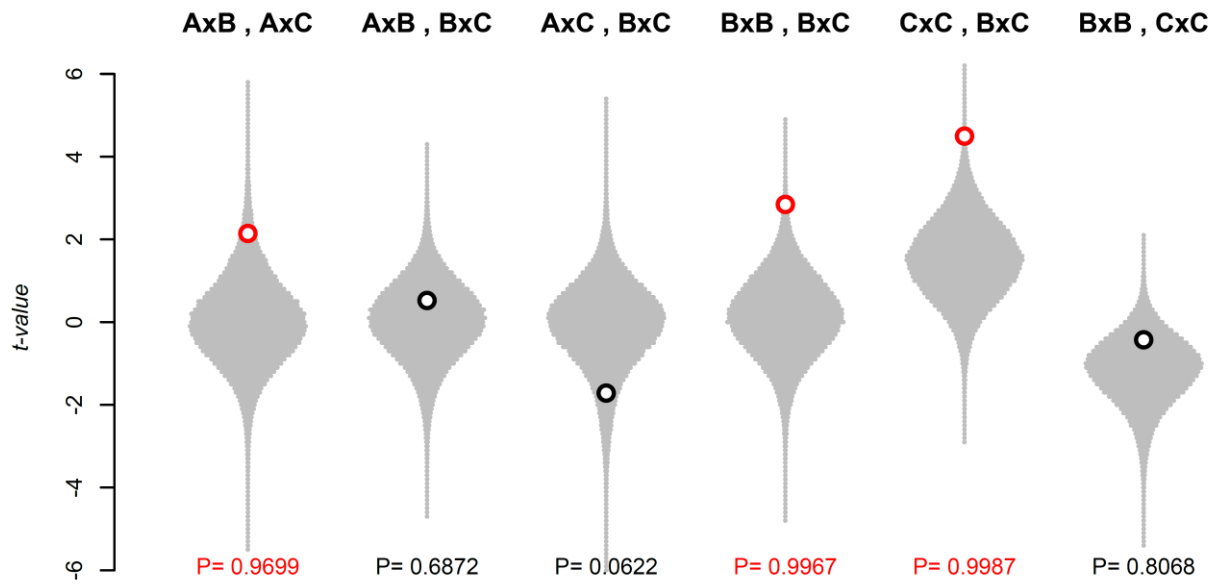

Figure S2: Violin plots showing the expected distribution (grey lines) of *t*-values from Welch t-tests comparisons of *S* values among categories of crosses (see Figure 3a).

The observed *t*-value is represented by a circle in each category, colored in red when lower ($P \leq 0.05$) or higher ($P \geq 0.95$) than expected. The expected distribution was obtained from Welch *t*-tests in 100,000 independent random permutations, by taking into account the non-independence of pairwise comparisons and the bias in strain usage (see Figure S1).

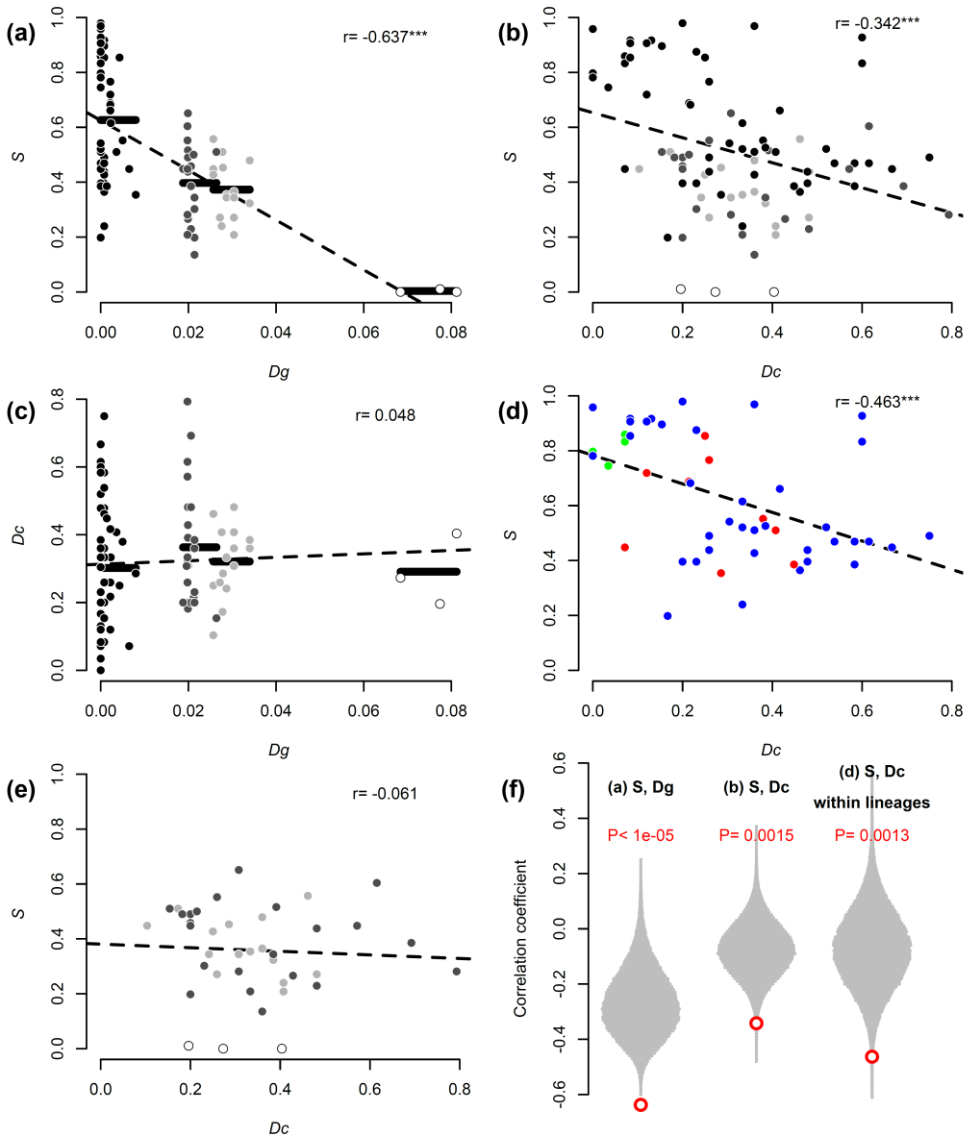

Figure S3: Spore survival (*S*) is negatively correlated with genetic divergence (*Dg*) and chromosomal divergence (*Dc*).

Raw measures of *S* as a function of (a) *Dg* and (b) *Dc*. Different colors symbolize different evolutionary categories: within lineage crosses (black); among American lineages B×C crosses (dark grey); among European and American A×B and A×C crosses (grey) and among species *S. paradoxus* × *S. cerevisiae* crosses (white). Horizontal black bars represent mean *S* value in each category. (c) Raw measures of *Dg* in function of *Dc*. (d) Raw measures of *S* for crosses performed among lineages only, as a function of *Dc*. Different colors symbolize different lineages: A (green); B (red) and C (blue). (e) Raw measures of *S* for crosses performed among lineages only, as a function of *Dc*. In each panel, the dotted

line represents the linear regression of *S* as a function of *Dc* of *Dg*. The significance of the correlation coefficient *r* (Pearson's test) is represented by following symbols: "***" (*p.value* <0.001), "**" (*p.value* <0.01) and "*" (*p.value* <0.05). (f) Violin plots showing the expected distribution (grey lines) of correlation coefficients for the regressions described in panels (a), (b) and (d). The observed coefficients (red circles) are lower than expected by chance ($P<0.05$). The expected distribution was obtained from Pearson's correlation tests in 100,000 independent random permutations by taking into account the non-independence of pairwise comparisons and the bias in strain usage (see Figure S1).

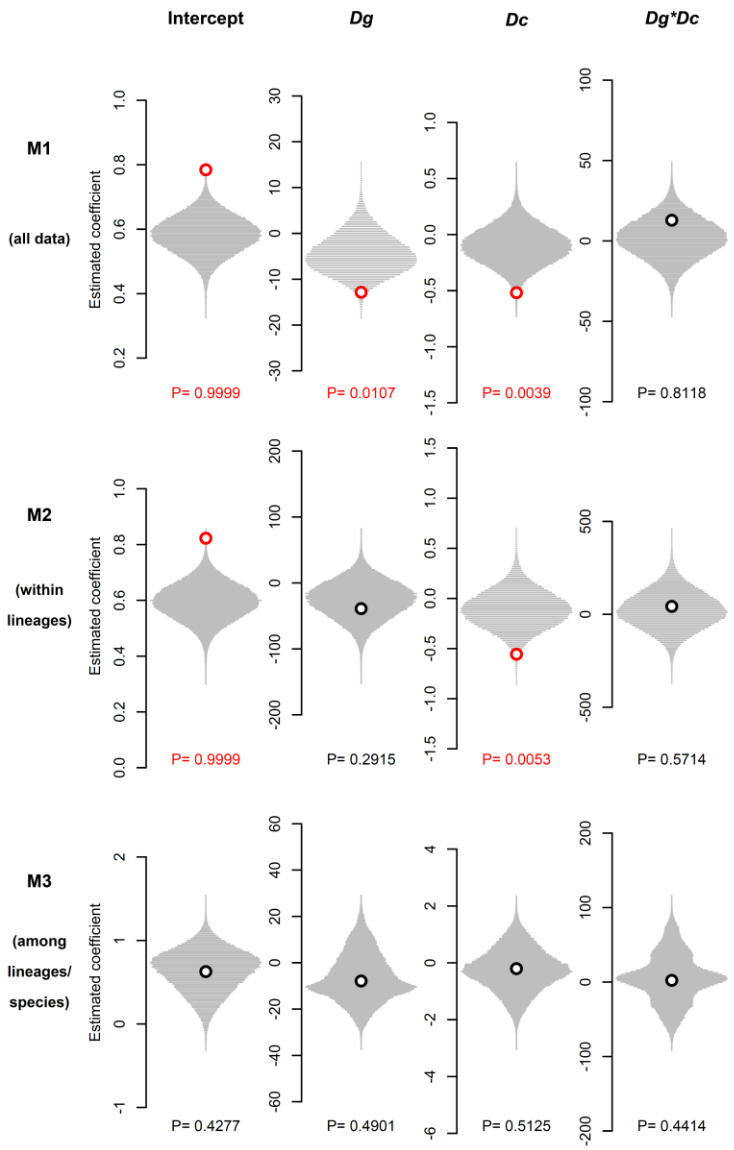

Figure S4: Violin plots showing the expected distribution (grey lines) of estimated coefficients from linear models (see Table S4).

For each model (in row: M1, M2 and M3) and each factor (in column: *Dg*, *Dc* and *Dg*Dc*), the observed coefficient is represented by a circle, colored in red when lower ($P \leq 0.05$) or higher ($P \geq 0.95$) than expected. The expected distribution was obtained from linear models in 100,000 independent random permutations, by taking into account the non-independence of pairwise comparisons and the bias in strain usage (see Figure S1).

Appendix 1: scripts used to generate figures from data.

```
#Chromosomal variation segregates within incipient species and correlates
with reproductive isolation.
#Guillaume Charron*, Jean-Baptiste Leducq* and Christian R Landry
#Supplementaty information: R script for statistical analyses

#LOADING DATA FILE

        setwd("/Users/jean-baptisteleducq/Dropbox/Souches
sauvages/ARTICLE_REPRODUCTIVE_ISOLATION/ResoumissionMOLECOL/Data")
        setwd("/Users/JBLED/Dropbox/Souches
sauvages/ARTICLE_REPRODUCTIVE_ISOLATION/ResoumissionMOLECOL/Data")

        data=read.table("Data.txt", header=TRUE)

#Description- St1: MATa strain; St2: MATalpha strain; Cross: categories
of cross according to species/Lineages; S: Spore survival; Dc:
Chromosomal divergence; Dg: Genetic divergence; Type: Categories of
crosses for graphical representation

#OPTIONAL: REMOVING CROSSES WITH S. cerevisiae (ScSp) and S. uvarum
(SuSp) DO NOT AFFECT THE MAIN RESULTS (DEFAULT: SuSp REMOVED IN THE
MANUSCRIPT FOR GRAPHICAL REPRESENTATION CONVENIENCE)

        data<-subset(data,data[,3]!="SuSp")
        #data<-subset(data,data[,3]!="ScSp")

        St1<-data$St1
        St2<-data$St2
        Cross<-data$Cross
        S<-data$S
        Dg<-data$Dg
        Dc<-data$Dc
        Type<-data$Type

#STATISTICAL ANALYSES ON OBSERVED DATA- Linear models (LM), Welch t-test
(WT) and correlation coefficients (CT)

        #GLOBAL LINEAR MODEL (M1)

                m=lm(S~Dg*Dc)
                M1<-summary(m)
                M1<-M1$coefficients
                M1<-round(M1,6)
                colnames(M1)<-c("Estimated coefficient","S.E.","t","p-
value")
                rownames(M1)<-c("M1:Intercept","M1:Dg","M1:Dc","M1:Dg*Dc")

        #LINEAR MODEL (M2): CROSSES WITHIN LINEAGES ONLY

                mi<-
lm(subset(S,Type=="intra")~subset(Dg,Type=="intra")*subset(Dc,Type=="intr
a"))
                M2<-summary(mi)
                M2<-M2$coefficients
```

```
                M2<-round(M2,6)
                colnames(M2)<-c("Estimated_coefficient","S.E.","t","p-
value")
                rownames(M2)<-c("M2:Intercept","M2:Dg","M2:Dc","M2:Dg*Dc")

        #LINEAR MODEL (M3): CROSSES AMONG LINEAGES/SPECIES ONLY

                mI<-
lm(subset(S,Type!="intra")~subset(Dg,Type!="intra")*subset(Dc,Type!="intr
a"))
                M3<-summary(mI)
                M3<-M3$coefficients
                M3<-round(M3,6)
                colnames(M3)<-c("Estimated coefficient","S.E.","t","p-
value")
                rownames(M3)<-c("M3:Intercept","M3:Dg","M3:Dc","M3:Dg*Dc")

                LM<-rbind(M1,M2,M3)

        #WELCH T-TESTS AMONG CATEGORIES OF CROSS

                WT<-t.test(subset(S,Cross=="AB"),subset(S,Cross=="AC"))
                WT.AB.AC<-c(WT$estimate,WT$statistic,WT$p.value)
                WT<-t.test(subset(S,Cross=="AB"),subset(S,Cross=="BC"))
                WT.AB.BC<-c(WT$estimate,WT$statistic,WT$p.value)
                WT<-t.test(subset(S,Cross=="AC"),subset(S,Cross=="BC"))
                WT.AC.BC<-c(WT$estimate,WT$statistic,WT$p.value)
                WT<-t.test(subset(S,Cross=="BB"),subset(S,Cross=="BC"))
                WT.BB.BC<-c(WT$estimate,WT$statistic,WT$p.value)
                WT<-t.test(subset(S,Cross=="CC"),subset(S,Cross=="BC"))
                WT.CC.BC<-c(WT$estimate,WT$statistic,WT$p.value)
                WT<-t.test(subset(S,Cross=="BB"),subset(S,Cross=="CC"))
                WT.BB.CC<-c(WT$estimate,WT$statistic,WT$p.value)

                WT<-round(rbind(WT.AB.AC, WT.AB.BC, WT.AC.BC, WT.BB.BC,
WT.CC.BC, WT.BB.CC),6)
                colnames(WT)<-c("Type1","Type2","t","p-value")

        #PEARSON'S CORRELATION TESTS

                #S against Dg (S uncorrected)

                        x<-S
                        y<-Dg
                        C<-cor.test(x,y)
                        CT.dg<-c(C$estimate,C$statistic,C$p.value)

                #S against Dc (S uncorrected)

                        y<-Dc
                        C<-cor.test(x,y)
                        CT.dc<-c(C$estimate,C$statistic,C$p.value)

                #S against Dg (S corrected by Dc: residuals from linear
model lm(S~Dc))

                        x<-lm(S~Dc)
```

```
                        x<-as.matrix(x$residuals)
                        y<-Dg
                        C<-cor.test(x,y)
                        CT.dg.corrected<-
c(C$estimate,C$statistic,C$p.value)

                #S against Dc (S corrected by Dg: residuals from linear
model lm(S~Dg))

                        x<-lm(S~Dg)
                        x<-as.matrix(x$residuals)
                        y<-Dc
                        C<-cor.test(x,y)
                        CT.dc.corrected<-
c(C$estimate,C$statistic,C$p.value)

                #Dg against Dc

                        x<-Dg
                        y<-Dc
                        C<-cor.test(x,y)
                        CT.dg.dc<-c(C$estimate,C$statistic,C$p.value)

        #PEARSON'S CORRELATION TESTS WITHIN LINEAGES ONLY

                #S against Dg (S uncorrected)

                        x<-subset(S,Type=="intra")
                        y<-subset(Dg,Type=="intra")
                        C<-cor.test(x,y)
                        CT.dg.intra<-c(C$estimate,C$statistic,C$p.value)

                #S against Dc (S uncorrected)

                        y<-subset(Dc,Type=="intra")
                        C<-cor.test(x,y)
                        CT.dc.intra<-c(C$estimate,C$statistic,C$p.value)

                #S against Dg (S corrected by Dc: residuals from linear
model lm(S~Dc))

                        x<-
lm(subset(S,Type=="intra")~subset(Dc,Type=="intra"))
                        y<-subset(Dg,Type=="intra")
                        x<-as.matrix(x$residuals)
                        C<-cor.test(x,y)
                        CT.dg.intra.corrected<-
c(C$estimate,C$statistic,C$p.value)

                #S against Dc (S corrected by Dg: residuals from linear
model lm(S~Dg))

                        x<-
lm(subset(S,Type=="intra")~subset(Dg,Type=="intra"))
                        y<-subset(Dc,Type=="intra")
                        x<-as.matrix(x$residuals)
                        C<-cor.test(x,y)
```

```
                        CT.dc.intra.corrected<-
c(C$estimate,C$statistic,C$p.value)

               #Dg against Dc

                        x<-subset(Dg,Type=="intra")
                        y<-subset(Dc,Type=="intra")
                        C<-cor.test(x,y)
                        CT.dg.dc.intra<-
c(C$estimate,C$statistic,C$p.value)

CT<-round(rbind(CT.dg, CT.dc, CT.dg.corrected,
CT.dc.corrected,CT.dg.intra, CT.dc.intra, CT.dg.intra.corrected,
CT.dc.intra.corrected,CT.dg.dc,CT.dg.dc.intra),6)
colnames(CT)<-c("Correlation_coefficient","t","p-value")

############## ############## ############## ##############
       ############## ##############

#STATISTICAL ANALYSES ON RANDOMIZED S VALUES (spore survival) TO
DETERMINE WETHER t ESTIMATED FROM OBSERVED DATA ARE SIGNIFICANTLY
DIFFERENT THAN THE EXPECTED DISTRIBUTION OF t VALUES, BY TAKING INTO
ACCOUNT THAT PAIRWISE COMPARISONS ARE NOT INDEPENDANT

#NOTE: The dataset is an incomplete matrix of crosses between 27 strains
(25 if removing Suva and Scer): not all crosses were realized, then
randomization cannot be performed between independent crosses.

       #FIGURE S!: EXEMPLE between four strains W, X, Y and Z; cross Y x
Z was not realized

                                Ex.k<-c(1:5)
                                Ex.St1<-c("X","X","W","W","Y")
                                Ex.St2<-c("Y","Z","Z","X","W")
                                Ex.S<-c(0.30,0.50,0.80,0.20,0.60)
                                Ex.Dg<-c(0.04,0.03,0.01,0.05,0.02)
                                Ex.Dc<-c(0.90,0.90,0.70,0.60,0.20)

                                par(bty="n",mfrow=c(1,2))
                                plot(Ex.Dg,Ex.S,pch=19,xlab="Dg",ylab="S")
                                abline(lm(Ex.S ~
Ex.Dg),lty=2,lwd=2,col="grey")
                                plot(Ex.Dc,Ex.S,pch=19,xlab="Dc",ylab="S")
                                abline(lm(Ex.S ~
Ex.Dc),lty=2,lwd=2,col="grey")

                        #CORRELATION TESTS ON OBSERVED VALUES: S is
significantly correlated with Dg (but not with Dc)


       rbind(cor.test(Ex.S,Ex.Dg)[c(1,3:4)],cor.test(Ex.S,Ex.Dc)[c(1,3:4
)])

#     statistic  p.value      estimate
#[1,] -15        0.0006431193 -0.9933993
#[2,] -0.4142706 0.7065146    -0.2326181
```

```
                #Values of S are randomized, WITHOUT TAKING INTO ACCOUNT
NON-INDEPENDANCE OF COMPARISONS

                                #NEW  CORRELATION TESTS ON RANDOMIZED
VALUES

                                #Do it 10,000 times

                                v<-c(1:10000)

                                T<-sapply(v,function(x){
                                        Ex.Sr<-sample(Ex.S)

                                        C<-cor.test(Ex.Sr,Ex.Dg)
                                        Ex.CT.dg<-
c(C$estimate,C$statistic,C$p.value)
                                        C<-cor.test(Ex.Sr,Ex.Dc)
                                        Ex.CT.dc<-
c(C$estimate,C$statistic,C$p.value)

                                        return(c(Ex.CT.dg, Ex.CT.dc))
                                })

                                par(bty="n",mfrow=c(1,2),mar=c(0,4,2,0))
                                h<-hist(T[1,],breaks=20,plot=FALSE)
                                plot(-10,-10,ylim=c(-1,1),xlim=c(-
1,1),main="S, Dg",xlab="",ylab="Correlation coefficient",xaxt="n")
                                segments(-
h$counts/max(h$counts),h$breaks,h$counts/max(h$counts),h$breaks,lwd=10,co
l="grey",lend=2)
                                P<-
length(subset(T[1,],T[1,]<=as.numeric(cor.test(Ex.S,Ex.Dg)[c(4)])))/max(v
)


        text(0,0.95,labels=ifelse(P==0,paste("P<",1/max(v)),paste("P=",ro
und(P,4))),col=ifelse(P<=0.05,"red",ifelse(P>=0.95,"red","black")),cex=1.
2)
                                par(new="T")

        plot(0,as.numeric(cor.test(Ex.S,Ex.Dg)[c(4)]),pch=21,type="p",cex
=1.5,ylim=c(-1,1),xlim=c(-
1,1),main="",xlab="",ylab="",xaxt="n",yaxt="n",col=ifelse(P<=0.05,"red",i
felse(P>=0.95,"red","black")),bg="white",lwd=2)

                                par(mar=c(0,0,2,4))

                                h<-hist(T[4,],breaks=20,plot=FALSE)
                                plot(-10,-10,ylim=c(-1,1),xlim=c(-
1,1),main="S, Dc",xlab="",ylab="",xaxt="n",yaxt="n")
                                segments(-
h$counts/max(h$counts),h$breaks,h$counts/max(h$counts),h$breaks,lwd=10,co
l="grey",lend=2)
                                P<-
length(subset(T[4,],T[4,]<=as.numeric(cor.test(Ex.S,Ex.Dc)[c(4)])))/max(v
)
```

```
        text(0,0.95,labels=ifelse(P==0,paste("P<",1/max(v)),paste("P=",ro
und(P,4))),col=ifelse(P<=0.05,"red",ifelse(P>=0.95,"red","black")),cex=1.
2)
                                par(new="T")

        plot(0,as.numeric(cor.test(Ex.S,Ex.Dc)[c(4)]),pch=21,type="p",cex
=1.5,ylim=c(-1,1),xlim=c(-
1,1),main="",xlab="",ylab="",xaxt="n",yaxt="n",col=ifelse(P<=0.05,"red",i
felse(P>=0.95,"red","black")),bg="white",lwd=2)


                                #Count of the number of times correlation
observed is lower or equal than expected


        length(subset(T[1,],T[1,]<=cor.test(Ex.S,Ex.Dg)[c(4)]))

        length(subset(T[4,],T[4,]<=cor.test(Ex.S,Ex.Dc)[c(4)]))

                                #       Dg                              Dc
                                #       66 (P = 0.0066)3489 (P = 0.3489)

                                #The correlation between S and Dg is
stronger (p = 0.0081) than expected among 10,000 randomizations

#Values of S are randomized, BY TAKING INTO ACCOUNT NON-INDEPENDANCE OF
COMPARISONS. Here, S values could not be permutated between crosses X x Y
and W x Z because they are independant

                                #First, randomize St1 and St2 for each
cross (row)

                                #Second, define column St1 as a filter to
randomize S values and randomize S values according to this filter: here,
S is randomized according to X in rows 1 and 4 and Z in rows 2 and 3

                                #Third, recalculted t values and do it
10,000 times

                                v<-c(1:10000)
                                St<-cbind(Ex.St1,Ex.St2)

                                T<-sapply(v,function(x){

                                        #1: permute St1 and St2

                                        strain<-sapply(Ex.k,function(x){
                                        kx<-sample(St[x,1:2])
                                        return(t(kx))
                                        })
                                        strain<-t(strain)

                                        datax<-
cbind(strain,Ex.S,Ex.Dg,Ex.Dc)
                                        datax <-datax[order(datax[,1]),]
```

```
                                        z<-sort(unique(datax[,c(1)]))

                                        #2: permute S value according to z

                                        Sx<-sapply(z,function(x){
                                        S1<-subset(datax[,3],datax[,1]==x)
                                        S1<-sample(S1)
                                        return(S1)
                                        })
                                        Sx<-matrix(unlist (Sx),nrow=1)

                                        #3: recalculate t values

                                        C<-
cor.test(as.numeric(Sx),as.numeric(datax[,4]))
                                        Ex.CT.dg<-
c(C$estimate,C$statistic,C$p.value)
                                        C<-
cor.test(as.numeric(Sx),as.numeric(datax[,5]))
                                        Ex.CT.dc<-
c(C$estimate,C$statistic,C$p.value)

                                        return(c(Ex.CT.dg, Ex.CT.dc))
                                })

                                #Count of the number of times correlation
observed is lower or equal than expected

        length(subset(T[1,],T[1,]<=cor.test(Ex.S,Ex.Dg)[c(4)]))

        length(subset(T[4,],T[4,]<=cor.test(Ex.S,Ex.Dc)[c(4)]))
                                #       Dg                              Dc
                                #       3038 (P = 0.3038)       6713 (P =
0.6713)

                                #The correlation between S and Dg is not
stronger (p = 0.3456) than expected among 10,000 randomizations, taking
into accound non-independence of pairwise comparisons

#####################
#Analyses on actual data#
#####################

#DEFINE THE NUMBER OF RANDOMIZATIONS

        v=100000
        v<-c(1:v)

        #Define a vector based on the number of crosses

        k<-c(1:length(data[,1]))

#START OF THE LOOP: CREATE A FILE WITH estimates (LM), t VALUES (WT) or
correlation coef (CT) CALCULATED FOR EACH PERMUTATION

        T1<-sapply(v,function(x){
```

```
                print(paste(round(100*x/max(v),3),"%"))

                #RANDOMIZE St1 and St2 from each cross (columns 1 and 2
from data)

                strain<-sapply(k,function(x){
                kx<-sample(data[x,1:2])
                return(t(kx))
                })
                strain<-t(strain)

                datax<-cbind(strain,data[,3:7])
                datax <-datax[order(datax[,1]),]

                z<-sort(unique(datax[,c(1)]))

                Sx<-sapply(z,function(x){
                S1<-subset(datax[,4],datax[,1]==x)
                S1<-sample(S1)
                return(S1)
                })
                Sx<-as.vector(matrix(unlist (Sx),nrow=1))
                Dg<-datax$Dg
                Dc<-datax$Dc
                Type<-datax$Type
                Cross<-datax$Cross

#STATISTICAL ANALYSES ON RANDOMIZED DATA- Only return estimates for
Linear models (LM), Welch t-test (WT) and correlation coefficients (CT)

        #GLOBAL LINEAR MODEL (M1)

                m=lm(Sx~Dg*Dc)
                M1<-summary(m)
                M1<-M1$coefficients[,1]

        #LINEAR MODEL (M2): CROSSES WITHIN LINEAGES ONLY

                mi<-
lm(subset(Sx,Type=="intra")~subset(Dg,Type=="intra")*subset(Dc,Type=="int
ra"))
                M2<-summary(mi)
                M2<-M2$coefficients[,1]

        #LINEAR MODEL (M3): CROSSES AMONG LINEAGES/SPECIES ONLY

                mI<-
lm(subset(Sx,Type!="intra")~subset(Dg,Type!="intra")*subset(Dc,Type!="int
ra"))
                M3<-summary(mI)
                M3<-M3$coefficients[,1]

                LM<-c(M1,M2,M3)

        #WELCH T-TESTS AMONG CATEGORIES OF CROSS

                WT<-t.test(subset(Sx,Cross=="AB"),subset(Sx,Cross=="AC"))
```

```
                WT.AB.AC<-WT$statistic
                WT<-t.test(subset(Sx,Cross=="AB"),subset(Sx,Cross=="BC"))
                WT.AB.BC<-WT$statistic
                WT<-t.test(subset(Sx,Cross=="AC"),subset(Sx,Cross=="BC"))
                WT.AC.BC<-WT$statistic
                WT<-t.test(subset(Sx,Cross=="BB"),subset(Sx,Cross=="BC"))
                WT.BB.BC<-WT$statistic
                WT<-t.test(subset(Sx,Cross=="CC"),subset(Sx,Cross=="BC"))
                WT.CC.BC<-WT$statistic
                WT<-t.test(subset(Sx,Cross=="BB"),subset(Sx,Cross=="CC"))
                WT.BB.CC<-WT$statistic

                WT<-c(WT.AB.AC, WT.AB.BC, WT.AC.BC, WT.BB.BC, WT.CC.BC,
WT.BB.CC)

        #PEARSON'S CORRELATION TESTS

                #S against Dg (S uncorrected)

                        x<-Sx
                        y<-Dg
                        C<-cor.test(x,y)
                        CT.dg<-C$estimate

                #S against Dc (S uncorrected)

                        y<-Dc
                        C<-cor.test(x,y)
                        CT.dc<-C$estimate

                #S against Dg (S corrected by Dc: residuals from linear
model lm(S~Dc))

                        x<-lm(Sx~Dc)
                        x<-as.matrix(x$residuals)
                        y<-Dg
                        C<-cor.test(x,y)
                        CT.dg.corrected<-C$estimate

                #S against Dc (S corrected by Dg: residuals from linear
model lm(S~Dg))

                        x<-lm(Sx~Dg)
                        x<-as.matrix(x$residuals)
                        y<-Dc
                        C<-cor.test(x,y)
                        CT.dc.corrected<-C$estimate

        #PEARSON'S CORRELATION TESTS WITHIN LINEAGES ONLY

                #S against Dg (S uncorrected)

                        x<-subset(Sx,Type=="intra")
                        y<-subset(Dg,Type=="intra")
                        C<-cor.test(x,y)
                        CT.dg.intra<-C$estimate
```

```
                #S against Dc (S uncorrected)

                        y<-subset(Dc,Type=="intra")
                        C<-cor.test(x,y)
                        CT.dc.intra<-C$estimate

                #S against Dg (S corrected by Dc: residuals from linear
model lm(S~Dc))

                        x<-
lm(subset(Sx,Type=="intra")~subset(Dc,Type=="intra"))
                        y<-subset(Dg,Type=="intra")
                        x<-as.matrix(x$residuals)
                        C<-cor.test(x,y)
                        CT.dg.intra.corrected<-C$estimate

                #S against Dc (S corrected by Dg: residuals from linear
model lm(S~Dg))

                        x<-
lm(subset(Sx,Type=="intra")~subset(Dg,Type=="intra"))
                        y<-subset(Dc,Type=="intra")
                        x<-as.matrix(x$residuals)
                        C<-cor.test(x,y)
                        CT.dc.intra.corrected<-C$estimate

                        CT<-c(CT.dg, CT.dc, CT.dg.corrected,
CT.dc.corrected,CT.dg.intra, CT.dc.intra, CT.dg.intra.corrected,
CT.dc.intra.corrected)

        return(c(LM,WT,CT))
        })

        ###For each test, return the number of times t observed is lower
or equal than expected

        z<-c(1:length(T1[,1]))

        #Vector of observed t-value

        TT<-c(LM[,1],WT[,3],CT[,1])

                LT<-sapply(z,function(x){
                P<-length(subset(T1[x,],T1[x,]<=TT[x]))
                return(P)
                })

        ###For each test, return the number of times t observed is higher
or equal than expected

                HT<-sapply(z,function(x){
                P<-length(subset(T1[x,],T1[x,]>=TT[x]))
                return(P)
                })

##Write tables of results
```

```
T1<-t(T1)
write.table(T1,file="T1.txt")
write.table(TT,file="TT.txt")

LM<-cbind(LM,LT[1:12]/max(v),HT[1:12]/max(v))
colnames(LM)<-c("Estimated_coefficient(E.C.)","S.E.","t","p-
value","P<=E.C.","P>=E.C.")
write.table(LM,file="Linear-Models")

WT<-cbind(WT,LT[13:18]/max(v),HT[13:18]/max(v))
colnames(WT)<-c("Mean1","Mean2","t","p-value","P<=t","P>=t")
write.table(WT,file="Welch-t-test")

CT<-cbind(CT,c(LT[19:26]/max(v),NA,NA),c(HT[19:26]/max(v),NA,NA))
colnames(CT)<-c("Correlation_coefficient","t","p-
value","P<=Coef.","P>=Coef.")
write.table(CT,file="Correlation-test")

############## ############## ############## ##############
       ############## ##############

##############FIGURES

setwd("/Users/jean-baptisteleducq/Dropbox/Souches
sauvages/ARTICLE_REPRODUCTIVE_ISOLATION/ResoumissionMOLECOL/Data")
setwd("/Users/JBLED/Dropbox/Souches
sauvages/ARTICLE_REPRODUCTIVE_ISOLATION/ResoumissionMOLECOL/Data")

data=read.table("Data.txt", header=T)
data<-subset(data,data[,3]!="SuSp")
#data<-subset(data,data[,3]!="ScSp")
T1=read.table("T1.txt", header=T)
LM=read.table("Linear-Models", header=T)
WT=read.table("Welch-t-test", header=T)
CT=read.table("Correlation-test", header=T)
TT<-c(LM[,1],WT[,3],CT[,1])
T1<-t(T1)
v=100000
v<-c(1:v)

St1<-data$St1
St2<-data$St2
Cross<-data$Cross
S<-data$S
Dg<-data$Dg
Dc<-data$Dc
Type<-data$Type

##############FIGURE 3: Comparison of S, Dc and Dg among groups and main
correlations

CAT<-
ifelse(Cross=="AA",1,ifelse(Cross=="BB",2,ifelse(Cross=="CC",3,ifelse(Cro
ss=="BC",4,ifelse(Cross=="AB",5,ifelse(Cross=="AC",6,7))))))

zones<-matrix(c(1,4,1,4,2,5,2,6,3,6,3,7),ncol=6)
layout(zones)
```

```
layout.show(max((zones)))
par(mar=c(0,0,0,0),bty="n", family="", ps=12)

plot(-10,-10,xlim=c(0,1),ylim=c(0,1),xlab="",xaxt="n",ylab="",yaxt="n")
text(0.01,0.95,labels="(a)",cex=1.5,font=2)
par(new="T",mar=c(4,5,0.5,0))
plot(-10,-
10,xlim=c(0,1),ylim=c(0,1),xlab="",xaxt="n",ylab="S",font.lab=3)
par(new="T")
boxplot(S~CAT, ylim=c(0,1.1*max(S)), ylab="", xlab="",xaxt="n",yaxt="n")

par(new="T",mar=c(0,5,0.5,0))
plot(-10,-10,xlim=c(0,7),ylim=c(0,1),xlab="",xaxt="n",ylab="",yaxt="n")
text(c(1:7)-
0.5,0.1,labels=c("AxA","BxB","CxC","BxC","AxB","AxC","ScxSp"),cex=1,font=
c(1,1,1,1,1,1,3))
text(c(2,4,5.5,7)-0.5,c(0.020, 0.020, 0.020,
0.020),labels=c("Lin.","Amer.","Cont.","Spec."))
segments(0.5,0.0625,2.5,0.0625)
segments(3,0.0625,4,0.0625)
segments(4.5,0.0625,5.5,0.0625)
segments(6,0.0625,7,0.0625)

par(mar=c(0,0,0,0))
plot(-10,-10,xlim=c(0,1),ylim=c(0,1),xlab="",xaxt="n",ylab="",yaxt="n")
text(0.01,0.95,labels="(b)",cex=1.5,font=2)
par(new="T",mar=c(4,5,0.5,0))
plot(-10,-
10,xlim=c(0,1),ylim=c(0,1),xlab="",xaxt="n",ylab="Dg",font.lab=3,yaxt="n"
)
par(new="T")
par(new="T",mar=c(4,5,0.5,0))
boxplot(Dg~CAT, ylim=c(0,1.1*max(Dg)), ylab="", xlab="",xaxt="n")
sapply(unique(CAT),function(x){
       text(x,max(subset(Dg,CAT==x))+0.005,labels=paste("(",length(subse
t(Dg,CAT==x)),")",sep=""))
})
par(new="T",mar=c(0,5,0.5,0))
plot(-10,-10,xlim=c(0,7),ylim=c(0,1),xlab="",xaxt="n",ylab="",yaxt="n")
text(c(1:7)-
0.5,0.1,labels=c("AxA","BxB","CxC","BxC","AxB","AxC","ScxSp"),cex=1,font=
c(1,1,1,1,1,1,3))
text(c(2,4,5.5,7)-0.5,c(0.020, 0.020, 0.020,
0.020),labels=c("Lin.","Amer.","Cont.","Spec."))
segments(0.5,0.0625,2.5,0.0625)
segments(3,0.0625,4,0.0625)
segments(4.5,0.0625,5.5,0.0625)
segments(6,0.0625,7,0.0625)

par(mar=c(0,0,0,0))
plot(-10,-10,xlim=c(0,1),ylim=c(0,1),xlab="",xaxt="n",ylab="",yaxt="n")
text(0.01,0.95,labels="(c)",cex=1.5,font=2)
par(new="T",mar=c(4,5,0.5,0))
boxplot(Dc~CAT, ylim=c(0,1.1*max(Dc)), ylab="", xlab="",xaxt="n")

par(new="T",mar=c(0,5,0.5,0))
```

```
plot(-10,-
10,xlim=c(0,7),ylim=c(0,1),xlab="",xaxt="n",ylab="Dc",yaxt="n",font.lab=3
)
text(c(1:7)-
0.5,0.1,labels=c("AxA","BxB","CxC","BxC","AxB","AxC","ScxSp"),cex=1,font=
c(1,1,1,1,1,1,3))
text(c(2,4,5.5,7)-0.5,c(0.020, 0.020, 0.020,
0.020),labels=c("Lin.","Amer.","Cont.","Spec."))
segments(0.5,0.0625,2.5,0.0625)
segments(3,0.0625,4,0.0625)
segments(4.5,0.0625,5.5,0.0625)
segments(6,0.0625,7,0.0625)

par(mar=c(0,0,0,0))
plot(-10,-10,xlim=c(0,1),ylim=c(0,1),xlab="",xaxt="n",ylab="",yaxt="n",
cex.lab=1)
text(0.01,0.95,labels="(d)",cex=1.5,font=2)

Sr.Dc<-lm(S~Dc)
Sr.Dc<-as.matrix(Sr.Dc$residuals)

Sr.Dg<-lm(S~Dg)
Sr.Dg<-as.matrix(Sr.Dg$residuals)

mDg.i<-min(subset(Dg,Type=="intra"))
MDg.i<-max(subset(Dg,Type=="intra"))
mS.i<-mean(subset(S,Type=="intra"))

mDg.I<-min(subset(Dg,Type=="inter"))
MDg.I<-max(subset(Dg,Type=="inter"))
mS.I<-mean(subset(S,Type=="inter"))

mDg.s<-min(subset(Dg,Type=="esp"))
MDg.s<-max(subset(Dg,Type=="esp"))
mS.s<-mean(subset(S,Type=="esp"))

mDg.c<-min(subset(Dg,Type=="cont"))
MDg.c<-max(subset(Dg,Type=="cont"))
mS.c<-mean(subset(S,Type=="cont"))

mS.i<-mean(subset(Sr.Dc,Type=="intra"))
mS.I<-mean(subset(Sr.Dc,Type=="inter"))
mS.s<-mean(subset(Sr.Dc,Type=="esp"))
mS.c<-mean(subset(Sr.Dc,Type=="cont"))

par(mar=c(5,5,0.5,1),new="T")
plot(c(mDg.i,MDg.i),c(mS.i,mS.i),type="l",lwd=5,xlim=c(min(Dg),max(Dg)),y
lim=c(min(Sr.Dc),max(Sr.Dc)),xlab="",xaxt="n",ylab="",yaxt="n",col="black
", cex.lab=1)
par(new="T")
plot(c(mDg.c,MDg.c),c(mS.c,mS.c),type="l",lwd=5,xlim=c(min(Dg),max(Dg)),y
lim=c(min(Sr.Dc),max(Sr.Dc)),xlab="",xaxt="n",ylab="",yaxt="n",col="black
", cex.lab=1)
par(new="T")
plot(c(mDg.I,MDg.I),c(mS.I,mS.I),type="l",lwd=5,xlim=c(min(Dg),max(Dg)),y
lim=c(min(Sr.Dc),max(Sr.Dc)),xlab="",xaxt="n",ylab="",yaxt="n",col="black
", cex.lab=1)
```

```
par(new="T")
plot(c(mDg.s,MDg.s),c(mS.s,mS.s),type="l",lwd=5,xlim=c(min(Dg),max(Dg)),y
lim=c(min(Sr.Dc),max(Sr.Dc)),xlab="",xaxt="n",ylab="",yaxt="n",col="black
", cex.lab=1)
abline(lm(Sr.Dc~Dg),lty=2,lwd=2)
par(new="T")
plot(Dg,Sr.Dc,xlim=c(min(Dg),max(Dg)),ylim=c(min(Sr.Dc),max(Sr.Dc)),pch=2
1,lwd=0.5,cex=1.2,bg=ifelse(Type=="intra","black",ifelse(Type=="esp","whi
te",ifelse(Type=="cont",rgb(0.3,0.3,0.3,1),rgb(0.7,0.7,0.7,1)))),col=ifel
se(Type=="esp","black","white"),ylab="residuals lm(S~Dc)",
cex.lab=1,font.lab=3)
text(0.8*max(Dg),0.95*max(Sr.Dc),labels=paste("r =
",round(CT[3,1],3),ifelse(CT[3,3]<=0.001,"***",ifelse(CT[3,3]<=0.01,"**",
ifelse(CT[3,3]<=0.05,"*",""))),sep=""),cex=1)


x=21
h<-hist(T1[x,],breaks=100,plot=FALSE)
par(bty="n",mar=c(4,4,2,2))
plot(-10,-
10,ylim=c(min(T1[21:22,],TT[21:22]),max(T1[21:22,],TT[21:22])),xlim=c(-
1,1),main="",xlab="",ylab="Correlation coefficient",xaxt="n")
segments(-
h$counts/max(h$counts),h$breaks,h$counts/max(h$counts),h$breaks,lwd=2,col
="grey")
P<-length(subset(T1[x,],T1[x,]<=TT[x]))/max(v)

text(0,max(T1[21:22,],TT[21:22]),labels=ifelse(P==0,paste("P<",1/max(v)),
paste("P=",round(P,4))),col="black",cex=1)
par(new="T")
plot(0,TT[x],pch=21,type="p",cex=1.5,ylim=c(min(T1[21:22,],TT[21:22]),max
(T1[21:22,],TT[21:22])),xlim=c(-
1,1),main="",xlab="",ylab="",xaxt="n",yaxt="n",col="black",bg="white",lwd
=2)

par(mar=c(0,0,0,0))
plot(-10,-10,xlim=c(0,1),ylim=c(0,1),xlab="",xaxt="n",ylab="",yaxt="n",
cex.lab=1)
text(0.01,0.95,labels="(e)",cex=1.5,font=2)
par(mar=c(5,5,0.5,1),new="T")
plot(-10,-
10,xlim=c(min(Dc),max(Dc)),ylim=c(min(Sr.Dg),max(Sr.Dg)),xlab="",xaxt="n"
,ylab="",yaxt="n", cex.lab=1)
par(new="T")
abline(lm(Sr.Dg~Dc),lty=2,lwd=2)
par(new="T")
plot(Dc,Sr.Dg,xlim=c(min(Dc),max(Dc)),ylim=c(min(Sr.Dg),max(Sr.Dg)),pch=2
1,lwd=0.5,cex=1.2,bg=ifelse(Type=="intra","black",ifelse(Type=="esp","whi
te",ifelse(Type=="cont",rgb(0.3,0.3,0.3,1),rgb(0.7,0.7,0.7,1)))),col=ifel
se(Type=="esp","black","white"),ylab="residuals lm(S~Dg)",
cex.lab=1,font.lab=3)
text(0.8*max(Dc),0.95*max(Sr.Dg),labels=paste("r =
",round(CT[4,1],3),ifelse(CT[4,3]<=0.001,"***",ifelse(CT[4,3]<=0.01,"**",
ifelse(CT[4,3]<=0.05,"*",""))),sep=""),cex=1)

x=22
h<-hist(T1[x,],breaks=100,plot=FALSE)
```

```
par(bty="n",mar=c(4,4,2,2))
plot(-10,-
10,ylim=c(min(T1[21:22,],TT[21:22]),max(T1[21:22,],TT[21:22])),xlim=c(-
1,1),main="",xlab="",ylab="Correlation coefficient",xaxt="n")
segments(-
h$counts/max(h$counts),h$breaks,h$counts/max(h$counts),h$breaks,lwd=2,col
="grey")
P<-length(subset(T1[x,],T1[x,]<=TT[x]))/max(v)

text(0,max(T1[21:22,],TT[21:22]),labels=ifelse(P==0,paste("P<",1/max(v)),
paste("P=",round(P,4))),col="black",cex=1)
par(new="T")
plot(0,TT[x],pch=21,type="p",cex=1.5,ylim=c(min(T1[21:22,],TT[21:22]),max
(T1[21:22,],TT[21:22])),xlim=c(-
1,1),main="",xlab="",ylab="",xaxt="n",yaxt="n",col="black",bg="white",lwd
=2)




#Figure S2 (Violin plots on t-values from Welch-tests after permutations)

par(bty="n",mfrow=c(1,7),mar=c(0,0,0,0))


CROSS<-c("AxB , AxC","AxB , BxC","AxC , BxC","BxB , BxC","CxC , BxC","BxB
, CxC")

par(mar=c(4,5,2,0))
plot(-10,-
10,ylim=c(min(T1[13:18,],TT[13:18]),max(T1[13:18,],TT[13:18])),xlim=c(-
1,1),main="",xlab="",ylab="t-value",xaxt="n",font.lab=3)

a<-c(13:18)

sapply(a,function(x){
        h<-hist(T1[x,],breaks=100,plot=FALSE)
        par(bty="n",mar=c(4,1,2,1))
        plot(-10,-
10,ylim=c(min(T1[13:18,],TT[13:18]),max(T1[13:18,],TT[13:18])),xlim=c(-
1,1),main=subset(CROSS[1:6],x==c(13:18)),xlab="",ylab="",xaxt="n",yaxt="n
",cex.lab=1.2)
        segments(-
h$counts/max(h$counts),h$breaks,h$counts/max(h$counts),h$breaks,lwd=2,col
="grey")
        P<-length(subset(T1[x,],T1[x,]<=TT[x]))/max(v)

        text(0,min(T1[13:18,],TT[13:18]),labels=ifelse(P==0,paste("P<",1/
max(v)),paste("P=",round(P,4))),col=ifelse(P<=0.05,"red",ifelse(P>=0.95,"
red","black")),cex=1)
        par(new="T")
        plot(0,TT[x],pch=21,type="p",cex=1.5,ylim=c(min(T1[13:18,],TT[13:
18]),max(T1[13:18,],TT[13:18])),xlim=c(-
1,1),main="",xlab="",ylab="",xaxt="n",yaxt="n",col=ifelse(P<=0.05,"red",i
felse(P>=0.95,"red","black")),bg="white",lwd=2)
})
```

```
#Figure S4 (Violin plots on t-values from Linear models after
permutations)

zones<-
matrix(c(1,6,6,6,11,11,11,16,16,16,2,7,7,7,12,12,12,17,17,17,3,8,8,8,13,1
3,13,18,18,18,4,9,9,9,14,14,14,19,19,19,5,10,10,10,15,15,15,20,20,20),nco
l=5)
layout(zones)
layout.show(max((zones)))
par(mar=c(1,1,1,1),bty="n", family="", ps=12)


plot(-10,-10,xlim=c(0,1),ylim=c(0,1),xaxt="n",yaxt="n",xlab="",ylab="")
plot(-10,-10,xlim=c(0,1),ylim=c(0,1),xaxt="n",yaxt="n",xlab="",ylab="")
text(0.5,0.2,labels="Intercept",cex=1.2,font=2)
plot(-10,-10,xlim=c(0,1),ylim=c(0,1),xaxt="n",yaxt="n",xlab="",ylab="")
text(0.5,0.2,labels="Dg",cex=1.2,font=4)
plot(-10,-10,xlim=c(0,1),ylim=c(0,1),xaxt="n",yaxt="n",xlab="",ylab="")
text(0.5,0.2,labels="Dc",cex=1.2,font=4)
plot(-10,-10,xlim=c(0,1),ylim=c(0,1),xaxt="n",yaxt="n",xlab="",ylab="")
text(0.5,0.2,labels="Dg*Dc",cex=1.2,font=4)

plot(-10,-10,xlim=c(0,1),ylim=c(0,1),xaxt="n",yaxt="n",xlab="",ylab="")
text(0.4,0.8,labels="M1",cex=1.2,font=2)
text(0.4,0.5,labels="(all data)",cex=1,font=2)
text(0.9,0.5,labels="Estimated coefficient",cex=1,srt=90)


a<-c(1:4)

sapply(a,function(x){
        h<-hist(T1[x,],breaks=100,plot=FALSE)
        par(bty="n",mar=c(1,1,1,1))
        plot(-10,-10,ylim=c(min(T1[x,],TT[x])-(max(T1[x,],TT[x])-
min(T1[x,],TT[x]))/2,max(T1[x,],TT[x])+(max(T1[x,],TT[x])-
min(T1[x,],TT[x]))/2),xlim=c(-1,1),main="",xlab="",ylab="",xaxt="n")
        segments(-
h$counts/max(h$counts),h$breaks,h$counts/max(h$counts),h$breaks,lwd=1,col
="grey")
        P<-length(subset(T1[x,],T1[x,]<=TT[x]))/max(v)

        text(0,min(T1[x,],TT[x])-(max(T1[x,],TT[x])-
min(T1[x,],TT[x]))/2,labels=ifelse(P==0,paste("P<",1/max(v)),paste("P=",r
ound(P,4))),col=ifelse(P<=0.05,"red",ifelse(P>=0.95,"red","black")),cex=1
)
        par(new="T")
        plot(0,TT[x],pch=21,type="p",cex=1.5,ylim=c(min(T1[x,],TT[x])-
(max(T1[x,],TT[x])-
```

```
min(T1[x,],TT[x]))/2,max(T1[x,],TT[x])+(max(T1[x,],TT[x])-
min(T1[x,],TT[x]))/2),xlim=c(-
1,1),main="",xlab="",ylab="",xaxt="n",yaxt="n",col=ifelse(P<=0.05,"red",i
felse(P>=0.95,"red","black")),bg="white",lwd=2)
})

plot(-10,-10,xlim=c(0,1),ylim=c(0,1),xaxt="n",yaxt="n",xlab="",ylab="")
text(0.4,0.8,labels="M2",cex=1.2,font=2)
text(0.4,0.55,labels="(within",cex=1,font=2)
text(0.4,0.45,labels="lineages)",cex=1,font=2)
text(0.9,0.5,labels="Estimated coefficient",cex=1,srt=90)

a<-c(5:8)

sapply(a,function(x){
        h<-hist(T1[x,],breaks=100,plot=FALSE)
        par(bty="n",mar=c(1,1,1,1))
        plot(-10,-10,ylim=c(min(T1[x,],TT[x])-(max(T1[x,],TT[x])-
min(T1[x,],TT[x]))/2,max(T1[x,],TT[x])+(max(T1[x,],TT[x])-
min(T1[x,],TT[x]))/2),xlim=c(-1,1),main="",xlab="",ylab="",xaxt="n")
        segments(-
h$counts/max(h$counts),h$breaks,h$counts/max(h$counts),h$breaks,lwd=1,col
="grey")
        P<-length(subset(T1[x,],T1[x,]<=TT[x]))/max(v)

        text(0,min(T1[x,],TT[x])-(max(T1[x,],TT[x])-
min(T1[x,],TT[x]))/2,labels=ifelse(P==0,paste("P<",1/max(v)),paste("P=",r
ound(P,4))),col=ifelse(P<=0.05,"red",ifelse(P>=0.95,"red","black")),cex=1
)
        par(new="T")
        plot(0,TT[x],pch=21,type="p",cex=1.5,ylim=c(min(T1[x,],TT[x])-
(max(T1[x,],TT[x])-
min(T1[x,],TT[x]))/2,max(T1[x,],TT[x])+(max(T1[x,],TT[x])-
min(T1[x,],TT[x]))/2),xlim=c(-
1,1),main="",xlab="",ylab="",xaxt="n",yaxt="n",col=ifelse(P<=0.05,"red",i
felse(P>=0.95,"red","black")),bg="white",lwd=2)
})

plot(-10,-10,xlim=c(0,1),ylim=c(0,1),xaxt="n",yaxt="n",xlab="",ylab="")
text(0.4,0.8,labels="M3",cex=1.2,font=2)
text(0.4,0.6,labels="(among",cex=1,font=2)
text(0.4,0.5,labels="lineages/",cex=1,font=2)
text(0.4,0.4,labels="species)",cex=1,font=2)
text(0.9,0.5,labels="Estimated coefficient",cex=1,srt=90)

a<-c(9:12)

sapply(a,function(x){
        h<-hist(T1[x,],breaks=100,plot=FALSE)
        par(bty="n",mar=c(1,1,1,1))
        plot(-10,-10,ylim=c(min(T1[x,],TT[x])-(max(T1[x,],TT[x])-
min(T1[x,],TT[x]))/2,max(T1[x,],TT[x])+(max(T1[x,],TT[x])-
min(T1[x,],TT[x]))/2),xlim=c(-1,1),main="",xlab="",ylab="",xaxt="n")
        segments(-
h$counts/max(h$counts),h$breaks,h$counts/max(h$counts),h$breaks,lwd=1,col
="grey")
        P<-length(subset(T1[x,],T1[x,]<=TT[x]))/max(v)
```

```
        text(0,min(T1[x,],TT[x])-(max(T1[x,],TT[x])-
min(T1[x,],TT[x]))/2,labels=ifelse(P==0,paste("P<",1/max(v)),paste("P=",r
ound(P,4))),col=ifelse(P<=0.05,"red",ifelse(P>=0.95,"red","black")),cex=1
)
        par(new="T")
        plot(0,TT[x],pch=21,type="p",cex=1.5,ylim=c(min(T1[x,],TT[x])-
(max(T1[x,],TT[x])-
min(T1[x,],TT[x]))/2,max(T1[x,],TT[x])+(max(T1[x,],TT[x])-
min(T1[x,],TT[x]))/2),xlim=c(-
1,1),main="",xlab="",ylab="",xaxt="n",yaxt="n",col=ifelse(P<=0.05,"red",i
felse(P>=0.95,"red","black")),bg="white",lwd=2)
})


##############FIGURE S3: other correlations

par(mfrow=c(3,2),bty="n", family="", ps=12,mar=c(0,0,0,0))
plot(-10,-10,xlim=c(0,1),ylim=c(0,1),xlab="",xaxt="n",ylab="",yaxt="n")
text(0.01,0.95,labels="(a)",cex=1.5,font=2)

mDg.i<-min(subset(Dg,Type=="intra"))
MDg.i<-max(subset(Dg,Type=="intra"))
mS.i<-mean(subset(S,Type=="intra"))
mDc.i<-mean(subset(Dc,Type=="intra"))

mDg.I<-min(subset(Dg,Type=="inter"))
MDg.I<-max(subset(Dg,Type=="inter"))
mS.I<-mean(subset(S,Type=="inter"))
mDc.I<-mean(subset(Dc,Type=="inter"))

mDg.s<-min(subset(Dg,Type=="esp"))
MDg.s<-max(subset(Dg,Type=="esp"))
mS.s<-mean(subset(S,Type=="esp"))
mDc.s<-mean(subset(Dc,Type=="esp"))

mDg.c<-min(subset(Dg,Type=="cont"))
MDg.c<-max(subset(Dg,Type=="cont"))
mS.c<-mean(subset(S,Type=="cont"))
mDc.c<-mean(subset(Dc,Type=="cont"))


par(mar=c(4,4.5,0.5,1),new="T")
plot(c(mDg.i,MDg.i),c(mS.i,mS.i),type="l",lwd=5,xlim=c(min(Dg),max(Dg)),y
lim=c(min(S),max(S)),xlab="",xaxt="n",ylab="",yaxt="n",col="black")
par(new="T")
plot(c(mDg.c,MDg.c),c(mS.c,mS.c),type="l",lwd=5,xlim=c(min(Dg),max(Dg)),y
lim=c(min(S),max(S)),xlab="",xaxt="n",ylab="",yaxt="n",col="black")
par(new="T")
plot(c(mDg.I,MDg.I),c(mS.I,mS.I),type="l",lwd=5,xlim=c(min(Dg),max(Dg)),y
lim=c(min(S),max(S)),xlab="",xaxt="n",ylab="",yaxt="n",col="black")
par(new="T")
plot(c(mDg.s,MDg.s),c(mS.s,mS.s),type="l",lwd=5,xlim=c(min(Dg),max(Dg)),y
lim=c(min(S),max(S)),xlab="",xaxt="n",ylab="",yaxt="n",col="black")
abline(lm(S~Dg),lty=2,lwd=2)
par(new="T")
```

```
plot(Dg,S,xlim=c(min(Dg),max(Dg)),ylim=c(min(S),max(S)),pch=21,lwd=0.5,ce
x=1.2,bg=ifelse(Type=="intra","black",ifelse(Type=="esp","white",ifelse(T
ype=="cont",rgb(0.3,0.3,0.3,1),rgb(0.7,0.7,0.7,1)))),col=ifelse(Type=="es
p","black","white"),font.lab=3)
text(0.8*max(Dg),0.95*max(S),labels=paste("r=
",round(CT[1,1],3),ifelse(CT[1,3]<=0.001,"***",ifelse(CT[1,3]<=0.01,"**",
ifelse(CT[1,3]<=0.05,"*",""))),sep=""))

par(mar=c(0,0,0,0))
plot(-10,-10,xlim=c(0,1),ylim=c(0,1),xlab="",xaxt="n",ylab="",yaxt="n")
text(0.01,0.95,labels="(b)",cex=1.5,font=2)
par(mar=c(4,4.5,0.5,1),new="T")
plot(-10,-
10,xlim=c(min(Dc),max(Dc)),ylim=c(min(S),max(S)),xlab="",xaxt="n",ylab=""
,yaxt="n")
par(new="T")
abline(lm(S~Dc),lty=2,lwd=2)
par(new="T")
plot(Dc,S,xlim=c(min(Dc),max(Dc)),ylim=c(min(S),max(S)),pch=21,lwd=0.5,ce
x=1.2,bg=ifelse(Type=="intra","black",ifelse(Type=="esp","white",ifelse(T
ype=="cont",rgb(0.3,0.3,0.3,1),rgb(0.7,0.7,0.7,1)))),col=ifelse(Type=="es
p","black","white"),font.lab=3)
text(0.8*max(Dc),1*max(S),labels=paste("r=
",round(CT[2,1],3),ifelse(CT[2,3]<=0.001,"***",ifelse(CT[2,3]<=0.01,"**",
ifelse(CT[2,3]<=0.05,"*",""))),sep=""))

par(mar=c(0,0,0,0))
plot(-10,-10,xlim=c(0,1),ylim=c(0,1),xlab="",xaxt="n",ylab="",yaxt="n")
text(0.01,0.95,labels="(c)",cex=1.5,font=2)
par(mar=c(4,4.5,0.5,1),new="T")
plot(c(mDg.i,MDg.i),c(mDc.i,
mDc.i),type="l",lwd=5,xlim=c(min(Dg),max(Dg)),ylim=c(min(Dc),max(Dc)),xla
b="",xaxt="n",ylab="",yaxt="n",col="black")
par(new="T")
plot(c(mDg.c,MDg.c),c(mDc.c,mDc.c),type="l",lwd=5,xlim=c(min(Dg),max(Dg))
,ylim=c(min(Dc),max(Dc)),xlab="",xaxt="n",ylab="",yaxt="n",col="black")
par(new="T")
plot(c(mDg.I,MDg.I),c(mDc.I,mDc.I),type="l",lwd=5,xlim=c(min(Dg),max(Dg))
,ylim=c(min(Dc),max(Dc)),xlab="",xaxt="n",ylab="",yaxt="n",col="black")
par(new="T")
plot(c(mDg.s,MDg.s),c(mDc.s,mDc.s),type="l",lwd=5,xlim=c(min(Dg),max(Dg))
,ylim=c(min(Dc),max(Dc)),xlab="",xaxt="n",ylab="",yaxt="n",col="black")
par(new="T")
plot(-10,-
10,xlim=c(min(Dg),max(Dg)),ylim=c(min(Dc),max(Dc)),xlab="",xaxt="n",ylab=
"",yaxt="n")
par(new="T")
abline(lm(Dc~Dg),lty=2,lwd=2)
par(new="T")
plot(Dg,Dc,xlim=c(min(Dg),max(Dg)),ylim=c(min(Dc),max(Dc)),pch=21,lwd=0.5
,cex=1.2,bg=ifelse(Type=="intra","black",ifelse(Type=="esp","white",ifels
e(Type=="cont",rgb(0.3,0.3,0.3,1),rgb(0.7,0.7,0.7,1)))),col=ifelse(Type==
"esp","black","white"),font.lab=3)
text(0.8*max(Dg),0.95*max(Dc),labels=paste("r=
",round(CT[9,1],3),ifelse(CT[9,3]<=0.001,"***",ifelse(CT[9,3]<=0.01,"**",
ifelse(CT[9,3]<=0.05,"*",""))),sep=""))
```

```
par(mar=c(0,0,0,0))
plot(-10,-10,xlim=c(0,1),ylim=c(0,1),xlab="",xaxt="n",ylab="",yaxt="n")
text(0.01,0.95,labels="(d)",cex=1.5,font=2)

par(mar=c(4,4.5,0.5,1),new="T")
plot(-10,-
10,xlim=c(min(Dc),max(Dc)),ylim=c(min(S),max(S)),xlab="",xaxt="n",ylab=""
,yaxt="n")
par(new="T")
abline(lm(subset(S,Type=="intra")~subset(Dc,Type=="intra")),lty=2,lwd=2)
par(new="T")
plot(subset(Dc,Type=="intra"),subset(S,Type=="intra"),xlim=c(min(Dc),max(
Dc)),ylim=c(min(S),max(S)),pch=21,lwd=0.5,cex=1.2,bg=ifelse(subset(Cross,
Type=="intra")=="AA","green",ifelse(subset(Cross,Type=="intra")=="BB","re
d","blue")),col="white",ylab="S",xlab="Dc",font.lab=3)
text(0.8*max(Dc),1*max(S),labels=paste("r=
",round(CT[6,1],3),ifelse(CT[6,3]<=0.001,"***",ifelse(CT[6,3]<=0.01,"**",
ifelse(CT[6,3]<=0.05,"*",""))),sep=""))



x<-subset(S,Type!="intra")
y<-subset(Dc,Type!="intra")
C<-cor.test(x,y)

par(mar=c(0,0,0,0))
plot(-10,-10,xlim=c(0,1),ylim=c(0,1),xlab="",xaxt="n",ylab="",yaxt="n")
text(0.01,0.95,labels="(e)",cex=1.5,font=2)

par(mar=c(4,4.5,0.5,1),new="T")
plot(-10,-
10,xlim=c(min(Dc),max(Dc)),ylim=c(min(S),max(S)),xlab="",xaxt="n",ylab=""
,yaxt="n")
par(new="T")
abline(lm(subset(S,Type!="intra")~subset(Dc,Type!="intra")),lty=2,lwd=2)
par(new="T")
plot(subset(Dc,Type!="intra"),subset(S,Type!="intra"),xlim=c(min(Dc),max(
Dc)),ylim=c(min(S),max(S)),pch=21,lwd=0.5,cex=1.2,bg=ifelse(subset(Type,T
ype!="intra")=="intra","black",ifelse(subset(Type,Type!="intra")=="esp","
white",ifelse(subset(Type,Type!="intra")=="cont",rgb(0.3,0.3,0.3,1),rgb(0
.7,0.7,0.7,1)))),col=ifelse(subset(Type,Type!="intra")=="esp","black","wh
ite"),ylab="S",xlab="Dc",font.lab=3)
text(0.8*max(Dc),0.95*max(S),labels=paste("r= ",round(C
$estimate,3),ifelse(C $p.value<=0.001,"***",ifelse(C
$p.value<=0.01,"**",ifelse(C $p.value<=0.05,"*",""))),sep=""))



par(mar=c(0,0,0,0))
plot(-10,-10,xlim=c(0,1),ylim=c(0,1),xlab="",xaxt="n",ylab="",yaxt="n")
text(0.01,0.95,labels="(f)",cex=1.5,font=2)

par(new="T",mar=c(1,5,0,1))

x=19
y=0
```

```
h<-hist(T1[x,],breaks=100,plot=FALSE)
plot(-10,-
10,ylim=c(min(T1[c(19:20,24),],TT[c(19:20,24)]),max(T1[c(19:20,24),],TT[c
(19:20,24)])),xlim=c(-1,8),main="",xlab="",ylab="Correlation
coefficient",xaxt="n",cex.lab=1)
segments(y-
h$counts/max(h$counts),h$breaks,y+h$counts/max(h$counts),h$breaks,lwd=2,c
ol="grey")
P<-length(subset(T1[x,],T1[x,]<=TT[x]))/max(v)

text(0+y,0.6*max(T1[c(19:20,24),],TT[c(19:20,24)]),labels=ifelse(P==0,pas
te("P<",1/max(v)),paste("P=",round(P,4))),col=ifelse(P<=0.05,"red",ifelse
(P>=0.95,"red","black")),cex=1)
text(0+y,0.9*max(T1[c(19:20,24),],TT[c(19:20,24)]),labels="(a) S,
Dg",cex=1,font=2)
par(new="T")
plot(0+y,TT[x],pch=21,type="p",cex=1.5,ylim=c(min(T1[c(19:20,24),],TT[c(1
9:20,24)]),max(T1[c(19:20,24),],TT[c(19:20,24)])),xlim=c(-
1,8),main="",xlab="",ylab="",xaxt="n",yaxt="n",col=ifelse(P<=0.05,"red",i
felse(P>=0.95,"red","black")),bg="white",lwd=2)

par(new="T")

x=20
y=3
h<-hist(T1[x,],breaks=100,plot=FALSE)
plot(-10,-
10,ylim=c(min(T1[c(19:20,24),],TT[c(19:20,24)]),max(T1[c(19:20,24),],TT[c
(19:20,24)])),xlim=c(-
1,8),main="",xlab="",ylab="",xaxt="n",yaxt="n",cex.lab=1)
segments(y-
h$counts/max(h$counts),h$breaks,y+h$counts/max(h$counts),h$breaks,lwd=2,c
ol="grey")
P<-length(subset(T1[x,],T1[x,]<=TT[x]))/max(v)

text(0+y,0.6*max(T1[c(19:20,24),],TT[c(19:20,24)]),labels=ifelse(P==0,pas
te("P<",1/max(v)),paste("P=",round(P,4))),col=ifelse(P<=0.05,"red",ifelse
(P>=0.95,"red","black")),cex=1)
text(0+y,0.9*max(T1[c(19:20,24),],TT[c(19:20,24)]),labels="(b) S,
Dc",cex=1,font=2)
par(new="T")
plot(0+y,TT[x],pch=21,type="p",cex=1.5,ylim=c(min(T1[c(19:20,24),],TT[c(1
9:20,24)]),max(T1[c(19:20,24),],TT[c(19:20,24)])),xlim=c(-
1,8),main="",xlab="",ylab="",xaxt="n",yaxt="n",col=ifelse(P<=0.05,"red",i
felse(P>=0.95,"red","black")),bg="white",lwd=2)

par(new="T")

x=24
y=6
h<-hist(T1[x,],breaks=100,plot=FALSE)
plot(-10,-
10,ylim=c(min(T1[c(19:20,24),],TT[c(19:20,24)]),max(T1[c(19:20,24),],TT[c
(19:20,24)])),xlim=c(-
1,8),main="",xlab="",ylab="",xaxt="n",yaxt="n",cex.lab=1)
```

```
segments(y-
h$counts/max(h$counts),h$breaks,y+h$counts/max(h$counts),h$breaks,lwd=2,c
ol="grey")
P<-length(subset(T1[x,],T1[x,]<=TT[x]))/max(v)

text(0+y,0.6*max(T1[c(19:20,24),],TT[c(19:20,24)]),labels=ifelse(P==0,pas
te("P<",1/max(v)),paste("P=",round(P,4))),col=ifelse(P<=0.05,"red",ifelse
(P>=0.95,"red","black")),cex=1)
text(0+y,max(T1[c(19:20,24),],TT[c(19:20,24)]),labels="(d) S,
Dc",cex=1,font=2)
text(0+y,0.8*max(T1[c(19:20,24),],TT[c(19:20,24)]),labels="within
lineages",cex=1,font=2)
par(new="T")
plot(0+y,TT[x],pch=21,type="p",cex=1.5,ylim=c(min(T1[c(19:20,24),],TT[c(1
9:20,24)]),max(T1[c(19:20,24),],TT[c(19:20,24)])),xlim=c(-
1,8),main="",xlab="",ylab="",xaxt="n",yaxt="n",col=ifelse(P<=0.05,"red",i
felse(P>=0.95,"red","black")),bg="white",lwd=2)
```

Appendix 2: Dataset used to generate figures.

St1	St2	Cross	S	Dc	Dg	Type
YPS744	BY4741	ScSp	0.0000	0.2730	0.0684	esp
YPS744	JRY8147	SuSp	0.0000	NA	0.1484	esp
LL2012_021	BY4741	ScSp	0.0104	0.1959	0.0774	esp
LL2012_028	JRY8147	SuSp	0.0000	NA	0.1566	esp
MSH-587-1	BY4741	ScSp	0.0000	0.4037	0.0812	esp
LL2012_018	JRY8147	SuSp	0.0208	NA	0.1534	esp
YPS744	UWOPS-79-140	AB	0.5570	0.4615	0.0257	inter
YPS744	YHBJ1	AB	0.5100	0.1724	0.0278	inter
YPS744	LL2012_021	AB	0.4530	0.2857	0.0279	inter
YPS744	YPS484	AB	0.4480	0.1034	0.0257	inter
YPS642	UWOPS-79-140	AB	0.4270	0.2500	0.0257	inter
LL2012_029	YPS642	AB	0.3540	0.3333	0.0287	inter
LL2012_029	YPS744	AB	0.3440	0.2413	0.0287	inter
YPS644	LL2012_028	AB	0.2710	0.2592	0.0272	inter
YPS642	YPS667	AC	0.4790	0.3600	0.0340	inter
YPS644	LL2011_009	AC	0.3650	0.3600	0.0304	inter
YPS744	MSH-587-1	AC	0.3440	0.3077	0.0304	inter
YPS695	YPS667	AC	0.3230	0.3846	0.0340	inter
YPS695	LL2011_011	AC	0.2710	0.4815	0.0304	inter
YPS642	LL2012_004	AC	0.2080	0.4074	0.0304	inter
YPS744	LL2012_016	AC	0.2400	0.4074	0.0276	inter
YPS484	YPS667	BC	0.6510	0.3077	0.0199	cont
UWOPS-91-202	MSH-587-1	BC	0.6040	0.6153	0.0198	cont
UWOPS-79-140	LL2011_009	BC	0.5160	0.3913	0.0199	cont
YPS667	YHBJ1	BC	0.5100	0.1538	0.0264	cont
YPS484	MSH-587-1	BC	0.4900	0.1999	0.0199	cont
LL2011_002	UWOPS-79-140	BC	0.4900	0.1818	0.0199	cont
YPS484	LL2011_011	BC	0.4380	0.4815	0.0199	cont
YPS484	LL2012_004	BC	0.2660	0.4285	0.0199	cont
LL2012_029	LL2011_006	BC	0.5000	0.2143	0.0212	cont
LL2012_018	LL2012_021	BC	0.4580	0.1999	0.0206	cont
LL2012_016	LL2012_021	BC	0.4480	0.1999	0.0188	cont
LL2011_005	LL2012_028	BC	0.3020	0.2307	0.0214	cont
YPS642	YPS744	AA	0.8590	0.0714	0.0000	intra
YPS644	YPS744	AA	0.8330	0.0714	0.0000	intra
YPS744	YPS744	AA	0.7970	0.0000	0.0000	intra
YPS695	YPS744	AA	0.7450	0.0345	0.0000	intra
UWOPS-79-140	LL2012_021	BB	0.8540	0.2500	0.0043	intra
YPS484	LL2012_021	BB	0.7660	0.2592	0.0022	intra
YPS484	UWOPS-79-140	BB	0.7190	0.1200	0.0022	intra
UWOPS-91-202	LL2012_021	BB	0.6880	0.2143	0.0022	intra
UWOPS-91-202	YHBJ1	BB	0.5520	0.3793	0.0050	intra
LL2012_029	LL2012_021	BB	0.5100	0.4074	0.0036	intra
YHBJ1	LL2012_028	BB	0.4479	0.0714	0.0065	intra
LL2012_029	YHBJ1	BB	0.3540	0.2857	0.0080	intra
LL2012_016	LL2012_018	CC	0.9170	0.0833	0.0008	intra
YPS667	YPS667	CC	0.9580	0.0000	0.0000	intra
LL2011_005	LL2012_016	CC	0.8540	0.0833	0.0008	intra
LL2011_005	LL2011_001	CC	0.1980	0.1667	0.0000	intra
MSH-587-1	YPS667	CC	0.6820	0.2174	0.0022	intra
LL2011_006	LL2012_011	CC	0.4900	0.2592	0.0008	intra
LL2011_006	LL2011_011	CC	0.4380	0.2592	0.0008	intra
LL2012_016	YPS667	CC	0.6150	0.3333	0.0023	intra

LL2012_004	LL2011_009	CC	0.5260	0.3846	0.0000	intra
YPS667	LL2011_009	CC	0.6610	0.4167	0.0022	intra
LL2012_016	LL2011_002	CC	0.4380	0.4783	0.0008	intra
LL2011_005	MSH-587-1	CC	0.3960	0.4783	0.0000	intra
LL2011_002	LL2012_004	CC	0.5210	0.5200	0.0000	intra
LL2012_016	LL2011_009	CC	0.4690	0.5833	0.0008	intra
LL2011_009	LL2011_011	CC	0.8330	0.6000	0.0000	intra
LL2011_005	LL2012_006	CC	0.4690	0.6153	0.0000	intra
LL2012_016	LL2011_004	CC	0.4900	0.7500	0.0008	intra
LL2011_006	LL2012_018	CC	0.8958	0.1538	0.0008	intra
LL2011_006	MSH-587-1	CC	0.4271	0.3600	0.0008	intra
LL2011_006	LL2011_002	CC	0.3958	0.1999	0.0008	intra
LL2011_006	LL2011_004	CC	0.3646	0.4615	0.0008	intra
LL2012_011	LL2012_018	CC	0.5104	0.3600	0.0000	intra
LL2012_011	MSH-587-1	CC	0.9167	0.1304	0.0000	intra
LL2012_011	LL2011_002	CC	0.9063	0.0833	0.0000	intra
LL2012_011	LL2011_004	CC	0.9271	0.6000	0.0000	intra
LL2011_004	LL2011_011	CC	0.9792	0.1999	0.0000	intra
LL2011_004	LL2012_004	CC	0.3958	0.2307	0.0000	intra
LL2011_004	LL2012_018	CC	0.3854	0.5833	0.0000	intra
LL2011_004	LL2011_001	CC	0.4479	0.6667	0.0000	intra
LL2012_006	LL2012_016	CC	0.4688	0.5385	0.0008	intra
LL2012_006	MSH-587-1	CC	0.9688	0.3600	0.0000	intra
LL2012_006	LL2011_002	CC	0.9063	0.1200	0.0000	intra
LL2012_006	LL2011_004	CC	0.8750	0.2307	0.0000	intra
LL2011_001	LL2012_016	CC	0.2396	0.3333	0.0008	intra
LL2011_001	MSH-587-1	CC	0.5208	0.3333	0.0000	intra
LL2011_001	LL2011_002	CC	0.5417	0.3043	0.0000	intra
LL2012_021	LL2011_011	BC	0.3854	0.6923	0.0206	cont
LL2012_028	MSH-587-1	BC	0.1979	0.1999	0.0214	cont
LL2012_028	LL2011_002	BC	0.1354	0.3600	0.0214	cont
LL2012_028	LL2011_004	BC	0.3438	0.3846	0.0214	cont
LL2012_028	LL2012_016	BC	0.2813	0.3077	0.0196	cont
LL2012_021	LL2012_004	BC	0.2292	0.4815	0.0206	cont
LL2012_018	LL2011_005	CC	0.7813	0.0000	0.0000	intra
UWOPS-91-202	LL2012_004	BC	0.2813	0.7931	0.0198	cont
UWOPS-91-202	LL2011_001	BC	0.2083	0.3333	0.0198	cont
UWOPS-91-202	LL2012_018	BC	0.5521	0.2592	0.0198	cont
UWOPS-91-202	LL2011_011	BC	0.4479	0.5714	0.0198	cont
UWOPS-91-202	LL2012_028	BB	0.3854	0.4483	0.0014	intra