Structural diversity and contrasted evolution of cytoplasmic genomes in flowering plants : a phylogenomic approach in Oleaceae

Celine van de Paer

Université de Toulouse

THÈSE

En vue de l'obtention du

DOCTORAT DE L'UNIVERSITÉ DE TOULOUSE

Délivré par :

Université Toulouse 3 Paul Sabatier (UT3 Paul Sabatier)

Présentée et soutenue par :

Céline VAN DE PAER

Le 19 décembre 2017

Titre :

Diversité structurelle et évolution contrastée des génomes cytoplasmiques des plantes à fleurs : une approche phylogénomique chez les Oleaceae

---

ED SEVAB : Écologie, biodiversité et évolution

Unité de recherche :

Laboratoire EDB UMR 5174

Directeur(s) de Thèse :

Guillaume BESNARD

Rapporteurs :

Yamama NACIRI
Xavier VEKEMANS

Autre(s) membre(s) du jury :

Jérôme DUMINIL
Émilie LECOMPTE
Christophe THEBAUD

Remerciements

Mes premiers remerciements s'adressent à mon directeur de thèse GUILLAUME BESNARD. Tout d'abord, merci Guillaume de m'avoir proposé ce sujet de thèse sur la famille des Oleaceae. Merci pour ton enthousiasme et ta passion pour la recherche qui m'ont véritablement portée pendant ces trois années. C'était un vrai plaisir de travailler à tes côtés. Moi qui étais focalisée sur les systèmes de reproduction chez les plantes, tu m'as ouvert à un nouveau domaine de la recherche tout aussi intéressant qui est l'évolution moléculaire (même si je suis loin de maîtriser tous les concepts...). Tu as toujours été bienveillant et à l'écoute, je t'en remercie. Je te remercie d'avoir été aussi présent et de m'avoir épaulée lorsque j'en avais besoin. J'ai aussi beaucoup apprécié ta rigueur scientifique. Je te remercie pour ton investissement en termes de temps et d'énergie et pour tes retours constructifs lors de l'écriture des papiers ou du manuscrit. Merci de m'avoir fait confiance. Je te remercie de m'avoir donné la possibilité de faire des expérimentions à Montpellier lors de ma première année de thèse et de m'avoir permis de travailler sur d'autres projets tels que celui des acacias et des aulnes. Merci également de m'avoir fait partager les petits moments de bonheur/malheur et les aléas de la vie de chercheur, quasi quotidiennement (sûrement parfois au grand désespoir de mes chers collègues de bureau). Et puis, une thèse, c'est aussi la rencontre entre un doctorant et un directeur de thèse, c'est la rencontre de deux personnes et c'est la rencontre de deux caractères. Les nôtres sont assez différents, mais c'est aussi la complémentarité de nos caractères qui a fait que ces trois ans se sont aussi bien passés. Alors une dernière fois merci Guillaume, ces trois ans ont été pour moi une des expériences les plus enrichissantes scientifiquement, mais aussi humainement. Et puis j'espère que tu continueras à me raconter tes petites histoires d'Oleaceae, de graminées, de tortues, de pigeons, ... bref, tes petites histoires de science!

***

Si je me suis lancée dans le projet fou de faire une thèse, ce n'est pas tout à fait par hasard, surtout une thèse sur les Oleaceae! Parfois, tout se joue à un détail …

Tout a commencé en 2011. Petite stagiaire de L2 au laboratoire de Lille (Evo-Eco-Paléo, qui s'appelait alors GEPV), je travaillais sur l'auto-fécondation chez la *Silene.* Une partie de mon travail consistait à peser des feuilles de *Silene* et à préparer des plaques d'extraction. Alors que je pesais mes petites feuilles de *Silene*, un certain Pierre est arrivé dans la salle de manip et m'a gentiment fait remarquer que ma balance était mal calibrée et qu'il fallait que je travaille un peu plus proprement. Un détail. Un détail, certes, mais un détail qui témoignait d'une rigueur scientifique et qui m'a donné envie de travailler avec ce Pierre, que je ne connaissais absolument pas. L'année suivante, c'est donc tout naturellement que j'ai décidé d'effectuer mon stage de L3 avec Pierre Saumitou-Laprade. J'ai d'abord travaillé sur la répartition racinaire d'un sol calaminaire (en extrayant à la pince quelques milliers de racines de blocs de terre), puis j'ai rapidement bifurqué sur les questions liées aux systèmes de reproduction chez les Oleaceae.

C'est ainsi que je tiens à remercier PIERRE SAUMITOU-LAPRADE. Merci Pierre de m'avoir fait partager ta passion pour la recherche et de m'avoir fait découvrir l'histoire étonnante et fascinante des systèmes de reproduction chez les Oleaceae. Merci pour ton écoute, ta rigueur scientifique et ta patience. Merci de m'avoir permis de faire des stages passionnants à tes côtés pendant plusieurs étés. Merci pour ces sorties terrains dans des lieux magnifiques : que ce soit dans la baie de Canche, dans le Platier d'Oye ou dans les Cévennes.

Merci pour ta bonne humeur, ta bienveillance et ton optimisme. Je remercie également PHILIPPE VERNET : travailler avec vous deux, Pierre et Philippe, a été particulièrement enrichissant pour moi. Je tiens aussi à remercier le laboratoire de Lille pour m'avoir accueillie en stages (obligatoires et volontaires) pendant plusieurs années.

Travailler sur les Oleaceae m'a également permis de me réconcilier avec les équations. Je tiens en effet à remercier SYLVAIN BILLIARD pour m'avoir ouvert au monde de la modélisation. Je souhaite à tout étudiant (souhaitant faire de la modélisation ou non) d'être encadré par une personne comme toi. Merci pour ton sens de la pédagogie, merci pour ta patience et pour le temps que tu as toujours pris pour expliquer, voire ré-expliquer, certains principes, même les plus simples. Merci pour ton écoute ; merci d'être ouvert à toutes les questions, même celles qui nous paraissent les plus bêtes et naïves. Merci pour ton savant dosage entre la part d'encadrement et la part de liberté et d'autonomie accordées à tes étudiants. Travailler à tes côtés a été vraiment stimulant scientifiquement et intellectuellement. Merci pour nos discussions diverses ; c'est toujours un plaisir de discuter avec toi de sujets scientifiques, ou simplement de parler de la vie en général. Merci Sylvain.

Ce sont donc ces différentes expériences scientifiques partagées avec Pierre, Philippe et Sylvain qui m'ont donné envie de faire de la recherche et qui m'ont amenée à faire cette thèse auprès de Guillaume; c'est pourquoi je tenais à les remercier à leur juste valeur.

***

Je remercie la REGION MIDI-PYRENEES et le LABEX TULIP d'avoir financé cette thèse.

Je remercie YAMAMA NACIRI, XAVIER VEKEMANS, JEROME DUMINIL, EMILIE LECOMPTE et CHRISTOPHE THEBAUD d'avoir accepté de faire partie de mon jury de thèse, et particulièrement YAMAMA NACIRI et XAVIER VEKEMANS d'avoir accepté d'être mes rapporteurs.

Je remercie les membres de mes comités de thèse pour leurs remarques et leurs suggestions constructives : merci à PIERRE-OLIVIER CHEPTOU, RAFAEL RUBIO DE CASAS, SYLVAIN BILLIARD, PIERRE SAUMITOU-LAPRADE, PHILIPPE VERNET, HERVE PHILIPPE et PASCAL-ANTOINE CHRISTIN.

Je remercie PIERRE-OLIVIER CHEPTOU de m'avoir permis de faire mes expérimentations au CEFE de Montpellier lors de ma première année de thèse. Par ailleurs, je remercie THIERRY MATHIEU et PAULINE DURBIN pour leurs soins apportés aux oliviers Laperrine situés dans le jardin expérimental du CEFE de Montpellier et d'avoir mis à ma disposition tout ce dont j'avais besoin lors de mes expérimentations. Je remercie JACQUES LEPART, PIERRE SAUMITOU-LAPRADE et PHILIPPE VERNET pour leur aide précieuse lors de ces expérimentations. Je remercie BENOIT DESPLANQUES pour son accueil dans son oliveraie et d'avoir mis à ma disposition ses oliviers pour effectuer mes manipulations.

Je remercie HERVE PHILIPPE pour son aide et pour toutes les discussions constructives et enrichissantes que nous avons eu sur les conséquences génomiques d'un potentiel «paternal leakage ».

Je remercie PASCAL-ANTOINE CHRISTIN pour son aide sur les analyses de sélection.

Je remercie MOHAMED MENSOUS, avec qui j'ai collaboré sur le papier de la diversité des plastomes chez les acacias.

Je remercie l'ensemble des membres du LABORATOIRE EDB pour l'atmosphère de travail stimulante qui y règne. Merci à SOPHIE MANZI, AMAIA IRIBAR-PELOZUELO et HELENE HOLOTA pour leur aide et leur investissement dans la partie séquençage. Merci à DOMINIQUE PANTALACCI, CATHERINE CAUNES, LESLIE LEFRANC, LINDA JALABERT et NICOLE HOMMET pour toute la partie administrative.

Je remercie tous les DOCTORANTS ET POST-DOC du labo (je ne me risquerai pas à tous vous nommer de peur d'en oublier) pour votre bonne humeur qui rend le quotidien plus agréable. Je remercie KEVIN pour son aide précieuse avec R. Un merci particulier à tous ceux qui ont partagé mon bureau : BORIS, LEA, ARTHUR, ISABELLE, MARINE et SEBASTIEN ; merci à tous pour votre bonne humeur et la bonne ambiance qui règne dans le bureau, merci de m'avoir supportée au quotidien, merci pour votre bienveillance, merci de m'avoir encouragée et rassurée dans les moments difficiles. Et désolée Marine, mais 2 ans dans le même bureau n'auront pas suffi pour me faire aimer le jazz (un jour, j'en suis sûre, j'aurai le déclic!).

Merci à SANDRA. C'était un plaisir de faire du sport avec toi pendant 1 an, on remet ça dès que Cacahuète a pointé le bout de son nez ! Merci pour nos discussions diverses et variées (qui ressemblent souvent à des « réunions Tupperware », avouons-le...). Merci pour ton soutien et tes encouragements dans la dernière ligne droite.

Merci à MARINA ; merci pour ton écoute et pour nos petites discussions de thésardes. Je te prie de m'excuser si durant ces trois dernières années je n'ai pas été suffisamment présente ou à l'écoute dans des moments où tu en aurais eu besoin.

Merci à JAN. J'aurais tellement de remerciements à t'adresser... Si je suis arrivée au bout de cette thèse, c'est aussi grâce à ton écoute, tes encouragements, ta patience et ta bienveillance. Tu m'as épaulée au quotidien. Merci de m'avoir supportée, encouragée et soutenue lors de mes (nombreux) moments de doutes. Tu m'as apporté l'équilibre dont j'avais besoin. J'ai été heureuse de partager l'aventure de cette thèse, ou plutôt de nos thèses, avec toi et j'espère encore partager de nombreuses aventures à tes côtés.

Enfin, je tiens tout particulièrement à remercier MES PARENTS. Bien que vous ne les connaissez pas personnellement, les noms de « Pierre », «Philippe », « Sylvain » ou « Guillaume » vous sont familiers depuis plus ou moins longtemps. Le monde de la recherche vous était inconnu, mais vous avez toujours fait l'effort de découvrir et de connaître mon univers et d'essayer de comprendre ce sur quoi je travaillais (en m'écoutant patiemment parler de « systèmes d'incompatibilités », de « génomes » ou de « publications »). Je vous en suis reconnaissante. Vous avez toujours été présents pour partager mes joies, pour me redonner confiance et m'encourager quand j'avais des doutes ou pour me réconforter dans les moments les plus difficiles. Pour tout cela, je vous remercie. Merci de m'avoir fait confiance et de m'avoir toujours soutenue, quels que soient mes choix. Merci d'avoir toujours été là pour moi.

SOMMAIRE

INTRODUCTION GENERALE

Les organismes vivants présentent des génomes très diversifiés en termes de taille, de structure, d'évolution et de fonction. Déterminer et comprendre les facteurs génétiques, écologiques et évolutifs qui gouvernent cette diversité génomique est un enjeu majeur en évolution moléculaire. Le génome des animaux, des plantes et des champignons est partitionné dans différents compartiments au sein de la cellule. Le noyau contient une large partie du génome et le reste du génome est contenu dans les mitochondries chez les eucaryotes mais aussi dans les chloroplastes chez les plantes. Les chloroplastes et les mitochondries sont les seuls organites (i.e. compartiments intracellulaires différenciés et spécialisés) à posséder leur propre génome. Les génomes nucléaire, mitochondrial et chloroplastique co-évoluent et interagissent afin de coordonner l'expression génique et assurer les fonctions métaboliques nécessaires au bon fonctionnement de la cellule. Chez les plantes, la dynamique structurelle et l'histoire évolutive concertée des génomes mitochondriaux et chloroplastiques sont encore mal connues. L'objectif de ce travail de thèse est d'étudier la diversité structurelle et l'évolution des génomes mitochondriaux et chloroplastiques chez une famille de plantes à fleurs, les Oleaceae, en combinant une approche génomique et phylogénétique. Bien qu'il existe d'autres organites que les chloroplastes et les mitochondries, le terme "organite" sera utilisé dans ce manuscrit pour désigner les chloroplastes et les mitochondries uniquement.

I/ Origines et évolution des organites chez les plantes à fleurs

1. Origines des organites

1.1. Apparition de la photosynthèse et de la respiration

La photosynthèse est un processus permettant la synthèse de molécules organiques complexes en utilisant de l'énergie lumineuse. La photosynthèse oxygénique entraîne la fixation du dioxyde de carbone (CO_2) et la libération d'oxygène (O_2). L'apparition de la photosynthèse est estimée entre 2,4 et 2,1 milliards d'années (Lyons *et al.* 2014 ; Soo *et al.* 2017). Jusqu'alors, l'oxygène était présent en quantité infime dans l'atmosphère terrestre (~0,0001%) et le métabolisme des organismes vivants était anaérobie (i.e. qui ne nécessite pas d'oxygène). Les premiers organismes photosynthétiques sont les cyanobactéries, appartenant au domaine des bactéries (Hohmann-Marriott & Blankenship 2011 ; Blankenship 2017 ; Soo *et al.* 2017). L'activité photosynthétique des cyanobactéries entraîna une augmentation significative de la concentration de l'oxygène dans l'atmosphère (Lyons *et al.* 2014). Certaines bactéries s'adaptèrent à ce changement de composition atmosphérique en utilisant l'oxygène dans leur métabolisme : c'est l'apparition de la respiration aérobie [i.e. qui nécessite de l'oxygène; Lyons *et al.* (2014) ; Soo *et al.* (2017)].

1.2. Origines des chloroplastes et des mitochondries par endosymbiose

Une endosymbiose est une relation durable et bénéfique entre deux organismes vivants, où l'un des deux organismes (appelé endosymbiote) est présent à l'intérieur de l'autre organisme (appelé hôte). Les chloroplastes et les mitochondries sont issus d'endosymbioses indépendantes (Fig. 1). Cependant, ces événements d'endosymbioses chloroplastique et mitochondriale sont fondamentalement similaires (Fig. 1), impliquant un hôte eucaryote (i.e. dont le noyau contenant l'ADN est séparé du reste de la cellule par une membrane) ou procaryote et un endosymbiote procaryote [i.e. une cellule dépourvue de noyau et de structures internes ; Margulis (1970)]. La théorie endosymbiotique de l'origine des organites stipule que l'endosymbiose mitochondriale a précédé l'endosymbiose chloroplastique (Soo *et al.* 2017).

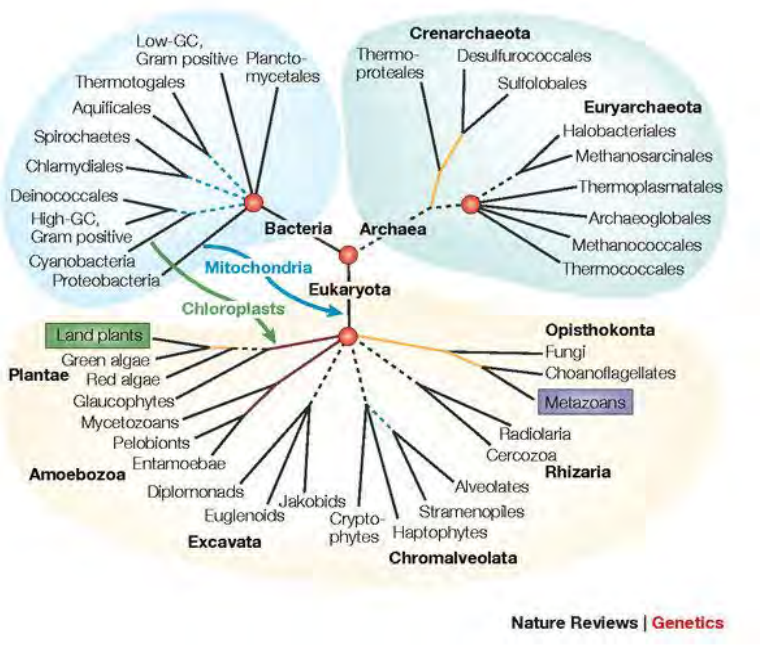


Fig. 1 Représentation schématique de l'arbre de la vie illustrant les relations phylogénétiques des trois domaines (Bacteria, Archaea et Eucaryota). Les événements d'endosymbioses à l'origine des chloroplastes et des mitochondries sont représentés par une flèche verte et une flèche bleue, respectivement. [Figure extraite de Delsuc *et al.* (2005)]

L'origine endosymbiotique des mitochondries (Martin & Mentel 2010 ; Gray & Archibald 2012). On distingue deux principales théories sur l'origine des mitochondries par endosymbiose, qui s'opposent sur la nature de l'hôte. La théorie dite traditionnelle suppose une endosymbiose entre un hôte eucaryote anaérobie (un protiste amitochondrial) et un endosymbiote aérobie de type bactérien (du groupe α-Proteobacteria ; Fig. 1). La mitochondrie (i.e. l'endosymbiote) permet alors de détoxifier l'oxygène de l'hôte en l'utilisant pour produire de l'énergie ; en contrepartie, la mitochondrie reçoit au sein de l'hôte la protection et les nutriments nécessaires à son fonctionnement. Cette théorie est remise en cause du fait qu'aucune lignée d'eucaryotes n'a pu être identifiée comme ayant divergé avant l'acquisition symbiotique de la mitochondrie. Une hypothèse inhérente à cette théorie est que l'ensemble des descendants issus de la lignée hôte s'est éteint, à l'exception de ceux porteurs

de mitochondries. Une théorie alternative postule que l'hôte impliqué dans l'endosymbiose de la mitochondrie était un procaryote de type archéobactérien aérobie facultatif (i.e. capable de vivre avec ou sans oxygène). Il est supposé que cet hôte archéobactérien était H_2-dépendant et aurait pu bénéficier de la production d'H_2 par la mitochondrie pour produire de l'énergie. Cette théorie suppose que les eucaryotes ont évolué à partir d'une endosymbiose mitochondriale dans un hôte procaryote, ce qui expliquerait l'ubiquité des mitochondries au sein des eucaryotes.

L'origine endosymbiotique des chloroplastes (Keeling 2010 ; Jensen & Leister 2014). La théorie de l'endosymbiose des chloroplastes (appelée aussi endosymbiose primaire) implique un hôte eucaryote et une cyanobactérie photosynthétique (Figs 1 et 2). Cette théorie, largement acceptée, repose notamment sur les larges similitudes entre les génomes des chloroplastes et des cyanobactéries. La plupart des reconstructions phylogénétiques regroupent les gènes chloroplastiques avec les gènes homologues (i.e. descendant d'un même gène ancestral) de cyanobactéries. Trois lignées d'eucaryotes ont dérivé de l'acquisition primaire d'une cyanobactérie : les glaucophytes, les algues rouges et les algues vertes (à partir desquelles les plantes terrestres ont dérivé ; Fig. 2). Les reconstructions phylogénétiques suggèrent que ces lignées dérivent d'un seul événement d'endosymbiose primaire, bien que différents événements d'endosymbioses rapprochés sont quasiment impossibles à distinguer d'un unique événement. La plupart des autres lignées d'algues sont issues d'endosymbioses secondaires impliquant un hôte eucaryote et une algue eucaryote photosynthétique (i.e. une algue rouge ou une algue verte ; Fig. 2). Ces événements d'endosymbioses secondaires ont eu lieu à de multiples occasions (entre deux et sept fois) au cours de l'évolution des algues (Delwiche 1999; Cavalier-Smith 2000).

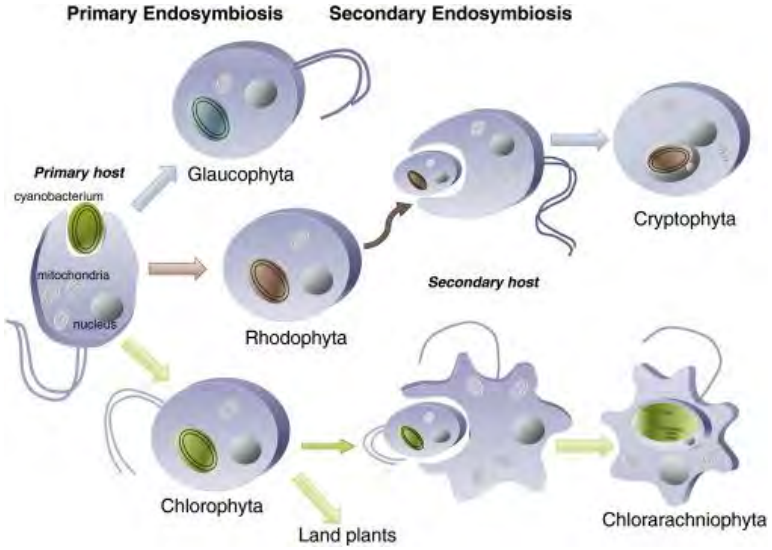


Fig. 2 Schéma des endosymbioses primaires et secondaires des chloroplastes. L'endosymbiose primaire implique un hôte eucaryote et une cyanobactérie photosynthétique et est à l'origine de trois lignées : les glaucophytes (Glaucophyta), les algues rouges (Rhodophyta) et les algues vertes (Chlorophyta, à partir desquelles les plantes terrestres ont dérivé). L'endosymbiose secondaire implique un hôte eucaryote et une algue eucaryote photosynthétique (une algue rouge ou une algue verte). [Figure extraite de Bölter & Soll (2016)]

2. Structure et évolution des génomes cytoplasmiques chez les plantes à fleurs

Compte tenu de leurs histoires évolutives similaires, les chloroplastes et les mitochondries partagent de nombreuses caractéristiques génomiques en commun. Les chloroplastes et les mitochondries possèdent en partie leur propre génome, relique des endosymbioses bactériennes à partir desquelles ils ont évolué. Suite aux événements d'endosymbioses, les chloroplastes et les mitochondries ont subi une réduction drastique de leurs génomes et la plupart de leurs gènes ont été transférés vers le génome nucléaire. Par ailleurs, les chloroplastes et les mitochondries ont suivi indépendamment des chemins évolutifs semblables, comme c'est le cas pour leurs stratégies d'expression génique et de réplication, leur mode de transmission ou encore certaines de leurs caractéristiques structurelles. Le génome nucléaire codant pour des protéines cytoplasmiques pourrait lier les génomes chloroplastiques et mitochondriaux et favoriser l'évolution convergente des deux organites. Néanmoins, on observe de nombreux traits divergents entre les chloroplastes et les mitochondries, tant au niveau de leur structure que de leur évolution.

2.1. Structure et organisation des génomes cytoplasmiques

Structure des génomes chloroplastiques (Wicke *et al.* 2011 ; Jansen & Ruhlman 2012 ; Tonti-Filippini *et al.* 2017). Les génomes chloroplastiques (appelés aussi plastomes) chez les plantes à fleurs sont hautement conservés en termes de structure, de taille et de contenu en gènes. Les plastomes sont généralement des génomes circulaires de ~140-160 kilobases (kb) et quadripartites incluant deux régions répétées et inversées (les "Inverted Repeats", notées IR) séparées par une petite et une grande régions en simple copie ["Small Single Copy" (SSC) et "Large Single Copy" (LSC) ; Fig. 3]. On peut néanmoins retrouver des espèces dont la taille du plastome est extrêmement réduite, notamment chez les plantes parasites non-photosynthétiques (appelées holoparasites). Le plus petit plastome connu, d'environ 11 kb, est celui de l'holoparasite africain *Pilostyles aethiopica* (Bellot & Renner 2015). A l'inverse, certaines espèces telles que *Pelargonium×hortorum* (Geraniaceae) possèdent des plastomes de tailles exceptionnellement larges [parfois supérieures à 200 kb; Chumley *et al.* (2006)]. Une extension des "Inverted Repeats" et des duplications partielles ou complètes de gènes peuvent entraîner une telle augmentation de la taille du plastome. En termes de contenu génique, le plastome chez les plantes à fleurs contient entre 100 et 120 gènes, codant principalement pour des protéines et des ARN [de transfert (ARNt) ou ribosomiques (ARNr)] impliqués dans la photosynthèse, la transcription et la translation.

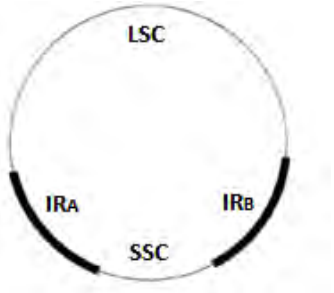


Fig. 3 Structure classique d'un plastome avec deux "Inverted Repeats" (IR_A et IR_B) séparées par une petite (SSC) et une grande (LSC) régions en simple copie. [Figure extraite de Khan *et al.* (2010)]

Structure des génomes mitochondriaux (Burger *et al.* 2003 ; Mower *et al.* 2012 ; Gualberto *et al.* 2014 ; Smith & Keeling 2015). Contrairement aux plastomes, les génomes mitochondriaux (appelés aussi mitogénomes) sont extrêmement variables en termes de structure, de taille et d'organisation au sein des plantes à fleurs. Le mitogénome des plantes présente une structure complexe, en sous-cercles d'ADN (Fig. 4). De nombreuses régions répétées (allant jusqu'à plusieurs kb) sont généralement présentes dans un mitogénome de plante. Des événements de recombinaison active entre ces régions répétées entraînent une structure multipartite du mitogénome, i.e. la formation de configurations multiples et interchangeables du mitogénome. La représentation physique du mitogénome en molécule circulaire, appelée "master-cercle", habituellement utilisée est donc factice et n'est qu'une séquence consensus résumant le contenu génomique de la mitochondrie. La taille du mitogénome des plantes à fleurs est extrêmement variable entre espèces, voire à l'échelle intra-spécifique. Le mitogénome des plantes présente une taille comprise entre ~200 kb chez *Brassica rapa* (Brassicaceae) et ~11 megabases (Mb) chez *Silene conica* (Caryophyllaceae). Ces grandes variations de taille des mitogénomes sont essentiellement dues aux nombreuses séquences répétées et aux régions non-codantes. Bien qu'extrêmement larges, les mitogénomes chez les plantes à fleurs ne contiennent qu'entre 50 et 60 gènes. Les gènes mitochondriaux codent pour des protéines, des ARNt et des ARNr impliqués dans la respiration et/ou la phosphorylation oxydative (qui permet la phosphorylation de l'ADP en ATP), la maturation des ARN, l'importation de protéines, la translation et occasionnellement la transcription. Les événements de recombinaison au sein du mitogénome sont à l'origine de nombreux réarrangements dans l'ordre des gènes.

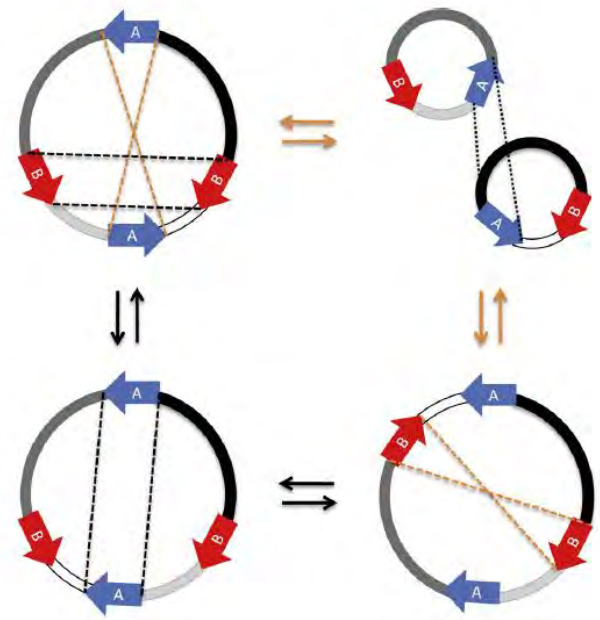


Fig. 4 Structure classique des mitogénomes de plantes en sous-molécules circulaires (structure multipartite). Les longues régions répétées (ici, représentées par des flèches rouges et bleues) favorisent la recombinaison intra-moléculaire, ce qui génère une structure du mitogénome en sous-cercles d'ADN. [Figure extraite de Gualberto *et al.* (2014)]

2.2. Évolution des génomes cytoplasmiques

Vitesse d'évolution des régions codantes (Wolfe *et al.* 1987 ; Palmer 1992 ; Zhu *et al.* 2016). Bien que le mitogénome des plantes à fleurs varie rapidement en termes de structure, la séquence nucléotidique mitochondriale évolue globalement très lentement. Le taux de substitutions synonymes (d_S) dans les gènes mitochondriaux est trois à quatre fois plus faible que dans les gènes chloroplastiques et près de 12 fois plus faible que dans le génome nucléaire. En revanche, le taux de substitutions non-synonymes (d_N) est similaire dans les gènes chloroplastiques et mitochondriaux. Les taux de substitutions nucléotidiques peuvent varier entre gènes selon leur position au sein du génome. Les gènes chloroplastiques localisés dans les régions en simple copie présentent un taux de substitutions synonymes ~3,7 fois plus élevé que les gènes situés dans les "Inverted Repeats". Néanmoins, il a été montré que trois lignées de plantes à fleurs (*Pelargonium*, *Plantago* et *Silene*) présentent un taux de substitutions synonymes légèrement plus élevé (de 1,4 à 2,1 fois) dans les gènes des "Inverted Repeats" que ceux situés dans les régions en simple copie (Zhu *et al.* 2016). Les taux de substitutions nucléotidiques peuvent être également très variables parmi les gènes et entre espèces (Jansen & Ruhlman 2012).

Séquences dupliquées et recombinaison homologue (Maréchal & Brisson 2010). La recombinaison homologue correspond à l'échange de matériel génétique entre des séquences nucléotidiques identiques ou similaires. La recombinaison homologue est impliquée dans la réparation des dommages subis par l'ADN et assure la stabilité des génomes. Un système actif de recombinaison homologue a été mis en évidence dans les plastomes et les mitogénomes chez les plantes à fleurs. Ces événements de recombinaison jouent un rôle important dans la dynamique de la structure du génome des organites et dans leur évolution.

- Dans le plastome des plantes, la recombinaison entre les "Inverted Repeats" génère deux types d'isoformes différant selon l'orientation des régions en simple copie. Mis à part les "Inverted Repeats", le plastome des plantes présente très peu de séquences répétées. Il a été montré que des répétitions suffisamment larges pour recombiner ont un effet déstabilisant délétère pour le plastome et seraient contre-sélectionnées.

- Le mitogénome des plantes présente un système de recombinaison particulièrement actif, du fait des nombreuses séquences répétées réparties dans le génome. Les larges répétitions (>1 kb) recombinent fréquemment au sein du mitogénome, entraînant la structure multipartite du génome. Les réarrangements du mitogénome sont souvent à l'origine de nouveaux cadres de lecture ouverts (dits ORFs : Open Reading Frames) codant pour des gènes chimériques. Ces gènes chimériques peuvent générer des incompatibilités cytoplasmiques telles que la stérilité mâle cytoplasmique (dite CMS : Cytoplasmic Male Sterility). Des recombinaisons homologues moins fréquentes ont lieu entre les petites répétitions du mitogénome et génèrent de petites molécules d'ADN appelées sublimons. Chez le millet perlé et le maïs, il a été montré que la recombinaison entre des petites répétitions mitochondriales peut entraîner la restauration de la fertilité chez des plantes initialement mâle-stériles (Schardl *et al.* 1985; Feng *et al.* 2009).

Transferts de séquences d'ADN entre organites (Wang *et al.* 2007 ; Xia *et al.* 2013 ; Sloan & Wu 2014). Les régions non-codantes du mitogénome des plantes à fleurs sont en partie composées d'ADN d'origine chloroplastique. Le mitogénome des plantes contient entre 1% (~4 kb) chez *Arabidopsis* et 11,5% (~113 kb) chez *Curcubita* de régions dérivées du plastome (Unseld *et al.* 1997 ; Alverson *et al.* 2010). L'ADN d'origine chloroplastique retrouvé dans le mitogénome peut provenir de différentes régions du plastome. Ces régions dérivées du plastome sont généralement des pseudogènes ou des séquences non-codantes, bien que des gènes codant pour des ARNt restent fonctionnels. Dans de rares cas, des séquences d'origine chloroplastique sont intégrées dans la région codante de gènes mitochondriaux fonctionnels. Le contenu en séquences d'origine chloroplastique peut fortement varier entre espèces, voire au niveau intra-spécifique. Les mitogénomes acquièrent et perdent fréquemment ces régions dérivées du plastome. Les séquences d'origine chloroplastique n'étant pas forcément orthologues (i.e. issues d'un ancêtre commun) au sein d'une même lignée (e.g. au sein d'une espèce, d'un genre ou d'une famille), il est souvent difficile d'éclaircir l'histoire évolutive de ces régions. Contrairement au mitogénome, le transfert de séquences mitochondriales vers le plastome se produit rarement. Des événements d'intégration de régions mitochondriales dans le plastome ont été reportés chez la carotte (Iorizzo *et al.* 2012), l'herbe aux perruches [*Asclepias syriaca* ; Straub *et al.* (2013)] et quelques graminées (Ma *et al.* 2015 ; Burke *et al.* 2016).

Transferts réciproques de séquences entre les génomes cytoplasmiques et le génome nucléaire. Alors que le transfert de gènes mitochondriaux vers le génome nucléaire a été décrit dans de nombreux cas chez les plantes à fleurs, l'intégration de gènes chloroplastiques dans le génome nucléaire n'a été que rarement reportée (Arthofer *et al.* 2010). Les rares cas de transfert de gènes chloroplastiques vers le génome nucléaire reportés concernent les gènes *rpl22* chez le pois (Gantt *et al.* 1991), *infA* chez *Arabidopsis* et le soja (Millen *et al.* 2001), *rpl32* chez le peuplier (Cusack & Wolfe 2007 ; Ueda *et al.* 2007) et *accD* chez le trèfle (Magee *et al.* 2010) et les Campanulaceae (Rousseau-Gueutin *et al.* 2013). Dans le mitogénome, le contenu génique est relativement stable parmi les lignées de plantes à fleurs. Seize gènes mitochondriaux semblent néanmoins être fréquemment perdus parmi les lignées d'angiospermes (Adams & Palmer 2003) : l'ensemble des 14 gènes ribosomiques et deux gènes codant pour la succinate déshydrogénase (*sdh3* et *sdh4*). Alors que la plupart des lignées basales des angiospermes possèdent les 40 gènes présents chez l'ancêtre commun des plantes à fleurs, certaines lignées [telles que *Allium* (l'oignon), *Lachnocaulon* et *Erodium*] ont perdu tous, ou presque tous, les 14 gènes ribosomiques ainsi que *sdh3* et *sdh4*. Ces lignées convergent vers une même réduction de leur contenu génique. On distingue trois raisons majeures de la perte de gènes au sein du mitogénome : (i) la fonction du gène n'est plus nécessaire et le gène est complètement perdu de la cellule (cet événement n'arrive que rarement) ; (ii) la "substitution de gènes", ou "remplacement de gènes", i.e. un gène préexistant est recruté pour assurer la fonction du gène perdu ; (iii) le gène perdu a été transféré vers le génome nucléaire en conservant sa fonction. Différentes études ont mis en évidence le transfert indépendant de 10 gènes ribosomiques et des gènes *cox2*, *sdh3* et *sdh4* du mitogénome vers le génome nucléaire au sein de différentes lignées d'angiospermes (Adams & Palmer 2003). Réciproquement, il a été montré que les régions intergéniques du

mitogénome des plantes à fleurs peuvent également être en partie composées de séquences dérivées du génome nucléaire. Le contenu en séquences dérivées du génome nucléaire est très variable entre espèces (Kubo & Mikami 2007). Chez les graminées, les séquences dérivées du génome nucléaire représentent 13,4% du mitogénome du riz (Notsu *et al.* 2002) et moins de 0,2% des mitogénomes du maïs et du blé (Notsu *et al.* 2002 ; Ogihara *et al.* 2005). Nécessitant le séquençage complet du génome nucléaire et du mitogénome d'une même espèce, peu de données sont actuellement disponibles quant au contenu en séquences nucléaires au sein des mitogénomes.

Transferts horizontaux vers les génomes cytoplasmiques (Bock 2009). Le transfert horizontal d'ADN correspond à l'acquisition de matériel génétique d'un autre organisme (pouvant appartenir à une autre espèce), sans en être son descendant. Le transfert horizontal de séquences mitochondriales entre plantes a été reporté chez plusieurs plantes à fleurs, alors qu'aucun cas d'échange d'ADN chloroplastique n'est actuellement connu. L'importante activité de recombinaison homologue des mitochondries chez les plantes et leur faculté de pouvoir fusionner sont deux caractéristiques qui peuvent faciliter l'échange horizontal d'ADN. La plupart des transferts horizontaux d'ADN mitochondrial concernent des plantes parasites ou épiphytes (i.e. poussant sur une autre plante en s'en servant comme support) et leur plante hôte. De larges régions d'ADN, voire le mitogénome complet, peuvent être transférées et intégrées dans le mitogénome de la plante hôte par fusion et recombinaison des deux mitogénomes. Bien que les transferts horizontaux d'ADN mitochondrial se passent généralement intra-genre, des cas de transferts horizontaux d'une angiosperme vers une fougère ou d'une mousse vers une angiosperme ont été reportés. L'angiosperme basal *Amborella trichopoda* (Amborellaceae) présente le plus grand nombre de transferts horizontaux détectés dans un mitogénome de plante (Bergthorsson *et al.* 2004 ; Rice *et al.* 2013). Cette espèce est un arbuste tropical généralement couvert de plantes épiphytes, dont diverses espèces de mousses (Fig. 5). Il a été montré que le mitogénome d'*A. trichopoda* de ~3,9 Mb a acquis de larges séquences mitochondriales d'algues, de mousses et de plusieurs angiospermes. Le mitogénome d'*A. trichopoda* contient notamment le mitogénome complet d'une mousse et de trois algues vertes. Par ailleurs, il est possible que les régions inter-géniques des mitogénomes de plantes à fleurs contiennent un grand nombre de régions intégrées horizontalement, mais leur identification reste encore difficile dans la mesure où peu de génomes sont encore complètement séquencés.

Fig. 5 Contexte écologique de transferts horizontaux d'ADN chez *Amborella* : **A** et **B.** Branches d'*Amborella* couvertes de plantes grimpantes (flèches vertes) et d'épiphytes, incluant des mousses, des hépatiques, des fougères et des angiospermes ; **C** à **F.** Feuilles et branches d'*Amborella* couvertes de lichens (**C** et **F**), d'hépatites (**D**) et de mousses (**E**). [Figure extraite de Rice *et al.* (2013)]

3. Transmission des organites chez les plantes à fleurs

L'ensemble des informations composant cette partie reposent sur la revue de Greiner *et al.* (2014).

3.1. Prédominance de la transmission uniparentale maternelle des organites

Chez la plupart des plantes à fleurs, la transmission des organites est uniparentale et maternelle. La transmission uniparentale des organites implique une ségrégation non-Mendélienne, avec l'absence de recombinaison. A la différence du génome nucléaire transmis par le pollen et par les graines, la valeur sélective du génome des organites transmis maternellement dépend directement de la qualité et de la quantité de graines produites. Par ailleurs, les génomes nucléaires et cytoplasmiques présentent généralement des structures génétiques de population différentes, reflétant des modes de dispersion contrastés (via le pollen et les graines versus les graines uniquement). Bien que la transmission uniparentale maternelle est dominante chez les angiospermes, le passage occasionnel ou régulier d'organites dans le pollen a été décrit chez de nombreuses espèces.

3.2. Forces évolutives entraînant l'apparition et le maintien d'une transmission uniparentale des organites

Une hypothèse proposée pour la dominance d'une transmission uniparentale des organites est l'évitement de conflits entre le génome des organites paternels et maternels, qui seraient délétères pour le génome nucléaire. L'une des incompatibilités cyto-nucléaires les plus connues est la stérilité mâle cytoplasmique, qui interfère avec le développement des anthères et du pollen pour rendre des individus hermaphrodites fonctionnellement femelles. Par ailleurs, le génome des organites provenant des deux parents ne subissant pas de recombinaison, ces derniers pourraient entrer en compétition directe. Un des deux génomes pourrait par exemple acquérir un avantage compétitif lui permettant de se répliquer plus rapidement que l'autre génome. La sélection naturelle favoriserait la transmission du génome le plus compétitif à la génération suivante. Une telle compétition entre organites serait désavantageuse pour le génome nucléaire. Un organite se répliquant plus rapidement que le génome nucléaire serait mal adapté à la cellule et entraînerait une réduction de la valeur sélective. Une autre possibilité de conflits cyto-nucléaires serait le cas où les organites paternels et maternels porteraient des allèles différents au même locus et que leur combinaison hétéroplasmique serait mal adaptée.

3.3. Forces évolutives entraînant la prédominance de la transmission maternelle des organites

La prédominance de la transmission maternelle des organites chez les plantes à fleurs pourrait être liée à un fardeau mutationnel plus élevé dans la cellule spermatique du pollen que dans l'ovule. Les organites du pollen mobile seraient davantage sujets à des dommages oxydatifs que les organites de l'ovule sessile. Par ailleurs, les cellules de la lignée germinale maternelle sont protégées du stress oxydatif par des tissus spécialisés. Enfin, le nombre de copies du

génome des organites est généralement plus faible dans la cellule spermatique du pollen que dans l'ovule. En conséquence, les génomes cytoplasmiques paternels seraient davantage soumis à la dérive génétique. Chez des espèces dont la transmission des organites est paternelle ou biparentale telles que *Pelargonium*, le melon ou la luzerne, il a été montré que le nombre de copies des génomes cytoplasmiques dans le pollen est plus élevé que chez les espèces à transmission maternelle, ce qui pourrait limiter la dérive génétique (Mogensen 1996; Wang *et al.* 2010). Par ailleurs, on peut supposer que la dispersion à faible distance des organites transmis par les graines peut favoriser l'adaptation locale, ce qui pourrait être bénéfique dans un contexte environnemental stable.

3.4. La transmission uniparentale maternelle des organites : un trait évolutivement instable

La transmission uniparentale des organites exclut toute recombinaison entre les organites provenant des deux parents. L'absence de recombinaison est susceptible d'entraîner une accumulation de mutations délétères au fil des générations [processus appelé cliquet de Muller, Muller (1964)]. Les mécanismes d'évitement d'un fardeau de mutations délétères chez les organites sont encore mal connus. Le passage occasionnel d'organites par la voie paternelle via le pollen pourrait favoriser la recombinaison entre les copies maternelles et paternelles des organites, et ainsi contrebalancer le cliquet de Muller. Du "paternal leakage" ou de la transmission biparentale ont été détectés chez de nombreuses espèces. Il apparaît que la transmission biparentale ait évolué indépendamment et de manière répétée au cours de l'évolution des plantes à fleurs. Néanmoins, les conséquences génomiques d'un "paternal leakage" étant peu connues, ces hypothèses nécessitent davantage d'études et de tests expérimentaux.

3.5. Une hypothèse consensus de l'évolution de la transmission uniparentale maternelle proposée par Greiner *et al.* (2014)

Dans leur revue, Greiner *et al.* (2014) proposent un modèle d'évolution de la transmission uniparentale maternelle des organites, en tenant en compte des différentes hypothèses théoriques et données expérimentales (Fig. 6). La transmission uniparentale des organites pourrait évoluer afin d'éviter la propagation d'éléments dits "égoïstes" (e.g. des organites se répliquant plus rapidement), qui entraîneraient de mauvaises adaptations ou des incompatibilités avec le génome nucléaire. Néanmoins, la transmission uniparentale est évolutivement instable du fait de l'accumulation de mutations délétères liée à l'absence de recombinaison. Afin de contrebalancer le cliquet de Muller, la recombinaison sexuelle pourrait avoir lieu via des événements épisodiques ou réguliers de transmission biparentale. A nouveau, la propagation d'éléments égoïstes favorisée par la transmission biparentale entraînerait un retour vers la transmission uniparentale des organites. La transmission uniparentale à prédominance maternelle pourrait être due à un fardeau mutationnel plus élevé chez les organites paternels, causé par des dommages oxydatifs et/ou par la dérive génétique. Ce fardeau de mutations délétères est d'autant plus important que le nombre de copies des génomes cytoplasmiques est faible dans la cellule spermatique du pollen.

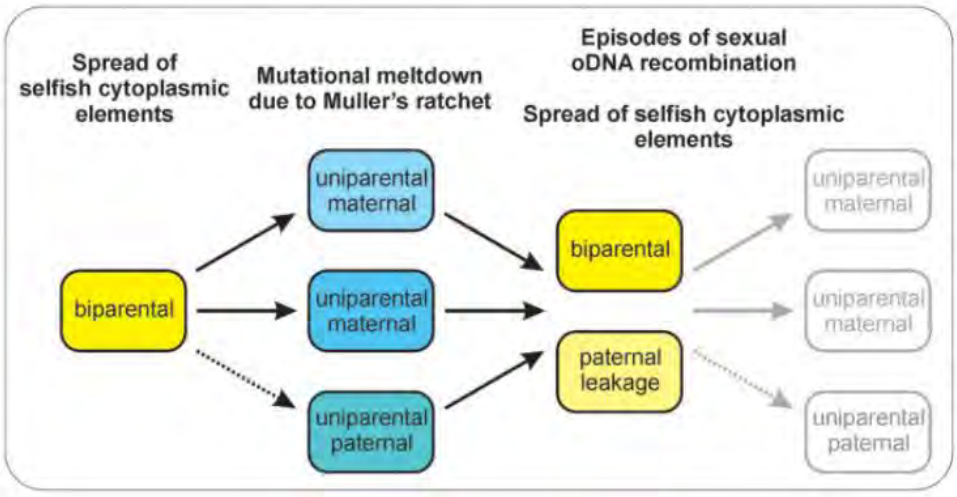


Fig. 6 Hypothèse consensus des pressions de sélection entraînant l'apparition et la perte répétées de la transmission uniparentale maternelle des organites chez les plantes à fleurs. [Figure extraite Greiner *et al.* (2014)]

Bien que partageant un certain nombre de caractéristiques communes, les données actuelles montrent que les génomes chloroplastiques et mitochondriaux présentent d'importantes différences structurelles et évolutives. Les mitogénomes des plantes à fleurs, en particulier, sont caractérisés par une large diversité de taille, de structure, de contenu génique et de substitutions nucléotidiques. Néanmoins, il est nécessaire d'accumuler davantage de données génomiques couvrant un plus large échantillonnage taxonomique. Par ailleurs, les conséquences génomiques d'un passage occasionnel ou régulier d'organites par la voie paternelle sont encore mal connues. Des approches phylogénétiques basées sur des données génomiques toujours plus abondantes pourraient apporter de réelles avancées dans la détermination des forces sélectives responsables de la diversité structurelle et de l'évolution des génomes cytoplasmiques.

II/ De la phylogénie à la phylogénomique

1. La notion de phylogénie

La notion de phylogénie repose sur la théorie de l'évolution développée par Charles Darwin (1859) dans *L'origine des espèces* qui stipule que l'ensemble des organismes vivants dérivent d'un ancêtre commun et sont génétiquement apparentés. Un arbre phylogénétique est une représentation graphique (sous forme d'arbre) et simplifiée des liens de parenté entre des espèces, des organismes ou des gènes et de leur évolution à partir d'un ancêtre commun. Une reconstruction phylogénétique repose sur l'identification de caractères homologues (i.e. descendant d'un ancêtre commun) partagés par différents organismes, d'une part, et sur l'application de méthodes d'inférence phylogénétique, d'autre part. Jusqu'à la moitié du XX^{ème} siècle, les reconstructions phylogénétiques se basaient sur des caractères homologues morpho-anatomiques. La découverte de la structure de l'ADN par Watson & Crick (1953) et le développement des techniques moléculaires dans les années 1970 ont révolutionné les méthodes de reconstructions phylogénétiques. Les reconstructions phylogénétiques se basent désormais sur des séquences d'ADN ou de protéines. Ceci permet notamment de limiter la subjectivité liée à l'identification de caractères morpho-anatomiques et d'augmenter considérablement le nombre de caractères homologues.

2. Séquençage d'ADN de nouvelle génération (NGS)

Le séquençage Sanger était la méthode de séquençage d'ADN la plus utilisée en biologie (Sanger *et al.* 1977), jusqu'à l'arrivée du séquençage de nouvelle génération (dit NGS : Next-Generation Sequencing). Le séquençage Sanger présente certaines limites. D'une part, il ne permet pas de séquencer des fragments d'ADN dont la longueur est supérieure à un kb. L'utilisation d'une telle technique est inenvisageable pour séquencer des génomes entiers de plusieurs milliers de paires de bases (pb), qui nécessiteraient de multiples séquençages. Une autre limite du séquençage Sanger est son coût. A environ 1 $ par kb, le séquençage d'un génome d'un gigabase (Gb) avec une couverture de séquençage de 10× couterait 10,000,000 $. Cela signifie que la plupart des laboratoires de recherche seraient contraints de séquencer des génomes à l'échelle de la mégabase.

Le séquençage de nouvelle génération (NGS), dit aussi séquençage à haut débit, a révolutionné la recherche en génomique, en médecine, en biologie moléculaire et évolutive ainsi que dans bien d'autres domaines. Les technologies NGS peuvent générer en parallèle des millions voire des milliards de petits fragments d'ADN d'une centaine de paires de bases en un seul cycle pour un coût relativement faible. Alors que le séquençage du génome humain (~3 Gb) avec la méthode Sanger a coûté 3 milliards de dollars et s'est effectuée entre 1990 et 2003 avec la collaboration de plusieurs laboratoires, il est désormais possible de séquencer le génome humain avec les techniques NGS en une journée pour un coût de 1,000 $. Les premières technologies NGS sont apparues en 2005 avec le séquençage Roche 454. Depuis, de nombreuses technologies NGS gagnant en efficacité et en rapidité pour un moindre coût ont vu le jour et se font concurrence sur le marché du séquençage. On distingue deux

méthodes majeures de séquençage : la méthode globale si on veut séquencer l'ensemble du génome et la méthode ciblée si on veut séquencer uniquement certaines régions du génome, par exemple uniquement les régions codantes. La technologie Illumina, qui utilise une méthode de séquençage globale du génome, est actuellement la technologie de séquençage la plus couramment utilisée sur le marché du séquençage à haut débit.

Principe du séquençage aléatoire avec la technologie Illumina (Metzker 2009). La technologie Illumina utilise un séquençage aléatoire (ou séquençage "shotgun") qui permet de séquencer des fragments d'ADN au hasard, qui sont ensuite assemblés pour reconstruire de longs fragments appelés contigs (dont la longueur est supérieure à 1 kb), voire des génomes entiers. Cette méthode de séquençage est dite aléatoire car elle ne permet pas de cibler des régions spécifiques à séquencer. Le séquençage s'effectue par fragmentation au hasard du génome en de multiples petites séquences d'ADN. La fragmentation est réalisée par sonication (i.e. des ultra-sons envoyés à une certaine fréquence découpent l'ADN en fragments de taille donnée) ou par des enzymes de restriction qui découpent l'ADN au niveau de certains motifs [e.g. RAD-Seq (Davey & Blaxter 2010) ou GBS (Elshire *et al.* 2011)]. Des adaptateurs sont ensuite fixés à chacune des extrémités des fragments, ce qui permet l'immobilisation des fragments sur un support solide par hybridation d'une amorce avec l'un des deux adaptateurs (Fig. 7A). La deuxième extrémité de chacun des fragments s'hybride ensuite avec l'amorce complémentaire pour former un pont. Le brin complémentaire au fragment est alors synthétisé. Cette opération sera répétée un grand nombre de fois afin d'obtenir plusieurs centaines de millions de fragments. Au cours du séquençage, l'ADN polymérase va incorporer des nucléotides associés à des fluorochromes spécifiques à chacun des quatre nucléotides (Fig. 7B). Les signaux lumineux émis par les fluorochromes sont détectés par le séquenceur et traduits en nucléotides. L'incorporation d'un nucléotide fluorescent dans la séquence est réversible. Ceci permet à la réaction de séquençage de continuer, jusqu'à l'incorporation d'un nouveau nucléotide fluorescent dans la séquence en cours de synthèse. Les fragments obtenus (appelés "reads") sont de petites tailles (quelques dizaines ou centaines de bases). L'étape suivante est l'assemblage qui consiste à reconstituer des contigs, voire le génome entier, à partir de ces petits fragments.

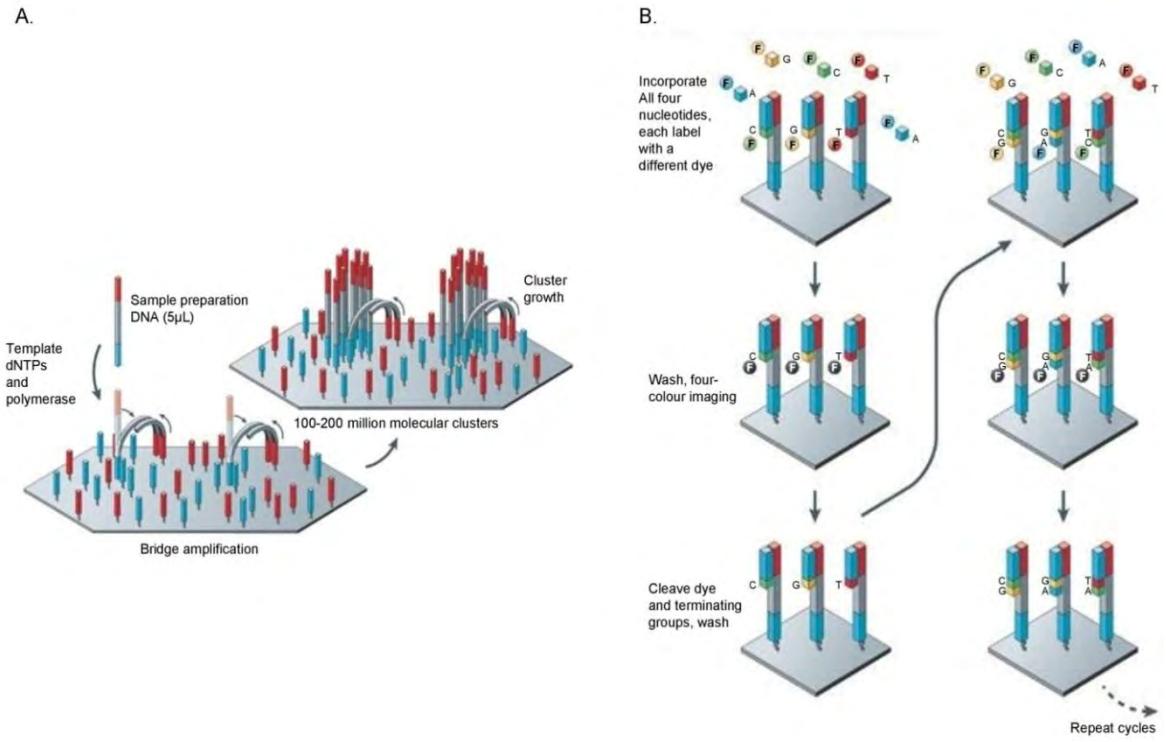


Fig. 7 Principe du séquençage par la technologie Illumina. **A.** Après fractionnement du génome en petits fragments d'ADN, chaque fragment (préalablement dénaturé en simple brin) fixe ses extrémités sur un support solide en s'hybridant avec des amorces, pour former un pont. Le brin complémentaire de chacun des fragments est alors synthétisé. **B.** La synthèse du brin complémentaire s'effectue en incorporant dans la séquence un des quatre nucléotides associés à des fluorochromes réversibles. Le fluorochrome associé à un nucléotide émet un signal lumineux, qui est directement détecté par le séquenceur et traduit en nucléotide. Le fluorochrome est ensuite clivé et la synthèse du fragment peut continuer. Au cycle suivant, un des quatre nouveaux nucléotides associés à des fluorochromes est incorporé à la séquence et le deuxième nucléotide est lu par le séquenceur. Les cycles de séquençage sont ainsi répétés, jusqu'à la synthèse complète du brin complémentaire au fragment séquencé. [Figure extraite de Metzker (2009)]

Assemblage. Le séquençage d'ADN génère une importante redondance de séquences, un nucléotide pouvant être séquencé plusieurs fois. La partie du génome ciblée, voire le génome entier, peut être séquencée aléatoirement *x* fois en de courts fragments, c'est ce qu'on appelle la profondeur de séquençage. De plus, les fragments obtenus peuvent se chevaucher du fait de la nature aléatoire du séquençage. C'est le chevauchement de deux fragments contigus qui permet de constituer un plus grand fragment (i.e. un contig). De chevauchement en chevauchement, les fragments s'assemblent en un contig de plus en plus long, jusqu'à obtenir la partie du génome ciblée, voire le génome complet. L'assemblage est d'autant plus facile que la profondeur de séquençage est élevée. On distingue deux approches d'assemblage : (i) l'assemblage avec référence, lorsque l'assemblage est initié par la projection des petits fragments contre une séquence (ou un génome) similaire de référence (e.g. d'une espèce proche), et (ii) l'assemblage *de novo*, utilisé quand aucune séquence similaire n'est disponible dans les bases de données. La combinaison de ces deux approches peut également être utilisée pour assembler une séquence.

Limites. L'une des limites majeures du séquençage à haut débit concerne l'étape d'assemblage du génome. La présence de régions répétées dans le génome est souvent à l'origine de cassures dans l'assemblage. Les régions répétées entraînent l'alignement de mêmes "reads" au niveau de différentes zones du génome. Ces "reads" ne s'alignent pas forcément à la bonne place dans le génome, ce qui peut entraîner une reconstruction erronée des zones adjacentes à ces répétitions. La difficulté à assembler ces zones répétées dépend essentiellement de la taille des "reads". L'assemblage de ces zones est plus aisé lorsque les "reads" comprennent la région répétée et une partie des séquences adjacentes à cette région. Néanmoins, d'autres stratégies d'assemblage restent à être développées afin d'assembler de larges régions répétées (e.g. transposons, duplications, gènes en tandem, répétitions >1 kb présents dans le mitogénome de certaines plantes).

Une autre limite concerne la profondeur de séquençage qui peut être insuffisante pour assembler certaines zones du génome. Une profondeur de séquençage minimale est requise afin de couvrir l'ensemble du génome et de distinguer une erreur de séquençage d'une substitution nucléotidique. Le coût du séquençage étant d'autant plus élevé que la profondeur de séquençage est grande, il convient d'estimer une profondeur de séquençage minimale et optimale. La détermination de la profondeur de séquençage minimale dépend de différents facteurs tels que le taux d'erreurs de la méthode de séquençage, les algorithmes d'assemblage utilisés, la longueur des "reads" ou encore le contenu en régions répétées du génome. Par ailleurs, la profondeur de séquençage dépend de la qualité et de la quantité d'ADN utilisé pour le séquençage. Le séquençage d'ADN de spécimens provenant d'herbiers ou de collections de musées peut être difficile du fait de la dégradation de l'ADN ou de la contamination avec d'autres organismes tels que des bactéries ou des moisissures.

Enfin, le séquençage haut débit a permis d'enrichir considérablement nos connaissances quant à la structure et l'évolution des génomes d'organites. Depuis le développement des technologies NGS en 2005, le nombre de plastomes complets disponibles dans les bases de données publiques (e.g. GenBank) ne cesse d'augmenter. En août 2017, on compte près de 1455 plastomes complets de plantes terrestres dans GenBank. Les assembleurs sont de plus en plus performants et spécialisés et permettent de reconstruire relativement facilement des molécules circulaires contenant des "Inverted Repeats" telles que les plastomes. Néanmoins, le mitogénome des plantes reste très compliqué à assembler, du fait de sa structure complexe. Il n'existe aujourd'hui aucune méthode d'assemblage des mitochondries de plantes standardisée et automatisée. Seuls 210 mitogénomes de plantes terrestres sont actuellement disponibles dans GenBank. La présence de nombreuses séquences répétées et de multiples régions dérivées du plastome requiert une intervention manuelle pour allonger et relier les contigs. L'assemblage manuel du mitogénome est cependant une étape laborieuse. Avec les nouvelles technologies de séquençage produisant des "reads" toujours plus longs, on peut imaginer que l'assemblage du mitogénome des plantes sera plus aisé dans les années futures.

3. Inférences phylogénétiques

3.1. Identification de sites homologues

La première étape des inférences phylogénétiques est l'identification de caractères homologues à partir des séquences d'ADN. En supposant une corrélation entre homologie et similarité, cette étape implique une comparaison des similarités entre les séquences des différents taxons. La comparaison entre séquences s'effectue en alignant les séquences entre elles. L'objectif de l'alignement de séquences est de déterminer la meilleure hypothèse d'homologie de chacun des sites nucléotidiques des séquences. Depuis la divergence avec leur ancêtre commun, les séquences ont subi des substitutions nucléotidiques, des délétions ou des insertions de séquences, des duplications voire des réarrangements de séquences. Le principe de l'alignement est d'ajouter des espaces dans les séquences d'ADN afin d'aligner les sites homologues des séquences (Fig. 8). On obtient une matrice dont les lignes correspondent aux séquences de chacun des taxons et les colonnes correspondent aux sites nucléotidiques homologues entre les différentes séquences. Chaque colonne contient les différents états de caractères du site homologue. L'alignement de sites homologues est d'autant plus difficile que les séquences d'ADN ont divergé.

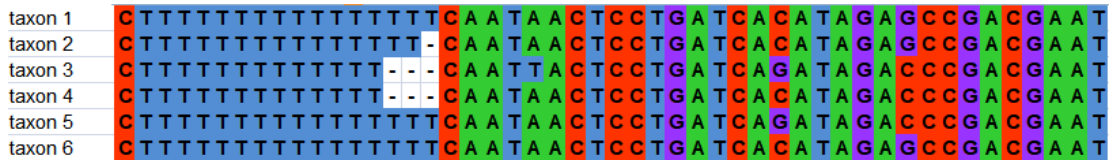


Fig. 8 Exemple d'un alignement de séquences de six taxons. Chaque colonne contient les différents états de caractères du site homologue. L'ajout d'espaces dans les séquences permet d'aligner les sites homologues des différentes séquences.

3.2. Méthodes d'inférence phylogénétique

La seconde étape d'une reconstruction phylogénétique consiste à appliquer une méthode d'inférence phylogénétique à partir de l'alignement de séquences. Le taux d'évolution des séquences n'étant pas forcément constant au cours du temps, le simple calcul des différences entre séquences ne permet pas toujours de déterminer leur degré de proximité phylogénétique. Différentes méthodes statistiques sont utilisées pour reconstruire l'arbre phylogénétique qui décrit au mieux l'évolution et les relations phylogénétiques de la matrice de séquences. On distingue quatre principales méthodes d'inférence phylogénétique : les méthodes basées sur les distances, la méthode de maximum de parcimonie, les méthodes de maximum de vraisemblance et les méthodes bayésiennes (Holder & Lewis 2003 ; Delsuc *et al.* 2005).

Méthodes basées sur les distances. Les méthodes basées sur les distances convertissent la matrice de séquences en une matrice de distances, représentant les distances évolutives entre chaque paire de séquences. Ces méthodes ne tiennent pas en compte des possibles cas

d'homoplasie, i.e. l'apparition indépendante de caractères similaires suite à des substitutions multiples. Ce type de méthode s'applique uniquement à des séquences ayant peu divergé.

Méthode de maximum de parcimonie. La méthode de maximum de parcimonie sélectionne l'arbre phylogénétique qui présente le moins de changements de caractères pour expliquer les séquences observées. Un des désavantages de cette méthode est qu'elle ne prend pas en compte les possibles variations du taux d'évolution le long des branches de l'arbre.

Méthodes de maximum de vraisemblance. Les méthodes de maximum de vraisemblance sélectionnent l'arbre qui a la probabilité la plus élevée de produire les séquences observées. Ces méthodes permettent d'incorporer des modèles d'évolution de séquences. Ce type de méthodes nécessite de lourds calculs informatiques, notamment pour évaluer la robustesse de l'arbre phylogénétique, ce qui peut être un facteur limitant pour de grands jeux de données (>100 séquences).

Ces trois types de méthodes permettent de générer un arbre phylogénétique optimal mais la robustesse des relations phylogénétiques entre taxons n'est pas évaluée. La méthode la plus courante pour évaluer statistiquement la robustesse des branches de l'arbre est le "bootstrap". Le "bootstrap" est une méthode de ré-échantillonnage qui permet de créer des pseudo-matrices de séquences de même taille que la matrice initiale. Il s'agit d'un tirage aléatoire avec remise des différentes positions de l'alignement, répété x fois. Pour chaque pseudo-matrice créée, un arbre phylogénétique est reconstruit. Le soutien d'une branche interne de l'arbre initial est estimé à partir du nombre de fois où cette branche interne est retrouvée dans les x arbres phylogénétiques reconstruits à partir des pseudo-matrices. Si la branche interne est retrouvée dans 100% des arbres, la valeur de "bootstrap" est de 100. Le calcul des valeurs de "bootstrap" étant relativement lent pour les grands jeux de données, les méthodes bayésiennes peuvent constituer une alternative. Les méthodes bayésiennes calculent directement les valeurs de "bootstrap" en estimant des distributions de probabilité postérieure.

Méthodes bayésiennes. Les méthodes bayésiennes sélectionnent l'arbre phylogénétique qui maximise la probabilité postérieure de l'arbre. La probabilité postérieure d'un arbre phylogénétique correspond à la probabilité que cet arbre soit vrai sachant les données (i.e. les séquences d'ADN). Alors que les méthodes de maximum de vraisemblance cherchent la probabilité d'observer des données D sachant un arbre phylogénétique T [$Pr(D|T)$], les méthodes bayésiennes cherchent la probabilité d'un arbre phylogénétique T sachant les données D [$Pr(T|D)$]. La probabilité postérieure $Pr(T|D)$, calculée selon le théorème de Bayes, est une fonction de la vraisemblance des données D sachant l'arbre T [$Pr(D|T)$] et de la probabilité *a priori* de l'arbre phylogénétique $Pr(T)$. Une des difficultés des méthodes bayésiennes est le choix *a priori* de la distribution de probabilité des différents paramètres. La spécification de la distribution *a priori* nécessite de prendre en compte l'ensemble des connaissances sur les paramètres avant le recueil des données (e.g. études antérieures). Une distribution uniforme (i.e. toutes les valeurs du paramètre sont possibles) est souvent spécifiée pour les paramètres où aucune connaissance *a priori* n'est disponible.

Les reconstructions phylogénétiques permettent d'éclairer l'histoire évolutive de groupes d'organismes vivants et de tester des hypothèses spatio-temporelles quant à la divergence de ces groupes. Par ailleurs, les arbres phylogénétiques peuvent être utilisés dans l'étude des changements de pressions de sélection exercées sur les génomes au cours de leur évolution.

4. Apports de la phylogénomique dans l'étude des génomes cytoplasmiques

4.1. La phylogénomique

La phylogénomique correspond à l'étude des relations phylogénétiques entre taxons à partir de données génomiques. Avant l'arrivée des technologies de séquençage à haut débit, la plupart des études phylogénétiques reposaient sur un ou plusieurs gènes. Il était souvent difficile de résoudre les relations phylogénétiques des taxons ayant divergé récemment ou ayant subi une radiation récente. Mais depuis l'émergence des technologies NGS, la phylogénomique permet d'étudier les relations de parenté entre espèces à partir de génomes complets (ou de la concaténation de multiples séquences d'ADN codantes ou non-codantes). L'analyse de génomes complets augmente considérablement le nombre de caractères informatifs, ce qui permet de clarifier le signal phylogénétique et de résoudre certaines polytomies. Par ailleurs, l'étude de la structure du génome (contenu génique, ordre des gènes, perte/gain d'introns, taille du génome, composition nucléotidique) peut permettre de résoudre certains problèmes phylogénétiques (Delsuc *et al.* 2005).

Les reconstructions phylogénomiques sont basées sur l'alignement multiple de séquences, i.e. alignement de plusieurs gènes ou régions orthologues. On distingue deux approches phylogénomiques d'alignement de séquences : l'approche super-matrice et l'approche super-arbre (Holder & Lewis 2003 ; Delsuc *et al.* 2005). L'approche super-matrice (la plus utilisée) consiste à concaténer plusieurs gènes orthologues pour créer une seule matrice, sur laquelle sont appliquées les méthodes phylogénétiques standard. Les séquences de gènes absents chez certains taxons sont notées comme des données manquantes. Les données manquantes n'interfèrent pas avec la reconstruction phylogénétique si elles représentent moins de 20-25% du jeu de données. La tolérance d'un certain seuil de données manquantes permet d'intégrer un grand nombre de taxons dans la matrice. La deuxième approche est l'approche super-arbre qui consiste à reconstruire un arbre individuellement pour chaque gène puis à combiner l'ensemble de ces arbres individuels en un seul super-arbre.

4.2. La phylogénomique à partir de données génomiques cytoplasmiques

La reconstruction d'arbres phylogénétiques à partir de données génomiques cytoplasmiques a permis d'éclairer l'histoire évolutive de nombreux groupes de plantes. Dans le cas d'une transmission maternelle des organites, l'utilisation de génomes cytoplasmiques dans les études moléculaires, et particulièrement dans les analyses phylogéographiques, présente plusieurs intérêts : (i) les génomes cytoplasmiques reflètent la dispersion des graines; (ii) leur taille de

population efficace étant divisée par deux par rapport aux génomes diploïdes, les génomes cytoplasmiques sont davantage soumis à la dérive génétique, et (iii) la dispersion des génomes cytoplasmiques par les graines est limitée, les graines étant généralement dispersées à des distances plus courtes que le pollen. Ces deux dernières caractéristiques conduisent généralement à des patrons de différentiation inter-population très marqués.

Les arbres phylogénétiques de groupes de plantes sont essentiellement basés sur le plastome et reflètent uniquement l'histoire évolutive de ce génome. Peu d'études se sont intéressées à l'évolution d'un groupe de plantes en utilisant des données mitogénomiques. La faible utilisation du mitogénome pour reconstruire des phylogénies moléculaires est due à son évolution lente et à sa complexité d'assemblage. Par ailleurs, les mitochondries et les chloroplastes sont transmis maternellement chez la plupart des plantes et sont supposés être en fort déséquilibre de liaison. Par conséquent, les mitochondries et chloroplastes pourraient partager la même histoire évolutive. Pourtant, l'intérêt potentiel des données mitogénomiques en phylogénie pourrait être sous-estimé. D'une part, le mitogénome pourrait apporter davantage de sites informatifs et permettre de résoudre certaines relations phylogénétiques non-résolues avec la seule utilisation du plastome. D'autre part, l'utilisation de données mitogénomiques pourrait être pertinente dans le cas d'une transmission contrastée des mitochondries et des chloroplastes. Aussi, des processus tels que des événements d'hybridation inter-spécifique peuvent créer une incongruence phylogénétique entre les génomes cytoplasmiques et le génome nucléaire, d'où l'intérêt de comparer les reconstructions phylogénétiques basées sur des marqueurs chloroplastiques, mitochondriaux et nucléaires.

4.3. La phylogénomique pour tester les changements de pressions de sélection au sein des génomes cytoplasmiques

La reconstruction de l'histoire évolutive de groupes de plantes à partir de données génomiques permet également d'étudier les changements de pressions de sélection qui s'opèrent au sein des génomes. L'approche la plus couramment utilisée pour caractériser les régimes de sélection impliqués dans l'évolution des gènes codant pour des protéines est l'estimation du ratio entre taux de substitutions non-synonymes et taux de substitutions synonymes (d_N/d_S, noté omega ω). Le ratio ω indique une sélection purifiante ($\omega < 1$), neutre ($\omega = 1$) ou positive ($\omega > 1$). L'estimation du ratio ω a notamment été utilisée sur des gènes chloroplastiques dans l'étude des changements de pressions de sélection au cours de l'évolution de la photosynthèse C_4 chez les Poaceae. Une évolution convergente du gène *rbcL* codant pour la large sous-unité de la RubisCO a été mise en évidence au sein des lignées de graminées à photosynthèse C_4 et plusieurs changements parallèles d'acides aminés ont pu être identifiés comme des déterminants majeurs de l'affinité pour le CO_2 et l'activité catalytique de l'enzyme (Christin *et al.* 2008). Par ailleurs, Piot *et al.* (2018) ont récemment montré qu'un tiers des gènes chloroplastiques sont sous sélection positive dans un clade de graminées (PACMAD), mais les pressions environnementales induisant ces changements sont encore mal identifiés. Une sélection positive a aussi été détectée chez plusieurs gènes chloroplastiques présentant une accélération du taux d'évolution dans certaines lignées. En particulier, une divergence élevée

associée à de la sélection positive a été détectée chez les gènes *clpP*, *ycf1*, *ycf2* et chez plusieurs gènes ribosomiques au sein de certaines lignées de *Silene* (Sloan *et al.* 2012), acacias (Mensous *et al.* 2017), Campanulaceae (Barnard-Kubow *et al.* 2014), Geraniaceae (Guisinger *et al.* 2008) et chez *Œnothera* (Erixon & Oxelman 2008).

Les nouvelles technologies de séquençage associées à des algorithmes de traitement de données génomiques toujours plus performants ont permis de réelles avancées dans l'étude de la diversité structurelle et l'évolution des génomes cytoplasmiques. Néanmoins, l'évolution concertée des génomes cytoplasmiques et la diversité structurelle des mitogénomes en particulier restent encore mal connues. Ce travail de thèse s'est intéressé à une famille de plantes à fleurs, les Oleaceae, qui présentent différents intérêts dans l'étude des génomes cytoplasmiques.

III/ Modèle d'étude : la famille de l'olivier, les Oleaceae

1. Caractéristiques générales

La famille des Oleaceae est une famille de plantes qui comprend plus de 600 espèces d'arbres et d'arbustes répartis en 25 genres (Wallander & Albert 2000 ; Green 2004). On y retrouve des espèces présentant un intérêt économique, culturel ou ornemental telles que l'olivier (*Olea europaea*), le frêne (*Fraxinus* spp.), le jasmin (*Jasminum* spp.), le forsythia (*Forsythia* spp.), le troène (*Ligustrum* spp.) ou encore le lilas (*Syringa* spp.). Les espèces d'Oleaceae sont distribuées sur l'ensemble des continents, à l'exception de l'Antarctique, et sont retrouvées dans des régions aussi bien tropicales que tempérées et à des altitudes très variables (Wallander & Albert 2000). Beaucoup d'espèces ou de lignées d'Oleaceae sont relativement difficiles à échantillonner du fait de leur rareté. C'est notamment le cas pour les espèces présentes dans des régions tropicales ou subtropicales isolées (e.g. Madagascar, Papouasie, Andes). En conséquence, les études phylogénétiques de la famille restent encore non exhaustives.

L'une des synapomorphies de la famille des Oleaceae est la présence de feuilles opposées (à l'exception de quelques *Noronhia* et *Jasminum*), simples ou composées. Les Oleaceae, et particulièrement la tribu des *Oleeae*, sont caractérisés par une grande diversité florale (Fig. 9), associée à divers modes de pollinisation [i.e. pollinisation par le vent (anémophile) ou par les insectes (entomophile), voire les deux (ambophile)] et de dispersion des graines (Wallander & Albert 2000). On y retrouve divers types de graines tels que des capsules, des schizocarpes, des samares, des baies ou encore des drupes. Les Oleaceae sont également caractérisés par une large diversité de systèmes de reproduction, allant de l'hermaphrodisme à la dioécie en passant par des états intermédiaires tels que l'androdioécie, la trioécie ou la polygamie.

Fig. 9 Aperçu de la diversité florale chez la famille des Oleaceae. **a.** *Noronhia densiflora*; **b.** *Fraxinus ornus*; **c.** *Jasminum nudiflorum*; **d.** *Ligustrum vulgaris*; **e.** *Forsythia* × *intermedia*; **f.** *Syringa vulgaris*. [**a.** : photographie de C. Birkinshaw extraite de Hong-Wa (2016); **b.** à **f.** : photographies libres de droits*]

2. Systématique

La famille des Oleaceae appartient à l'ordre des Lamiales. A partir de données moléculaires, Wallander & Albert (2000) ont révisé la systématique des Oleaceae en divisant la famille en cinq tribus : les *Myxopyreae*, *Fontanesieae*, *Forsythieae* et *Jasmineae* qui constituent les "lignées diploïdes", et les *Oleeae* qui sont issus d'un ancêtre allopolyploïde [i.e. paléopolyploïde : ancien polyploïde dont le génome ségrége comme un diploïde; Taylor (1945)]. Cet événement de polyploïdie est associé à une duplication du nombre de chromosomes, passant de $x = 11$-14 chez les "lignées diploïdes" à $x = 23$ chez les *Oleeae*. La tribu des *Oleeae* comprend près de la moitié des espèces d'Oleaceae et se divise en quatre sous-tribus: *Ligustrinae*, *Schreberinae*, *Fraxininae* et *Oleinae*.

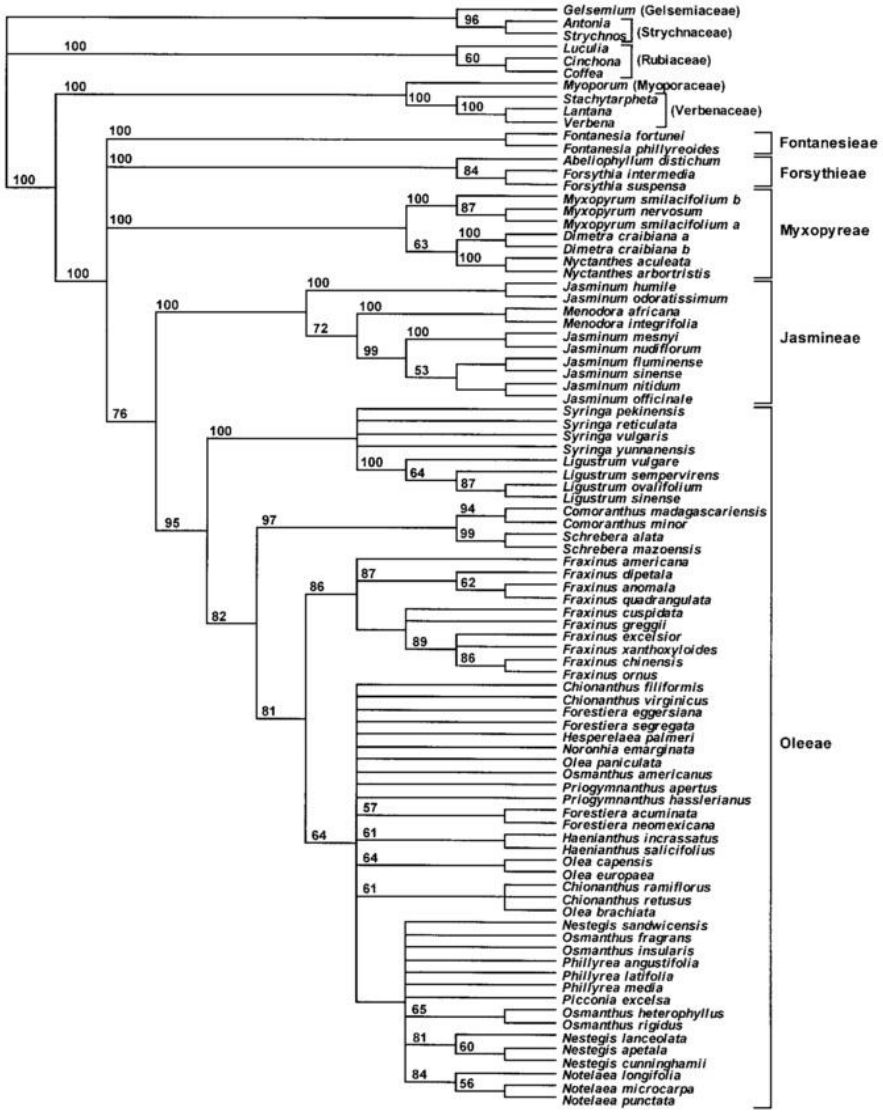
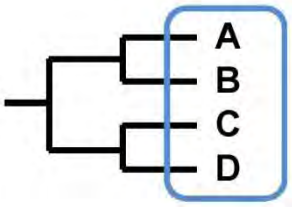


Fig. 10 Arbre phylogénétique de la famille des Oleaceae reconstruite par Wallander & Albert (2000), basé sur deux régions chloroplastiques non-codantes (l'intron de *rps16* et l'espace inter-génique *trnL-trnF*). [Figure extraite de Wallander & Albert (2000)]

NOTIONS DE GROUPE MONOPHYLETIQUE, POLYPHYLETIQUE ET PARAPHYLETIQUE

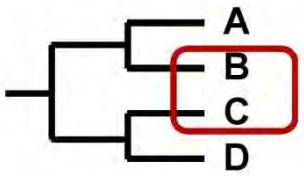
Groupe monophylétique :

comprend tous les descendants d'un ancêtre commun



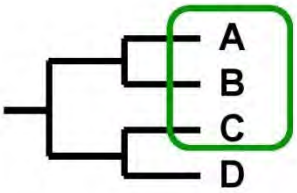
Groupe polyphylétique :

ne comprend pas tous les descendants d'un ancêtre commun



Groupe paraphylétique :

comprend tous les descendants d'un ancêtre commun
sauf une lignée
(cas spécifique d'un groupe polyphylétique)



Les travaux de Wallander & Albert (2000), basés sur deux marqueurs chloroplastiques non-codants (l'intron de *rps16* et l'espace inter-génique *trnL-trnF*), ont permis de clarifier la systématique des Oleaceae à l'échelle des tribus et des sous-tribus (Fig. 10). Néanmoins, la plupart des relations phylogénétiques intra-génériques ne sont pas résolues. Plusieurs études utilisant des marqueurs nucléaires (ETS et ITS) et/ou chloroplastiques se sont attachées à résoudre les relations phylogénétiques au sein de différents genres tels que *Olea* (Besnard *et al.* 2009), *Noronhia* et *Chionanthus* (Hong-Wa & Besnard 2013, 2014), *Fraxinus* (Jeandroz *et al.* 1997 ; Wallander 2008 ; Arca *et al.* 2012 ; Hinsinger *et al.* 2013, 2014), *Syringa* et *Ligustrum* (Kim & Jansen 1998 ; Li *et al.* 2012 ; Lendvay *et al.* 2016) et *Forestiera*, *Priogymnanthus* et *Hesperelaea* (Zedane *et al.* 2016). Ces différentes études ont montré que certains genres (*Olea*, *Chionanthus*, *Syringa* ou *Osmanthus*) ne sont pas des groupes monophylétiques et nécessiteraient une révision taxonomique. Par ailleurs, des genres tels que *Nestegis*, *Notelaea*, *Haenianthus* ou *Picconia* ne sont que très peu étudiés.

Les travaux de thèse de Zedane (2016) ont permis d'éclairer les relations phylogénétiques au sein des différents genres de la tribu des *Oleeae*. Zedane et ses collaborateurs ont reconstruit une phylogénie chloroplastique des *Oleeae* basée sur les plastomes complets de 82 accessions couvrant l'ensemble des genres d'*Oleeae*. La phylogénie

obtenue est fortement soutenue avec des valeurs de soutien supérieures à 90% pour la plupart des nœuds. Ces analyses ont permis de confirmer la monophylie des quatre sous-tribus composant la tribu des *Oleeae* (*Ligustrinae*, *Schreberinae*, *Fraxininae* et *Oleinae*). Par ailleurs, de nombreuses paraphylies et polyphylies ont été détectées à l'échelle du genre, comme c'est le cas pour les genres *Chionanthus*, *Osmanthus*, *Nestegis, Olea* ou encore *Syringa*.

3. Transmission des organites chez les Oleaceae

La transmission des chloroplastes et des mitochondries chez les Oleaceae est peu connue. Néanmoins, une transmission maternelle stricte des organites a été mise en évidence dans le genre *Syringa* et chez l'olivier, alors qu'une potentielle transmission biparentale est supposée chez deux lignées indépendantes (les *Ligustrinae* et les *Jasmineae*).

3.1. Transmission maternelle des organites chez le lilas commun et chez l'olivier

Corriveau & Coleman (1988) ont détecté par microscopie à fluorescence une transmission maternelle des organites chez *Syringa vulgaris*. Aucun chloroplaste, ni aucune mitochondrie n'a été observé dans la cellule germinative des grains de pollen de *S. vulgaris.* Liu *et al.* (2004) ont confirmé ces observations et ont également détecté par microscopie à fluorescence une transmission maternelle des organites chez six autres accessions de *Syringa* appartenant aux séries *Syringa* et *Pinnatifoliae* (sous-genre *Syringa*). Par ailleurs, Besnard *et al.* (2000) ont montré à partir de croisements que les organites sont transmis maternellement dans le complexe de l'olivier (*Olea europaea*) et qu'il existe un fort déséquilibre de liaison entre les polymorphismes chloroplastiques et mitochondriaux chez cette espèce (Besnard *et al.* 2002). Une stérilité mâle cytoplasmique a également été mise en évidence chez une des lignées (lignée E3) d'*O. e.* subsp. *europaea* (Besnard *et al.* 2000).

3.2. Transmission potentiellement biparentale occasionnelle chez deux lignées indépendantes : les *Ligustrinae* et les *Jasmineae*

Liu *et al.* (2004) ont observé par microscopie à fluorescence la présence de chloroplastes et de mitochondries dans la cellule germinative de grains de pollen de 15 espèces de la sous-tribu *Ligustrinae* (*Ligustrum* et *Syringa* appartenant au sous-genre *Ligustrina* et aux séries *Villosae* et *Pubescentes* du sous-genre *Syringa*). Par une technique d'immuno-marquage aux particules d'or, ils ont détecté une diminution du nombre de mitochondries et une augmentation du nombre de chloroplastes au cours de la maturation de la cellule germinative du pollen. Des mitochondries et des chloroplastes ont également été observés dans la cellule germinative mature de grains de pollen de deux espèces de jasmin phylogénétiquement éloignés ; *Jasminum officinale* et *J. nudiflorum* (tribu des *Jasmineae* ; Sodmergen *et al.* 1998). Ces observations laissent supposer le passage occasionnel de chloroplastes dans les grains de pollen de certaines espèces de *Ligustrinae* et de chloroplastes et de mitochondries dans les grains de pollen de *Jasminum*.

IV/ Plan de la thèse

Chez les plantes à fleurs, les chloroplastes et les mitochondries sont les seuls organites à posséder leur propre génome. Les génomes chloroplastiques et mitochondriaux ont été réduits drastiquement depuis leur origine endosymbiotique mais ont conservé des gènes qui jouent des rôles essentiels dans le métabolisme des cellules. Bien que sous contrôle du génome nucléaire, les génomes chloroplastiques et mitochondriaux présentent une large diversité structurelle et des dynamiques évolutives variées. Le génome mitochondrial, en particulier, possède des caractéristiques génomiques complexes et diversifiées. Le séquençage à haut débit a permis d'élargir nos connaissances sur la diversité structurelle des mitogénomes et sur leur évolution. Néanmoins, aucune méthode standardisée et automatisée n'a été jusqu'à présent développée pour assembler les mitogénomes des plantes à fleurs, du fait de leur structure complexe. C'est une raison pour laquelle peu de mitogénomes complets sont aujourd'hui disponibles dans les banques de données publiques. Il est donc nécessaire d'enrichir les données mitogénomiques afin de mieux appréhender leur diversité et leur dynamique évolutive.

Dans un premier temps, cette étude s'intéresse à la diversité structurelle et à l'évolution des génomes cytoplasmiques au sein de la famille des Oleaceae, et plus particulièrement à l'échelle d'une espèce, l'olivier (*Olea europaea*). Une première étape de ce travail a été d'assembler les premiers mitogénomes complets d'Oleaceae, sur six accessions d'*O. europaea* et deux autres taxa d'*Oleinae* utilisés comme groupes externes : *Chionanthus rupicolus* et un genre éteint*, Hesperelaea*.

Le CHAPITRE I décrit l'assemblage et la structure du mitogénome d'*Hesperelaea*. Le genre *Hesperelaea* est monotypique et a été collecté une seule fois, il y a 140 ans sur l'île de Guadalupe (Mexique). Nous montrons que les données de séquençage shotgun peuvent permettre de reconstruire des mitogénomes de plantes, même à partir d'ADN très dégradé provenant d'herbiers.

Dans le CHAPITRE II, nous avons tout d'abord décrit la diversité structurelle et l'évolution des mitogénomes chez les *Oleinae*. L'espèce *O. europaea* est particulièrement intéressante pour l'étude des mitogénomes du fait qu'une des lignées (la lignée E3) est associée à une stérilité mâle cytoplasmique. En comparant les mitogénomes d'*O. europaea*, nous avons cherché à savoir si il existait des variations de taille et d'organisation des mitogénomes à l'échelle intra-spécifique. Nous avons également tenté d'identifier des cadres de lecture ouverts (ORFs) spécifiques à la lignée E3, susceptibles d'être impliqués dans la stérilité mâle. Dans un second temps, nous avons testé l'apport de l'utilisation de données mitogénomiques dans la résolution des relations phylogénétiques à l'échelle d'une espèce, *O. europaea*, puis nous avons finalement testé la congruence des reconstructions phylogénétiques d'Oleaceae basées sur des plastomes complets et sur des régions mitochondriales codantes et non-codantes (i.e. CDS et introns).

Ce travail de thèse s'est ensuite intéressé aux conséquences génomiques d'une transmission paternelle occasionnelle des organites ("paternal leakage") chez les Oleaceae

(CHAPITRE III). A partir de croisements ou d'observations au microscope à fluorescence, du "paternal leakeage" a été décrit chez quelques espèces. Il est supposé que le "paternal leakeage" entraînerait de la recombinaison sexuelle, ce qui permettrait de réduire le fardeau de mutations délétères des génomes cytoplasmiques. Ces hypothèses n'ont été que rarement testées et les conséquences génomiques du "paternal leakeage" restent peu connues. La famille des Oleaceae semble être un excellent cas d'étude puisque des cas d'une potentielle transmission des organites par la voie paternelle ont été décrits chez deux lignées indépendantes : les *Jasmineae* et les *Ligustrinae*. Nous nous sommes demandé si le passage occasionnel d'organites par la voie mâle peut entraîner des réarrangements du génome ou des variations de taux d'évolution. Nous avons également cherché à déterminer si le passage occasionnel d'organites dans le pollen peut modifier les forces sélectives qui s'exercent sur les gènes cytoplasmiques.

En marge de cette thèse, j'ai aussi collaboré avec Mohamed Mensous, doctorant à l'université d'Ouargla (Algérie), afin d'étudier la diversité et l'évolution des plastomes chez les acacias sahariens. Dans ce cadre, j'ai analysé les changements de pressions de sélection chez une autre lignée (tribu Ingeae + *Acacia* s.s.) présentant une accélération du taux d'évolution et une extension des "Inverted Repeats". Cette étude a fait l'objet d'un article présenté en Annexe IV :

Mensous M*, Van de Paer C*, Manzi S *et al.* (2017) Diversity and evolution of plastomes in Saharan mimosoids: potential use for phylogenetic and population genetic studies. *Tree Genetics & Genomes*, **13**, 48. (*: co-premiers auteurs)

Références

Adams KL, Palmer JD (2003) Evolution of mitochondrial gene content: gene loss and transfer to the nucleus. *Molecular Phylogenetics and Evolution*, **29**, 380–395.

Alverson AJ, Wei X, Rice DW *et al.* (2010) Insights into the evolution of mitochondrial genome size from complete sequences of *Citrullus lanatus* and *Cucurbita pepo* (Cucurbitaceae). *Molecular Biology and Evolution*, **27**, 1436–1448.

Arca M, Hinsinger DD, Cruaud C *et al.* (2012) Deciduous trees and the application of universal DNA barcodes: a case study on the circumpolar *Fraxinus*. *PLoS ONE*, **7**, e34089.

Arthofer W, Schueler S, Steiner FM, Schlick-Steiner BC (2010) Chloroplast DNA-based studies in molecular ecology may be compromised by nuclear-encoded plastid sequence. *Molecular Ecology*, **19**, 3853–3856.

Barnard-Kubow KB, Sloan DB, Galloway LF (2014) Correlation between sequence divergence and polymorphism reveals similar evolutionary mechanisms acting across multiple timescales in a rapidly evolving plastid genome. *BMC Evolutionary Biology*, **14**, 268.

Bellot S, Renner SS (2015) The plastomes of two species in the endoparasite genus *Pilostyles* (Apodanthaceae) each retain just five or six possibly functional genes. *Genome Biology and Evolution*, **8**, 189–201.

Bergthorsson U, Richardson AO, Young GJ, Goertzen LR, Palmer JD (2004) Massive horizontal transfer of mitochondrial genes from diverse land plant donors to the basal angiosperm *Amborella*. *Proceedings of the National Academy of Sciences of the United States of America*, **101**, 17747–17752.

Besnard G, Khadari B, Baradat P, Bervillé A (2002) Combination of chloroplast and mitochondrial DNA polymorphisms to study cytoplasm genetic differentiation in the olive complex (*Olea europaea* L.). *Theoretical and Applied Genetics*, **105**, 139–144.

Besnard G, Khadari B, Villemur P, Bervillé A (2000) Cytoplasmic male sterility in the olive (*Olea europaea* L.). *Theoretical and Applied Genetics*, **100**, 1018–1024.

Besnard G, Rubio De Casas R, Christin PA, Vargas P (2009) Phylogenetics of *Olea* (Oleaceae) based on plastid and nuclear ribosomal DNA sequences: tertiary climatic shifts and lineage differentiation times. *Annals of Botany*, **104**, 143–160.

Blankenship RE (2017) How Cyanobacteria evolved. *Science*, **355**, 1372–1374.

Bock R (2009) The give-and-take of DNA : horizontal gene transfer in plants. *Trends in Plant Science*, **15**, 11–22.

Bölter B, Soll J (2016) Once upon a time – chloroplast protein import research from infancy to future challenges. *Molecular Plant*, **9**, 798–812.

Burger G, Gray MW, Lang BF (2003) Mitochondrial genomes: anything goes. *Trends in Genetics*, **19**, 709–716.

Burke S V, Wysocki WP, Zuloaga FO *et al.* (2016) Evolutionary relationships in panicoid grasses based on plastome phylogenomics (Panicoideae; Poaceae). *BMC Plant Biology*, **16**, 140.

Cavalier-Smith T (2000) Membrane heredity and early chloroplast evolution. *Trends in Plant Science*, **5**, 174–182.

Christin PA, Salamin N, Muasya AM *et al.* (2008) Evolutionary switch and genetic convergence on *rbcL* following the evolution of C_4 photosynthesis. *Molecular Biology and Evolution*, **25**, 2361–2368.

Chumley TW, Palmer JD, Mower JP *et al.* (2006) The complete chloroplast genome sequence of *Pelargonium* x *hortorum*: organization and evolution of the largest and most highly rearranged chloroplast genome of land plants. *Molecular Biology and Evolution*, **23**, 2175–2190.

Corriveau JL, Coleman AW (1988) Rapid screening method to detect potential biparental inheritance of plastid DNA and results for over 200 angiosperm species. *American Journal of Botany*, **75**, 1443–1458.

Cusack BP, Wolfe KH (2007) When gene marriages don't work out: divorce by subfunctionalization. *Trends in Genetics*, **23**, 270–272.

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

Davey JL, Blaxter MW (2010) RADseq: Next-generation population genetics. *Briefings in Functional Genomics*, **9**, 416–423.

Delsuc F, Brinkmann H, Philippe H (2005) Phylogenomics and the reconstruction of the tree of life. *Nature Reviews. Genetics*, **6**, 361–375.

Delwiche CF (1999) Tracing the thread of plastid diversity through the tapestry of life. *The American Naturalist*, **154**, S164–S177.

Elshire RJ, Glaubitz JC, Sun Q *et al.* (2011) A robust, simple genotyping-by-sequencing (GBS) approach for high diversity species. *PLoS ONE*, **6**, 1–10.

Erixon P, Oxelman B (2008) Whole-gene positive selection, elevated synonymous substitution rates, duplication, and indel evolution of the chloroplast *clpP1* gene. *PLoS ONE*, **3**.

Feng X, Kaur AP, MacKenzie SA, Dweikat IM (2009) Substoichiometric shifting in the fertility reversion of cytoplasmic male sterile pearl millet. *Theoretical and Applied Genetics*, **118**, 1361–1370.

Gantt JS, Baldauf SL, Calie PJ, Weeden NF, Palmer

JD (1991) Transfer of *rpl22* to the nucleus greatly preceded its loss from the chloroplast and involved the gain of an intron. *The EMBO Journal*, **10**, 3073–3078.

Gray MW, Archibald J (2012) Origins of mitochondria and plastids. In: *Genomics of chloroplasts and mitochondria* (eds Bock R, Knoop V), pp. 1–30. Springer Science & Business Media.

Green P (2004) Flowering plants - Dicotyledons. In: *The families and genera of vascular plants*, pp. 296–306. Springer Berlin Heidelberg, New-York.

Greiner S, Sobanski J, Bock R (2014) Why are most organelle genomes transmitted maternally? *BioEssays*, **37**, 80–94.

Gualberto JM, Mileshina D, Wallet C *et al.* (2014) The plant mitochondrial genome: dynamics and maintenance. *Biochimie*, **100**, 107–120.

Guisinger MM, Kuehl J V, Boore JL, Jansen RK (2008) Genome-wide analyses of Geraniaceae plastid DNA reveal unprecedented patterns of increased nucleotide substitutions. *Proceedings of the National Academy of Sciences of the United States of America*, **105**, 18424–18429.

Hinsinger DD, Basak J, Gaudeul M *et al.* (2013) The phylogeny and biogeographic history of ashes (*Fraxinus*, Oleaceae) highlight the roles of migration and vicariance in the diversification of temperate trees. *PLoS ONE*, **8**, e80431.

Hinsinger DD, Gaudeul M, Couloux A, Bousquet J, Frascaria-Lacoste N (2014) The phylogeography of Eurasian *Fraxinus* species reveals ancient transcontinental reticulation. *Molecular Phylogenetics and Evolution*, **77**, 223–237.

Hohmann-Marriott MF, Blankenship RE (2011) Evolution of photosynthesis. *Annual Review of Plant Biology*, **62**, 515–548.

Holder M, Lewis PO (2003) Phylogeny estimation: traditional and Bayesian approaches. *Nature Reviews Genetics*, **4**, 275–284.

Hong-Wa C (2016) *A taxonomic revision of the genus Noronhia Stadtm. ex Thouars (Oleaceae) in Madagascar and the Comoro Islands* (M Callmander, P Lowry, Eds,). Boissiera 70.

Hong-Wa C, Besnard G (2013) Intricate patterns of phylogenetic relationships in the olive family as inferred from multi-locus plastid and nuclear DNA sequence analyses: a close-up on *Chionanthus* and *Noronhia* (Oleaceae). *Molecular Phylogenetics and Evolution*, **67**, 367–378.

Hong-Wa C, Besnard G (2014) Species limits and diversification in the Madagascar olive (*Noronhia*, Oleaceae). *Botanical Journal of the Linnean Society*, **174**, 141–161.

Iorizzo M, Senalik D, Szklarczyk M *et al.* (2012) De novo assembly of the carrot mitochondrial genome using next generation sequencing of whole genomic DNA provides first evidence of DNA transfer into an angiosperm plastid genome. *BMC Plant Biology*, **12**, 1–17.

Jansen RK, Ruhlman TA (2012) Plastid genomes of seed plants. In: *Genomes of chloroplasts and mitochondria* (eds Bock R, Knoop V), pp. 103–126. Springer Science & Business Media.

Jeandroz S, Roy A, Bousquet J (1997) Phylogeny and phylogeography of the circumpolar genus *Fraxinus* (Oleaceae) based on internal transcribed spacer sequences of nuclear ribosomal DNA. *Molecular Phylogenetics and Evolution*, **7**, 241–251.

Jensen PE, Leister D (2014) Chloroplast evolution, structure and functions. *F1000Prime Reports*, **6**, 1–14.

Keeling PJ (2010) The endosymbiotic origin, diversification and fate of plastids. *Philosophical Transactions of the Royal Society B: Biological Sciences*, **365**, 729–748.

Khan A, Khan IA, Asif H, Azim MK (2010) Current trends in chloroplast genome research. *African Journal of Biotechnology*, **9**, 3494–3500.

Kim KJ, Jansen RK (1998) A chloroplast DNA phylogeny of lilacs (*Syringa*, Oleaceae): plastome groups show a strong correlation with crossing groups. *American Journal of Botany*, **85**, 1338–1351.

Kubo T, Mikami T (2007) Organization and variation of angiosperm mitochondrial genome. *Physiologia Plantarum*, **129**, 6–13.

Lendvay B, Kadereit JW, Westberg E *et al.* (2016) Phylogeography of *Syringa josikaea* (Oleaceae): early Pleistocene divergence from East Asian relatives and survival in small populations in the Carpathians. *Biological Journal of the Linnean Society*, **119**, 689–703.

Li J, Goldman-Huertas B, DeYoung J, Alexander J (2012) Phylogenetics and diversification of *Syringa* inferred from nuclear and plastid DNA sequences. *Castanea*, **77**, 82–88.

Liu Y, Cui H, Zhang Q, Sodmergen (2004) Divergent potentials for cytoplasmic inheritance within the genus *Syringa*. A new trait associated with speciogenesis. *Plant Physiology*, **136**, 2762–2770.

Lyons TW, Reinhard CT, Planavsky NJ (2014) The rise of oxygen in Earth's early ocean and atmosphere. *Nature*, **506**, 307–315.

Ma PF, Zhang YX, Guo ZH, Li DZ (2015) Evidence for horizontal transfer of mitochondrial DNA to the plastid genome in a bamboo genus. *Scientific Reports*, **5**, 11608.

Magee AM, Aspinall S, Rice DW *et al.* (2010) Localized hypermutation and associated gene losses in legume chloroplast genomes. *Genome Research*, **20**, 1700–1710.

Maréchal A, Brisson N (2010) Recombination and the maintenance of plant organelle genome stability. *New Phytologist*, **186**, 299–317.

Margulis L (1970) *Origin of eukaryotic cells: evidence*

and research implications for a theory of the origin and evolution of microbial, plant and animal cells on the precambrian Earth. Yale University Press.

Martin W, Mentel M (2010) The origin of mitochondria. *Nature Education*, **3**, 58.

Mensous M, Van de Paer C, Manzi S *et al.* (2017) Diversity and evolution of plastomes in Saharan mimosoids: potential use for phylogenetic and population genetic studies. *Tree Genetics and Genomes*, **13**, 48.

Metzker ML (2009) Sequencing technologies — the next generation. *Nature Reviews Genetics*, **11**, 31–46.

Millen RS, Olmstead RG, Adams KL *et al.* (2001) Many parallel losses of *infA* from chloroplast DNA during angiosperm evolution with multiple independent transfers to the nucleus. *The Plant Cell*, **13**, 645–658.

Mogensen HL (1996) The hows and whys of cytoplasmic inheritance in seed plants. *American Journal of Botany*, **83**, 383–404.

Mower JP, Sloan DB, Alverson AJ (2012) Plant mitochondrial genome diversity: the genomics revolution. In: *Plant Genome Diversity*, pp. 123–144. Springer Vienna.

Muller HJ (1964) The relation of recombination to mutational advance. *Mutation Research/Fundamental and Molecular Mechanisms of Mutagenesis*, **1**, 2–9.

Notsu Y, Masood S, Nishikawa T *et al.* (2002) The complete sequence of the rice (*Oryza sativa* L.) mitochondrial genome: frequent DNA sequence acquisition and loss during the evolution of flowering plants. *Molecular Genetics and Genomics*, **268**, 434–445.

Ogihara Y, Yamazaki Y, Murai K *et al.* (2005) Structural dynamics of cereal mitochondrial genomes as revealed by complete nucleotide sequencing of the wheat mitochondrial genome. *Nucleic Acids Research*, **33**, 6235–6250.

Palmer JD (1992) Mitochondrial DNA in plant systematics: applications and limitations. In: *Molecular Systematics of Plants* (eds Soltis P, Soltis D, Doyle J), pp. 36–49. Chapman and Hall, New York.

Piot A, Hackel J, Christin PA, Besnard G (2018) One third of the plastid genes evolved under positive selection in PACMAD grasses. *Planta*, **247**, 255–266.

Rice DW, Alverson AJ, Richardson AO *et al.* (2013) Horizontal transfer of entire genomes via mitochondrial fusion in the angiosperm *Amborella*. *Science*, **342**, 1468–1473.

Rousseau-Gueutin M, Huang X, Higginson E *et al.* (2013) Potential functional replacement of the plastidic acetyl-CoA carboxylase subunit (*accD*) gene by recent transfers to the nucleus in some angiosperm lineages. *Plant Physiology*, **161**, 1918–1929.

Sanger F, Nicklen S, Coulson A (1977) DNA sequencing with chain-terminating inhibitors. *Proceedings of the National Academy of Sciences of the United States of America*, **74**, 5463–5467.

Schardl CL, Pring DR, Fauron C-R, Lonsdale DM (1985) Mitochondrial DNA rearrangements resulting in fertile revertants of S-type male sterile maize. *Cell*, **43**, 361–368.

Sloan DB, Alverson AJ, Wu M, Palmer JD, Taylor DR (2012) Recent acceleration of plastid sequence and structural evolution coincides with extreme mitochondrial divergence in the angiosperm genus *Silene*. *Genome Biology and Evolution*, **4**, 294–306.

Sloan DB, Wu Z (2014) History of plastid DNA insertions reveals weak deletion and AT mutation biases in angiosperm mitochondrial genomes. *Genome Biology and Evolution*, **6**, 3210–3221.

Smith DR, Keeling PJ (2015) Mitochondrial and plastid genome architecture: reoccurring themes, but significant differences at the extremes. *Proceedings of the National Academy of Sciences of the United States of America*, **112**, 201422049.

Sodmergen, Bai HH, He JX *et al.* (1998) Potential for biparental cytoplasmic inheritance in *Jasminum officinale* and *Jasminum nudiflorum*. *Sexual Plant Reproduction*, **11**, 107–112.

Soo R, Hemp J, Parks D, Fischer W, Hugenholtz P (2017) On the origins of oxygenic photosynthesis and aerobic respiration in Cyanobacteria. *Science*, **1440**, 1436–1440.

Straub SCK, Cronn RC, Edwards C, Fishbein M, Liston A (2013) Horizontal transfer of DNA from the mitochondrial to the plastid genome and its subsequent evolution in milkweeds (Apocynaceae). *Genome Biology and Evolution*, **5**, 1872–1885.

Taylor H (1945) Cyto-taxonomy and phylogeny of the Oleaceae. *Brittonia*, **5**, 337–367.

Tonti-Filippini J, Nevill PG, Dixon K, Small I (2017) What can we do with 1000 plastid genomes? *Plant Journal*, **90**, 808–818.

Ueda M, Fujimoto M, Arimura SI *et al.* (2007) Loss of the *rpl32* gene from the chloroplast genome and subsequent acquisition of a preexisting transit peptide within the nuclear gene in *Populus*. *Gene*, **402**, 51–56.

Unseld M, Marienfeld JR, Brandt P, Brennicke A (1997) The mitochondrial genome of *Arabidopsis thaliana* contains 57 genes in 366,924 nucleotides. *Nature Genetics*, **15**, 57–61.

Wallander E (2008) Systematics of *Fraxinus* (Oleaceae) and evolution of dioecy. *Plant Systematics and Evolution*, **273**, 25–49.

Wallander E, Albert V (2000) Phylogeny and classification of Oleaceae based on *rps16* and

trnL-F sequence data. *American Journal of Botany*, **87**, 1827–1841.

Wang D, Wu YW, Shih ACC *et al.* (2007) Transfer of chloroplast genomic DNA to mitochondrial genome occurred at least 300 MYA. *Molecular Biology and Evolution*, **24**, 2040–2048.

Wang DY, Zhang Q, Liu Y *et al.* (2010) The levels of male gametic mitochondrial DNA are highly regulated in angiosperms with regard to mitochondrial inheritance. *The Plant Cell Online*, **22**, 2402–2416.

Watson J, Crick F (1953) Molecular structure of nucleic acids. *Nature*, **171**, 737–738.

Wicke S, Schneeweiss GM, DePamphilis CW, Müller KF, Quandt D (2011) The evolution of the plastid chromosome in land plants: gene content, gene order, gene function. *Plant Molecular Biology*, **76**, 273–297.

Wolfe KH, Li WH, Sharp PM (1987) Rates of nucleotide substitution vary greatly among plant mitochondrial, chloroplast, and nuclear DNAs. *Proceedings of the National Academy of Sciences of the United States of America*, **84**, 9054–9058.

Xia J, Lu J, Wang ZX *et al.* (2013) Pollen limitation and Allee effect related to population size and sex ratio in the endangered *Ottelia acuminata* (Hydrocharitaceae): implications for conservation and reintroduction. *Plant Biology*, **15**, 376–383.

Zedane L (2016) Phylogenetic hypothesis of the *Oleeae* tribe (Oleaceae). Diversification and molecular evolution patterns in plastid and nuclear ribosomal DNA (PhD thesis). Université Toulouse 3 Paul Sabatier.

Zedane L, Hong-Wa C, Murienne J *et al.* (2016) Museomics illuminate the history of an extinct, paleoendemic plant lineage (*Hesperelaea*, Oleaceae) known from an 1875 collection from Guadalupe Island, Mexico. *Biological Journal of the Linnean Society*, **117**, 44–57.

Zhu A, Guo W, Gupta S, Fan W, Mower JP (2016) Evolutionary dynamics of the plastid inverted repeat: the effects of expansion, contraction, and loss on substitution rates. *New Phytologist*, **209**, 1747–1756.

*Licences des photographies utilisées Fig. 9. *Fraxinus ornus* : Amada44 [cc-by-2.0]; *Jasminum nudiflorum* : FlickrLickr[cc-by-2.0]; *Ligustrum vulgaris* : redagainPatti [cc-by-nd-2.0]; *Forsythia* × *intermedia* : Dan Smith [cc-by-sa-2.0]; *Syringa vulgaris* : pixel2013 [cc0 1.0].

CHAPITRE I

Mitogenomics of *Hesperelaea*, an extinct genus of Oleaceae

VAN DE PAER C., HONG-WA C., JEZIORSKI C., BESNARD G.

Gene, 2016, **594**: 197-202

Gene 594 (2016) 197–202

Contents lists available at ScienceDirect

Gene

journal homepage: www.elsevier.com/locate/gene

Research paper

Mitogenomics of *Hesperelaea*, an extinct genus of Oleaceae

Céline Van de Paer [a,*], Cynthia Hong-Wa [b], Céline Jeziorski [c,1], Guillaume Besnard [a,*]

[a] CNRS, Université de Toulouse, ENFA, UMR5174 EDB (Laboratoire Évolution & Diversité Biologique), 118 Route de Narbonne, 31062 Toulouse, France
[b] Claude E. Phillips Herbarium, Delaware State University, 1200 N Dupont Hwy Dover, DE 19901-2277, USA
[c] GeT-PlaGe, Genotoul, INRA Auzeville, 31326 Castanet-Tolosan, France

ARTICLE INFO

Article history:
Received 15 July 2016
Accepted 2 September 2016
Available online 4 September 2016

Keywords:
De novo assembly
Herbarium specimen
Hesperelaea palmeri
Lamiales
Mitochondrial DNA
Olive
Phylogeny

ABSTRACT

The recent developments in high-throughput DNA sequencing allowed major advances in organelle genomics. Assembly of mitochondrial genomes (hereafter mitogenomes) in higher plants however remains a challenge due to their large size and the presence of plastid-derived regions and repetitive sequences. In this study, we reconstructed the first mitogenome of Oleaceae using a herbarium specimen of the extinct genus *Hesperelaea* collected in 1875. Paired-end reads produced with the HiSeq technology (shotgun) in a previous study were reused. With an approach combining reference-guided and de novo assembly, we obtained a circular molecule of 658,522 bp with a mean coverage depth of 35×. We found one large repeat (ca. 8 kb) and annotated 46 protein-coding genes, 3 rRNA genes and 19 tRNA genes. A phylogeny of Lamiales mitogenomes confirms Oleaceae as sister to a group comprising Lamiaceae, Phyrmaceae and Gesneriaceae. The *Hesperelaea* mitogenome has lower rates of synonymous and non-synonymous substitution compared to *Nicotiana tabacum* than other available mitogenomes of Lamiales. To conclude, we show that mitogenome reconstruction in higher plants is possible with shotgun data, even from poorly preserved DNA extracted from old specimens. This approach offers new perspectives to reconstruct plant phylogenies from mitochondrial markers, and to develop functional mitogenomics in Oleaceae.

1. Introduction

Recent advances in high-throughput DNA sequencing, so-called Next Generation Sequencing (NGS), have revolutionized phylogenetics by generating massive genomic resources (Harrison and Kidner, 2011). In particular, the genome skimming approach (i.e. low coverage shotgun sequencing) allowed the assembly of the most represented genomic regions such as the nuclear ribosomal cluster and cytoplasmic genomes (e.g. Cronn et al., 2008; Straub et al., 2012; Besnard et al., 2013; Malé et al., 2014). In plants, genome skimming has been mainly used to generate chloroplast genomes (hereafter plastomes) and nearly complete nuclear ribosomal sequences (Straub et al., 2012), but the method can also be used to assemble sequences from the mitochondrial genome (hereafter mitogenome) (e.g. Ma et al., 2012; Bock et al., 2014; Malé et al., 2014). In addition, the genome skimming method was successfully applied to low-quality DNA extracted from herbarium specimens, which are generally not amenable to classical PCR (e.g. Staats et al., 2013; Besnard et al., 2014; Bakker et al., 2016; Zedane et al., 2016). This approach thus offers the possibility to include in phylogenetic analyses rare and extinct species from natural history collections.

Both plastomes and mitogenomes are maternally inherited in numerous plant groups, in particular in angiosperms, but there are a few exceptions (e.g. Corriveau and Coleman, 1988; Birky, 1995). Polymorphism of the plastome has been extensively used in phylogeographic studies, in contrast with the mitogenome (e.g. Ruhlman and Jansen, 2014). The general lack of interest in plant mitogenomes for phylogenetic analyses mainly stems from their supposed low sequence variation (Wolfe et al., 1987; Smith and Keeling, 2015). Consequently, the assembly of plant mitochondrial sequences from shotgun data was reported only in a few studies (e.g. Ma et al., 2012; Bock et al., 2014; Malé et al., 2014). In addition, reconstructing full mitogenome sequences is challenging due to the presence of both long duplications and plastid-derived regions (Mower et al., 2012b; Smith and Keeling, 2015). These characteristics are lineage-specific and make the assembly of plant mitogenome from short sequence reads very complex, with no standard procedure (as proposed for reconstructing plastomes; Straub et al., 2012; Bakker et al., 2016). Yet, mitogenomic variation in plants may be valuable in investigating not only phylogenetic relationships among species, but also in studying patterns of substitution (e.g. Zhu

Abbreviations: bp, base pair; d_N, non-synonymous substitution rate; d_S, synonymous substitution rate; GC content, Guanine and Cytosine content; GTR + G, General Time Reversible + Gamma; mitogenome, mitochondrial genome; NGS, Next Generation Sequencing; plastome, plastid or chloroplast genome; RFLPs, Restriction Fragment Length Polymorphisms; rRNA, ribosomal RNA; SD, standard deviation; tRNA, transfer RNA.

* Corresponding authors.
E-mail addresses: vandepaer.celine@gmail.com (C. Van de Paer), guillaume.besnard@univ-tlse3.fr (G. Besnard).
[1] Present address: INRA-CNRGV, INRA Auzeville, 31326 Castanet-Tolosan, France.

http://dx.doi.org/10.1016/j.gene.2016.09.007
0378-1119/

et al., 2014) or the genetic determinism of maternally inherited traits such as the cytoplasmic male sterility (e.g. Liu et al., 2011; Mower et al., 2012a; Hiroshi et al., 2014; Touzet and Meyer, 2014).

Here, we describe the assembly of the first mitogenome of Oleaceae (Lamiales). We applied our approach to degraded DNA extracted from an old herbarium specimen collected 141 years ago on Guadalupe Island, Mexico. This specimen represents the only known remain of the genus *Hesperelaea*, which is now considered extinct (Watson, 1876; Zedane et al., 2016). Our study demonstrates that shotgun sequencing data, even from poorly preserved DNA, can be used to generate mitochondrial DNA datasets. Molecular analysis of available mitogenomes of Lamiales reveals deep structural reorganizations and heterogeneous evolutionary rates between mitochondrial genes. Finally, we discuss the implication of our results for further studies on the Oleaceae functional mitogenomics and phylogenetics.

2. Material and methods

2.1. Study model and DNA sequencing

Hesperelaea palmeri A. Gray belongs to the tribe Oleeae (Wallander and Albert, 2000). A recent phylogenetic study indicates that it was a paleoendemic lineage of an American group that includes *Forestiera* and *Priogymnanthus* (Zedane et al., 2016). The complete plastome of this species has been sequenced previously using a genome skimming approach (Zedane et al., 2016). Here, we reused these shotgun data produced with the HiSeq technology (Illumina Inc., 2010), consisting of 10,694,511 paired-end reads of 100 bp (Zedane et al., 2016). The mean size of inserts (DNA fragments used for library construction) was estimated at ca. 120 bp, reflecting that the DNA extracted from a herbarium leaf is highly fragmented (for more details see Zedane et al., 2016). We first removed duplicates (identical reads) with FastUniq (Xu et al., 2012). These duplicates were likely resulting from library amplification before sequencing (see Zedane et al., 2016). Second, overlapping paired-end reads were merged using BBMerge [BBtools implemented in Geneious v. 9.0.5 (Biomatters Ltd., 2015)]. Read merging allowed us to align reads and their reverse complements to create a single consensus read. This step also allowed improving the sequence quality of overlapping parts. In cases of non-overlapping paired-end reads, reads were not merged and were used independently. We thus obtained 9,620,010 merged reads (with a mean length of 120 bp, $SD = 27.6$) and 2,011,824 unmerged reads (with a length of 100 bp).

2.2. Mitogenome assembly

We used an approach combining reference-guided and de novo assembly. To discriminate plastid and mitochondrial sequences, merged and unmerged reads were first mapped simultaneously to the plastome of *H. palmeri* [GenBank: NC_025787] and to the mitogenome of *Boea hygrometrica* (Lamiales: Gesneriaceae, [NC_016741]) using Geneious. Due to large genomic rearrangements relative to the mitochondrial reference genome, we obtained 76 non-overlapping contigs that had >500 bp (up to 3459 bp). These contigs were then extended independently using Geneious by mapping all merged and unmerged reads to each of them with series of 100 iterations. Coverage depth was relatively constant across the genome but we observed peaks of coverage corresponding to duplicated and plastid-derived regions: (i) in long duplicated regions, the depth-coverage was, as expected, twice as high as in single-copy regions (ca. 70× vs. 35×; see below); (ii) in plastid-derived regions, plastid reads mapped to their homologous mitochondrial regions, leading to an increase of the coverage depth. In the first case, as mentioned by Mower et al. (2012a), some reads contiguous to the duplicated regions allowed different possible paths of assembly, reflecting several alternative conformations. We thus arbitrarily resolved the path of assembly when we encountered long repeated sequences (e.g. Cahill et al., 2010). In the second case, as the sequencing depth of the plastome was higher than that of the mitogenome (ca. 300× vs. 35×), we resolved these regions manually (using Geneious) by identifying mitochondrial reads taking into account polymorphisms (i.e. single nucleotide polymorphisms and indels) with the lowest coverage.

During the process of assembly, we regularly checked whether the extended contigs could be joined, taking into account the sequence similarity between their terminal regions using a de novo analysis in Geneious with a minimum overlap of 100 bp. In this manner, several circular DNA sequences can be obtained. Duplicated regions (with higher sequencing depth; see above) that are shared by such circular sequences allowed their integration in a "master circle". Even if the master circle could not be an accurate representation of the mitochondrial structure in vivo due to alternative genomic conformations (e.g. Mower et al., 2012b; Warren et al., 2016), it is a convenient mitochondrial consensus sequence representing the genomic content. A last verification of the assembly was performed by mapping all paired-end reads simultaneously to the plastome of *H. palmeri* and to the mitochondrial consensus sequence to check for abnormal variation of coverage. The average coverage depth of the mitogenome was estimated on non-plastid-derived regions.

2.3. Identification of genes, ORFs, duplicated regions and plastid-derived regions

Coding-protein genes and rRNA were annotated with Geneious by transferring the annotations of available Lamiales mitogenomes of *Boea hygrometrica*, *Mimulus guttatus*, *Ajuga reptans* and *Salvia miltiorrhiza* (GenBank accession numbers given in Table 1). Start and stop codons were manually adjusted. Annotation of tRNA was performed with tRNAscanSE (http://lowelab.ucsc.edu/tRNAscanSE). Additional open reading frames (ORFs) with minimum length of 300 bp were detected with Geneious.

Table 1
Comparison of genomic content (sequence length in base pairs and percent of the complete mitogenome sequence given in brackets; GC content in percent) among the mitogenomes of Lamiales and *Nicotiana tabacum* (Solanales). Annotations of the mitogenomes were extracted from GenBank. To estimate the plastid-derived content of the mitogenome of each species, we used their corresponding complete plastome available in GenBank as query. For *A. reptans* and *M. guttatus* whose complete plastomes were not available, we used the complete plastome of *Lavandula angustifolia* (Lamiaceae) and *Lindenbergia philippensis* (Orobanchaceae), respectively.

	Hesperelaea palmeri (KX545367)[a]	*Boea hygrometrica* (NC_016741)	*Mimulus guttatus* (NC_018041)	*Salvia miltiorrhiza* (NC_023209)	*Ajuga reptans* (NC_023103)	*Nicotiana tabacum* (KR780036)
Genome size (bp)	658,522	510,519	525,671	499,236	352,069	430,863
Protein coding genes	35,118 (5.33)	31,272 (6.13)	28,518 (5.58)	58,647 (11.75)	27,339 (7.76)	28,587 (6.63)
tRNA genes	1449 (0.22)	2056 (0.40)	1836 (0.35)	1670 (0.33)	1294 (0.37)	1754 (0.41)
rRNA genes	5485 (0.83)	2154 (0.42)	5449 (1.04)	8493 (1.10)	5179 (1.47)	5404 (1.25)
Introns	23,500 (3.57)	15,834 (3.10)	23,261 (4.43)	22,385 (4.48)	28,242 (8.02)	22,218 (5.16)
Non-annotated sequences	592,970 (90.05)	459,203 (89.95)	464,090 (88.29)	408,041 (81.73)	290,015 (82.37)	372,900 (86.55)
Repeats content	35,361 (5.37)	15,350 (3.01)	95,606 (18.19)	37,582 (7.53)	4186 (1.19)	52,683 (12.23)
Plastid-like sequences	57,617 (8.75)	51,305 (10.05)	33,406 (6.35)	34,897 (6.99)	4418 (1.25)	9095 (2.11)
GC content (%)	44.5	43.3	45.1	44.4	45.1	45.0

[a] GenBank accession number.

Repeats with a 40-bp minimum length and with at least 95% identity were identified using VMATCH (Kurtz, 2016). Identification of plastid-derived regions was then carried out using BLASTN with the *H. palmeri* plastome as a query sequence. We listed plastid-derived regions with a minimum length of 100 bp and with at least 80% of identity with the *H. palmeri* plastome.

2.4. Phylogenetic analyses and estimation of substitution rates

A phylogeny of Lamiales including *Hesperelaea* was reconstructed based on mitochondrial gene sequences, in order to compare the evolutionary rates of these genes in this plant group. We used the four available Lamiales mitogenomes (see above) with the mitogenome of *Nicotiana tabacum* (Solanales) as outgroup. From each mitogenome, we extracted nucleotide sequences for all protein-coding genes and introns. We excluded genes that were either absent or present as pseudogenes in several mitogenomes. Each gene and intron was aligned individually using MUSCLE (Edgar, 2004) with default parameters. The alignments of protein-coding genes were manually adjusted taking into account their amino-acid translation. All aligned protein-coding genes and introns were then concatenated. Maximum-Likelihood (ML) analyses were performed using RAxML v.8 (Stamatakis, 2014) in

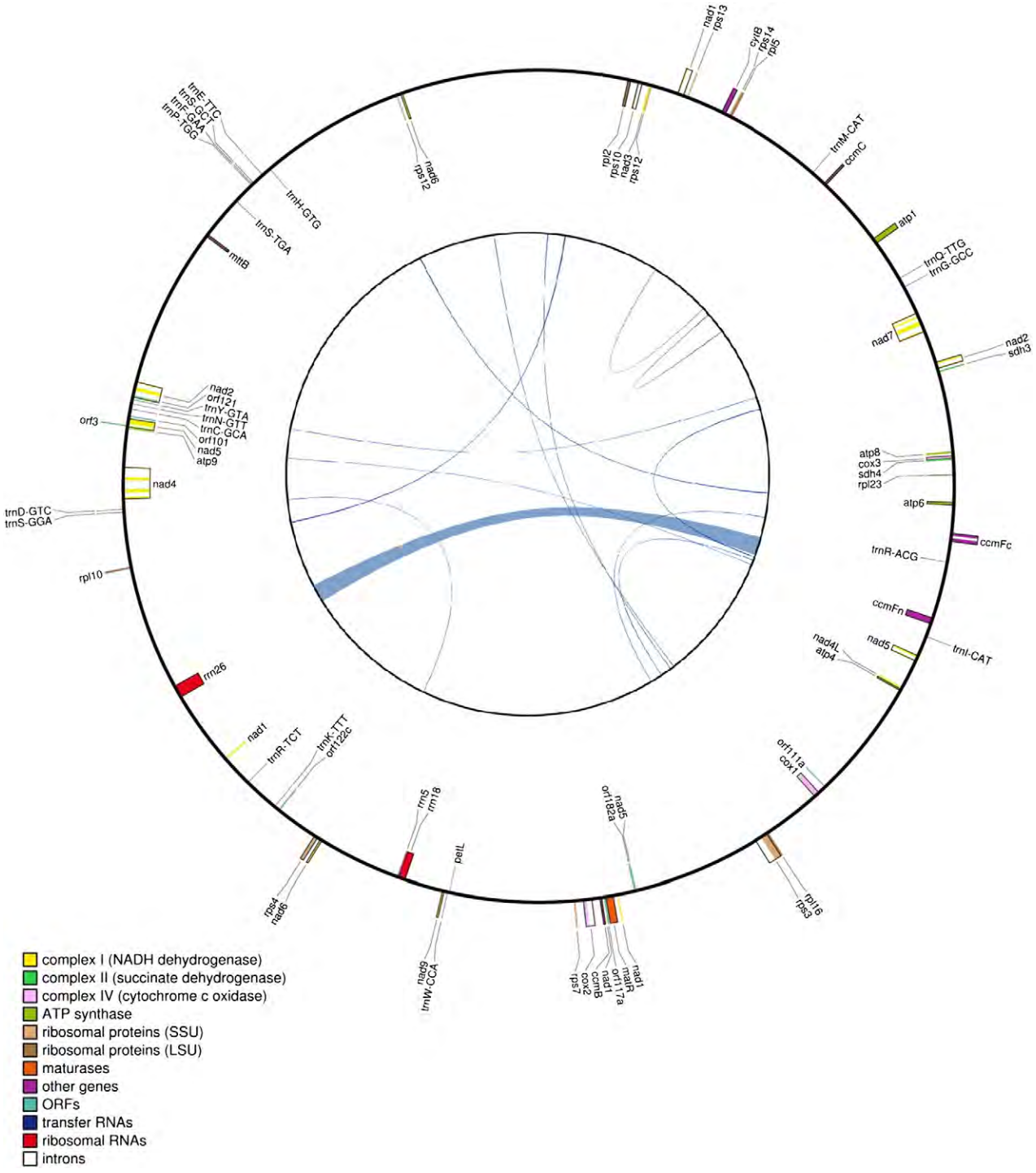


Fig. 1. Master circle of the *Hesperelaea palmeri* mitogenome (658,522 bp). The outer circle represents the gene content, with protein-coding genes, rRNA and tRNA genes. The inner circle represents repeated sequences with a minimum length of 100 bp in the mitogenome. The map was performed with OGDRAW v1.2 (Lohse et al., 2007) and CIRCOS (Krzywinski et al., 2009).

the CIPRES platform. We applied the GTR + G model and the rapid bootstrap algorithm with 1000 iterations. We visualized the best likelihood phylogenetic tree using FigTree v1.4.2 (Rambaut, 2014). Non-synonymous and synonymous substitution rates (d_N and d_S) were calculated individually on each coding-protein gene using *codeml* in PAML v.4.8 (Yang, 2007). These gene divergence rates were estimated for the five Lamiales mitogenomes by comparing each of them to the mitogenome of *N. tabacum* as a reference (outgroup).

3. Results and discussion

3.1. Mitogenome assembly, content and annotation

Our approach allowed us to reconstruct a master sequence of the *H. palmeri* mitogenome (EMBL accession no: KX545367; mitogenome length: 658,522 bp; GC content: 44.50%). About 5% of this genome is composed of duplicated regions, with one large repeat of 7975 bp and 151 small repeats ranging from 40 to 496 bp (Tables 1 and S1). We identified 21 plastid-derived regions with a length > 100 bp and with a minimum identity of 80% with the *H. palmeri* plastome [covering in total 57,617 bp (8.75%) of the mitogenome; Tables 1 and S2]. Fifteen of these regions (54,996 bp) have an identity of >96% with the plastome, and two of them are part (1722 bp) of the long duplicated region. We thus estimated that about 41% of the plastome (130,108 bp when excluding one inverted repeat) have a highly homologous region in the mitogenome. The existence of sequences with variable identity with the plastome (96–99% vs. 80–84%; Table S2) suggests at least two rounds of integration of plastid sequences into the mitogenome. The analysis of the evolution (e.g. multiple integrations, degeneration) of such genomic regions however requires an extended sample of Oleaceae mitogenomes (e.g. Hao and Palmer, 2009). Excluding these plastid-derived regions, the average coverage depth of the *H. palmeri* mitogenome was estimated at 34.6× ($SD = 11$).

The largest part of the *H. palmeri* mitogenome corresponds to non-annotated sequences (592,970 bp). Annotated gene regions represent 65,552 bp (9.95%) and comprise 46 protein-coding genes with introns, three rRNA genes and 19 tRNA genes (Fig. 1; Tables 1 and S3). Two copies of *nad6* were detected in the *H. palmeri* mitogenome, but with different start codons (i.e. ATG or GTG). We also identified a partially duplicated copy of *rps12*; it is very close to its paralogous functional copy but appears to be truncated at the end of the sequence as similarly reported in other genes of plant mitogenomes (e.g. *rps3* duplicates in *Gossypium*; Tang et al., 2015). Lastly, we identified 171 ORFs with lengths ranging from 300 to 1000 bp and 14 additional ORFs ranging from 1000 to 2757 bp, for a total of 185 ORFs detected (Table S4). The longest ORFs (>1000 bp) were not observed in other Lamiales mitogenomes and may not be functional.

The *H. palmeri* mitogenome is larger ~659 kb than that of other Lamiales (ranging from ~352 to 526 kb; Table 1) but within the range reported for mitogenomes of angiosperms [from ~200 kb in *Brassica rapa* (Brassicaceae) to ~11 Mb in *Silene conica* (Caryophyllaceae); Smith and Keeling, 2015]. Gene content in Lamiales mitogenomes varies between species (from 54 in *B. hygrometrica* to 71 in *S. miltiorrhiza*; Table S3) due to the transfer of mitochondrial genes to the nuclear genome (e.g. ribosomal protein genes; Adams and Palmer, 2003). Variable gene content, however, does not explain the large mitogenomic size difference in Lamiales (Table 1). As in most angiosperms, the mitogenome of *H. palmeri* has a complex organization with a large portion of non-coding regions, plastid-derived regions and repeated sequences (see above), and the difference of genome size in Lamiales appears to be due to a large variation in these non-annotated regions (Table 1). In higher plant mitogenomes, numerous repeated sequences found in the non-coding regions could promote recombination events and lead to sequence duplication (Satoh et al., 2006). Degenerated duplicated sequences could diverge though continuous reshuffling and contribute to the extension of the genome size. This could explain mitogenome size variation in related species. The amount of plastid-like sequences in the *H. palmeri* mitogenome also partly explains its larger size relative to that of the other Lamiales available so far (Table 1).

3.2. Phylogenetic position of H. palmeri in Lamiales

A phylogeny of Lamiales including a Solanales as outgroup was reconstructed for six species with a concatenated alignment of 28 mitochondrial protein-coding genes and 13 introns (with a total of 56,448 aligned bp; Fig. 2). We excluded 18 protein-coding genes (*ccmC*, *petL*, *rpl2*, *rpl5*, *rpl10*, *rpl16*, *rpl23*, *rps7*, *rps10*, *rps14*, *sdh3*, *sdh4*, *orf101*, *orf111a*, *orf117a*, *orf121*, *orf122c* and *orf182a*) from the analysis due to their absence or their presence as a pseudogene in at least two species. The best ML tree (scored under the GTR + G model) was well supported with all bootstrap values reaching 100%. *Hesperelaea palmeri* appears to have a basal position in the Lamiales (Fig. 2). Such a phylogenetic placement of Oleaceae (*Hesperelaea*) as sister to a clade including Gesneriaceae (*Boea*), Phrymaceae (*Mimulus*) and Lamiaceae (*Ajuga*, *Salvia*) is congruent with other phylogenies reconstructed on cytoplasmic genes (Wortley et al., 2005; Refulio-Rodriguez and Olmstead, 2014). The branch leading to *Hesperelaea* is however shorter than that of other Lamiales [especially *Ajuga* (Zhu et al., 2014)], suggesting lower rate of evolution in the Oleaceae lineage compared to other families within this plant order (see below).

3.3. Substitution rates in Lamiales mitogenomes

Non-synonymous and synonymous substitution rates for protein-coding genes of Lamiales species were estimated against those of *N. tabacum*. Being absent in the mitogenome of *N. tabacum*, *atp8* was excluded from the analysis that was thus performed on 27 genes (Figs. 3 and S1). As shown by Zhu et al. (2014), protein-coding genes of *A. reptans* evolve particularly quickly compared to the other Lamiales and substitution rates have an exceptional variability among genes of the *A. reptans* mitogenome (Fig. S1). Compared to the other Lamiales, *H. palmeri* diverges slowly from *N. tabacum*, with non-synonymous site divergence (d_N) ranging from 0.005 to 0.392 and synonymous site divergence (d_S) ranging from 0.008 to 1.098 (Fig. S1). Non-synonymous and synonymous substitution rates strongly vary among genes in the mitogenome (Fig. 3). As with other Lamiales, genes coding for ATP synthase subunits and for ribosomal proteins in the *H. palmeri* mitogenome appear to have higher substitution rates, especially *atp4*, *atp6* and *atp9* (Figs. 3 and S1). This may be due to either episodic positive selection or a relaxed selective pressure compared to other protein-coding genes, but a larger sample of species is necessary to investigate in detail the past changes in evolutionary rates of these genes.

Biological explanations of such a variation in substitution rates across mitochondrial genes remain unknown (e.g. Mower et al.,

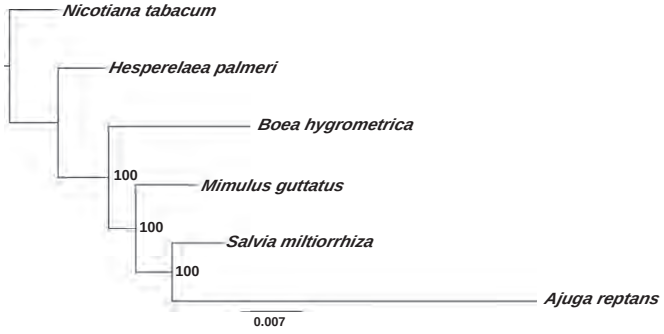


Fig. 2. Best likelihood phylogenetic tree of the Lamiales obtained with RAxML v.8 (Stamatakis, 2014) with a GTR + G model. We used a concatenated alignment of 28 mitochondrial protein-coding genes and 13 introns. We applied the rapid bootstrap algorithm with 1000 iterations. The phylogenetic tree was rooted with *Nicotiana tabacum* (Solanales).

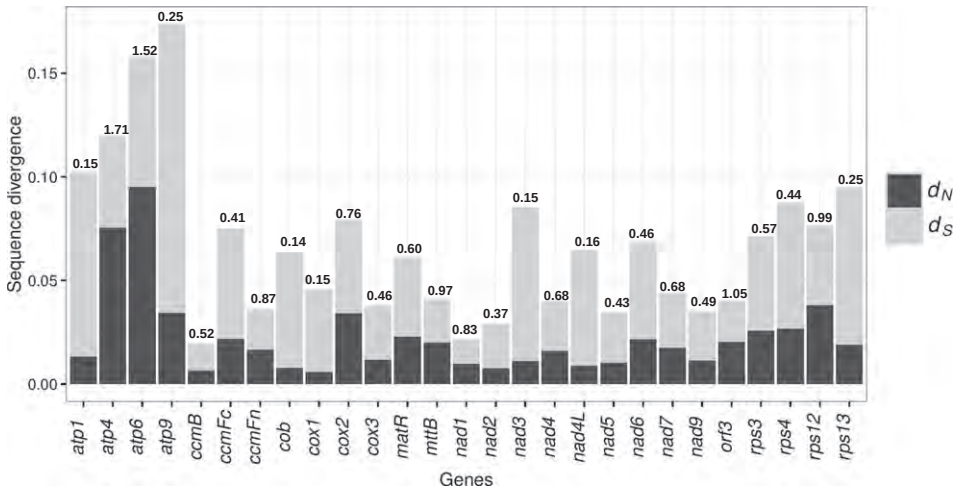


Fig. 3. Substitution rates among 27 protein-coding genes in *Hesperelaea palmeri*. For each protein-coding gene, non-synonymous (d_N) and synonymous (d_S) rates were calculated against *Nicotiana tabacum* using PAML (Yang, 2007). The value over each bar corresponds to the d_N/d_S ratio (ω) for each gene. For a comparative analysis among Lamiales species, see Fig. S1 in Supplementary material.

2007). A possible reason could be the co-evolution between the nuclear and the mitochondrial genomes. Enzyme complexes of oxidative phosphorylation (OXPHOS) and ribosomal protein subunits are encoded by genes located in both the nucleus and the mitochondria. In the case of cyto-nuclear co-evolution, coordinated substitutions between mitochondrial and nuclear proteins could lead to an increase in amino acid substitutions or an elevated d_N/d_S ratio (Rand et al., 2004; Dowling et al., 2008). This co-evolution between the nucleus and the mitochondria could thus influence the evolutionary dynamics of the mitochondrial genome.

4. Conclusions and perspectives

Our approach shows that the assembly of mitogenomes or long mitochondrial sequences is feasible with shotgun data, even with low-quality DNA extracted from herbarium samples. The assembly of mitochondrial genomes from Illumina short sequence reads remains, however, a challenge due to the length of the mitogenome, plastid-derived regions and repetitive sequences. Indeed, the reconstruction of a master circle is not fully automatized and thus time-consuming (e.g. Cahill et al., 2010; Malé et al., 2014).

The reconstruction of mitogenomes or partial mitochondrial sequences (i.e. genes and introns) could be an important step to develop mtDNA phylogenetics in Oleaceae, even from herbarium material. The possibility to include rare or extinct species in phylogenies thanks to museum specimens is especially relevant to the Oleaceae family, within which several species are rare and/or only known from herbaria (e.g. Hong-Wa and Besnard, 2014; Zedane et al., 2016). Mitochondrial markers are rarely used to reconstruct the evolutionary history of higher plants due to low substitution rate in coding sequences (Gualberto et al., 2014). In addition, phylogenies of plastid and mitochondrial DNAs may both reflect seed dispersal and be mostly congruent when they are inherited maternally (e.g. Olson and McCauley, 2000; Besnard et al., 2002). However, biparental inheritance or occasionally paternal transmission (i.e. paternal leakage) of organelles has been reported in some plant species (Greiner et al., 2015). In these cases, chloroplast and mitochondrial DNAs could reflect different evolutionary histories of plants and be complementary (e.g. Govindarajulu et al., 2015). For instance, such incongruence between chloroplast and mitochondrial phylogenies has been reported in Pinaceae, in which organelles have contrasting inheritance modes (Xiang et al., 2015). A comparison of chloroplast and mitochondrial phylogenies of Oleaceae may thus be useful to test for possible paternal leakage of organelles during Oleaceae evolution (e.g. Sodmergen et al., 1998; Li et al., 2004) and also to analyze potential heterogeneous patterns of nucleotide substitutions in plastomes and mitogenomes.

Furthermore, previous studies based on Restriction Fragment Length Polymorphisms (RFLPs) and full or partial mitochondrial DNA sequencing have shown that the mitogenome structure can vary even at the species level (e.g. Smith and Keeling, 2015). The study of this intraspecific variation has allowed detecting genes putatively involved in cytoplasmic male sterility (e.g. Liu et al., 2011; Mower et al., 2012a; Hiroshi et al., 2014; Touzet and Meyer, 2014). In Oleaceae, three cytoplasmic lineages were identified with RFLPs in the Mediterranean olive tree (Besnard et al., 2000). One of these lineages (i.e. Lineage E3; formerly known as CCK-MCK) has been shown to be associated to a male sterility phenotype (*ms2*), for which a nucleo-cytoplasmic genetic determinism was demonstrated (Besnard et al., 2000). The comparative analysis of olive mitogenomes may thus be helpful to identify putative genes involved in this male sterility phenotype (C. Van de Paer et al., in prep.).

Acknowledgments

This work was funded by the Regional Council Midi-Pyrenees (AAP 13053637, 2014-EDB-UT3-DOCT) and the LABEX entitled TULIP managed by Agence Nationale de la Recherche (ANR-10-LABX-0041). We also thank Jan Hackel for helpful comments, and the Missouri Botanical Garden (MO) for permission to extract DNA from a leaf of *Hesperelaea*.

Appendix A. Supplementary data

Supplementary data to this article can be found online at http://dx.doi.org/10.1016/j.gene.2016.09.007.

References

Adams, K.L., Palmer, J.D., 2003. Evolution of mitochondrial gene content: gene loss and transfer to the nucleus. Mol. Phylogenet. Evol. 29, 380–395.
Bakker, F.T., Lei, D., Yu, J., Mohammadin, S., Wei, Z., van de Kerke, S., Gravendeel, B., Nieuwenhuis, M., Staats, M., Alquezar-Planas, D.E., Holmer, R., 2016. Herbarium genomics: plastome sequence assembly from a range of herbarium specimens using an Iterative Organelle Genome Assembly pipeline. Biol. J. Linn. Soc. 117, 33–43.
Besnard, G., Khadari, B., Villemur, P., Bervillé, A., 2000. Cytoplasmic male sterility in the olive (*Olea europaea* L.). Theor. Appl. Genet. 100, 1018–1024.
Besnard, G., Khadari, B., Baradat, P., Bervillé, A., 2002. Combination of chloroplast and mitochondrial DNA polymorphisms to study cytoplasm genetic differentiation in the olive complex (*Olea europaea* L.). Theor. Appl. Genet. 105, 139–144.
Besnard, G., Christin, P.A., Malé, P.J., Coissac, E., Ralimanana, H., Vorontsova, M.S., 2013. Phylogenomics and taxonomy of Lecomtelleae (Poaceae), an isolated, early diverging panicoid tribe from Madagascar. Ann. Bot. 112, 1057–1066.

Besnard, G., Christin, P.A., Malé, P.J., Coissac, E., Lhuillier, E., Lauzeral, C., Vorontsova, M.S., 2014. From museums to genomics: old herbarium specimens shed light on a C_3 to C_4 transition. J. Exp. Bot. 65, 6711–6721.
Biomatters Ltd., 2015. Geneious v. 9.0.5. Available at http://www.geneious.com.
Birky Jr., C.W., 1995. Uniparental inheritance of mitochondrial and chloroplast genes: mechanisms and evolution. Proc. Natl. Acad. Sci. U. S. A. 92, 11331–11338.
Bock, D.G., Kane, N.C., Ebert, D.P., Rieseberg, L.H., 2014. Genome skimming reveals the origin of the Jerusalem Artichoke tuber crop species: neither from Jerusalem nor an artichoke. New Phytol. 201, 1021–1030.
Cahill, M.J., Köser, C.U., Ross, N.E., Archer, J.A., 2010. Read length and repeat resolution: exploring prokaryote genomes using next-generation sequencing technologies. PLoS ONE 5, e11518.
Corriveau, J.L., Coleman, A.W., 1988. Rapid screening method to detect potential biparental inheritance of plastid DNA and results for over 200 angiosperms. Am. J. Bot. 75, 1443–1458.
Cronn, R., Liston, A., Parks, M., Gernandt, D.S., Shen, R., Mockler, T., 2008. Multiplex sequencing of plant chloroplast genomes using Solexa sequencing-by-synthesis technology. Nucleic Acids Res. 36, e122.
Dowling, D.K., Friberg, U., Lindell, J., 2008. Evolutionary implications of non-neutral mitochondrial genetic variation. Trends Ecol. Evol. 23, 546–554.
Edgar, R.C., 2004. Muscle: a multiple sequence alignment method with reduced time and space complexity. BMC Bioinformatics 5, 113.
Govindarajulu, R., Parks, M., Tennessen, J.A., Liston, A., Ashman, T.L., 2015. Comparison of nuclear, plastid, and mitochondrial phylogenies and the origin of wild octoploid strawberry species. Am. J. Bot. 102, 544–554.
Greiner, S., Sobanski, J., Bock, R., 2015. Why are most organelle genomes transmitted maternally? BioEssays 37, 80–94.
Gualberto, J.M., Mileshina, D., Wallet, C., Niazi, A.K., Weber-Lotfi, F., Dietrich, A., 2014. The plant mitochondrial genome: dynamics and maintenance. Biochimie 100, 107–120.
Hao, W., Palmer, J.D., 2009. Fine-scale mergers of chloroplast and mitochondrial genes create functional, transcompartmentally chimeric mitochondrial genes. Proc. Natl. Acad. Sci. U. S. A. 106, 16728–16733.
Harrison, N., Kidner, C.A., 2011. Next-generation sequencing and systematics: what can a billion base pairs of DNA sequence data do for you? Taxon 60, 1552–1566.
Hiroshi, Y., Yoshiyuki, T., Toru, T., 2014. Complete mitochondrial genome sequence of black mustard (*Brassica nigra*; BB) and comparison with *Brassica oleracea* (CC) and *Brassica carinata* (BBCC). Genome 57, 577–582.
Hong-Wa, C., Besnard, G., 2014. Species limits and diversification in the Madagascar olive (*Noronhia*, Oleaceae). Bot. J. Linn. Soc. 174, 141–161.
Illumina Inc., 2010. HiSeq™ 2000 sequencing system. http://www.illumina.com/documents/products/datasheets/datasheet_hiseq2000.pdf.
Krzywinski, M., Schein, J., Birol, I., Connors, J., Gascoyne, R., Horsman, D., Jones, S.J., Marra, M.A., 2009. Circos: an information aesthetic for comparative genomics. Genome Res. 19, 1639–1645.
Kurtz, S., 2016. The Vmatch large scale sequence analysis software. Available at http://vmatch.de.
Li, Y., Cui, H., Zhang, Q., Sodmergen, 2004. Divergent potentials for cytoplasmic inheritance within the genus *Syringa*. A new trait associated with speciogenesis. Plant Physiol. 136, 2762–2770.
Liu, H., Cui, P., Zhan, K., Lin, Q., Zhuo, G., Guo, X., Ding, F., Yang, W., Liu, D., Hu, S., Yu, J., Zhang, A., 2011. Comparative analysis of mitochondrial genomes between a wheat K-type cytoplasmic male sterility (CMS) line and its maintainer line. BMC Genomics 12, 163.
Lohse, M., Drechsel, O., Bock, R., 2007. OrganellarGenomeDRAW (OGDRAW): a tool for the easy generation of high-quality custom graphical maps of plastid and mitochondrial genomes. Curr. Genet. 52, 267–274.
Ma, P.F., Guo, Z.H., Li, D.Z., 2012. Rapid sequencing of the bamboo mitochondrial genome using Illumina technology and parallel episodic evolution of organelle genomes in grasses. PLoS ONE 7, e30297.
Malé, P.J., Bardon, L., Besnard, G., Coissac, E., Delsuc, F., Engel, J., Lhuillier, E., Scotti-Saintagne, C., Tinaut, A., Chave, J., 2014. Genome skimming by shotgun sequencing helps resolve the phylogeny of a pantropical tree family. Mol. Ecol. Resour. 14, 966–975.
Mower, J.P., Touzet, P., Gummow, J.S., Delph, L.F., Palmer, J.D., 2007. Extensive variation in synonymous substitution rates in mitochondrial genes of seed plants. BMC Evol. Biol. 7, 1.
Mower, J.P., Case, A.L., Floro, E.R., Willis, J.H., 2012a. Evidence against equimolarity of large repeat arrangements and a predominant master circle structure of the mitochondrial genome from a monkeyflower (*Mimulus guttatus*) lineage with cryptic CMS. Genome Biol. Evol. 4, 670–686.
Mower, J.P., Sloan, D.B., Alverson, A.J., 2012b. Plant mitochondrial genome diversity: the genomics revolution. In: Wendel, J.F., Greilhuber, J., Dolezel, J., Leitch, I.J. (Eds.), Plant Genome Diversity Plant Genomes: Their Residents, and Their Evolutionary Dynamics vol. 1. Springer, Vienna, pp. 123–144.
Olson, M.S., McCauley, D.E., 2000. Linkage disequilibrium and phylogenetic congruence between chloroplast and mitochondrial haplotypes in *Silene vulgaris*. Proc. R. Soc. Lond. B 267, 1801–1808.
Rambaut, A., 2014. FigTree 1.4.2. Available from http://tree.bio.ed.ac.uk.
Rand, D.M., Haney, R.A., Fry, A.J., 2004. Cytonuclear coevolution: the genomics of cooperation. Trends Ecol. Evol. 19, 645–653.
Refulio-Rodriguez, N.F., Olmstead, R.G., 2014. Phylogeny of Lamiidae. Am. J. Bot. 101, 287–299.
Ruhlman, T.A., Jansen, R.K., 2014. The plastid genomes of flowering plants. In: Maliga, P. (Ed.), Chloroplast Biotechnology: Methods and ProtocolsMethods in Molecular Biology vol. 1132. Springer, New York, pp. 3–38.
Satoh, M., Kubo, T., Mikami, T., 2006. The Owen mitochondrial genome in sugar beet (*Beta vulgaris* L.): possible mechanisms of extensive rearrangements and the origin of the mitotype-unique regions. Theor. Appl. Genet. 113, 477–484.
Smith, D.R., Keeling, P.J., 2015. Mitochondrial and plastid genome architecture: reoccurring themes, but significant differences at the extremes. Proc. Natl. Acad. Sci. U. S. A. 112, 10177–10184.
Sodmergen, Bai, H.H., He, J.X., Kuroiwa, H., Kawano, S., Kuroiwa, T., 1998. Potential for biparental cytoplasmic inheritance in *Jasminum officinale* and *Jasminum nudiflorum*. Sex. Plant Reprod. 11, 107–112.
Staats, M., Erkens, R.H.J., van de Vossenberg, B., Wieringa, J.J., Kraaijeveld, K., Stielow, B., Geml, J., Richardson, J.E., Bakker, F.T., 2013. Genomic treasure troves: complete genome sequencing of herbarium and insect museum specimens. PLoS ONE 8, e69189.
Stamatakis, A., 2014. RAxML version 8: a tool for phylogenetic analysis and post-analysis of large phylogenies. Bioinformatics 30, 1312–1313.
Straub, S.C.K., Parks, M., Weitemier, K., Fishbein, M., Cronn, R.C., Liston, A., 2012. Navigating the tip of the genomic iceberg: next-generation sequencing for plant systematics. Am. J. Bot. 99, 349–364.
Tang, M., Chen, Z., Grover, C.E., Wang, Y., Li, S., Liu, G., Ma, Z., Wendel, J.F., Hua, J., 2015. Rapid evolutionary divergence of *Gossypium barbadense* and *G. hirsutum* mitochondrial genomes. BMC Genomics 16, 1.
Touzet, P., Meyer, E.H., 2014. Cytoplasmic male sterility and mitochondrial metabolism in plants. Mitochondrion 19, 166–171.
Wallander, E., Albert, V.A., 2000. Phylogeny and classification of Oleaceae based on *rps16* and *trn*L-F sequence data. Am. J. Bot. 87, 1827–1841.
Warren, J.M., Simmons, M.P., Wu, Z., Sloan, D.B., 2016. Linear plasmids and the rate of sequence evolution in plant mitochondrial genomes. Genome Biol. Evol. 8, 364–374.
Watson, S., 1876. Botanical contributions. I. On the flora of Guadalupe Island, Lower California. Proc. Am. Acad. Arts Sci. 11, 105–112.
Wolfe, K.H., Li, W.H., Sharp, P.M., 1987. Rates of nucleotide substitution vary greatly among plant mitochondrial, chloroplast and nuclear DNA. Proc. Natl. Acad. Sci. U. S. A. 84, 9054–9058.
Wortley, A.H., Rudall, P.J., Harris, D.J., Scotland, R.W., 2005. How much data are needed to resolve a difficult phylogeny? Case study in Lamiales. Syst. Biol. 54, 697–709.
Xiang, Q.P., Wei, R., Shao, Y.Z., Yang, Z.Y., Wang, X.Q., Zhang, X.C., 2015. Phylogenetic relationships, possible ancient hybridization, and biogeographic history of *Abies* (Pinaceae) based on data from nuclear, plastid, and mitochondrial genomes. Mol. Phylogenet. Evol. 82, 1–14.
Xu, H., Luo, X., Qian, J., Pang, X., Song, J., Qian, G., Chen, J., Chen, S., 2012. FastUniq: a fast de novo duplicates removal tool for paired short reads. PLoS ONE 7, e52249.
Yang, Z., 2007. PAML 4: phylogenetic analysis by maximum likelihood. Mol. Biol. Evol. 24, 1586–1591.
Zedane, L., Hong-Wa, C., Murienne, J., Jeziorski, C., Baldwin, B.G., Besnard, G., 2016. Museomics illuminate the history of an extinct, paleoendemic plant lineage (*Hesperelaea*, Oleaceae) known from an 1875 collection on Guadalupe Island, Mexico. Biol. J. Linn. Soc. 117, 44–57.
Zhu, A., Guo, W., Jain, K., Mower, J.P., 2014. Unprecedented heterogeneity in the synonymous substitution rate within a plant genome. Mol. Biol. Evol. 31, 1228–1236.

CHAPITRE II

Prospects on the evolutionary mitogenomics of plants: a case study on the olive family (Oleaceae)

VAN DE PAER C., BOUCHEZ O., BESNARD G.

Molecular Ecology Resources, 2017, 1-17

Received: 19 June 2017 | Revised: 14 November 2017 | Accepted: 18 November 2017

DOI: 10.1111/1755-0998.12742

RESOURCE ARTICLE

WILEY MOLECULAR ECOLOGY RESOURCES

Prospects on the evolutionary mitogenomics of plants: A case study on the olive family (Oleaceae)

Céline Van de Paer[1] | Olivier Bouchez[2] | Guillaume Besnard[1]

[1]CNRS, Université de Toulouse, ENSFEA, IRD, UMR5174 EDB (Laboratoire Évolution & Diversité Biologique), Toulouse, France

[2]INRA, US 1426, GeT-PlaGe, Genotoul, Castanet-Tolosan, France

Correspondence
Céline Van de Paer and Guillaume Besnard, CNRS, Université de Toulouse, ENFA, UMR5174 EDB (Laboratoire Évolution & Diversité Biologique), Toulouse, France.
Emails: vandepaer.celine@gmail.com; guillaume.besnard@univ-tlse3.fr

Funding information
Agence Nationale de la Recherche, Grant/Award Number: ANR-10-LABX-0041; Investissement d'Avenir grant from the Agence Nationale de la Recherche, Grant/Award Number: CEBA: ANR-10-LABX-25-01; Regional Council Midi-Pyrénées, Grant/Award Number: AAP 13053637, 2014-EDB-UT3-DOCT; PhyloAlps project

Abstract

The mitogenome is rarely used to reconstruct the evolutionary history of plants, contrary to nuclear and plastid markers. Here, we evaluate the usefulness of mitochondrial DNA for molecular evolutionary studies in Oleaceae, in which cases of cytoplasmic male sterility (CMS) and of potentially contrasted organelle inheritance are known. We compare the diversity and the evolution of mitochondrial and chloroplast genomes by focusing on the olive complex and related genera. Using high-throughput techniques, we reconstructed complete mitogenomes (ca. 0.7 Mb) and plastomes (ca. 156 kb) for six olive accessions and one *Chionanthus*. A highly variable organization of mitogenomes was observed at the species level. In olive, two specific chimeric genes were identified in the mitogenome of lineage E3 and may be involved in CMS. Plastid-derived regions (*mtpt*) were observed in all reconstructed mitogenomes. Through phylogenetic reconstruction, we demonstrate that multiple integrations of *mtpt* regions have occurred in Oleaceae, but *mtpt* regions shared by all members of the olive complex derive from a common ancestor. We then assembled 52 conserved mitochondrial gene regions and complete plastomes of ten additional accessions belonging to tribes *Oleeae*, *Fontanesieae* and *Forsythieae*. Phylogenetic congruence between topologies based on mitochondrial regions and plastomes suggests a strong disequilibrium linkage between both organellar genomes. Finally, while phylogenetic reconstruction based on plastomes fails to resolve the evolutionary history of maternal olive lineages in the Mediterranean area, their phylogenetic relationships were successfully resolved with complete mitogenomes. Overall, our study demonstrates the great potential of using mitochondrial DNA in plant phylogeographic and metagenomic studies.

KEYWORDS
cytoplasmic male sterility, genome skimming, metagenomics, mitogenome, *mtpt*, *Olea europaea*, *Oleinae*, phylogeny, plastome

1 | INTRODUCTION

Molecular phylogenetics is a fundamental tool to study the evolutionary history of living organisms, the assembly of communities and the evolution of genes and genomes (e.g., Avise, 2009; Paterson, Freeling, Tang, & Wang, 2010; Pillar & Duarte, 2010). Different genomic regions can be used for phylogenetic reconstruction and/or DNA barcoding-based identification, but these often produce conflicting signals as a consequence of incomplete ancestral lineage sorting and reticulate evolution (including hybridization and horizontal gene transfer), as well as other more complex processes such as molecular convergence (due to selection or mutational biases) and heterogeneity of evolutionary rates among lineages (e.g., Christin, Besnard, Edwards, & Salamin, 2012; Delsuc, Brinkmann, & Philippe,

2005; Wendel & Doyle 1998). Confronting phylogenetic signals from cytoplasmic and nuclear genomic compartments is thus essential to understand the evolutionary history of species and their genomes. In plants, while ribosomal DNA has been extensively used as a nuclear phylogenetic marker (e.g., Álvarez & Wendel, 2003; Nieto Feliner & Rosselló, 2007), plastid genes, and since recently, full chloroplast genomes (hereafter plastomes), are now the markers of choice for reconstructing phylogenies and/or species identification (Coissac, Hollingsworth, Lavergne, & Taberlet, 2016; Hollingsworth et al., 2009; Li et al., 2014).

Unlike chloroplast and nuclear genes, polymorphism in the mitochondrial genome (hereafter mitogenome) is not frequently used to reconstruct phylogenies and phylogeographies or for DNA barcoding in higher plants, which stands in contrast with studies on animals (e.g., Donnelly et al., 2017; Duminil, 2014; Govindarajulu, Parks, Tennessen, Liston, & Ashman, 2015; Mower, Sloan, & Alverson, 2012; Qiu et al., 2006). This is due to three main reasons: First, mitochondrial genes of plants usually evolve slowly compared to those of the plastome (three to four times slower) or the nuclear genome (~12 times slower; Wolfe, Li, & Sharp, 1987; Palmer & Herbon, 1988; Palmer, 1992). The rate of nucleotide substitutions in coding regions of a plant's mitogenome has been estimated to be around 100 times lower than in animal mitogenomes and four to six times lower than in fungal mitogenomes (Aguileta et al., 2014; Nabholz, Glémin, & Galtier, 2009; Wolfe et al., 1987). However, studies have shown that substitution rates can be highly variable among mitochondrial genes and between species (Mower, Touzet, Gummow, Delph, & Palmer, 2007; Zhu, Guo, Jain, & Mower, 2014). Second, both organellar genomes (i.e., mitogenome and plastome) are usually inherited maternally (e.g., Corriveau & Coleman, 1988; Reboud & Zeyl, 1994). Consequently, mitochondrial DNA (mtDNA) and chloroplast or plastid DNA (ptDNA) polymorphisms are supposedly in strong linkage disequilibrium and may thus reveal similar phylogenetic patterns reflecting seed dispersal (e.g., Desplanque et al., 2000; Olson & McCauley, 2000), making the combined analysis of both organellar genomes not necessarily useful (although potentially informative for demographic analyses). Differential inheritance of mitochondria and chloroplasts, however, can reveal divergent evolutionary histories of cytoplasmic genomes. Third, the complete assembly of plant mitogenomes remains challenging due to their complex and variable structure (Smith & Keeling, 2015). For example, in the angiosperm order Lamiales, only six complete mitogenomes versus 72 plastomes were available on GenBank as of April 2017.

Despite the above-mentioned features that discourage the analysis of mitogenomes, comparative studies of both organellar genomes should be relevant to reveal their potentially contrasted evolutionary histories and structural dynamics. First, although maternal transmission of organelles prevails in plants (e.g., Corriveau & Coleman, 1988; Reboud & Zeyl, 1994), the inheritance of chloroplasts is predominantly paternal in conifers and a few angiosperms (e.g., *Medicago sativa*, *Actinidia* spp.; Rusche, Mogensen, Zhu, & Smith, 1995; Testolin & Cipriani, 1997). Paternal leakage (i.e., occasional paternal transmission of plastids and/or mitochondria via pollen) has also been reported in distantly related groups such as Geraniaceae (Weihe, Apitz, Pohlheim, Salinas-Hartwig, & Börner, 2009), Caryophyllaceae (Pearl, Welch, & McCauley, 2009) or Solanaceae (Svab & Maliga, 2007; Thyssen, Svab, & Maliga, 2012). Contrasted inheritance of cytoplasmic genomes can lead to incongruent phylogenetic signal from ptDNA and mtDNA, as in Actinidiaceae (Chat, Jáuregui, Petit, & Nadot, 2004) and Pinaceae (Xiang et al., 2015). Second, unlike plastomes, plant mitogenomes are highly variable in size and structural organization, ranging from ~200 kb in *Brassica rapa* (Brassicaceae) to ~11 Mb in *Silene conica* (Caryophyllaceae; Smith & Keeling, 2015). Much of this variation is due to a large proportion of noncoding and repeated sequences (Gualberto et al., 2014; Smith & Keeling, 2015). Numerous repeated sequences promote recombination and generate multiple possible configurations of mitogenomes (Lonsdale, Brears, Hodge, Melville, & Rottmann, 1988; Palmer & Herbon, 1988). Rearrangements are often at the origin of new open reading frames (ORFs), notably chimeric genes, which can generate nuclear–mitochondrial incompatibilities such as cytoplasmic male sterility (CMS; Newton, 1988; Hanson & Bentolila, 2004). Mitogenomes of higher plants also contain a variable portion of plastid-derived regions (*mtpt*), ranging from 1% (~4 kb) to 11.5% (~113 kb) in *Arabidopsis* and *Curcubita*, respectively (Alverson et al., 2010; Marienfeld, Unseld, & Brennicke, 1999; Unseld, Marienfeld, Brandt, & Brennicke, 1997). Unravelling the evolutionary history of *mtpt* regions can be challenging due to their recurrent losses and gains among species (Hao & Palmer, 2009; Sloan & Wu, 2014). Moreover, it has been shown that *mtpt* regions and their plastid homologues are subject to frequent gene conversion (Hao, Richardson, Zheng, & Palmer, 2010; Sloan & Wu, 2014), potentially blurring the phylogenetic signal deduced from those regions (Hao et al., 2010; Sloan & Wu, 2014). Although tRNAs of plastid origin appear to retain their functionality, most *mtpt* regions are noncoding or pseudogenes (Hao & Palmer, 2009; Notsu et al., 2002), sometimes integrated in coding sequences of functional mitochondrial genes (Wang, Rousseau-Gueutin, & Timmis, 2012). In contrast, the transfer of mtDNA to the plastome has been reported in some rare cases only, such as in carrot (Iorizzo et al., 2012), milkweed (Straub, Cronn, Edwards, Fishbein, & Liston, 2013) and a few Poaceae (Burke et al., 2016; Ma, Zhang, Guo, & Li, 2015).

The number of studies that compare the diversity of both organellar genomes at the family level remains limited (e.g., Smith & Keeling, 2015), and more investigations are necessary to understand the evolution of organellar genomes in various plant groups. Here, we report a study on the olive family (Lamiales: Oleaceae), which encompasses more than 600 species in 25 genera (Green 2004). The family has a worldwide distribution, occurring in tropical, subtropical and temperate climates (Stevens 2001). Several species are of major economic interest, such as the Mediterranean olive (*Olea europaea* L.) and the ash tree (*Fraxinus excelsior* L.), while numerous other Oleaceae are widely used as ornamentals or fragrant species (i.e., the genera *Chionanthus*, *Forsythia*, *Fraxinus*, *Jasminum*, *Ligustrum*, *Osmanthus* and *Syringa*). Taxa of Oleaceae are currently classified in five tribes, namely *Fontanesieae*, *Forsythieae*, *Jasmineae*, *Myxopyreae* and *Oleeae* (Wallander & Albert, 2000). *Oleeae*, including the olive, is the most speciose lineage, with four subtribes currently recognized:

Ligustrinae, *Fraxininae*, *Schreberinae* and *Oleinae* (Wallander & Albert, 2000). Phylogenetic trees of Oleaceae have been reconstructed with chloroplast and nuclear ribosomal DNA markers (e.g., Hong-Wa & Besnard, 2013; Wallander & Albert, 2000; Zedane et al., 2016). Some nodes have yet to be resolved, while phylogenetic signals from nuclear and chloroplast DNA markers were partly incongruent. This calls for additional phylogenetic investigations into the evolutionary history of this plant family. In addition, while strict maternal inheritance and male sterility with a nucleo-cytoplasmic determinism have been reported in olive (Besnard, Khadari, Villemur, & Bervillé, 2000), potential paternal leakage of organelles has been suggested for two unrelated lineages [*Jasminum* spp. (Sodmergen et al., 1998), and *Ligustrinae* (Liu, Cui, & Zhang, 2004)]. Oleaceae is thus an excellent study case to compare plastomes and mitogenomes, as their evolution may be potentially affected by contrasted inheritance modes of organelles and nucleo-cytoplasmic interactions.

In this study, we first aim to investigate the variation and evolutionary dynamics of mitogenome organization at the species level (i.e., the olive complex). We then evaluate the usefulness of mtDNA for molecular evolutionary studies at different taxonomic levels (from tribe to species). The organelle inheritance of accessions used here is assumed to be uniparental and maternal. As a consequence, we expected congruence between topologies deduced from mtDNA and ptDNA regions. To address these objectives, we develop a methodology to simultaneously characterize the diversity and evolution of plastomes and mitogenomes in Oleaceae. We first use high-throughput sequencing to assemble organellar genomes of Mediterranean olive lineages and another *Oleinae* species. The assembly of mitogenomes allows us to examine differences in size, gene content and structure of this organellar genome within the same species and between related Oleaceae genera. In olive, a CMS phenotype has been specifically associated to lineage E3 (Besnard et al., 2000). Here, we identify ORFs specific to this lineage as putatively determining the CMS. We then analyse ptDNA integration in mitogenomes of *Oleinae* through phylogenetic reconstruction from the plastid-derived sequences (*mtpt* regions). Finally, we compared nucleotide diversity and phylogenetic topologies based on conserved mtDNA regions and complete plastomes assembled from distantly related species of Oleaceae. The combined use of full plastomes and mitogenomes to resolve phylogenetic relationships between maternal lineages of the Mediterranean olive is also tested. The utility of the mitochondrial DNA in plant phylogeographic and metagenomic studies is briefly discussed.

2 | MATERIALS AND METHODS

2.1 | Plant material and DNA sequencing

We first aimed to investigate mitogenome variation at the species level in *Olea europaea*. We thus focused on the wild olive tree because cytoplasmic lineages have been previously described within the *O. europaea* complex (Besnard, Rubio de Casas, & Vargas, 2007). Two main clades have been identified: (i) the Mediterranean/North African clade (Cp-II) that includes lineages E1, E2, E3 and M, and (ii) the *cuspidata* clade (Cp-I) which extends from Southern Asia to Tropical Africa and includes lineages C1, C2 and A. Each lineage shows a well-delimited geographic distribution (Besnard et al., 2007). Sequence variation in the olive plastomes was not, however, sufficient to resolve relationships between lineages of clade Cp-II (Besnard et al., 2013). Chloroplasts and mitochondria are maternally inherited in olive, and as expected, the analysis of their polymorphisms allowed congruent identification of cytoplasmic lineages (Besnard, Khadari, Baradat, & Bervillé, 2002). Interestingly, a CMS phenotype (*ms2*) is specifically associated with lineage E3 (formerly named CCK-MCK; Besnard et al., 2000). We here analysed six accessions of *O. europaea* to represent five cytoplasmic lineages of the olive complex (Table 1; Besnard et al., 2007; Besnard, Terral, & Cornille, 2018): three accessions of subsp. *europaea* ["Stavrovouni Monastery 11" (Cp-II, lineage E1, sublineage *e*, haplotype 4; hereafter E1-*e*.4), "Vallée du Fango 5" (Cp-II, lineage E2, haplotype 1; hereafter E2.1), "Oeiras 1" (Cp-II, lineage E3, haplotype 1; hereafter E3.1)], one of subsp. *laperrinei* ("Adjelella 10"; Cp-II, lineage E1, sublineage *l*1, haplotype 1; hereafter E1-*l*1.1), one of subsp. *guanchica* ("La Gomera 10"; Cp-II, lineage M; sublineage *g*1, haplotype 3; hereafter M-*g*1.3) and one of subsp. *cuspidata* ("Menagesha Forest 14"; Cp-I, lineage C2, haplotype 13; hereafter C2.13).

To evaluate the phylogenetic utility of mtDNA on the family level, and especially in the olive subtribe (*Oleinae*), we then extended our sample to represent more distantly related lineages of Oleaceae (Table 1). In addition to *Hesperelaea palmeri* (for which organellar genomes have been already released on GenBank; Van de Paer, Hong-Wa, Jeziorski, & Besnard, 2016; Zedane et al., 2016), we included nine species representative of different *Oleeae* lineages, plus one representative of two other Oleaceae tribes: *Fontanesia fortunei* (*Fontanesieae*) and *Forsythia* × *intermedia* (*Forsythieae*). Four *Oleeae* samples were herbarium specimens collected between 2001 and 2012 (Table 1). The eight remaining samples were extracted from recent silica-dried leaves.

For all studied samples, total genomic DNA was extracted from ca. 5 mg of dried leaves. Each sample was ground in a 2-ml tube containing three tungsten beads with a TissueLyser (Qiagen Inc., Texas). The BioSprint 15 DNA Plant Kit (Qiagen Inc.) was then used to extract DNA, which was eluted in 200 μl AE buffer. Shotgun sequencing was performed using a genome skimming approach as described in Zedane et al. (2016). DNA extracted from herbarium specimens was highly fragmented and thus not sonicated before library construction. Sequencing libraries were multiplexed (24 per flow cell) and sequenced with an Illumina machine (i.e., HiSeq 2000, 2500 or 3000; Table S1). Duplicated reads were removed with FastUniq (Xu et al., 2012), and overlapping paired-end reads were then merged using BBMerge [BBTOOLS implemented in GENEIOUS v. 9.0.5 (http://www.geneious.com; Kearse et al., 2012)].

2.2 | Assembly, annotation of organellar DNA and identification of *mtpt* regions

For all accessions, we first reconstructed full plastomes as described by Zedane et al. (2016). Read mapping against the obtained

TABLE 1 List of plant accessions analysed in our study. GenBank accession number of complete plastomes and mitogenomes is indicated. For the six olive accessions, the cytoplasmic DNA lineage and haplotypes (according to Besnard et al., 2007, 2018) are given in square brackets

		GenBank no	
Tribe, subtribe	**Species, and name and origin of analysed individuals**	**Plastome**	**Mitogenome**
Oleeae, Oleinae	*Olea europaea* L. subsp. *europaea*		
	- Stavrovouni Monastery 11, Cyprus [Cp-II, E1-*e*.4]	HF558645	MG372117
	- Vallée du Fango 5, Corsica, France [Cp-II, E2.1]	MG255762	MG372118
	- Oeiras 1, Lisbon, Portugal [Cp-II, E3.1]	MG255763	MG372119
	Olea europaea L. subsp. *laperrinei* (Batt. & Trab.) Cif.		
	- Adjelella 10, Hoggar, Algeria [Cp-II, E1-*l*1.1]	MG255765	MG372121
	Olea europaea L. subsp. *guanchica* P. Vargas et al.		
	- La Gomera 10, Canary Islands [Cp-II, M-*g*1.3]	MG255764	MG372120
	Olea europaea L. subsp. *cuspidata* (Wall. ex G. Don) Cif.		
	- Menagesha Forest 14, Ethiopia [Cp-I, C2.13]	MG255760	MG372116
	Chionanthus rupicolus (Lingelsh.) Kiew		
	- W. Takeuchi et al. 15149 [K] (2001), Papua New Guinea	MG255753	MG372115
	Chionanthus parkinsonii (Hutch.) Bennet & Raizada		
	- M. Van de Bult 1222 [M] (2012), Thailand	MG255752	—[a]
	Forestiera isabelae Hammel & Cornejo		
	- de Hammel & Perez 24248 [K] (2006), Costa Rica	MG255755	—[a]
	Hesperelaea palmeri A. Gray		
	- E. Palmer 81 [MO] (1875), Guadalupe Island, Mexico	LN515489	KX545367
	Nestegis apetala (Vahl) L.A.S. Johnson		
	- Kepos 45351 [K], New Zealand	MG255758	—[a]
	Noronhia lowryi C. Hong-Wa		
	- J. Razanatsoa 550 [TAN], Madagascar	MG255759	—[a]
	Olea exasperata Jacq.		
	- A. Costa 1 [MPU], South Africa	MG255766	—[a]
Oleeae, Fraxininae	*Fraxinus ornus* L.		
	- coll. B2-Cefe (Montpellier), France	MG255757	—[a]
Oleeae, Ligustrinae	*Syringa vulgaris* L.		
	- G. Besnard sn [MPU], ornamental, Europe	MG255768	—[a]
Oleeae, Schreberinae	*Schrebera arborea* A. Chev.		
	- R. Letouzey 5136 [K] (2005), Cameroon	MG255767	—[a]
Fontanesieae	*Fontanesia fortunei* Carrière		
	- T. Jossberger 1829 [BONN], China	MG255754	—[a]
Forsythieae	*Forsythia* × *intermedia* Zabel		
	- G. Besnard sn [MPU], ornamental, E. Asia	MG255756	—[a]

Voucher samples deposited at the herbaria of Kew [K], Munich [M], Bonn [BONN], Montpellier [MPU], Missouri [MO], Pretoria [PET] and Antananarivo [TAN].

[a]Accessions only used for the assembly of 52 conserved mitochondrial DNA regions (Table S6).

consensus sequence was also performed using GENEIOUS v. 9.0.5 (Biomatters Ltd., 2015) to check assembly quality and coverage depth. Then, we assembled full mitogenomes for the six *O. europaea* accessions and *Chionanthus rupicolus* in order to investigate their structural organization and evolution within the olive species and in related genera (i.e., *Chionanthus* and *Hesperelaea*). Mitogenomes have only in part or not at all been assembled in current Oleaceae genome-sequencing projects (Cruz et al., 2016; Sollars et al., 2017; Unver et al., 2017) due to their structural complexity, especially the presence of *mtpt* regions. Here, we followed the procedure described by Van de Paer et al. (2016) to assemble complete mitogenomes. The *H. palmeri* mitogenome [KX545367] was initially used as a reference to generate mitochondrial DNA contigs that were then used to assemble complete mitogenomes. In this process, *mtpt* regions were manually resolved by identifying mitochondrial reads based on polymorphisms (i.e., single nucleotide polymorphisms and

indels) and their lower sequencing depth compared to their plastid homologues (for the detailed procedure, see Van de Paer et al., 2016). The potential presence of another homologous nuclear copy does not interfere with *mtpt* assembly as coverage is much higher for the mitogenome compared to the nuclear genome (by at least 20 times; J.K. Olofsson et al. in preparation). Because no automated approach of full plant mitogenome assembly based on shotgun data is currently available, we then recovered selected mtDNA regions for our Oleaceae-wide sample. We focused on 52 conserved mtDNA regions (i.e., 36 protein-coding genes and 16 introns, for a total of ca. 55 kb) that were assembled for the remaining 10 species by mapping merged and unmerged reads against the annotated mitogenome of *H. palmeri* [KX545367] using GENEIOUS.

Annotation of plastomes was performed using GENEIOUS by transferring the annotations of an *O. europaea* plastome [FN996972]. In the same manner, we transferred the annotations of the *H. palmeri* mitogenome to annotate coding protein and rRNA mitochondrial genes and introns. Start and stop codons of protein-coding genes were manually adjusted. We identified tRNA genes in mitogenomes using ARAGORN (Laslett & Canback, 2004). Within each complete mitogenome, we also identified repeats with a 40-bp minimum length and with at least 95% identity using RepeatFinder (Volfovsky, Haas, & Salzberg, 2001). Physical maps of mitogenomes were drawn with OGDRAW v.1.2 (Lohse, Drechsel, & Bock, 2007) and CIRCOS (Krzywinski et al., 2009).

We finally analysed the integration of plastid DNA in the mitogenomes (*mtpt* regions) of all *O. europaea* accessions, *C. rupicolus* and *H. palmeri*. We also analysed *Boea hygrometrica* (Lamiales: Gesneriaceae), for which both mitogenome and plastome are available on GenBank [NC_016741 and NC_016468, respectively]. For each mitogenome, we detected *mtpt* regions using BLASTN with the corresponding plastome as a query sequence. We retained *mtpt* regions with a 100-bp minimum length and at least 85% identity with a plastid region. We then mapped all *mtpt* regions detected in the nine accessions against the E1-*e*.4 plastome.

2.3 | Identification of mitochondrial genes specific to the E3 olive lineage

As mentioned above, one cytoplasmic lineage (E3) is associated with a type of CMS in the olive (*mp2*; Besnard et al., 2000). The responsible cytoplasmic gene or genes remain unknown, but are supposed to be located in the mitogenome (Besnard et al., 2000). To identify candidate mitochondrial genes that may determine this CMS, we first identified genomic regions specific to the E3.1 mitogenome using the progressiveMauve algorithm (Darling, Mau, & Perna, 2010). All ORFs (open reading frames) with a minimum length of 150 bp were identified within these genomic regions as done by Mower, Case, Floro, and Willis (2012). ORFs were then compared to the identified genes in the E3.1 mitogenome using BLASTN (Altschul, Gish, Miller, Myers, & Lipman, 1990) with a cut-off *e*-value of 0.001 to check for homology with known genes and detect chimeric genes.

2.4 | Nucleotide diversity within plastid and mitochondrial genomes of olive and *Oleinae*

Nucleotide diversity (π; Nei & Li, 1979) was measured with MEGA (Tamura, Stecher, Peterson, Filipski, & Kumar, 2013) as an average number of substitutions per site between two homologous sequences. We computed π on plastid and mitochondrial sequence alignments, independently for protein-coding genes and noncoding regions. We also distinguished plastid regions from the single copies [i.e., long and short single copies (LSC and SSC, respectively)] and inverted repeats (IR). The standard error of π was estimated by bootstrapping on sequences (100 replicates). π was investigated at two taxonomical levels. First, we estimated π for the olive complex (species level), considering one member of each lineage (i.e., E1-*e*.4, E2.1, E3.1, M-*g*1.3 and C2.13). π of *mtpt* regions (that are considered orthologous in *O. europaea*; see results) and their plastid homologue regions was also compared. We then calculated π for protein-coding and noncoding regions at the subtribe level considering eight *Oleinae* species (i.e., *C. parkinsonii*, *C. rupicolus*, *For. isabelae*, *H. palmeri*, *Ne. apetala*, *No. lowryi*, *O. exasperata* and *O. e. europaea* E1-*e*.4).

2.5 | Phylogenetic analyses

Phylogenetic analyses were conducted to address three main issues. First, we investigated the evolution and origin of mtDNA regions showing high homology with plastomes (*mtpt* regions) in the olive subtribe (*Oleinae*). Second, we tested for the congruence of phylogenetic signal (i.e., topology and node support) from mitochondrial and chloroplast DNA at the tribe level (*Oleeae*). Finally, phylogenetic relationships between cytoplasmic lineages of the Mediterranean olive were investigated for the first time based on both full plastomes and full mitogenomes.

Plastid integration in mitogenomes of the olive subtribe (*Oleinae*): Phylogenetic signal from *mtpt* regions (detected in all *Oleinae*) and their plastid homologues may inform on approximately when plastid regions were integrated into the mitogenome. We thus investigated the relative placement of each *mtpt* region and its plastid homologue in phylogenies of *Oleeae* with *Boea hygrometrica* as out-group. The nine *mtpt* regions shared by the eight accessions of *Oleinae* (see results) were individually aligned with their homologous plastid regions (extracted from the plastome of 18 Oleaceae and *Boea hygrometrica*) using MUSCLE (Edgar, 2004). Based on these sequence alignments, we then reconstructed nine maximum-likelihood (ML) phylogenetic trees using RAxML v. 8 (Stamatakis, 2014) on the CIPRES platform (Miller, Pfeiffer, & Schwartz, 2010). We applied the GTR + G model and the rapid bootstrap algorithm with 1,000 iterations. We used FigTree v. 1.4.2 (Rambaut, 2014) to visualize the ML trees.

Plastid and mitochondrial DNA phylogenies of *Oleeae*: We investigated phylogenetic relationships between 16 species representative of the main *Oleeae* lineages and the two Oleaceae species belonging to *Fontanesieae* and *Forsythieae* used as out-groups (Table 1). Two phylogenetic trees were inferred (i) from complete plastomes, and (ii) from mitochondrial protein-coding genes and introns. Complete

plastomes of all 18 species were aligned using MUSCLE with default parameters. For the mitochondrial alignment, each assembled region (protein-coding gene or intron) was aligned individually using MUSCLE with default parameters. All aligned mitochondrial regions were then concatenated into one matrix. The alignments were manually refined, particularly in regions with short inversions (<100 bp). ML trees were reconstructed on both alignment data sets using RAxML as described above.

Phylogenetic relationships in the olive complex: lastly, phylogenetic trees were reconstructed for the olive complex using complete plastome and/or mitogenome sequences. First, we aligned the six complete *O. europaea* plastomes generated in this study, using MUSCLE with default parameters. We also performed a second olive plastome alignment including seven additional *O. europaea* accessions from GenBank (Table S2) and *O. woodiana* as out-group to check the impact of plant sampling on the topology. Second, an alignment of the six *O. europaea* mitogenomes was constructed. Conserved regions between highly rearranged genomes were detected and aligned using the progressiveMauve algorithm (Darling et al., 2010). Finally, the alignments of the six olive plastomes and mitogenomes generated in our study were also concatenated into one single matrix. ML phylogenetic trees were reconstructed on each data set using RAxML as described above.

3 | RESULTS

Shotgun sequencing produced between 3.2 and 58.4 million paired-end reads per accession (Table S1). As expected, the mean size of inserts was lower for DNA extracted from herbarium specimens (118 to 202 bp) than for DNA extracted from silica-dried samples (233 to 419 bp; Table S1). Using these data, we assembled 17 complete plastomes and seven complete mitogenomes (six for *O. europaea*). In addition, 52 conserved mitochondrial regions (genes and introns) were assembled for ten Oleaceae accessions. The lowest sequencing depth of mtDNA and ptDNA regions was observed in *Schrebera arborea* (i.e., 17.1× and 64×, respectively; Table S1). In each analysed sample, the sequencing depth of the mitogenome was always lower than that of the plastome, with a ratio between 4% (*Fo. fortunei*) and 65% (*C. parkinsonii*). Note that this ratio was higher for herbarium specimens (ranging from 10% to 65%; mean = 28.2%) than for silica-dried samples (ranging from 4% to 23%; mean = 10.6%), but we did not statistically test this difference due to the small sampling size.

3.1 | Mitogenome size and composition

The mitogenomes of *O. europaea* and *C. rupicolus* were assembled as single circular consensus sequences (i.e., master circle), with a length ranging from 667,135 (*C. rupicolus*) to 769,995 bp (*O. e. europaea* E3.1; Figure 1) and a GC content of 44.7%–44.8% (Table 2). While both mitogenomes belonging to lineage E1 (E1-*l*1.1 and E1-*e*.4) had a very similar structure, mitogenome organization was highly variable among olive cytoplasmic lineages (Figures S1 and S2).

Gene annotation with the *H. palmeri* mitogenome as reference identified 37 distinct protein-coding genes, three rRNA genes and 12 tRNA genes in all mitogenomes, except in the E3.1 olive mitogenome whose complete *sdh3* gene was missing (Tables 2 and S3). A partially duplicated copy of *ccmC* was also detected in E2.1, E3.1, M-*g*1.3 and C2.13 olive mitogenomes (noted Ψ-*ccmC*; Table S3). Several protein-coding and tRNA genes were found in duplicated regions and thus annotated two or three times (Table S3; Figures 1 and S1). Repeat content was highly variable among mitogenomes. While the mitogenome of *C. rupicolus* had only 3.85% repetitive sequences with no repeat >1 kb, the repeat content of E2.1 olive represented 20.38% of the mitogenome and contained two large repeats >10 kb (with one reaching ~25 kb; Figure S1, Tables 2 and S4).

The amount of *mtpt* regions in the mitogenomes of *O. europaea* was relatively constant between lineages (~4.7% of the mitogenome; Table 3), but lower than in *C. rupicolus* (5.2%), *H. palmeri* (8.5%) and *B. hygrometrica* (10.1%). Four conserved plastid-derived genes (*rpl23*, *petL*, *petN* and *psbD*; Table S3) were additionally found in *mtpt* regions of *O. europaea*. *petL* was, however, annotated as a pseudogene in the E3.1 mitogenome (Figure 1) as its coding sequence was longer than expected due to a degenerated stop codon. We only found the plastid-derived gene *petN* in *C. rupicolus* (Table S3). We also detected ten and eight plastid-derived tRNA genes in the *O. europaea* and *C. rupicolus* mitogenomes, respectively (Table S3).

Three genomic regions, ranging from 1,453 to 5,278 bp, were found to be specific to the E3.1 mitogenome. Two chimeric ORFs [*orf_E3a* (upstream of *cox3*) and *orf_E3b* (upstream of *atp6*)] were detected within these regions (Figure S3). Both are chimeric genes since *orf_E3a* contains a fragment of *sdh3* (271 bp; see above), while *orf_E3b* contains 1,225 bp of the plastid-derived gene *rbcL*. We also detected seven short ORFs of unidentified origin with length ranging from 162 to 270 bp (data not shown).

3.2 | Nucleotide diversity within organellar genomes among olive and *Oleinae*

A higher nucleotide diversity was observed in noncoding regions compared to protein-coding genes in ptDNA and mtDNA both within the olive complex and across *Oleinae* (Table 4). Within the plastome, lower nucleotide diversity in inverted repeats (IR) than in single copies was revealed for both coding and noncoding regions. Globally, nucleotide diversity was lower in mtDNA than in ptDNA. Within *Oleinae*, we estimated that mitochondrial sequences evolved three to five times slower than LSC and SSC plastid sequences (Table 4). However, *mtpt* regions of *O. europaea* showed higher nucleotide diversity (0.076, *SE* = 0.012) than their plastid homologous regions (0.024, *SE* = 0.009; Tables 4 and S5).

3.3 | Integration of *mtpt* regions in the olive mitogenome

The total length of *mtpt* regions ranged from 31,905 (in *O. e.* ssp. *cuspidata*) to 55,513 bp (in *B. hygrometrica*), covering 20.6%–35.2%

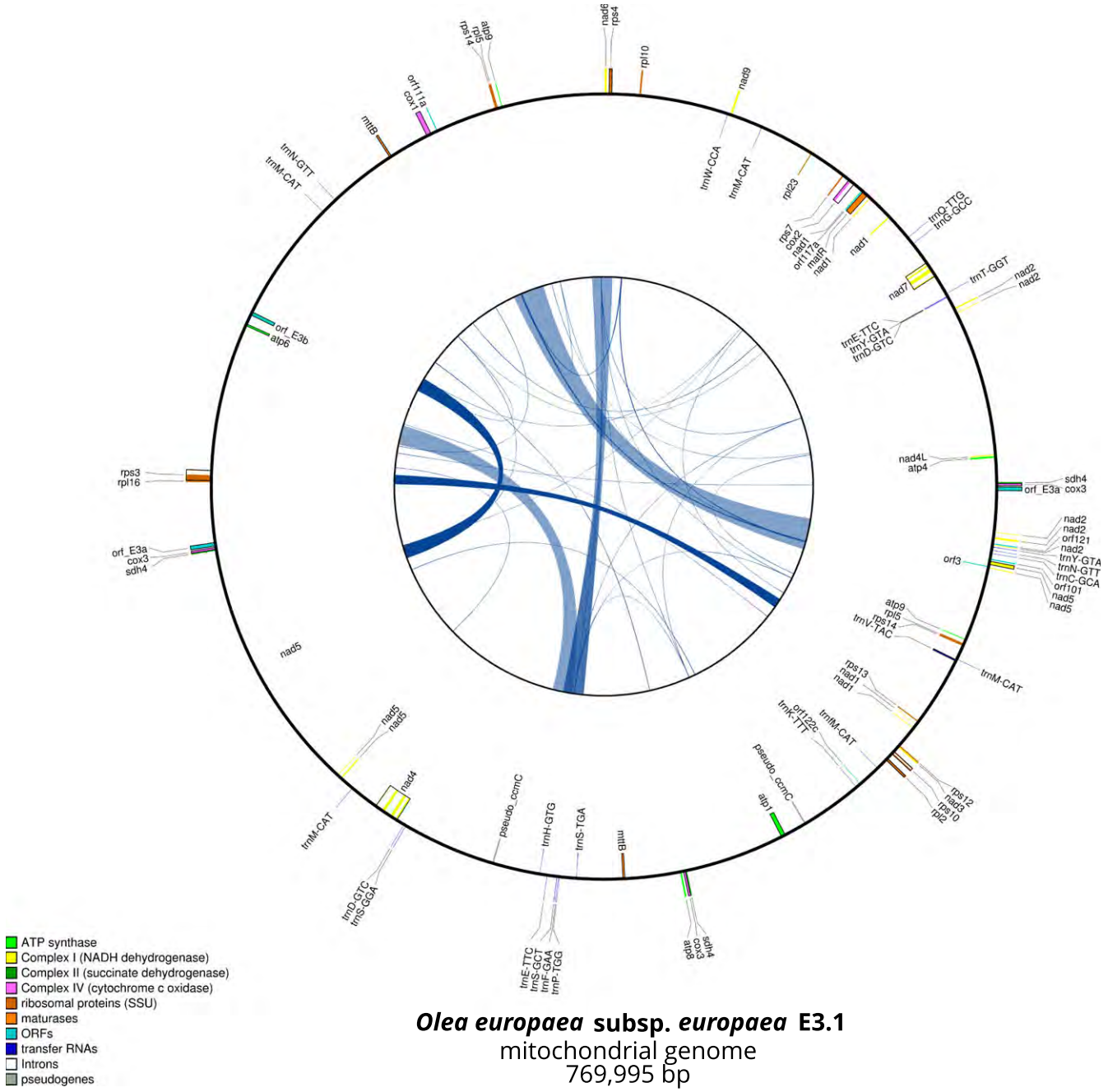


FIGURE 1 Master circle of olive mitogenome E3.1 (associated with CMS; Besnard et al., 2000). The outer circle represents the gene content, with protein-coding genes, rRNA and tRNA genes. The inner circle represents repeated sequences with a minimum length of 100 bp in the mitogenome. The large repeats >10 kb are represented in light blue colour. The maps were performed with OGDRAW v1.2 (Lohse et al., 2007) and CIRCOS (Krzywinski et al., 2009)

of the plastome (Figure 2). These regions had origins in both single copies (LSC and SSC) and the IR. Most boundaries of *mtpt* regions were shared between *O. europaea* accessions, while the boundaries of *mtpt* regions of *H. palmeri* and *C. rupicolus* were partially shared with those of *O. europaea* (Figure 2). In contrast, the *mtpt* regions detected in *B. hygrometrica* were differentially distributed along the E1-*e*.4 plastome compared to Oleaceae accessions (Figure 2).

We identified nine *mtpt* regions shared by all studied *Oleeae* accessions with length ranging from 495 to 3,198 bp (Figure 2). Six of these *mtpt* regions were also partly present in the mitogenome of *B. hygrometrica* (regions 1, 2, 3, 4, 6 and 9; Figure 2). ML phylogenies were reconstructed from each *mtpt* segment and their homologous regions extracted from the plastomes of *B. hygrometrica* and Oleaceae (Figures 3 and S4). In the nine topologies, *mtpt* sequences

TABLE 2 Comparison of genomic content among the mitogenomes of *Olea europaea* and *Chionanthus rupicolus* (sequence length in base pairs and per cent of the complete mitogenome sequence given in parentheses; GC content in per cent). We used the corresponding plastome of each accession as query to estimate the plastid-derived content of the mitogenomes

	O. e. laperrinei E1-*l*1.1	*O. e. europaea* E1-*e*.4	*O. e. europaea* E2.1	*O. e. guanchica* M-*g*1.3	*O. e. europaea* E3.1	*O. e. cuspidata* C2.13	*Chionanthus rupicolus*
Genome size (bp)	710,688	710,737	737,665	694,631	769,995	710,338	667,135
Protein-coding genes	33,810 (4.76)	33,834 (4.76)	36,633 (4.97)	31,929 (4.60)	35,643 (4.63)	33,387 (4.70)	31,764 (4.76)
Protein-coding genes-*mtpt*	1,530 (0.22)	1,530 (0,22)	1,434 (0.19)	1,530 (0.22)	1,530 (0.20)	1,530 (0.22)	90 (0.01)
tRNA genes	1,077 (0.15)	1,077 (0.15)	1,077 (0.15)	1,078 (0.16)	1,077 (0.14)	1,077 (0.15)	1,075 (0.16)
tRNA genes-*mtpt*	901 (0.13)	901 (0.13)	904 (0.12)	904 (0.13)	904 (0.12)	902 (0.13)	759 (0.11)
rRNA genes	5,491 (0.77)	5,491 (0.77)	5,491 (0.74)	5,491 (0.79)	5,491 (0.71)	5,491 (0.77)	5,486 (0.82)
Introns	23,464 (3.30)	23,468 (3.30)	23,460 (3.18)	20,214 (2.91)	23,462 (3.05)	23,461 (3.30)	23,387 (3.51)
Nonannotated sequences	644,415 (90.67)	644,434 (90.67)	667,509 (90.49)	632,541 (91.06)	700,402 (90.96)	644,018 (90.66)	604,574 (90.62)
Repeats content	76,758 (10.80)	76,750 (10.80)	150,369 (20.38)	35,874 (5.16)	130,744 (16.98)	87,343 (12.30)	25,704 (3.85)
Plastid-like sequences	33,698 (4.7)	33,696 (4.7)	33,211 (4.5)	32,545 (4.7)	37,560 (4.9)	31,905 (4.5)	34,905 (5.2)
GC content (%)	44.7	44.7	44.7	44.7	44.7	44.7	44.8

TABLE 3 Comparison of plastid-derived regions (*mtpt*) content among the mitogenomes of *Olea europaea*, *Chionanthus rupicolus*, *Hesperelaea palmeri* and *Boea hygrometrica* (number of *mtpt* regions, total length in base pairs, accumulative length of *mtpt* regions against length of the mitogenome and the plastome, in per cent)

Taxon	Number of *mtpt*	Total length	Amount of the mitogenome (%)	Amount of the plastome (%)
Boea hygrometrica	57	55,513	10.1	35.2
Hesperelaea palmeri	20	55,761	8.5	35.8
Chionanthus rupicolus	22	34,905	5.2	27.0
O. e. cuspidata C2.13	18	31,905	4.5	20.6
O. e. europaea E3.1	23	37,560	4.9	29.1
O. e. guanchica M-*g*1.3	21	32,545	4.7	25.2
O. e. europaea E2.1	21	33,211	4.5	25.7
O. e. europaea E1-*e*.4	20	33,696	4.7	21.8
O. e. laperrinei E1-*l*1.1	20	33,698	4.7	26.1

of olive accessions clustered together, suggesting they have derived from a common ancestor. In contrast, the phylogenetic placement of *Oleinae mtpt* sequences relative to their homologous plastid regions was not congruent between segments (Figures 3 and S4), indicating that they were not all integrated into the mitogenome at the same time: (i) first, one topology (*mtpt6*) supported the distinction of *mtpt* sequences of *B. hygrometrica* and Oleaceae from their plastid homologous regions (Figure 3b). This suggests that this *mtpt* region integrated in the mitogenome before the divergence between Gesneriaceae and Oleaceae; (ii) second, three topologies (regions corresponding to *mtpt1*, *mtpt3* and *mtpt8*) showed an *mtpt* clade of *Oleinae* sister to plastid homologous regions of *Oleeae* + *Fontanesieae* + *Forsythieae*, supporting the integration of these *mtpt* regions before the divergence of these three tribes; (iii) third, three topologies (regions corresponding to *mtpt4*, *mtpt5* and *mtpt7*) showed an *mtpt* clade of *Oleinae* sister to plastid homologous regions of *Oleinae*, supporting the integration of these *mtpt* regions before the diversification of the investigated taxa of *Oleinae* (but after the divergence of subtribe *Oleeae*; Figure 3a); and finally (iv) two topologies (regions corresponding to *mtpt2* and *mtpt9*) did not support a common origin of all *mtpt* regions of *Oleinae*, suggesting potential multiple integrations of these *mtpt* regions in the mitogenome, but this remains tentative due to the low node support.

3.4 | Reconstructing the phylogeny of *Oleeae* from plastid and mitochondrial DNA data

In our study, we produced an alignment of 18 complete plastomes (134,130 bp; after excluding one IR). In this alignment, we identified 6,572 variable sites of which 1,735 were parsimony-informative. In addition, 52 conserved mitochondrial DNA regions were aligned for the same 18 taxa (Table S6). A concatenated alignment of 55,301 bp was obtained. In this mtDNA data set, we detected 776 variable sites of which 212 were parsimony-informative.

Topologies of both plastid and mitochondrial ML phylogenies were congruent, except for two nodes with low support (Figure 4). The topology of the phylogeny derived from complete plastomes had >95% bootstrap support at most nodes (Figure 4a). Only intraspecific relationships between the olive lineages E1, E2 and M remained unresolved. The topology derived from mitochondrial regions was overall less supported, with bootstrap values inferior to 80% at seven nodes (Figure 4b). Both topologies differed in the placement of the clade including *N. apetala* and *C. rupicolus* that was strongly supported as sister to *Olea* (100%) in the ptDNA phylogeny

TABLE 4 Estimates of the nucleotide diversity (π in %; Nei & Li, 1979) among the olive complex and subtribe *Oleinae* for mitochondrial DNA (mtDNA) and plastid DNA (ptDNA) regions. The number in parentheses indicates the number of accessions used to measure π. First, π was estimated on five lineages of the olive complex (i.e., E1, E2, E3, M and C2) independently for protein-coding genes and noncoding regions, distinguishing regions in the single copies (SSC and LSC) and the inverted repeats (IR). It was also estimated for shared DNA regions [i.e., plastid-derived regions detected in the mitogenome (*mtpt* regions) and their homologous regions in the plastomes]. Second, π was calculated on protein-coding and noncoding regions at the subtribe level considering eight *Oleinae* species (i.e., *C. parkinsonii*, *C. rupicolus*, *For. isabelae*, *H. palmeri*, *Ne. apetala*, *No. lowryi*, *O. exasperata* and *O. e. europaea* E1-*e*.4). The standard error of π was estimated by bootstrapping on sequences (100 replicates)

		π	
Clade	**DNA region type**	**mtDNA**	**ptDNA**
Olea europaea (5)	Protein-coding genes	0.022 ± 0.003	0.035 ± 0.005
	Protein-coding genes-IR	—	0.00 ± 0.00
	Noncoding regions	0.041 ± 0.001	0.088 ± 0.008
	Noncoding regions-IR	—	0.021 ± 0.010
	Shared DNA regions (*mtpt*)	0.076 ± 0.012	0.024 ± 0.009
Oleinae (8)	Protein-coding genes	0.159 ± 0.009	0.532 ± 0.021
	Protein-coding genes-IR	—	0.124 ± 0.017
	Noncoding regions	0.188 ± 0.014	0.936 ± 0.026
	Noncoding regions-IR	—	0.133 ± 0.022

but not resolved in the mtDNA phylogeny. The placement of Mediterranean olive lineages (E1, E2, E3 and M) was not resolved either (bs < 80%) in the mtDNA phylogeny (Figure 4b).

3.5 | Phylogenetic relationships between cytoplasmic lineages of the Mediterranean olive

We finally focused on the Mediterranean olive. Twenty-one parsimony-informative sites were detected in the plastome alignment including six olive accessions (out of a total of 130,242 bp; after excluding one IR; Table S7). These sites are distributed throughout the LSC and SSC of the plastome, in both coding and noncoding regions (12 and 9, respectively). Six of 12 substitutions located in plastid genes are nonsynonymous (Table S7). For the mitogenome, we aligned 26 conserved regions with a length ranging from 4,724 to 68,630 bp. We detected 156 parsimony-informative sites in the concatenated alignment (out of a total of 747,352 aligned nucleotides; Table S8). These sites are mainly located in the noncoding regions of the mitogenome (150 sites/156), including *mtpt* regions (7 sites). Of the six substitutions located in mitochondrial genes, five are nonsynonymous (Table S8).

The tree topology based on ptDNA was poorly resolved (as in the analysis based on the 18 Oleaceae accessions; see above), only supporting close relationships between the E1-*l*1 and E1-*e* sublineages (Figure 5a). Adding further *O. europaea* accessions from GenBank did not increase node support (Figure S5). In contrast, the topology of both phylogenies based on mtDNA and ptDNA + mtDNA was congruent and recovered lineage E2 as sister to E1, while lineage E3 was sister to other Mediterranean lineages with a strong support (94%–95%; Figure 5b,c).

4 | DISCUSSION

Our study demonstrates the possibility of using genome skimming data to assemble both complete plastome and mitochondrial genomic regions, as already shown by Malé et al. (2014). Assembly of full mitogenomes was possible but time-consuming, and we thus developed a methodology to recover selected, conserved mitochondrial regions (genes and introns) for phylogenetic reconstruction. Plastid-derived mitochondrial regions (*mtpt*) are a potential issue in reconstructing some genomic regions, but mtDNA sequences always had lower read coverage than ptDNA regions. However, the mitogenome/plastome sequencing depth ratio was highly variable and tended to be higher in herbarium specimens. The potential reasons for this pattern need to be investigated with a broader specimen sampling. It could be due to a generally variable ratio of plastids to mitochondria in our samples and/or differential degradation of plastid and mitochondrial sequences,

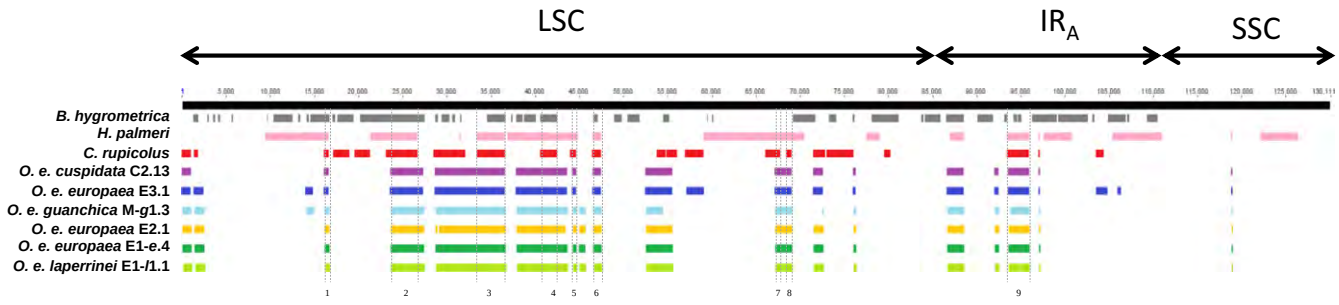


FIGURE 2 Position of each *mtpt* homologue within the E1-*e*.4 plastome. We mapped *mtpt* regions detected in the mitogenomes of all accessions of *Olea europaea*, *Chionanthus rupicolus*, *Hesperelaea palmeri* and *Boea hygrometrica* against the E1-*e*.4 plastome using GENEIOUS. The nine *mtpt* regions shared by all *Oleeae* are delimited with dotted grey lines and numbered from 1 to 9. The positions of the inverted repeat A (IR$_A$) and the small and long single copies (SSC and LSC, respectively) are indicated

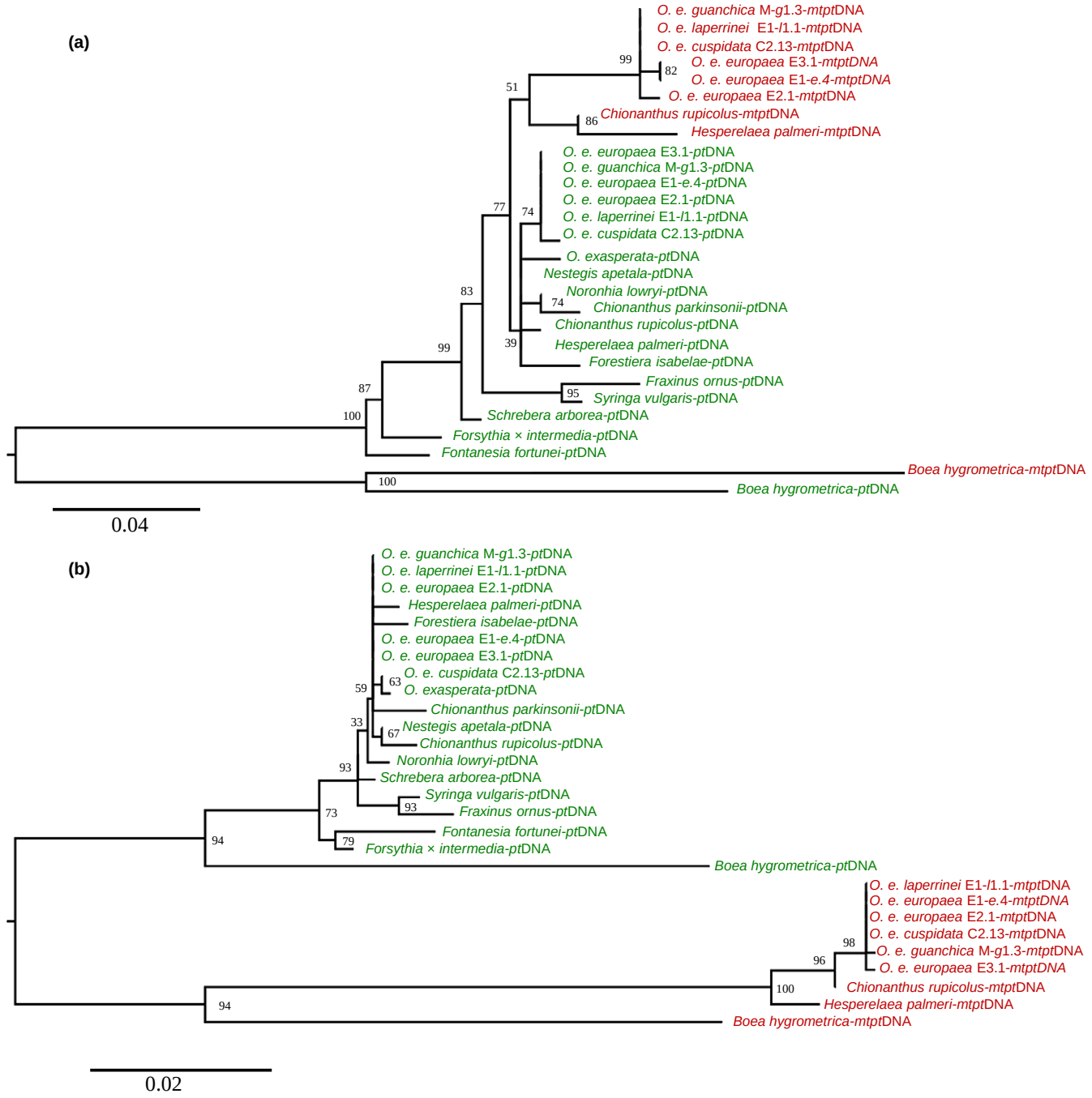


FIGURE 3 Maximum-likelihood phylogenetic trees reconstructed on two *mtpt* segments [(a) *mtpt4* and (b) *mtpt6*] and their homologous plastid regions extracted from plastomes of *Boea hygrometrica* (Gesneriaceae), 16 accessions belonging to tribe *Oleeae* and two accessions belonging to *Fontanesieae* and *Forsythieae*. The *mtpt* regions are represented in red and their corresponding homologous plastid sequences in green. Both phylogenetic trees were reconstructed with RAxML v.8 (Stamatakis, 2014) using a GTR + G model, applying the rapid bootstrap algorithm with 1,000 iterations. We rooted the trees at mid-point

potentially inducing some sequencing bias (Bakker et al., 2016; Besnard et al., 2014; Staats et al., 2011).

4.1 | Structural diversity of mitogenomes in the olive complex

While complete plastid and nuclear genomes of the olive tree were available prior to our study (Besnard, Hernández, Khadari, Dorado, & Savolainen, 2011; Cruz et al., 2016; Mariotti, Cultrera, Díez, Baldoni, & Rubini, 2010; Unver et al., 2017), mitochondrial genomic resources of *O. europaea* were still lacking. We here present six newly assembled, complete mitogenomes of *O. europaea*. In accordance with previous results based on restriction fragment length polymorphisms (Besnard et al., 2002), we found no rearrangement within lineage E1 (sublineages *e* and *l*1) but a large variation in the mitogenome structure among cytoplasmic lineages of the olive tree (Figure S2). The

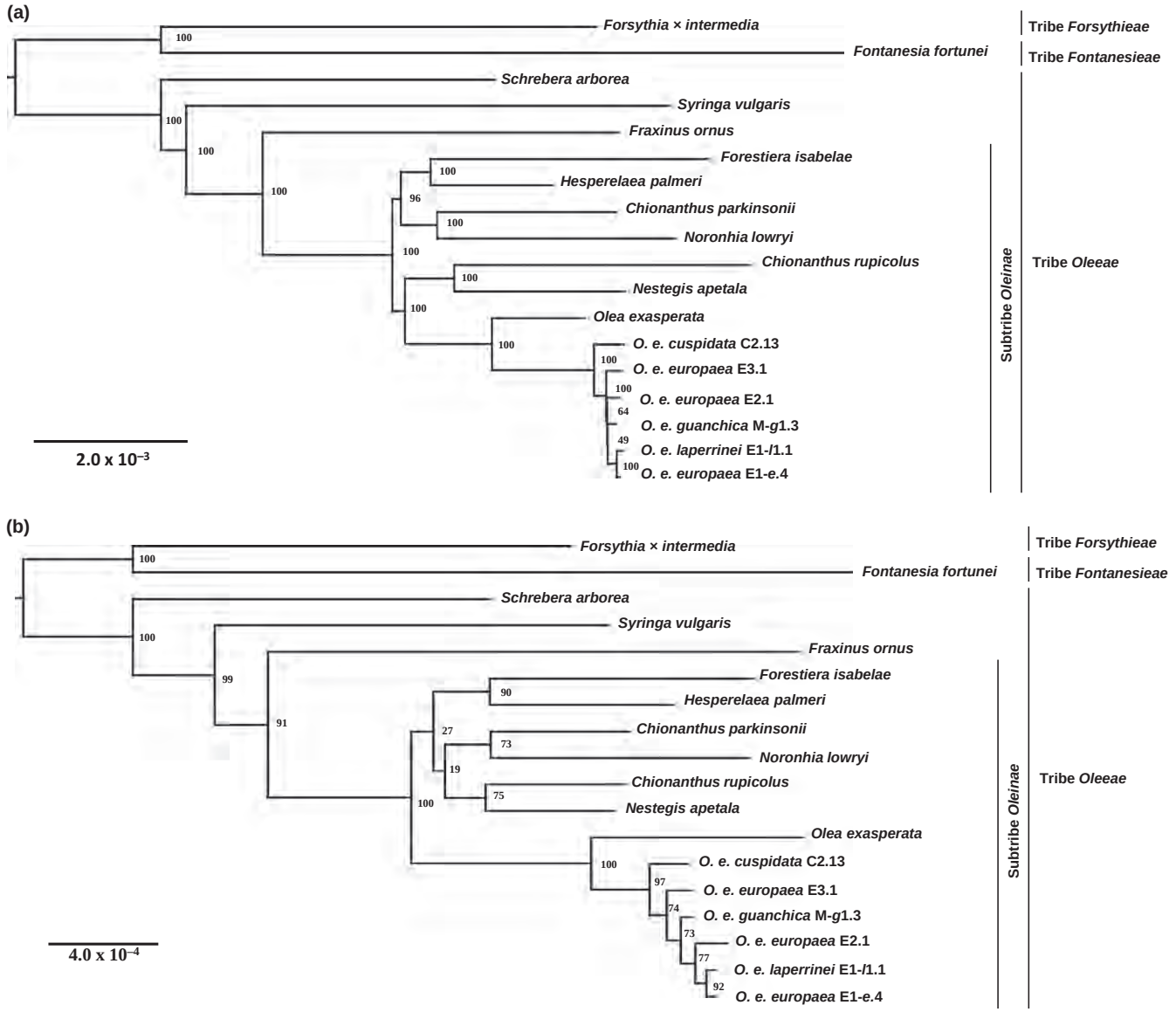


FIGURE 4 Maximum-likelihood phylogenetic trees of the *Oleeae* obtained with RAxML v.8 (Stamatakis, 2014) with the GTR + G model. We reconstructed two phylogenetic trees of *Oleeae* using *Fontanesia fortunei* and *Forsythia* × *intermedia* to root the trees: (a) based on complete plastomes, and (b) based on 52 mitochondrial protein-coding genes and introns

variation in size among mitogenomes is mainly explained by different proportions of repeated sequences (Table 2). Lineage E3 has the largest mitogenome of the olive complex (Figure 1), with the highest content of noncoding regions and repeats >10 kb. Contrary to the other accessions, the E3.1 mitogenome has lost *sdh3* (coding for a subunit of the succinate dehydrogenase), a gene known to have been recurrently transferred to the nuclear genome in higher plants (Adams & Palmer, 2003). Interestingly, two chimeric ORFs (one including a *sdh3* segment) were also found to be specific to the E3.1 mitogenome (Figures 1 and S3). Such genes, and particularly the chimeric ORF including a portion of *sdh3*, could be involved in the olive CMS (Besnard et al., 2000). Moreover, among all mitochondrial genes here assembled for the 18 Oleaceae accessions, *sdh3* is the only gene missing in one accession. The CMS associated with lineage E3 could thus be associated with a dysfunction of *sdh3*, normally involved in the mitochondrial electron transport chain (Huang & Millar, 2013). Furthermore, a *sdh3* pseudogene has also been suggested as a candidate involved in a CMS type of *Nicotiana tabacum* (Zhang, Li, Jia, Li, & Feng, 2013). The chimeric ORF including a portion of *sdh3* thus appears as a good candidate for CMS in olive, but testing this hypothesis requires further analyses, in particular of gene expression and its potential consequence for pollen development and abortion (Hanson & Bentolila, 2004).

4.2 | Insertion of plastid sequences in mitogenomes of the *Oleinae* subtribe

A high proportion of *Oleinae* mitogenomes (ca. 5%–10%; Table 3) consists of plastid-derived sequences (*mtpt*) as similarly reported in numerous plants (e.g., Alverson et al., 2010; Marienfeld et al., 1999;

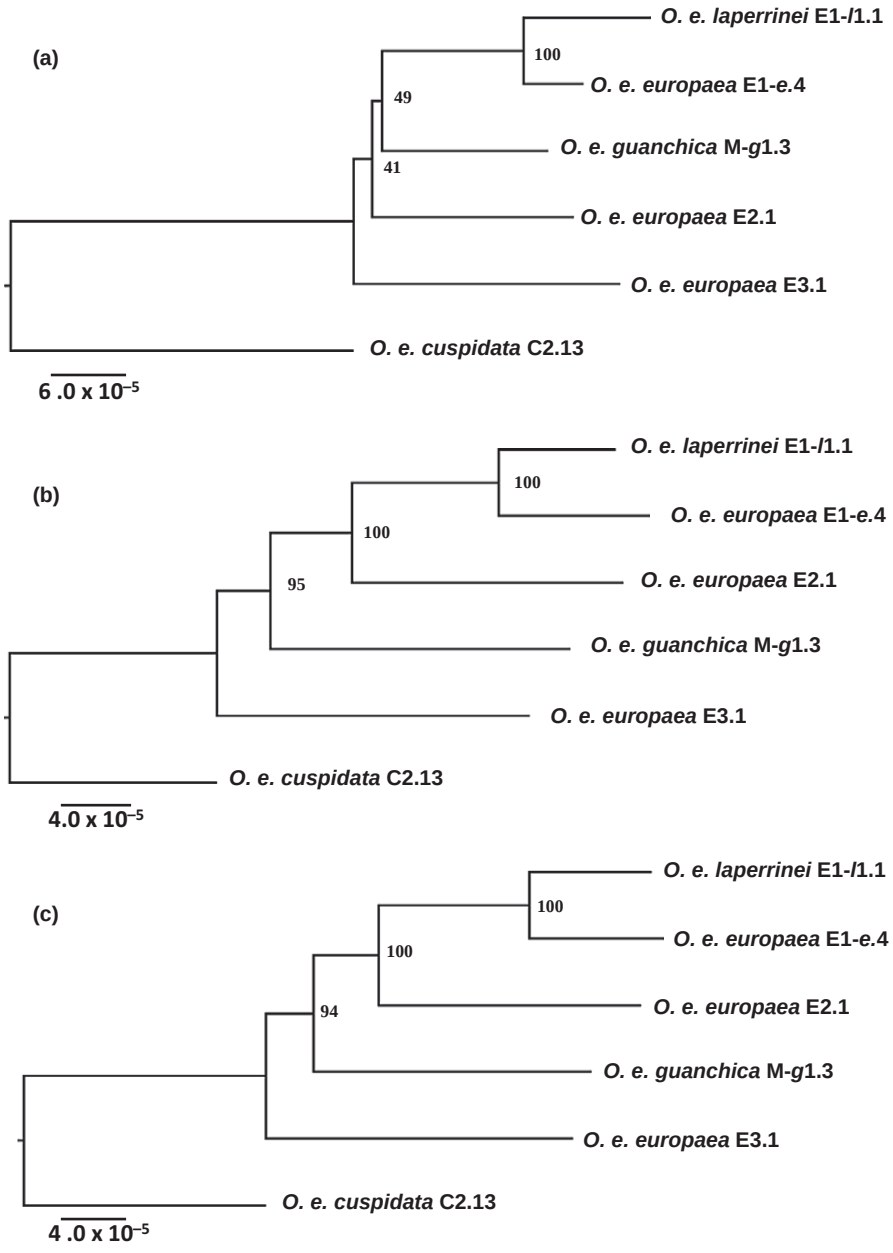


FIGURE 5 Maximum-likelihood phylogenetic trees of the olive complex obtained with RAxML v.8 (Stamatakis, 2014) with a GTR + G model. We applied the rapid bootstrap algorithm with 1,000 iterations. We reconstructed three phylogenetic trees of the six newly sequenced accessions of *O. europaea* based on (a) complete plastomes, (b) conserved mitochondrial regions and (c) the concatenation of complete plastomes and conserved mitochondrial regions. We used *O. e. cuspidata* (lineage C2) as outgroup to root the trees

Unseld et al., 1997). These regions cover a large part (>20%) of the plastome (Figure 2). While most *mtpt* regions show the same boundaries between the six olive accessions, nine *mtpt* regions are shared between *Olea*, *Hesperelaea* and *Chionanthus* (Figure 2). Phylogenies individually reconstructed from each *mtpt* and their homologous regions support several rounds of integration of plastid sequences in ancestral mitogenomes of *Oleinae* (Figures 3 and S4). One of them (*mtpt6*, which includes a putative functional tRNA gene, *trnS-GGA*) was probably integrated before the divergence between Gesneriaceae and Oleaceae. Such recurrent integrations of plastid sequences in the mitogenome have also been reported in other plant families (Hao & Palmer, 2009; Sloan & Wu, 2014). The use of *mtpt* sequences for the reconstruction of plant phylogenies has thus to be taken with caution, and their orthology needs to be confirmed beforehand. Our phylogenies strongly suggest that all *mtpt* regions of the olive complex derived from a common ancestor, but this was not confirmed within the *Oleinae* subtribe for two *mtpt* regions (i.e., *mtpt2* and *mtpt9*; Figure S4). Surprisingly, *mtpt* regions of the olive tree show higher nucleotide diversity than their plastid homologous regions, but only seven parsimony-informative sites were detected in those regions. This faster evolution in *mtpt* regions may at least partially reflect a relaxation of selective pressures on these sequences that mainly contain nonfunctional genes after their integration to the mitogenome (Sloan & Wu, 2014).

4.3 | Comparative phylogenetic reconstruction of *Oleeae* based on plastid and mitochondrial genomes

Due to the difficulty in reconstructing full mitogenomes and the possibility that some mitochondrial noncoding regions may not be

orthologous between distantly related taxa, we decided to reconstruct phylogenies of the *Oleeae* tribe using only conserved mitochondrial regions (i.e., protein-coding genes and introns). Fifty-two sequences (covering ca. 55 kb) were thus assembled from an annotated reference as reported in other studies (Govindarajulu et al., 2015; Malé et al., 2014; Zhang, Wen, & Zimmer, 2015). We then compared phylogenetic topologies based on conserved mitochondrial sequences and full plastomes and observed high congruence between these analyses, although some nodes were not resolved in the mtDNA tree (Figure 4). This suggests a strong linkage disequilibrium between DNA polymorphisms of both organellar genomes, as a consequence of their common maternal inheritance (Besnard et al., 2002; Liu et al., 2004). Our results also suggest that our ptDNA and mtDNA data may be appropriate for comparative phylogenetic studies on other Oleaceae lineages with potential biparental inheritance of organelles (Liu et al., 2004; Sodmergen et al., 1998). However, as similarly reported in other studies (Palmer & Herbon, 1988; Wolfe et al., 1987), such mitochondrial conserved regions evolved three to five times slower than plastid sequences in *Oleinae* and such a low variation could be insufficient to resolve certain nodes of the phylogeny (Figure 4b).

4.4 | Phylogenetic relationships between cytoplasmic olive lineages resolved using mitochondrial polymorphisms

Until now, phylogenetic relationships of Mediterranean plastid lineages have never been resolved despite repeated efforts (Besnard et al., 2002, 2007, 2011, 2013; Médail, Quézel, Besnard, & Khadari, 2001), even using complete plastomes sequences. A rapid radiation of North African olive plastid lineages (E1, E2, E3, and M) in the Late Neogene may explain this difficulty (Besnard et al., 2013). Here, we were still unable to resolve phylogenetic relationships between the four Mediterranean cytoplasmic lineages based on new ptDNA sequences (Figures 5a and S5), due to the low number of informative sites in the plastome (Besnard et al., 2011). Yet, we resolved the phylogenetic tree (Figure 5b) using the complete mitogenome that contains 7.4 times more informative sites than the full plastome. This was unexpected, as the mitogenome is generally believed less informative than the plastome. In the tree topology, lineage E2 is sister to E1 as reported by Médail et al. (2001) based on the analysis of restriction fragment length polymorphisms of the mtDNA. The position of lineage E3 as sister to all North African lineages is also strongly supported. Overall, the topology indicates that Mediterranean olive lineages may have rapidly diversified in the Western Mediterranean Basin first (lineages E2, E3 and M, including Macaronesia) before colonizing the Eastern Mediterranean and Saharan Africa (lineage E1). Although other species complexes share a circum-Mediterranean distribution (e.g., *Laurus*, *Myrtus*, *Phillyrea*, *Pinus*, *Quercus*), such a colonization from west to east during the Late Neogene has never been suggested until now for other tree species based on phylogeographic reconstructions (Migliore, Baumel, Juin, & Médail, 2012; Nieto Feliner, 2014; Rodríguez-Sánchez, Guzmán, Valido, Vargas, & Arroyo, 2009; Thompson, 2005).

5 | CONCLUSION AND PERSPECTIVES

Overall, our study confirms that the combined use of mtDNA and ptDNA is useful to investigate the evolutionary history of plants, opening new avenues of research in particular in the olive family. First, differential inheritance of organelles and nuclear–cytoplasmic incompatibilities have been reported in Oleaceae (Besnard et al., 2000; Liu et al., 2004; Sodmergen et al., 1998), and phylogenetic reconstruction with tests of selection may be performed on both organellar genomes to decipher the contrasted evolutionary histories and potential shifts in selective pressures on mitochondrial and/or chloroplast genes (e.g., Mensous et al., 2017; Zhu et al., 2014). Second, mtDNA polymorphisms may also be used in olive phylogeography. Here, the rapid radiation of maternal lineages of the Mediterranean olive was resolved for the first time with the use of mtDNA polymorphisms. This was potentially due to a higher number of selectively neutral informative sites in mtDNA (151 sites; when excluding the five nonsynonymous sites) than in ptDNA (15 sites; when excluding the six nonsynonymous sites), but a faster coalescence time of the mitogenome compared to the plastome (for instance due to a lower relative number of mitogenomes in reproductive cells) may also explain the better resolved topologies. Finally, the different genome compartments may be differently affected by postmortem damages, and their relative utility for ancient DNA analysis still needs to be tested. Our mitogenomic data are thus a new source of informative polymorphisms to investigate the spread of domesticated olive lineages (Besnard et al., 2013), in particular with metagenomic approaches to analyse subfossil remains from archaeological sites.

ACKNOWLEDGEMENTS

C.V.D.P. and G.B. are members of the Laboratoire Evolution & Diversité Biologique (EDB) part of the LABEX TULIP managed by the Agence Nationale de la Recherche (ANR-10-LABX-0041). We also acknowledge an Investissement d'Avenir grant from the Agence Nationale de la Recherche (CEBA: ANR-10-LABX-25-01). C.V.D.P. and G.B. were funded by the Regional Council Midi-Pyrénées (AAP 13053637, 2014-EDB-UT3-DOCT). This work was performed in collaboration with the GeT core facility, Toulouse, France (http://get.genotoul.fr), and was supported by France Génomique National infrastructure, funded as part of "Investissement d'avenir" program managed by Agence Nationale pour la Recherche (contract ANR-10-INBS-09). This study also received support from the PhyloAlps project, and we thank H. Holota and A. Iribar. L. Csiba and E. Kepos (Jodrell Laboratory) provided DNA extracts from accessions of the living collection of Royal Botanic Gardens (Kew, Richmond). We are also grateful to D. Stadie (Eisleben), M.S. Vorontsova and D. Goyder (Royal Botanical Gardens, Kew), T. Haevermans and M. Gaudeul (Muséum National d'Histoire Naturelle de Paris), H. Esser (Munich Botanical Gardens), J. Razanatsoa (Parc de Tsimbazaza, Antananarivo), P. Saumitou-Laprade and P. Vernet (Evo-Eco-Paléo Lille), T. Josseberger (Botanische Gärten der Universität Bonn) for providing samples for our study. We thank Freek T. Bakker, Pascal-Antoine Christin, Jan Hackel, Jill K. Olofsson and Hervé Philippe for helpful suggestions.

DATA ACCESSIBILITY

In the NCBI nucleotide database, all newly assembled chloroplast genomes, mitogenomes and mitochondrial DNA phylogenetic markers are available under the accession numbers provided in Tables 1 and S6.

AUTHOR'S CONTRIBUTIONS

C.V.D.P. and G.B. designed the study. O.B. and G.B. generated the sequence data. C.V.D.P. and G.B. analysed the data, interpreted the results and wrote the manuscript.

ORCID

Céline Van de Paer http://orcid.org/0000-0003-2207-3721

REFERENCES

Adams, K. L., & Palmer, J. D. (2003). Evolution of mitochondrial gene content: Gene loss and transfer to the nucleus. *Molecular Phylogenetics and Evolution*, *29*, 380–395. https://doi.org/10.1016/S1055-7903(03)00194-5

Aguileta, G., de Vienne, D. M., Ross, O. N., Hood, M. E., Giraud, T., Petit, E., & Gabaldón, T. (2014). High variability of mitochondrial gene order among fungi. *Genome Biology and Evolution*, *6*, 451–465. https://doi.org/10.1093/gbe/evu028

Altschul, S. F., Gish, W., Miller, W., Myers, E. W., & Lipman, D. J. (1990). Basic local alignment search tool. *Journal of Molecular Biology*, *215*, 403–410. https://doi.org/10.1016/S0022-2836(05)80360-2

Álvarez, I., & Wendel, J. F. (2003). Ribosomal ITS sequences and plant phylogenetic inference. *Molecular Phylogenetics and Evolution*, *29*, 417–434. https://doi.org/10.1016/S1055-7903(03)00208-2

Alverson, A. J., Wei, X., Rice, D. W., Stern, D. B., Barry, K., & Palmer, J. D. (2010). Insights into the evolution of mitochondrial genome size from complete sequences of *Citrullus lanatus* and *Cucurbita pepo* (Cucurbitaceae). *Molecular Biology and Evolution*, *27*, 1436–1448. https://doi.org/10.1093/molbev/msq029

Avise, J. C. (2009). Phylogeography: Retrospect and prospect. *Journal of Biogeography*, *36*, 3–15. https://doi.org/10.1111/j.1365-2699.2008.02032.x

Bakker, F. T., Lei, D., Yu, J., Mohammadin, S., Wei, Z., van de Kerke, S., … Holmer, R. (2016). Herbarium genomics: plastome sequence assembly from a range of herbarium specimens using an Iterative Organelle Genome Assembly pipeline. *Biological Journal of the Linnean Society*, *117*, 33–43. https://doi.org/10.1111/bij.12642

Besnard, G., Christin, P. A., Malé, P. J., Lhuillier, E., Lauzeral, C., Coissac, E., & Vorontsova, M. S. (2014). From museums to genomics: Old herbarium specimens shed light on a C_3 to C_4 transition. *Journal of Experimental Botany*, *65*, 6711–6721. https://doi.org/10.1093/jxb/eru395

Besnard, G., Hernández, P., Khadari, B., Dorado, G., & Savolainen, V. (2011). Genomic profiling of plastid DNA variation in the Mediterranean olive tree. *BMC Plant Biology*, *11*, 80. https://doi.org/10.1186/1471-2229-11-80

Besnard, G., Khadari, B., Baradat, P., & Bervillé, A. (2002). Combination of chloroplast and mitochondrial DNA polymorphisms to study cytoplasm genetic differentiation in the olive complex (*Olea europaea* L.). *Theoretical and Applied Genetics*, *105*, 139–144.

Besnard, G., Khadari, B., Navascués, M., Fernández-Mazuecos, M., El Bakkali, A., Arrigo, N., … Savolainen, V. (2013). The complex history of the olive tree: From Late Quaternary diversification of Mediterranean lineages to primary domestication in the northern Levant. *Proceedings of the Royal Society of London B: Biological Sciences*, *280*, 20122833. https://doi.org/10.1098/rspb.2012.2833

Besnard, G., Khadari, B., Villemur, P., & Bervillé, A. (2000). Cytoplasmic male sterility in the olive (*Olea europaea* L.). *Theoretical and Applied Genetics*, *100*, 1018–1024. https://doi.org/10.1007/s001220051383

Besnard, G., Rubio de Casas, R., & Vargas, P. (2007). Plastid and nuclear DNA polymorphism reveals historical processes of isolation and reticulation in the olive tree complex (*Olea europaea* L.). *Journal of Biogeography*, *34*, 736–752. https://doi.org/10.1111/j.1365-2699.2006.01653.x

Besnard, G., Terral, J., & Cornille, A. (2018). On the origins and domestication of the olive: A review and perspectives. *Annals of Botany*, *121*, in press. https://doi.org/10.1093/aob/mcx145

Biomatters Ltd. (2015). *GENEIOUS v. 9.0.5*. Retrieved from http://www.geneious.com

Burke, S. V., Wysocki, W. P., Zuloaga, F. O., Craine, J. M., Pires, J. C., Edger, P. P., … Duvall, M. R. (2016). Evolutionary relationships in panicoid grasses based on plastome phylogenomics (Panicoideae; Poaceae). *BMC Plant Biology*, *16*, 140. https://doi.org/10.1186/s12870-016-0823-3

Chat, J., Jáuregui, B., Petit, R. J., & Nadot, S. (2004). Reticulate evolution in kiwifruit (*Actinidia*, Actinidiaceae) identified by comparing their maternal and paternal phylogenies. *American Journal of Botany*, *91*, 736–747. https://doi.org/10.3732/ajb.91.5.736

Christin, P. A., Besnard, G., Edwards, E. J., & Salamin, N. (2012). Effect of genetic convergence on phylogenetic inference. *Molecular Phylogenetics and Evolution*, *62*, 921–927. https://doi.org/10.1016/j.ympev.2011.12.002

Coissac, E., Hollingsworth, P. M., Lavergne, S., & Taberlet, P. (2016). From barcodes to genomes: Extending the concept of DNA barcoding. *Molecular Ecology*, *25*, 1423–1428. https://doi.org/10.1111/mec.13549

Corriveau, J. L., & Coleman, A. W. (1988). Rapid screening method to detect potential biparental inheritance of plastid DNA and results for over 200 angiosperm species. *American Journal of Botany*, *75*, 1443–1458. https://doi.org/10.2307/2444695

Cruz, F., Julca, I., Gómez-Garrido, J., Loska, D., Marcet-Houben, M., Cano, E., … Gut, M. (2016). Genome sequence of the olive tree, *Olea europaea*. *GigaScience*, *5*, 29. https://doi.org/10.1186/s13742-016-0134-5

Darling, A. E., Mau, B., & Perna, N. T. (2010). progressiveMauve: Multiple genome alignment with gene gain, loss and rearrangement. *PLoS ONE*, *5*, e11147. https://doi.org/10.1371/journal.pone.0011147

Delsuc, F., Brinkmann, H., & Philippe, H. (2005). Phylogenomics and the reconstruction of the tree of life. *Nature Reviews. Genetics*, *6*, 361–375. https://doi.org/10.1038/nrg1603

Desplanque, B., Viard, F., Bernard, J., Forcioli, D., Saumitou-Laprade, P., Cuguen, J., & Van Dijk, H. (2000). The linkage disequilibrium between chloroplast DNA and mitochondrial DNA haplotypes in *Beta vulgaris* ssp. *maritima* (L.): The usefulness of both genomes for population genetic studies. *Molecular Ecology*, *9*, 141–154. https://doi.org/10.1046/j.1365-294x.2000.00843.x

Donnelly, K., Cottrell, J., Ennos, R. A., Vendramin, G. G., A'Hara, S., King, S., … Cavers, S. (2017). Reconstructing the plant mitochondrial genome for marker discovery: A case study using Pinus. *Molecular Ecology Resources*, *17*, 943–954. https://doi.org/10.1111/1755-0998.12646

Duminil, J. (2014). Mitochondrial genome and plant taxonomy. In P. Besse (Ed.), *Molecular plant taxonomy. Methods in molecular biology (methods and protocols)* (pp. 121–140). Totowa, NJ: Humana Press.

Edgar, R. C. (2004). MUSCLE: Multiple sequence alignment with high accuracy and high throughput. *Nucleic Acids Research*, *32*, 1792–1797. https://doi.org/10.1093/nar/gkh340

Govindarajulu, R., Parks, M., Tennessen, J. A., Liston, A., & Ashman, T. L. (2015). Comparison of nuclear, plastid, and mitochondrial phylogenies and the origin of wild octoploid strawberry species. *American Journal of Botany*, *102*, 544–554. https://doi.org/10.3732/ajb.1500026

Green, P. S. (2004). Oleaceae. In K. Kubitzki, & J. W. Kadereit (Eds.), *The families and genera of vascular plants. Flowering plants, dicotyledons*, Vol. *VII* (pp. 296–306). New York, NY: Springer. https://doi.org/10.1007/978-3-642-18617-2

Gualberto, J. M., Mileshina, D., Wallet, C., Niazi, A. K., Weber-Lotfi, F., & Dietrich, A. (2014). The plant mitochondrial genome: Dynamics and maintenance. *Biochimie*, *100*, 107–120. https://doi.org/10.1016/j.biochi.2013.09.016

Hanson, M. R., & Bentolila, S. (2004). Interactions of mitochondrial and nuclear genes that affect male gametophyte development. *The Plant Cell*, *16*, S154–S169. https://doi.org/10.1105/tpc.015966

Hao, W., & Palmer, J. D. (2009). Fine-scale mergers of chloroplast and mitochondrial genes create functional, transcompartmentally chimeric mitochondrial genes. *Proceedings of the National Academy of Sciences of the United States of America*, *106*, 16728–16733. https://doi.org/10.1073/pnas.0908766106

Hao, W., Richardson, A. O., Zheng, Y., & Palmer, J. D. (2010). Gorgeous mosaic of mitochondrial genes created by horizontal transfer and gene conversion. *Proceedings of the National Academy of Sciences of the United States of America*, *107*, 21576–21581. https://doi.org/10.1073/pnas.1016295107

Hollingsworth, M. L., Andra Clark, A. L., Forrest, L. L., Richardson, J., Pennington, R., Long, D. G., … Hollingsworth, P. M. (2009). Selecting barcoding loci for plants: Evaluation of seven candidate loci with species-level sampling in three divergent groups of land plants. *Molecular Ecology Resources*, *9*, 439–457. https://doi.org/10.1111/j.1755-0998.2008.02439.x

Hong-Wa, C., & Besnard, G. (2013). Intricate patterns of phylogenetic relationships in the olive family as inferred from multi-locus plastid and nuclear DNA sequence analyses: A close-up on *Chionanthus* and *Noronhia* (Oleaceae). *Molecular Phylogenetics and Evolution*, *67*, 367–378. https://doi.org/10.1016/j.ympev.2013.02.003

Huang, S., & Millar, A. H. (2013). Succinate dehydrogenase: The complex roles of a simple enzyme. *Current Opinion in Plant Biology*, *16*, 344–349. https://doi.org/10.1016/j.pbi.2013.02.007

Iorizzo, M., Grzebelus, D., Senalik, D., Szklarczyk, M., Spooner, D., & Simon, P. (2012). Against the traffic: The first evidence for mitochondrial DNA transfer into the plastid genome. *Mobile Genetic Elements*, *2*, 261–266. https://doi.org/10.4161/mge.23088

Kearse, M., Moir, R., Wilson, A., Stones-Havas, S., Cheung, M., Sturrock, S., … Thierer, T. (2012). Geneious Basic: An integrated and extendable desktop software platform for the organization and analysis of sequence data. *Bioinformatics*, *28*, 1647–1649. https://doi.org/10.1093/bioinformatics/bts199

Krzywinski, M., Schein, J., Birol, I., Connors, J., Gascoyne, R., Horsman, D., … Marra, M. A. (2009). Circos: An information aesthetic for comparative genomics. *Genome Research*, *19*, 1639–1645. https://doi.org/10.1101/gr.092759.109

Laslett, D., & Canback, B. (2004). ARAGORN, a program to detect tRNA genes and tmRNA genes in nucleotide sequences. *Nucleic Acids Research*, *32*, 11–16. https://doi.org/10.1093/nar/gkh152

Li, X., Yang, Y., Henry, R. J., Rossetto, M., Wang, Y., & Chen, S. (2014). Plant DNA barcoding: From gene to genome. *Biological Reviews*, *90*, 157–166.

Liu, Y., Cui, H., & Zhang, Q. (2004). Divergent potentials for cytoplasmic inheritance within the genus *Syringa*. A new trait associated with speciogenesis. *Plant Physiology*, *136*, 2762–2770. https://doi.org/10.1104/pp.104.048298

Lohse, M., Drechsel, O., & Bock, R. (2007). OrganellarGenomeDRAW (OGDRAW): A tool for the easy generation of high-quality custom graphical maps of plastid and mitochondrial genomes. *Current Genetics*, *52*, 267–274. https://doi.org/10.1007/s00294-007-0161-y

Lonsdale, D. M., Brears, T., Hodge, T. P., Melville, S. E., & Rottmann, W. H. (1988). The plant mitochondrial genome: Homologous recombination as a mechanism for generating heterogeneity. *Philosophical Transactions of the Royal Society B: Biological Sciences*, *319*, 149–163. https://doi.org/10.1098/rstb.1988.0039

Ma, P. F., Zhang, Y. X., Guo, Z. H., & Li, D. Z. (2015). Evidence for horizontal transfer of mitochondrial DNA to the plastid genome in a bamboo genus. *Scientific Reports*, *5*, 11608. https://doi.org/10.1038/srep11608

Malé, P. J., Bardon, L., Besnard, G., Coissac, E., Delsuc, F., Engel, J., … Chave, J. (2014). Genome skimming by shotgun sequencing helps resolve the phylogeny of a pantropical tree family. *Molecular Ecology Resources*, *14*, 966–975.

Marienfeld, J., Unseld, M., & Brennicke, A. (1999). The mitochondrial genome of *Arabidopsis* is composed of both native and immigrant information. *Trends in Plant Science*, *4*, 495–502. https://doi.org/10.1016/S1360-1385(99)01502-2

Mariotti, R., Cultrera, N. G. M., Díez, C. M., Baldoni, L., & Rubini, A. (2010). Identification of new polymorphic regions and differentiation of cultivated olives (*Olea europaea* L.) through plastome sequence comparison. *BMC Plant Biology*, *10*, 211. https://doi.org/10.1186/1471-2229-10-211

Médail, F., Quézel, P., Besnard, G., & Khadari, B. (2001). Systematics, ecology and phylogeographic significance of *Olea europaea* L. ssp. *maroccana* (Greuter & Burdet) P. Vargas *et al.*, a relictual olive tree in south-west Morocco. *Botanical Journal of the Linnean Society*, *137*, 249–266.

Mensous, M., Van de Paer, C., Manzi, S., Bouchez, O., Baâli-Cherif, D., & Besnard, G. (2017). Diversity and evolution of plastomes in Saharan mimosoids: Potential use for phylogenetic and population genetic studies. *Tree Genetics and Genomes*, *13*, 48. https://doi.org/10.1007/s11295-017-1131-2

Migliore, J., Baumel, A., Juin, M., & Médail, F. (2012). From Mediterranean shores to central Saharan mountains: Key phylogeographical insights from the genus *Myrtus*. *Journal of Biogeography*, *39*, 942–956. https://doi.org/10.1111/j.1365-2699.2011.02646.x

Miller, M. A., Pfeiffer, W., & Schwartz, T. (2010). *Creating the CIPRES Science Gateway for inference of large phylogenetic trees*. Proceedings of the Gateway Computing Environments Workshop (GCE), 1–8.

Mower, J. P., Case, A. L., Floro, E. R., & Willis, J. H. (2012). Evidence against equimolarity of large repeat arrangements and a predominant master circle structure of the mitochondrial genome from a monkeyflower (*Mimulus guttatus*) lineage with cryptic CMS. *Genome Biology and Evolution*, *4*, 670–686. https://doi.org/10.1093/gbe/evs042

Mower, J. P., Sloan, D. B., & Alverson, A. J. (2012). Plant mitochondrial genome diversity: The genomics revolution. In J. F. Wendel, J. Greilhuber, J. Dolezel, & I. J. Leitch (Eds.), *Plant genome diversity* (pp. 123–144). Vienna, Austria: Springer. https://doi.org/10.1007/978-3-7091-1130-7

Mower, J. P., Touzet, P., Gummow, J. S., Delph, L. F., & Palmer, J. D. (2007). Extensive variation in synonymous substitution rates in mitochondrial genes of seed plants. *BMC Evolutionary Biology*, *7*, 135. https://doi.org/10.1186/1471-2148-7-135

Nabholz, B., Glémin, S., & Galtier, N. (2009). The erratic mitochondrial clock: Variations of mutation rate, not population size, affect mtDNA diversity across birds and mammals. *BMC Evolutionary Biology*, *9*, 54. https://doi.org/10.1186/1471-2148-9-54

Nei, M., & Li, W. H. (1979). Mathematical model for studying genetic variation in terms of restriction endonucleases. *Proceedings of the National Academy of Sciences of the United States of America*, *76*, 5269–5273. https://doi.org/10.1073/pnas.76.10.5269

Newton, K. J. (1988). Plant mitochondrial genomes: Organization, expression and variation. *Annual Review of Plant Physiology and Plant Molecular Biology*, *39*, 503–532. https://doi.org/10.1146/annurev.pp.39.060188.002443

Nieto Feliner, G. (2014). Patterns and processes in plant phylogeography in the Mediterranean Basin. *Perspectives in Plant Ecology, Evolution and Systematics*, *16*, 265–278. https://doi.org/10.1016/j.ppees.2014.07.002

Nieto Feliner, G., & Rosselló, J. A. (2007). Better the devil you know? Guidelines for insightful utilization of nrDNA ITS in species-level evolutionary studies in plants. *Molecular Phylogenetics and Evolution*, *44*, 911–919. https://doi.org/10.1016/j.ympev.2007.01.013

Notsu, Y., Masood, S., Nishikawa, T., Kubo, N., Akiduki, G., Nakazono, M., … Kadowaki, K. (2002). The complete sequence of the rice (*Oryza sativa* L.) mitochondrial genome: Frequent DNA sequence acquisition and loss during the evolution of flowering plants. *Molecular Genetics and Genomics*, *268*, 434–445. https://doi.org/10.1007/s00438-002-0767-1

Olson, M. S., & McCauley, D. E. (2000). Linkage disequilibrium and phylogenetic congruence between chloroplast and mitochondrial haplotypes in *Silene vulgaris*. *Proceedings of the Royal Society of London B: Biological Sciences*, *267*, 1801–1808. https://doi.org/10.1098/rspb.2000.1213

Palmer, J. D. (1992). Mitochondrial DNA in plant systematics: Applications and limitations. In P. Soltis, D. Soltis, & J. Doyle (Eds.), *Molecular systematics of plants* (pp. 36–49). New York, NY: Chapman and Hall. https://doi.org/10.1007/978-1-4615-3276-7

Palmer, J. D., & Herbon, L. A. (1988). Plant mitochondrial DNA evolves rapidly in structure, but slowly in sequence. *Journal of Molecular Evolution*, *28*, 87–97. https://doi.org/10.1007/BF02143500

Paterson, A. H., Freeling, M., Tang, H., & Wang, X. (2010). Insights from the comparison of plant genome sequences. *Annual Review of Plant Biology*, *61*, 349–372. https://doi.org/10.1146/annurev-arplant-042809-112235

Pearl, S. A., Welch, M. E., & McCauley, D. E. (2009). Mitochondrial heteroplasmy and paternal leakage in natural populations of *Silene vulgaris*, a gynodioecious plant. *Molecular Biology and Evolution*, *26*, 537–545. https://doi.org/10.1093/molbev/msn273

Pillar, V. D., & Duarte, L. S. (2010). A framework for metacommunity analysis of phylogenetic structure. *Ecology Letters*, *13*, 587–596. https://doi.org/10.1111/j.1461-0248.2010.01456.x

Qiu, Y. L., Li, L., Hendry, T. A., Li, R., Taylor, D. W., Issa, M. J., … White, A. M. (2006). Reconstructing the basal angiosperm phylogeny: Evaluating information content of mitochondrial genes. *Taxon*, *55*, 837–856. https://doi.org/10.2307/25065680

Rambaut, A. (2014). *FigTree 1.4.2*. Retrieved from http://tree.bio.ed.ac.uk/

Reboud, X., & Zeyl, C. (1994). Organelle inheritance in plants. *Heredity*, *72*, 132–140. https://doi.org/10.1038/hdy.1994.19

Rodríguez-Sánchez, F., Guzmán, B., Valido, A., Vargas, P., & Arroyo, J. (2009). Late Neogene history of the laurel tree (*Laurus* L., Lauraceae) based on phylogeographical analyses of Mediterranean and Macaronesian populations. *Journal of Biogeography*, *36*, 1270–1281. https://doi.org/10.1111/j.1365-2699.2009.02091.x

Rusche, M. L., Mogensen, H. L., Zhu, T., & Smith, S. E. (1995). The zygote and proembryo of alfalfa: Quantitative, three-dimensional analysis and implications for biparental plastid inheritance. *Protoplasma*, *189*, 88–100. https://doi.org/10.1007/BF01280294

Sloan, D. B., & Wu, Z. (2014). History of plastid DNA insertions reveals weak deletion and AT mutation biases in angiosperm mitochondrial genomes. *Genome Biology and Evolution*, *6*, 3210–3221. https://doi.org/10.1093/gbe/evu253

Smith, D. R., & Keeling, P. J. (2015). Mitochondrial and plastid genome architecture: Reoccurring themes, but significant differences at the extremes. *Proceedings of the National Academy of Sciences of the United States of America*, *112*, 201422049.

Sodmergen, H., Bai, H., He, J. X., Kuroiwa, H., Kawano, S., & Kuroiwa, T. (1998). Potential for biparental cytoplasmic inheritance in *Jasminum officinale* and *Jasminum nudiflorum*. *Sexual Plant Reproduction*, *11*, 107–112.

Sollars, E. S., Harper, A. L., Kelly, L. J., Sambles, C. M., Ramirez-Gonzalez, R. H., Swarbreck, D., … Worswick, G. (2017). Genome sequence and genetic diversity of European ash trees. *Nature*, *541*, 212–216.

Staats, M., Cuenca, A., Richardson, J. E., Vrielink-van Ginkel, R., Petersen, G., Seberg, O., & Bakker, F. T. (2011). DNA damage in plant herbarium tissue. *PLoS ONE*, *6*, e28448. https://doi.org/10.1371/journal.pone.0028448

Stamatakis, A. (2014). RAxML version 8: A tool for phylogenetic analysis and post-analysis of large phylogenies. *Bioinformatics*, *30*, 1312–1313. https://doi.org/10.1093/bioinformatics/btu033

Stevens, P. F. (2001). *Angiosperm phylogeny website*. Retrieved from http://www.mobot.org/MOBOT/research/APweb/

Straub, S. C. K., Cronn, R. C., Edwards, C., Fishbein, M., & Liston, A. (2013). Horizontal transfer of DNA from the mitochondrial to the plastid genome and its subsequent evolution in milkweeds (Apocynaceae). *Genome Biology and Evolution*, *5*, 1872–1885. https://doi.org/10.1093/gbe/evt140

Svab, Z., & Maliga, P. (2007). Exceptional transmission of plastids and mitochondria from the transplastomic pollen parent and its impact on transgene containment. *Proceedings of the National Academy of Sciences of the United States of America*, *104*, 7003–7008. https://doi.org/10.1073/pnas.0700063104

Tamura, K., Stecher, G., Peterson, D., Filipski, A., & Kumar, S. (2013). MEGA6: Molecular evolutionary genetics analysis version 6.0. *Molecular Biology and Evolution*, *30*, 2725–2729. https://doi.org/10.1093/molbev/mst197

Testolin, R., & Cipriani, G. (1997). Paternal inheritance of chloroplast DNA and maternal inheritance of mitochondrial DNA in the genus *Actinidia*. *Theoretical and Applied Genetics*, *94*, 897–903. https://doi.org/10.1007/s001220050493

Thompson, J. D. (2005). *Plant evolution in the Mediterranean*. Oxford, UK: Oxford University Press. https://doi.org/10.1093/acprof:oso/9780198515340.001.0001

Thyssen, G., Svab, Z., & Maliga, P. (2012). Exceptional inheritance of plastids via pollen in *Nicotiana sylvestris* with no detectable paternal mitochondrial DNA in the progeny. *The Plant Journal*, *72*, 84–88. https://doi.org/10.1111/j.1365-313X.2012.05057.x

Unseld, M., Marienfeld, J. R., Brandt, P., & Brennicke, A. (1997). The mitochondrial genome of *Arabidopsis thaliana* contains 57 genes in 366,924 nucleotides. *Nature Genetics*, *15*, 57–61. https://doi.org/10.1038/ng0197-57

Unver, T., Wu, Z., Sterck, L., Turktas, M., Lohaus, R., Li, Z., … Llorens, C. (2017). Genome of wild olive and the evolution of oil biosynthesis. *Proceedings of the National Academy of Sciences of the United States of America*, *114*, E9413–E9422. https://doi.org/10.1073/pnas.1708621114

Van de Paer, C., Hong-Wa, C., Jeziorski, C., & Besnard, G. (2016). Mitogenomics of *Hesperelaea*, an extinct genus of Oleaceae. *Gene*, *594*, 197–202. https://doi.org/10.1016/j.gene.2016.09.007

Volfovsky, N., Haas, B. J., & Salzberg, S. L. (2001). A clustering method for repeat analysis in DNA sequences. *Genome Biology*, *2*, research0027-1.

Wallander, E., & Albert, V. (2000). Phylogeny and classification of Oleaceae based on *rps16* and *trnL-F* sequence data. *American Journal of Botany*, *87*, 1827–1841. https://doi.org/10.2307/2656836

Wang, D., Rousseau-Gueutin, M., & Timmis, J. N. (2012). Plastid sequences contribute to some plant mitochondrial genes. *Molecular Biology and Evolution*, *29*, 1707–1711. https://doi.org/10.1093/molbev/mss016

Weihe, A., Apitz, J., Pohlheim, F., Salinas-Hartwig, A., & Börner, T. (2009). Biparental inheritance of plastidial and mitochondrial DNA and hybrid

variegation in *Pelargonium*. *Molecular Genetics and Genomics*, *282*, 587–593. https://doi.org/10.1007/s00438-009-0488-9

Wendel, J. F., & Doyle, J. J. (1998). Phylogenetic incongruence: Window into genome history and molecular evolution. In D. E. Soltis, P. S. Soltis, & J. J. Doyle (Eds.), *Molecular systematics of plants II: DNA sequencing* (pp. 265–296). New York, NY: Springer. https://doi.org/10.1007/978-1-4615-5419-6

Wolfe, K. H., Li, W. H., & Sharp, P. M. (1987). Rates of nucleotide substitution vary greatly among plant mitochondrial, chloroplast, and nuclear DNAs. *Proceedings of the National Academy of Sciences of the United States of America*, *84*, 9054–9058. https://doi.org/10.1073/pnas.84.24.9054

Xiang, Q. P., Wei, R., Shao, Y. Z., Yang, Z. Y., Wang, X. Q., & Zhang, X. C. (2015). Phylogenetic relationships, possible ancient hybridization, and biogeographic history of *Abies* (Pinaceae) based on data from nuclear, plastid, and mitochondrial genomes. *Molecular Phylogenetics and Evolution*, *82*, 1–14. https://doi.org/10.1016/j.ympev.2014.10.008

Xu, H., Luo, X., Qian, J., Pang, X., Song, J., Qian, G., … Chen, S. (2012). FastUniq : A fast de novo duplicates removal tool for paired short reads. *PLoS ONE*, *7*, e52249. https://doi.org/10.1371/journal.pone.0052249

Zedane, L., Hong-Wa, C., Murienne, J., Jeziorski, C., Baldwin, B. G., & Besnard, G. (2016). Museomics illuminate the history of an extinct, paleoendemic plant lineage (*Hesperelaea*, Oleaceae) known from an 1875 collection from Guadalupe Island, Mexico. *Biological Journal of the Linnean Society*, *117*, 44–57. https://doi.org/10.1111/bij.12509

Zhang, C., Li, F., Jia, X., Li, Z., & Feng, Q. (2013). Bioinformatics analysis of mitochondrial genes *sdh3* and *sdh4* of cytoplasmic male sterility in tobacco. *Chinese Tobacco Science*, *3*, 42–47.

Zhang, N., Wen, J., & Zimmer, E. A. (2015). Congruent deep relationships in the grape family (Vitaceae) based on sequences of chloroplast genomes and mitochondrial genes via genome skimming. *PLoS ONE*, *10*, e0144701. https://doi.org/10.1371/journal.pone.0144701

Zhu, A., Guo, W., Jain, K., & Mower, J. P. (2014). Unprecedented heterogeneity in the synonymous substitution rate within a plant genome. *Molecular Biology and Evolution*, *31*, 1228–1236. https://doi.org/10.1093/molbev/msu079

SUPPORTING INFORMATION

Additional Supporting Information may be found online in the supporting information tab for this article.

How to cite this article: Van de Paer C, Bouchez O, Besnard G. Prospects on the evolutionary mitogenomics of plants: A case study on the olive family (Oleaceae). *Mol Ecol Resour*. 2017;00:1–17. https://doi.org/10.1111/1755-0998.12742

CHAPITRE III

Genomic consequences of occasional biparental transmission of organelles in two lineages of the olive family (Oleaceae)

VAN DE PAER C., BESNARD G.

In preparation

Abstract

In most of flowering plants, the inheritance of DNA-containing organelles (i.e. chloroplasts and mitochondria) is uniparental, and usually maternal. However, occasional transmission of organelles via pollen (i.e. paternal leakage) has been reported in numerous angiosperm lineages. The genomic consequences of paternal leakage on the evolutionary trajectory of chloroplastic and mitochondrial genomes (hereafter plastomes and mitogenomes, respectively) remain unclear. A link between accelerations of the evolutionary rates with genomic rearrangements of organellar genomes and paternal leakage seems, however, to occur. In this study, we investigated the genomic consequences of paternal leakage in the Oleaceae family, in which both strict maternal inheritance and potential paternal leakage of organelles occur. Two independent lineages, the *Jasmineae* and a clade of *Ligustrinae* (*LVP*), exhibit potential transmission of organellar genomes via pollen. We highlighted large extensions/contractions of the Inverted Repeats (IR) in plastomes of both lineages. In particular, the plastomes of *Jasmineae* have multiple indels and have lost the gene *accD*. In addition, we detected an acceleration of the evolutionary rate in plastomes of these two lineages. While genes coding for subunits of photosynthetic complexes appear to be unaffected by change events of the evolutionary rate, an acceleration of the non-synonymous substitution rate (d_N) was detected for 42 and 19% of chloroplastic genes in *Jasmineae* and the *LVP* clade, respectively. We found that positive selection occurred on ten genes in *Jasmineae*, including *ycf1*, *ycf2*, and *clpP*. Although mitochondria were observed in pollen of *Jasmineae*, we do not detect clear evidences for changes in the evolutionary rates of mitochondrial genes in this lineage.

Key words: *accD*, *clpP*, mitochondria, paternal leakage, phylogeny, plastid, positive selection

Introduction

The genome of eukaryotes is partitioned into different compartments in the cell. While the main part of the genome is located in the nucleus, some organelles (i.e. mitochondria and chloroplasts in plants) also contain their own genome, relic of bacterial endosymbioses (Keeling 2010; Martin & Mentel 2010; Gray & Archibald 2012; Jensen & Leister 2014). Following the endosymbiosis events, mitochondrial genome (hereafter mitogenome) and chloroplastic genome (hereafter plastome) were drastically reduced and most of their genes were transferred to the nuclear genome (Bock & Timmis 2008). Nuclear and organellar genomes have also co-evolved, especially due to their coordinated gene expression (mainly controlled by the nuclear genome) to ensure metabolic functions necessary for the proper functioning of the cell (Greiner & Bock 2013). Organellar genomes thus followed similar evolutionary pathways, for instance in their inheritance mode, and their strategies of gene expression and replication (Greiner & Bock 2013). An uniparental inheritance (usually maternal) also constrains the evolution of organellar genomes due to an absence of recombination. However, plastomes and mitogenomes of plants also generally show contrasted evolutionary rates and show distinct features particularly in their structural organization.

In flowering plants, the mitogenomes usually have a complex structure in subgenomic circles (i.e. multipartite structure; Gualberto *et al.* 2014). The numerous repeats [up to several kilobases (kb)] distributed in the mitogenome promote recombination events, leading to multiple and interchangeable configurations of the mitogenome. Moreover, mitogenomes of plants are very large and variable in size between species, or even at the intra-specific level. The size of angiosperm mitogenomes ranges from ~200 kb in *Brassica rapa* (Brassicaceae) to ~11 megabases (Mb) in *Silene conica* [Caryophyllaceae; Smith & Keeling (2015)]. These large variations of size are generally due to variable content of repeated sequences and non-coding regions in the mitogenome. The mitogenomes of plants contain approximately 30–37 protein-coding genes, 15-21 tRNA genes and three rRNA genes (Clifton *et al.* 2004; Kubo *et al.* 2011). In contrast, the size, structure and organization of plastomes are typically very conserved in terms of size, structure and gene content. The typical plastome of angiosperms contains 66-82 protein-coding genes, 29-32 tRNA genes and four rRNA genes, and is usually represented as a circular molecule of ~150 kb in length (Jansen & Ruhlman 2012). Its structure is quadripartite including two long inverted, repeated regions (the Inverted Repeats, IR) separated by two single-copy regions [one small and one large region referred as SSC and LSC, respectively; Wicke *et al.* (2011)]. Gene conversion between the IR contributes to the stability of the plastomes (Khakhlova & Bock 2006).

In terms of sequence evolution, the rate of synonymous substitutions in mitochondrial genes is particularly low, being three to four times lower than in plant plastomes, 10–20 times lower than in plant nuclear genome and 50–100 times lower than in mammalian mitogenomes (Wolfe *et al.* 1987; Palmer & Herbon 1988). Deviations of these usual structural and evolutionary features have, however, been reported both in mitogenomes and plastomes. An

extreme acceleration of the evolutionary rate has been described in organellar genomes of several lineages of flowering plants:

- Lineages of *Silene* (Sloan *et al.* 2012), *Plantago* (Cho *et al.* 2004), *Ajuga* (Zhu *et al.* 2014) and Geraniaceae [particularly *Pelargonium* (Parkinson *et al.* 2005; Park *et al.* 2017)] exhibit extreme accelerated rates of mitochondrial DNA (mtDNA) sequence evolution. In Geraniaceae, an increase of the substitution rate has been reported in at least two independent lineages, while reversions towards a low substitution rate have also occurred at least three times in the family (Parkinson *et al.* 2005). The acceleration of the substitution rate is usually homogeneous throughout the whole mitogenome, except for few cases such as the highly variable gene *atp9* in most of *Silene* species (Sloan *et al.* 2009).
- Similarly, accelerations of the chloroplastic DNA (cpDNA) evolution have been reported in numerous lineages. Contrary to mitogenomes, only a portion of plastid genes exhibits high substitution rates. Genes coding for subunits of photosynthetic complexes (e.g. *atp*, *pet*, *psa*, *psb*, *ndh*, *rbcL*) appear to be unaffected by change events of the evolutionary rate. On the other hand, *accD*, *clpP*, *ycf1* and *ycf2* have an accelerated evolutionary rate in numerous lineages such as *Plantago*, *Medicago*, *Silene*, Campanulaceae, Geraniaceae and acacias (Haberle *et al.* 2008; Sloan *et al.* 2012; Gurdon & Maliga 2014; Zhu *et al.* 2016; Mensous *et al.* 2017; Park *et al.* 2017). In particular, an extremely acceleration of the substitution rate with positive selection was detected on *clpP* in lineages of *Silene*, *Oenothera, Geranium* and acacias (Erixon & Oxelman 2008; Sloan *et al.* 2012; Mensous *et al.* 2017; Park *et al.* 2017). In *Lathyrus odoratus* (Fabaceae), the evolutionary rate of *ycf4* is also exceptionally high (Magee *et al.* 2010). These extreme accelerations of the substitution rates in ptDNA genes are usually correlated with genomic rearrangements of plastomes such as expansions/contractions of the IR, losses of genes and a high content of repeated sequences (Jansen *et al.* 2007; Weng *et al.* 2013). First, large expansions of the IR have been reported, for instance, in plastomes of *Pelargonium* (Chumley *et al.* 2006), *Plantago* (Zhu *et al.* 2016), Campanulaceae (Cosner *et al.* 1997) and acacias (Mensous *et al.* 2017; Wang *et al.* 2017), while large contractions and even the complete loss of the IR have been described in *Erodium* [Geraniaceae, Guisinger *et al.* (2011)] and many Fabaceae (Palmer *et al.* 1987; Lavin & Doyle 1990). These large expansions and reductions of the IR resulted in the move of genes between the IR and the single-copy regions (Zhu *et al.* 2016). The synonymous substitution rate being ~3.4 fold lower in the IR compared to single-copy regions, the relocation of genes to the IR usually leads to a decrease of the substitution rate (Zhu *et al.* 2016). However, the lineages of *Pelargonium*, *Plantago* and *Silene* exhibit genes in the IR with a synonymous substitution rate higher (1.4- to 2.1-fold) than genes in single-copy regions (Zhu *et al.* 2016). Second, genes are often lost in lineages with accelerated evolutionary rate (for a review, see Jansen & Ruhlman 2012). Seven independent losses of *accD* are documented in the flowering plants, including lineages of Geraniaceae, *Trachelium* (Campanulaceae), *Trifolium* (Fabaceae) and *Jasminum* (Oleaceae). *clpP* has also been lost in several lineages such as *Trachelium*, Actinidiaceae or some Geraniaceae. Among the numerous gene losses reported in angiosperms, a gene relocation towards the nuclear genome has been demonstrated for *rpl22* in pea (Gantt *et al.* 1991), *infA* in *Arabidopsis* and soybean

(Millen *et al.* 2001), *rpl32* in poplar (Cusack & Wolfe 2007; Ueda *et al.* 2007) and *accD* in *Trifolium* and Campanulaceae (Magee *et al.* 2010; Rousseau-Gueutin *et al.* 2013). Lastly, plastomes of lineages with an extreme substitution rate acceleration (e.g. Geraniaceae, Campanulaceae, *Plantago*) appear to contain more repeated regions than those of other angiosperms and to be highly rearranged [with multiple indels, inversions and gene duplications; Cosner *et al.* (1997, 2004); Chumley *et al.* (2006); Lee *et al.* (2007); Guisinger *et al.* (2011); Weng *et al.* (2013); Zhu *et al.* (2016)].

It is generally acknowledged that the generation time (e.g. perennial versus annual plants), population size and speciation rate are involved in changes of the evolutionary rate (Laroche *et al.* 1997; Smith & Donoghue 2008). However, these factors cannot explain extreme accelerations of the substitution rate in organellar genomes. Localized hypermutation regions could be responsible of high variations in the synonymous rate (Parkinson *et al.* 2005; Magee *et al.* 2010; Zhu *et al.* 2016). Repeated DNA breakage and repair could lead to a hotspot of mutations (Guisinger *et al.* 2008, 2011; Yang *et al.* 2008). Sloan *et al.* (2009) assumed that the extreme divergence of *atp9* in the mitogenome of *Silene* could result from recombination between multiple copies of the gene, potentially associated with paternal leakage (i.e. occasional transmission of mitochondria via pollen). An association between accelerated substitution rates with large genomic rearrangements and potential biparental inheritance was also reported (Jansen & Ruhlman 2012).

Although the inheritance of organelles is predominantly maternal, paternal leakage has been observed in numerous lineages, both for mitochondria and/or chloroplasts (Corriveau & Coleman 1988; Zhang *et al.* 2003). The maternal inheritance of organelles could lead to the accumulation of the deleterious mutations (called Muller's ratchet) due to the absence of recombination (Muller 1964; Greiner *et al.* 2014). It is assumed that paternal leakage could allow to counteract the Muller's ratchet (Greiner *et al.* 2014). The genomic consequences of the occasional transmission of organelles via pollen are, however, unclear. Paternal leakage could promote recombination between maternal and paternal organellar genomes (Greiner *et al.* 2014). While recombination between mitogenomes of plants have been demonstrated, recombination events between plastomes is less frequent, particularly due to the absence of chloroplastic fusion in the zygote (Greiner *et al.* 2014). Instead, gene conversion between cpDNA fragments could be a more likely process (Goulding *et al.* 1996). DNA repair processes via gene conversion could generate errors in DNA synthesis because the recombinational DNA repair by double-strand-break (DSB) may have an higher error rate than the other repair processes as shown, for instance, in yeast (Strathern *et al.* 1995). In addition, the DSB of one IR followed by the recombinational repair against the complementary IR could result in large expansions of the IR (Goulding *et al.* 1996).

A possible scenario of an accelerated substitution rate is that gene conversion by DSB between maternal and paternal organellar copies could generate errors in DNA synthesis. Moreover, the possibility of heteroplasmy with recombination and/or gene conversion between organelles of both parents could relax some constraints and, for instance, allow DNA replication enzymes to make more errors. Both these potentially error-prone processes should

lead to an increase of the mutation rate. Several changes in selective pressures could result from a rise of the substitution rate: (*i*) a reinforcement of the purifying selection in order to eliminate deleterious mutations, (*ii*) a relaxation of the selection due to a new evolutionary context (i.e. organelle inheritance), and (*iii*) positive selection allowing deleterious mutations to be compensated by other mutations. This last process could provide the opportunity to reach new evolutionary optimums. Selection could operate in order to promote the organellar genomes compatible with the nuclear genome (Weng *et al.* 2013). Biparental inheritance could allow to counteract the incompatibilities between the nuclear and organellar genomes (Zhang & Sodmergen 2010). These expectations have, however, never been demonstrated and a first step to test these assumptions is to detect a correlation between paternal leakage and acceleration of the substitution rate.

Oleaceae is an excellent study system to investigate the link between paternal leakage and accelerations of the substitution rate with genomic rearrangements in organellar genomes, because both maternal inheritance and paternal leakage of organelles have been observed in this family. Maternal inheritance of both organellar genomes was documented in the olive complex (*Olea europaea*) based on the genetic characterization of cpDNA and mtDNA polymorphisms in progenies of controlled crosses (Besnard *et al.* 2000). In addition, Besnard *et al.* (2002) highlighted a strong linkage disequilibrium between these polymorphisms (despite intra-population diversity) also suggesting their strict maternal inheritance in the olive. A maternal transmission of chloroplasts and mitochondria has also been observed by microscopy in seven accessions of *Syringa* belonging to subgenus *Syringa* series *Syringa* and *Pinnatifoliae* (Corriveau & Coleman 1988; Liu *et al.* 2004). In contrast, potential paternal leakage has been reported in two lineages. Firstly, Liu *et al.* (2004) documented potential paternal leakage of the chloroplasts but a strict maternal inheritance of the mitochondria in 15 species of subtribe *Ligustrinae* (i.e. 12 *Syringa* species belonging to subgenus *Ligustrina* and subgenus *Syringa* series *Villosae* and *Pubescentes,* and three species of *Ligustrum*). These species belong to a monophyletic lineage of *Ligustrinae* meaning that the switch to potential paternal leakage has occurred in the ancestor of this lineage (Liu *et al.* 2004). Secondly, chloroplasts and mitochondria have also been observed in pollen grains of two distantly related species of *Jasminum* (*J. officinale* and *J. nudiflorum*) suggesting a paternal leakage of both organellar genomes in tribe *Jasmineae* (Sodmergen *et al.* 1998).

In this study, we investigated the genomic consequences of a paternal leakage in Oleaceae. We reconstructed phylogenies of 100 accessions covering the lineages of Oleaceae, based on complete plastomes and on coding and non-coding mtDNA regions. We detected changes in the evolutionary rate, both in cpDNA and mtDNA, by estimating rates of non-synonymous and synonymous substitutions and testing the violation of the strict molecular clock. We then performed tests of positive selection in *Jasmineae* and a clade of *Ligustrinae* for each chloroplastic and mitochondrial genes, in order to test for changes of selective pressures.

Materials and methods

Taxon sampling

In this study, we analyzed 100 species representative of all tribes and sub-tribes of Oleaceae as defined by Wallander & Albert (2000). Firstly, we re-used 90 complete plastomes of Oleaceae generated by Olofsson *et al.* (*in prep.*). Compared to the taxon sampling used in Olofsson *et al.* (*in prep.*), we removed six accessions (*Forestiera angustifolia*, *Nestegis cunninghamii*, *Olea europaea* subsp. *europaea* E1-*e*.4, *O. e.* subsp. *europaea* E2.1, *O. e.* subsp. *guanchica* M-*g*1.3 and *Phillyrea angustifolia*) in order to have one accession per species. Secondly, we used the complete plastome [NC_015401] and mtDNA contigs of *Fraxinus excelsior* generated by Sollars *et al.* (2017). Lastly, we added nine species belonging to tribes *Jasmineae*, *Forsythieae* and *Fontanesieae,* including *Nyctanthes arbor-tristis* (tribe *Myxopyreae*) as an outgroup (Table 1). The position of tribe *Myxopyreae* as sister to all other Oleaceae tribes was reported by Wallander & Albert (2000) and confirmed by Zedane (2016). The nine additional samples were either collected from the field or herbarium collections.

Table 1. List of plant accessions sequenced in our study.

Tribe, subtribe	Species, and name and origin of analyzed individuals
Jasmineae	*Jasminum fluminense* Vell. - O. Maurin 273 [PET], South Africa *Jasminum fruticans* L. - CEFE (J. Thompson), G. Besnard 2016, France *Jasminum kitchingii* Baker -F. Rakotonasolo, sn [TAN], Madagascar *Jasminum polyanthum* French. - Hervé Gryta, sn, ornamental from Asia
Fontanesieae	*Fontanesia fortunei* Carrière - T. Jossberger 1829 [BONN], China
Forsythieae	*Forsythia giraldiana* Lingelsh. - Fliegner et al. SICH1137; coll QBGII-1992.238 [QBG], China *Forsythia* × *intermedia* Zabel - G. Besnard [MPU], ornamental from E. Asia *Forsythia* × *mandschurica* Uyeki - 18-40-1968 (Montreal BG, JFRU21), China
Myxopyreae	*Nyctanthes arbor-tristis* L. - C. Parma 9713 [M], India

Voucher samples deposited at the herbaria of Kew [K], Munich [M], Bonn [BONN], Montpellier [MPU], Missouri [MO], Pretoria [PET] and Antananarivo [TAN].

Organellar DNA sequencing and assembly

Total genomic DNA of the nine species was extracted from ca. 5 mg of fresh or dried leaves. Each sample was ground in a 2-ml tube containing three tungsten beads with a TissueLyser (Qiagen Inc., Texas). The BioSprint 15 DNA Plant Kit (Qiagen Inc.) was then used to extract DNA, which was eluted in 200 µl of the AE buffer. We performed shotgun sequencing using a genome skimming approach as described in Zedane *et al.* (2016). Twenty four DNA libraries were multiplexed per flow cell and they were sequenced with an Illumina machine (i.e. HiSeq 2000, 2500 or 3000).

We reconstructed full plastomes of the nine species as described in Zedane *et al.* (2016). A verification of the assembly was performed by mapping all paired-reads against the obtained consensus sequence using GENEIOUS V.9.0.5 (Kearse *et al.* 2012) to check assembly quality and coverage depth. We annotated plastomes using GENEIOUS by transferring the annotations of an *O. europaea* plastome [FN996972]. Start and stop codons of protein-coding genes were manually adjusted. Then, we assembled 52 mtDNA regions (i.e. 36 protein-coding genes and 16 introns) for each of the 100 accessions. For each accession, we mapped all paired-end reads against the annotated mitogenome of *H. palmeri* [KX545367] using GENEIOUS following the method described in Van de Paer *et al.* (2017). To retrieve mtDNA regions of *F. excelsior,* we mapped assembled mtDNA genes and introns of *F. angustifolia* against the 26 mtDNA contigs of the library BATG-0.5-CLCbioSSPACE generated by Sollars *et al.* (2017) .

Phylogenetic analyses

We performed two phylogenetic analyses based on the complete plastomes and on the concatenation of the 52 newly assembled mtDNA regions. We aligned the plastomes and mtDNA regions using MUSCLE (Edgar 2004) with default parameters. Maximum likelihood (ML) analyses were then performed using RAxML (Stamatakis 2014) in the CIPRES platform (Miller *et al.* 2010). We applied the GTR+G model and the rapid bootstrap algorithm with 1,000 iterations.

Estimation of substitution rates and GC content

We extracted nucleotide sequences for all protein-genes from each plastome. Each chloroplastic and mitochondrial gene was aligned individually using MUSCLE with default parameters. The gene alignments were manually refined taking into account their amino-acid translation. We calculated non-synonymous and synonymous substitution rates (d_N and d_S) using *codeml* implemented in PAML v.4.9 (Yang 2007) for (*i*) each chloroplastic and mitochondrial gene individually, and (*ii*) each concatenated dataset of genes classified per functional category. We estimated d_N and d_S for each accession against the outgroup (*Nyctanthes*).

As a paternal leakage was detected in *Jasmineae* (Sodmergen *et al.* 1998) and a monophyletic lineage of *Ligustrinae* [hereafter '*LVP*': *Ligustrum* spp., plus *Syringa* subgenus *Syringa* series *Villosae* (*Sy. yunnanensis*) and *Pubescentes* (*Sy. microphylla*); Liu *et al.* 2004],

we tested for a difference in the substitution rates between these two clades and the other accessions of the tree. To compare the synonymous and the non-synonymous substitution rates between *Jasmineae*, *LVP* and the other accessions of Oleaceae (noted in the analyses '*Jasmineae*', '*LVP*' and '*Others*', respectively), we performed a phylogenetic ANOVA between the three groups for each gene and each functional category with the function *phylANOVA* in the R package *phytools* (Garland *et al.* 1993; Revell 2012). We used the previous phylogenetic topology based on the complete plastomes to perform the analyses. We assigned *Syringa vulgaris* to '*Others*' as strict maternal inheritance of chloroplasts was observed in this species (Corriveau & Coleman 1998; Liu *et al.* 2004).

We estimated the GC content [GC] on six datasets: (*i*) the complete plastomes and the concatenated coding and non-coding mtDNA regions, (*ii*) the concatenated chloroplastic and mitochondrial coding genes, separately, and (*iii*) the third position of codons of chloroplastic and mitochondrial genes, separately. For each of these three cases, we performed the estimations on (*i*) all 100 species of Oleaceae, (*ii*) the Oleaceae species without *Jasmineae* and *LVP*, (*iii*) the *Jasmineae*, and (*iv*) the *LVP* clade.

Models of molecular clock

In order to highlight different evolutionary rates in cpDNA and mtDNA between both clades *Jasmineae* and *LVP* and the other accessions of Oleaceae, we compared models of molecular clock using *baseml* implemented in PAML. First, we tested the violation of the strict molecular clock among the Oleaceae lineages without clades *Jasmineae* and *LVP*. We compared for each gene the strict molecular clock, which assumes a single rate for all lineages in the tree (noted *clock1*), and the relaxed molecular clock, which allows all branches to evolve independently in the tree (noted *clock0*). Second, we tested the assumption of rate-change events across both clades *Jasmineae* and *LVP* in the Oleaceae tree. We compared the strict and relaxed molecular clocks tested on the whole Oleaceae tree with the local molecular clock, which allows a distinct evolutionary rate to some lineages while all the other lineages evolved with a same evolutionary rate (noted *clock2*). To test the local molecular clock model, we assigned a priori changes of evolutionary rates in the Oleaceae tree: (*i*) on *Jasmineae* (noted *clock2_Jasmineae*), (*ii*) on the *LVP* clade (noted *clock2_LVP*), and (*iii*) on both *Jasmineae* and the *LVP* clade (noted *clock2_Jasmineae_LVP*). The molecular clock model that fits the data best was obtained given the lowest absolute Akaike Information Criterion (AIC). We considered that models were not significantly different if the difference between their AIC was inferior to six ($\Delta AIC < 6$). We performed these analyses using the previous phylogenetic topology based on complete plastomes. To test the relaxed molecular clock (i.e. *clock0*), we used the unrooted cpDNA tree.

Positive selection tests

We analyzed changes of selective pressures in *Jasmineae* and *LVP* for each chloroplastic and mitochondrial gene. We performed the analyses with *codeml* implemented in PAML and using the phylogenetic topology based on complete plastomes. For each gene, we tested three codon models, which use the non-synonymous versus synonymous ratio ($\omega = d_N / d_S$; omega). The ratio ω allows to evaluate the selective pressures at the protein level, indicating purifying

($\omega < 1$), relaxed ($\omega = 1$) or positive ($\omega > 1$) selection. The first model M1a is a site model allowing ω to vary among codons (model = 0, NSSites = 1). In this model, codons are either under relaxed selection ($\omega = 1$) or under purifying selection ($\omega < 1$; Yang *et al.* 2000). Second, the model A is a branch-site model, which allows ω to vary among both codons and branches of the phylogenetic tree (model = 2, NSSites = 2, fix_omega = 0; Zhang *et al.* 2005). Branch-site models allow distinct evolutionary rates along particular branches (called foreground branches) while a constant rate is assumed for all other branches of the tree (called background branches). The branch-site models need to assign a priori foreground branches for which positive selection is expected. In the model A, codons can be attributed to four types of selection: (*i*) purifying selection on the whole tree ($\omega < 1$), (*ii*) relaxed selection on the whole tree ($\omega = 1$), (*iii*) purifying selection on background branches ($\omega_0 < 1$) and positive or relaxed selection on foreground branches ($\omega_2 \geq 1$), and (*iv*) relaxed selection on background branches ($\omega_1 = 1$) and positive or relaxed selection on foreground branches ($\omega_2 \geq 1$). The last model A' is the same as model A, except that positive selection on foreground branches is replaced by relaxed selection ($\omega_2 = 1$; model = 2, NSSites = 2, fix_omega = 1, omega = 1; Zhang *et al.* 2005). Individually for each lineage (i.e. *Jasmineae* and *LVP*), we tested two hypotheses: (*i*) positive selection operates on the internal branch ancestral to the clade (hypothesis H1; Fig. 1); and (*ii*) positive selection operates on all branches of the clade (hypothesis H2; Fig. 1). We removed the *LVP* clade from the phylogenetic topology to perform selection tests on *Jasmineae*; and similarly, we removed *Jasmineae* to perform selection tests on *LVP*. To identify the model that fits data best, we choose for each lineage the model with the lowest AIC between the five models (model M1a; models A and A' under the hypothesis H1; models A and A' under the hypothesis H2). We considered that the models were not significantly different if the difference between their AIC was inferior to six ($\Delta AIC < 6$).

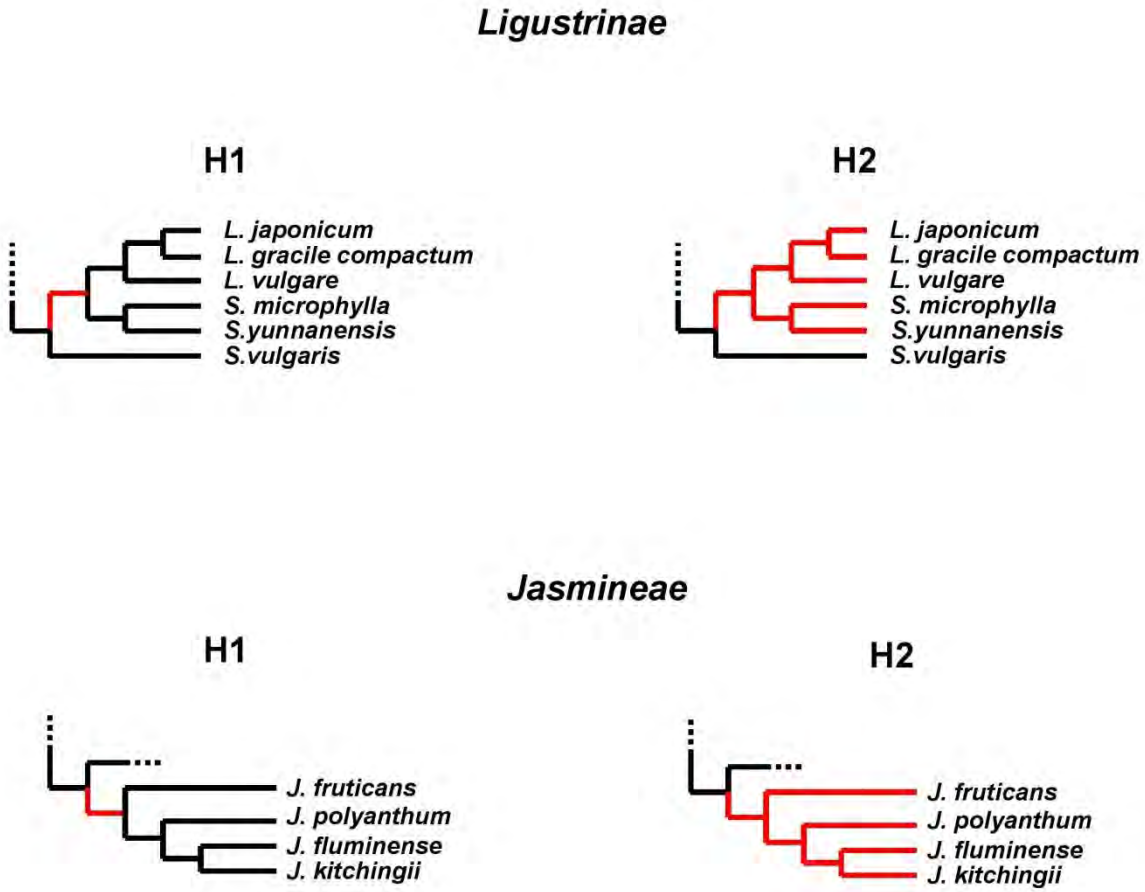


Fig. 1. Alternative hypotheses used for positive selection tests. For the two tested hypotheses, foreground branches of the phylogenetic tree are indicated in red. In hypothesis H1, positive selection has occurred only on the internal branch ancestral to the *Jasmineae* and *LVP* clades; while in hypothesis H2, positive selection has occurred on all branches of the *Jasmineae* and *LVP* clades.

Results

Content and structure of organellar genomes

First, we sequenced and assembled complete plastomes of nine accessions of Oleaceae. All the newly assembled plastomes and those generated by Olofsson *et al.* (*in prep*.) contained 80 protein-coding genes, except for those of *Jasminum*. As previously reported in *J. officinale* and *J. nudiflorum* (Lee *et al.* 2007), *accD* was absent from the four *Jasminum* plastomes that we reconstructed. Lastly, we assembled between 34 and 36 mtDNA protein-coding genes and between 14 and 16 mtDNA introns (Table S1). We excluded the mtDNA regions with a too low depth coverage, which did not allow us to reconstructed the complete mtDNA sequence.

The GC content was ~37% and ~46% across the complete plastomes and the concatenated mtDNA fragments, respectively; and was ~29% and ~37% at the third position of codons of concatenated chloroplastic and mitochondrial genes, respectively (Table 2). We did not detect any differences of the GC content between the *Jasmineae* and *LVP* clades and the other accessions of Oleaceae, whatever in cpDNA or mtDNA (Table 2).

Table 2. The GC contents [GC] estimated on all Oleaceae species (noted 'All'), on Oleaceae species without the *Jasmineae* and *LVP* clades (noted 'Others'), on the *Jasmineae* clade (noted '*Jasmineae*'), on the *LVP* clade (*Ligustrinae* excluding *Sy. vulgaris*). We estimated GC contents on several datasets of cpDNA and mtDNA.

	All	Others	*Jasmineae*	*LVP*
cpDNA				
[GC] on the complete plastomes	36.8	36.7	36.6	37.1
[GC] on all concatenated genes	37.8	37.8	38.0	38.2
[GC] on the 3rd position of codons	29.0	29.0	29.3	29.6
mtDNA				
[GC] on the concatenation of genes and introns	46.2	46.3	46.2	46.2
[GC] on all concatenated genes	42.3	42.2	42.1	42.2
[GC] on the 3rd position of codons	37.1	37.1	36.7	36.9

We observed rearrangements in plastomes of *Jasmineae* and *Ligustrinae*. The four *Jasminum* are highly rearranged with large inversions and multiple insertions/deletions, as previously described in other *Jasminum* species by Lee *et al.* (2007). In particular, we observed expansion or contraction of the Inverted Repeat (IR; Fig. S1). The IR of *J. polyanthum* exhibits both an extension of 4,662 bp into the SSC resulting in two complete copies of *ycf1* in IR and a large reduction of 18,889 bp at the LSC/IR boundary, leading to the incorporation of 13 genes (including seven protein-coding genes, five tRNA genes and one rRNA gene) in the LSC (Fig. S1A). In *J. kitchingii* and *J. fluminense*, we also detected an

expansion of the IR (of 4,695 bp and 4,708 bp, respectively) with a shift of *ycf1* from the SSC to the IR (Fig. S1A). In *Ligustrinae*, we observed an extension (of ~6.5 kb) of the IR in all studied species, except in *Sy. vulgaris* that is sister to the *LVP* clade (Fig. S1B). This extension involved a shift of *ycf1*, *rps15* and *ndhH* from the SSC to the IR. We also detected a IR extension of 918 bp at the IR/SSC boundary in *Priogymnanthus hasslerianus* (*Oleinae*).

Phylogenetic analyses

We performed an alignment of the 100 complete plastomes (179,063 bp, after excluding one IR), of which 12,718 (~7%) out of 24,550 (~14%) variable sites are parsimony-informative. In addition, we aligned the 52 coding and non-coding mtDNA regions (58,585 bp). We detected 1,024 (~2%) parsimony-informative substitutions out of 2,095 (~4%) variable sites.

Parsimony tree topologies based on cpDNA and mtDNA were mostly congruent at the genus level, except for the slowly supported nodes (< 70%; Fig. 2). The cpDNA topology was strongly supported with > 95% bootstrap values at most nodes (Fig. 2A). Only phylogenetic relationships between closely related species were slowly supported (< 70%) in five cases. In contrast, the mtDNA topology was overall weakly supported, with a lot of bootstrap values inferior to 70% (Fig. 2B). In the cpDNA phylogeny, we observed long branches for *Jasmineae* and the *LVP* clade. The liana species *C. macrobotrys* has also evolved more rapidly than the other tree species of *Oleinae*. In the mtDNA phylogeny, long branches were observed for *Jasmineae* and for a pair of *Noronhia* species (*N. intermedia* and *N. clarinerva*).

Fig. 2. ML phylogenetic trees of the Oleaceae obtained with RAxML v.8 (Stamatakis 2014) with the GTR + G model. We reconstructed cpDNA and mtDNA phylogenetic trees based on complete plastomes (Fig. 2.A) and on the concatenated non-coding and coding mtDNA regions (Fig. 2.B), respectively. Bootstrap supports between 70% and 99% are represented with a black point and those inferior to 70% are represented with a gray point. Bootstrap values equal to 100% are not represented on phylogenies.

B

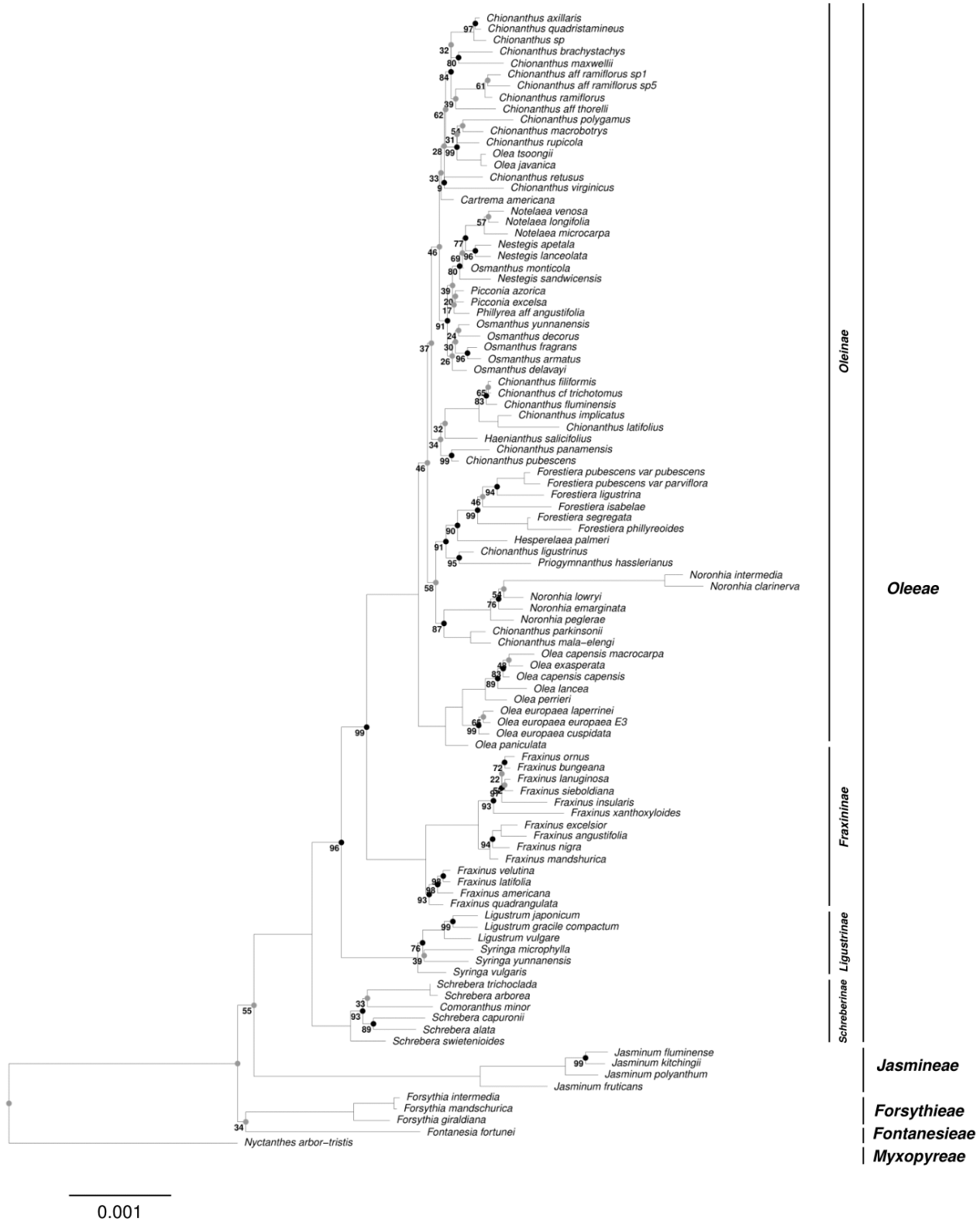

Chionanthus axillaris
Chionanthus quadristamineus
Chionanthus sp
Chionanthus brachystachys
Chionanthus maxwellii
Chionanthus aff ramiflorus sp1
Chionanthus aff ramiflorus sp5
Chionanthus ramiflorus
Chionanthus aff thorelii
Chionanthus polygamus
Chionanthus macrobotrys
Chionanthus rupicola
Olea tsoongii
Olea javanica
Chionanthus retusus
Chionanthus virginicus
Cartrema americana
Notelaea venosa
Notelaea longifolia
Notelaea microcarpa
Nestegis apetala
Nestegis lanceolata
Osmanthus monticola
Nestegis sandwicensis
Picconia azorica
Picconia excelsa
Phillyrea aff angustifolia
Osmanthus yunnanensis
Osmanthus decorus
Osmanthus fragrans
Osmanthus armatus
Osmanthus delavayi
Chionanthus filiformis
Chionanthus cf trichotomus
Chionanthus fluminensis
Chionanthus implicatus
Chionanthus latifolius
Haenianthus salicifolius
Chionanthus panamensis
Chionanthus pubescens
Forestiera pubescens var pubescens
Forestiera pubescens var parviflora
Forestiera ligustrina
Forestiera isabelae
Forestiera segregata
Forestiera phillyreoides
Hesperelaea palmeri
Chionanthus ligustrinus
Priogymnanthus hasslerianus
Noronhia intermedia
Noronhia clarinerva
Noronhia lowryi
Noronhia emarginata
Noronhia peglerae
Chionanthus parkinsonii
Chionanthus mala−elengi
Olea capensis macrocarpa
Olea exasperata
Olea capensis capensis
Olea lancea
Olea perrieri
Olea europaea laperrinei
Olea europaea europaea E3
Olea europaea cuspidata
Olea paniculata
Fraxinus ornus
Fraxinus bungeana
Fraxinus lanuginosa
Fraxinus sieboldiana
Fraxinus insularis
Fraxinus xanthoxyloides
Fraxinus excelsior
Fraxinus angustifolia
Fraxinus nigra
Fraxinus mandshurica
Fraxinus velutina
Fraxinus latifolia
Fraxinus americana
Fraxinus quadrangulata
Ligustrum japonicum
Ligustrum gracile compactum
Ligustrum vulgare
Syringa microphylla
Syringa yunnanensis
Syringa vulgaris
Schrebera trichoclada
Schrebera arborea
Comoranthus minor
Schrebera capuronii
Schrebera alata
Schrebera swietenioides
Jasminum fluminense
Jasminum kitchingii
Jasminum polyanthum
Jasminum fruticans
Forsythia intermedia
Forsythia mandschurica
Forsythia giraldiana
Fontanesia fortunei
Nyctanthes arbor−tristis
Oleinae
Oleeae
Fraxininae
Ligustrinae
Schreberinae
Jasmineae
Forsythieae
Fontanesieae
Myxopyreae
0.001

Fig. 2. End

Estimation of sequence divergence

We estimated non-synonymous and synonymous rates (d_N and d_S) on the 80 cpDNA protein-coding genes (Fig. S2A). We found a significant difference between '*Others*' and '*Jasmineae*' and/or '*LVP*' for d_N in 35 genes and for d_S in 17 genes ($P < 0.05$). These genes are distributed throughout the plastome and do not belong to a particular functional group. It appears that d_N does not differ significantly between the three groups for the genes belonging to the functional categories *atp*, *pet*, *psa* and *psb* (except for *atpA* and *psbB*). A significant increase of d_N and d_S was detected between '*Jasmineae*' and '*Others*' for 34 genes (out of 35) and 13 genes (out of 17), respectively ($P < 0.05$). Similarly, a significant increase of d_N and d_S was detected between '*LVP*' and '*Others*' for 15 genes (out of 35) and 8 genes (out of 17), respectively ($P < 0.05$). Among these genes, d_N was higher in both '*Jasmineae*' and '*LVP*' than '*Others*' for 13 genes (out of 35; $P < 0.05$), while d_S was higher in both '*Jasmineae*' and '*LVP*' than *'Others'* for 5 genes (out of 17; $P < 0.05$). An extreme increase of d_N and d_S was observed on *clpP* in both '*Jasmineae*' and '*LVP*' compared to '*Others*', and particularly between '*Jasmineae*' ($\bar{d}_N = 0.4852$; $\bar{d}_S = 0.3070$) and '*Others*' ($\bar{d}_N$ 0.0047); $\bar{d}_S = 0.0859$). Similarly, we estimated d_N and d_S on the 36 mtDNA genes and we only detected a significant increase of d_S between '*Jasmineae*' and '*Others*' in *mttB* (Fig. S2B).

We defined eight functional categories for chloroplastic genes [ATP synthase (*atp*), NADH dehydrogenase (*ndh*), cytochrome b_6f (*pet*), photosystem I (*psa*), photosystem II (*psb*), RNA polymerase (*rpo*), and small and large ribosomal subunits (*rps* and *rpl*)] and seven functional categories for mitochondrial genes [ATP synthase (*atp*), cytochrome *c* maturation (*ccm*), cytochrome *c* oxydase (cox), NADH dehydrogenase (*nad*), succinate dehydrogenase (*sdh*), and small and large ribosomal subunits (*rps* and *rpl*)]. We detected a significant increase of d_N and d_S between '*Jasmineae*' and '*Others*' for five and six functional categories of chloroplastic genes, respectively (Fig. S3A). Similarly, we found a significant increase of d_N and d_S between '*LVP*' and '*Others*' for three and two functional categories of chloroplastic genes, respectively. No significant difference has been detected between '*Jasmineae*', '*LVP*' and '*Others*' for functional categories of mitochondrial genes (Fig. S3B).

Models of molecular clocks

The test of strict molecular clock violation among the Oleaceae lineages without *Jasmineae* and *LVP* revealed that the hypothesis of a strict molecular clock (i.e. a constant evolutionary rate among lineages of the tree) is validated for all chloroplastic genes, except for two genes (*rpoC2* and *ycf1*; Table S2.A). The assumption of a relaxed molecular clock, which assumes more than a single rate on the whole tree, fitted better to these two genes **than** the strict molecular clock. All the mitochondrial genes validated the assumption of a single rate among the lineages of the tree, except *atp1* and *atp6*, which followed the hypothesis of a relaxed molecular clock among the Oleaceae lineages without *Jasmineae* and *LVP* (Table S2.B). We performed the same test for *atp1* and *atp6* after removing *N. intermedia* and *N. clarinerva* (for which an acceleration of the evolutionary rate occurred), but we also found that the relaxed molecular clock fitted better to these two genes.

By testing the assumption of more than a single rate across the Oleaceae tree including *Jasmineae* and *LVP*, we found that the local molecular clock fitted to 54 (71%) cpDNA genes (Table S3.A). We detected 23 genes for which a distinct rate is assumed for both clades *Jasmineae* and *LVP*, 29 genes for which a distinct rate is assumed to *Jasmineae* (or at least to *Jasmineae*, in the case of a non-significant difference between *clock2_Jasmineae_LVP* and *clock2_Jasmineae*) and two genes (*accD* and *ycf15*) for which a distinct rate is assumed to *LVP* (or at least to *LVP*, in the case of a non-significant difference between *clock2_Jasmineae_LVP* and *clock2_LVP*). We found that four genes followed the relaxed molecular clock. The molecular clock models were not significantly different for the other genes and did not allow us to conclude. For mitochondrial genes, only *mttB* followed the hypothesis of a distinct rate for both *Jasmineae* and *LVP* and three genes (*cox1*, *cytB* and *rps3*) followed the hypothesis of a distinct rate for *Jasmineae* (or at least for *Jasmineae*; Table S3.B). We found that the relaxed molecular clock fitted to *atp1* and *atp6*. The molecular clock models were not significantly different for the other genes.

Positive selection

We tested the occurrence of positive selection in *Jasmineae* and *LVP* for the 80 chloroplastic genes (Table S4.A). The model A under the hypothesis H2, which assumes that positive selection occurred on all branches of the clade (i.e. *Jasmineae* or *LVP*), was significantly better than other models for nine genes (*clpP*, *rbcL*, *rpoB*, *rpoC1*, *rps7*, *rps11*, rps12, *rps18* and *ycf1*) in *Jasmineae* (Tables 3 and S4.A). The model A under the hypothesis H1, which assumes that positive selection occurred on the internal branch of the clade, was significantly better than other models for *atpF* in *Jasmineae* and for *matK* in *LVP* (Tables 3 and S4.A). We also found that models A under both hypotheses H1 and H2 were not significantly different from each other for *ycf2* in *Jasmineae* but were significantly better than the other models. We thus could assume that positive selection occurred at least on the internal branch of *Jasmineae* for this gene. The models were not significantly different for the other genes, which did not allow us to conclude. We performed similar analyses for the 36 mitochondrial genes (Table S4.B). Model A under the hypothesis H2 was significantly better than other models for *cytB* and *nad5* in *Jasmineae* (Tables 3 and S4.B). We also found that models A under both hypotheses H1 and H2 were not significantly different from each other for *ccmFc* and *rps3* in *Jasmineae* but were significantly better than the other models. The models for the other genes were not significantly different.

Table 3. List of genes for which the model A (assuming a positive selection on foreground branches) was significantly better than other models under the hypothesis H1, which assumes a potential positive selection on the internal branch of the clade, or the hypothesis H2, which assumes a potential positive selection on all branches of the clade. For *ycf2*, *ccmFc* and *rps3*, both models A were not significantly different under the hypotheses H1 and H2, which assumes a positive selection on at least the internal branch of the clade.

	gene	clade	hypothesis	model
cpDNA				
	atpF	*Jasmineae*	H1	A
	clpP	*Jasmineae*	H2	A
	matK	*LVP*	H1	A
	rbcL	*Jasmineae*	H2	A
	rpoB	*Jasmineae*	H2	A
	rpoC1	*Jasmineae*	H2	A
	rps7	*Jasmineae*	H2	A
	rps11	*Jasmineae*	H2	A
	rps12	*Jasmineae*	H2	A
	rps18	*Jasmineae*	H2	A
	ycf1	*Jasmineae*	H2	A
	ycf2	*Jasmineae*	H1/H2	A
mtDNA				
	ccmFc	*Jasmineae*	H1/H2	A
	cytB	*Jasmineae*	H2	A
	nad5	*Jasmineae*	H2	A
	rps3	*Jasmineae*	H1/H2	A

Discussion

Plastomes showing deep rearrangements and faster evolution have been reported in angiosperm species with paternal leakage (Jansen *et al.* 2007; Jansen & Ruhlman 2012). To test for possible switches in the evolutionary trajectory linked to potential biparental inheritance of organelles, the Oleaceae family is an interesting case study, since both strict maternal inheritance and paternal leakage of chloroplasts and/or mitochondria have been detected in this tree lineage. Here, the structural features and substitution rates were compared in plastomes and mitochondrial genes between lineages with an a priori strict inheritance of organelles (that however needs to be documented in most species) and two lineages (*Jasmineae* and the clade *LVP* of *Ligustrinae*) with potential paternal leakage. We also tested the changes in selective pressures occurring in these two latter lineages.

Variation of the structural features in plastomes of Jasmineae *and a clade of* Ligustrinae

As already reported by Lee *et al.* (2007), the plastomes of *Jasmineae* are highly rearranged with large inversions and multiple indels. In our study, an extension of the IR into the SSC (~5 kb) was detected in three species out four, as similarly revealed in *J. nudiflorum* (Lee *et al.* 2007). In addition, both an extension of the IR into the SSC and a large contraction (~19 kb) of the IR at the LSC/IR boundary were detected in the plastome of one species (*J. polyanthum*). Plastomes of *J. fluminense* and *J. kitchingii* sharing the IR extension of *J. polyanthum*, we assume that this extension is ancestral to the divergence of these three species. Interestingly, the *LVP* clade of *Ligustrinae* with potential paternal leakage of chloroplasts also exhibits an extension (~6 kb) of the IR into the SSC. The expansions of the IR in *Jasmineae* and *Ligustrinae* are not of the same extent and should thus be independent events. A third lineage showing a shorter extension (~1 kb) of the IR into the SSC was also detected in *Priogymnanthus* (*Oleinae*). While most small IR extensions (< 100bp) in land plants involve a shift of the IR/LSC boundary (Goulding *et al.* 1996; Zhu *et al.* 2016), both IR extension events in Oleaceae consist in an extension of the IR into the SSC. An extension of the IR with a shift of the IR/SSC boundary has also been described in acacias (Mensous *et al.* 2017; Wang *et al.* 2017), *Plantago* (Zhu *et al.* 2016), *Trachelium* (Cosner *et al.* 1997) and *Silene* (Sloan & Wu 2014) and successive extensions and contractions of the IR occurred in *Pelargonium* (Chumley *et al.* 2006) and Apiales (Downie & Jansen 2015). Here, it is noticeable that the two lineages with possible paternal leakage (*Jasmineae* and the clade *LVP* of *Ligustrinae*) also show deep structural reorganizations of plastomes suggesting that transitory heteroplasmy have favored plastomic rearrangements. Goulding *et al.* (1996) proposed a model involving a double-strand break (DSB) to explain an IR expansion at the boundary IR/SSC in *Nicotiana*. The DSB of one IR would be followed by strand invasion and gene conversion, which could lead to large expansions of the IR. The different features of the IR expansions observed in angiosperms could, however, result from various mechanisms still unknown (Zhu *et al.* 2016).

The gene content was conserved among plastomes of Oleaceae except *accD* that is absent from our *Jasmineae* plastomes as previously reported by Lee *et al.* (2007) in this lineage. This gene encodes for the chloroplastic subunit of the acetyl-CoA carboxylase

(ACCase), which is an essential enzyme implicated in the fatty acid synthesis (Magnusson *et al.* 1993). Plants contain two ACCase forms: an eukaryotic form and a prokaryotic form. While genes encoding for the eukaryotic-type ACCase are located in the nuclear genome, the prokaryotic-type ACCase comprises four subunits encoded by three nuclear genes (*accA*, *accB* and *accC*) and *accD* in the plastome (Wakil *et al.* 1983; Magnuson *et al.* 1993; Kode *et al.* 2005). The latter gene has been lost in multiple lineages of angiosperms (Jansen & Ruhlman 2012), but its functional transfer towards the nuclear genome has been demonstrated in Campanulaceae (Rousseau-Gueutin *et al.* 2013) and *Trifolium* (Magee *et al.* 2010).

Acceleration of the evolutionary rate of chloroplastic genes in Jasmineae *and the LVP clade of* Ligustrinae

While the substitution rate of chloroplastic genes is usually slow with a low variation among lineages, our study reveals an acceleration of the evolutionary rate in plastomes of *Jasmineae* and the clade *LVP* of *Ligustrinae*. Again, the acceleration of the evolutionary rate in *Ligustrinae* occurred after the divergence from *Sy. vulgaris*, thus in apparent link to genomic reorganizations and potential paternal leakage mentioned above for this lineage. We thus hypothesize that transient heteroplasmy with possible gene conversion/recombination could make more allowable DNA polymerase errors and in turn change the selective pressures on some chloroplastic genes. While an acceleration of the non-synonymous substitution rate (d_N) is detected for more than one third (42%) of chloroplastic genes in *Jasmineae*, this d_N acceleration is less pronounced in *Ligustrinae* (concerning 19% of chloroplastic genes). According the phylogeny, the event responsible of the acceleration of the evolutionary rate seems to be more ancient in *Jasmineae* than in the clade *LVP* of *Ligustrinae*, which could explain the lower number of genes exhibiting a significant change of d_N in *Ligustrinae*. A similar increase of d_N in chloroplastic genes has been described in Geraniaceae (Guisinger *et al.* 2008; Park *et al.* 2017), *Silene* (Sloan *et al.* 2012) and acacias (Mensous *et al.* 2017). This acceleration of d_N could reflect a relaxation of the purifying selection or adaptive selection for chloroplastic genes which are involved in various functions, but most affected genes are not involved in the photosynthesis metabolism. Considering that the reading frame of fast evolving genes is probably maintained over a few millions of years in *Jasmineae* and the clade *LVP* of *Ligustrinae*, we consider that genes are still functional. One process could result in such deep changes of selective patterns. The acceleration of the mutation rate could lead to the fixation of slightly deleterious mutations that need to be compensated by other mutations (i.e. involving epistatic interactions) positively selected (e.g. “Red Queen” hypothesis). In this case, evidence of positive selection should be detected.

Positive selection in chloroplastic genes

Ten genes were significantly affected by positive selection in *Jasmineae*. These genes include *ycf1*, *ycf2*, and *clpP* which have also evolved quickly in multiple lineages (Haberle *et al.* 2008; Sloan *et al.* 2012; Gurdon & Maliga 2014; Zhu *et al.* 2016; Mensous *et al.* 2017; Park *et al.* 2017). The exact function of the highly diverging *ycf1* and *ycf2* remains however unclear, but they are essential for the cell survival, being probably involved in the cellular metabolism or in cellular structure (Drescher *et al.* 2000). More interestingly, *clpP* (also known as *clpP1*) exhibits an extreme acceleration of d_N and d_S in *Jasmineae* (and also, at a lower extent, in the *LVP* clade). This gene encodes for a serine protease, which is a subunit of the *Clp* protease (Maurizi *et al.* 1990). The chloroplastic gene *clpP* has an essential function in the cell survival and growth (Ramundo *et al.* 2014). A link between the paternal leakage and a rapid evolution of *clpP* (including a loss in *Actinidia*) seems to occur (Jansen & Ruhlman 2012; Wang *et al.* 2016), but it has been shown that the gene expression levels of plastid-targeted *Clp* proteins are very low or absent in the mature pollen suggesting that *clpP* does not determine the potential paternal inheritance of chloroplasts (Olinares *et al.* 2011). A possible experiment to test a possible role of *clpP* in the plastid paternal leakage should be to compare the expression levels of the gene at different stages of the pollen development between lineages with differential inheritance mode of chloroplasts (maternal vs. biparental).

Slow evolution of mitochondrial genes

The presence of mitochondria in the mature pollen of *Jasmineae* suggested a potential paternal leakage of mitochondria in this lineage (Sodmergen *et al.* 1998). However, our study does not reveal clear evidence for evolutionary trajectory changes of mitochondrial genes in *Jasmineae*. Although a change of the global substitution rate occurred for three mtDNA genes (*cox1*, *cytB*, *rps3*) in *Jasmineae* and for *mttB* in *Jasmineae* and the *LVP* clade of *Ligustrinae*, their d_N and d_S are too low to conclude to a significant acceleration of the evolutionary rate for these genes. Only *mttB* exhibits a significant increase of d_S in *Jasmineae*. The function of *mttB* is unclear and it probably encodes for a protein translocase (Mower *et al.* 2012). Four genes (*ccmFc*, *cytB*, *nad5* and *rps3*) are, however, detected to be under positive selection in *Jasmineae*. Besides this, the phylogenetic reconstruction of the Oleaceae based on coding and non-coding mtDNA regions also revealed long branches for a pair of *Noronhia* species (*N. intermedia* and *N. clarinerva*). These two species were sampled in the same forest patch, located in a diversity hotspot of genus *Noronhia* in northern Madagascar, where tens of species co-occur (Hong-Wa 2016). We can just hypothesize that frequent hybridizations between sympatric species may influence some nucleo-cytoplasmic genome interactions and in turn could result in an accelerated divergence rates of mitochondrial genes.

Conclusions and perspectives

Our study highlights an acceleration of the evolutionary rate in the plastomes of *Jasmineae* and a clade of *Ligustrinae*, in which potential paternal leakage occurs. Most chloroplastic genes of *Jasmineae* are highly divergent. Contrary to plastomes, the mitogenome of both lineages evolves slowly and does not seem to experience a significant increase of the evolutionary rate. Our results add evidences of a link between paternal leakage and genomic rearrangements with a faster evolution of chloroplastic genes (particularly *accD*, *clpP*, *ycf1* and *ycf2*). However, the cause and effect relationship between these processes are still unclear. More investigations are necessary to define the role of gene conversion between maternal and paternal copies of organellar genomes on the substitution rate change and on genomic rearrangement in these lineages. To detect adaptive selection on genes, further analyses could be also considered using mutation-selection models (Rodrigue & Lartillot 2016). In addition, the mechanisms of chloroplast degradation in the case of a maternal inheritance are poorly understood in plants. The confirmation by controlled crosses of a real paternal inheritance of organelles are thus necessary in the cases where we suspect a paternal leakage only based on microscopic observations. Also, the detection of potential paternal leakage of organellar genomes should be investigated in a larger taxonomic sampling in Oleaceae. Last, the interactions between the nuclear and organellar genomes remain little known. With the expansion of new NGS technologies producing long reads with high sequencing coverage, we can also imagine that more nuclear genome will be soon sequenced, and that the study of the interactions between genomes will be more accessible.

References

Besnard G, Khadari B, Baradat P, Bervillé A (2002) Combination of chloroplast and mitochondrial DNA polymorphisms to study cytoplasm genetic differentiation in the olive complex (*Olea europaea* L.). *Theoretical and Applied Genetics*, **105**, 139–144.

Besnard G, Khadari B, Villemur P, Bervillé A (2000) Cytoplasmic male sterility in the olive (*Olea europaea* L.). *Theoretical and Applied Genetics*, **100**, 1018–1024.

Bock R, Timmis JN (2008) Reconstructing evolution: gene transfer from plastids to the nucleus. *BioEssays*, **30**, 556–566.

Cho Y, Mower JP, Qiu YL, Palmer JD (2004) Mitochondrial substitution rates are extraordinarily elevated and variable in a genus of flowering plants. *Proceedings of the National Academy of Sciences of the United States of America*, **101**, 17741–17746.

Chumley TW, Palmer JD, Mower JP *et al.* (2006) The complete chloroplast genome sequence of *Pelargonium* x *hortorum*: organization and evolution of the largest and most highly rearranged chloroplast genome of land plants. *Molecular Biology and Evolution*, **23**, 2175–2190.

Clifton SW, Minx P, Fauron CM *et al.* (2004) Sequence and comparative analysis of the maize NB mitochondrial genome. *Plant Physiology*, **136**, 3486–3503.

Corriveau JL, Coleman AW (1988) Rapid screening method to detect potential biparental inheritance of plastid DNA and results for over 200 angiosperm species. *American Journal of Botany*, **75**, 1443–1458.

Cosner ME, Jansen RK, Palmer JD, Downie SR (1997) The highly rearranged chloroplast genome of *Trachelium caeruleum* (Campanulaceae): multiple inversions, inverted repeat expansion and contraction, transposition, insertions/deletions, and several repeat families. *Current Genetics*, **31**, 419–429.

Cosner ME, Raubeson LA, Jansen RK (2004) Chloroplast DNA rearrangements in Campanulaceae: phylogenetic utility of highly rearranged genomes. *BMC Evolutionary Biology*, **4**, 27.

Cusack BP, Wolfe KH (2007) When gene marriages don't work out: divorce by subfunctionalization. *Trends in Genetics*, **23**, 270–272.

Downie SR, Jansen RK (2015) A comparative analysis of whole plastid genomes from the Apiales: expansion and contraction of the Inverted Repeat, mitochondrial to plastid transfer of DNA, and identification of highly divergent noncoding regions. *Systematic Botany*, **40**, 336–351.

Drescher A, Ruf S, Jr TC, Carrer H, Bock R (2000) The two largest chloroplast genome-encoded open reading frames of higher plants are essential genes. *The Plant Journal*, **22**, 97–104.

Edgar RC (2004) MUSCLE: Multiple sequence alignment with high accuracy and high throughput. *Nucleic Acids Research*, **32**, 1792–1797.

Erixon P, Oxelman B (2008) Whole-gene positive selection, elevated synonymous substitution rates, duplication, and indel evolution of the chloroplast *clpP1* gene. *PLoS ONE*, **3**.

Gantt JS, Baldauf SL, Calie PJ, Weeden NF, Palmer JD (1991) Transfer of *rpl22* to the nucleus greatly preceded its loss from the chloroplast and involved the gain of an intron. *The EMBO Journal*, **10**, 3073–3078.

Garland T, Dickerman AW, Janis CM, Jones JA (1993) Phylogenetic analysis of covariance by computer simulation. *Systematic Biology*, **42**, 265–292.

Goulding SE, Olmstead RG, Morden CW, Wolfe KH (1996) Ebb and flow of the chloroplast inverted repeat. *Molecular and General Genetics*, **252**, 195–206.

Gray MW, Archibald J (2012) Origins of mitochondria and plastids. In: *Genomics of chloroplasts and mitochondria* (eds Bock R, Knoop V), pp. 1–30. Springer Science & Business Media.

Greiner S, Bock R (2013) Tuning a ménage à trois: co-evolution and co-adaptation of nuclear and organellar genomes in plants. *BioEssays*, **35**, 354–365.

Greiner S, Sobanski J, Bock R (2014) Why are most organelle genomes transmitted maternally? *BioEssays*, **37**, 80–94.

Gualberto JM, Mileshina D, Wallet C *et al.* (2014) The plant mitochondrial genome: dynamics and maintenance. *Biochimie*, **100**, 107–120.

Guisinger MM, Kuehl J V, Boore JL, Jansen RK (2008) Genome-wide analyses of Geraniaceae plastid DNA reveal unprecedented patterns of increased nucleotide substitutions. *Proceedings of the National Academy of Sciences of the United States of America*, **105**, 18424–18429.

Guisinger MM, Kuehl J V, Boore JL, Jansen RK (2011) Extreme reconfiguration of plastid genomes in the angiosperm family Geraniaceae: rearrangements, repeats, and codon usage. *Molecular Biology and Evolution*, **28**, 583–600.

Gurdon C, Maliga P (2014) Two distinct plastid genome configurations and unprecedented intraspecies length variation in the *accD* coding region in *Medicago truncatula*. *DNA Research*, **21**, 417–427.

Haberle RC, Fourcade HM, Boore JL, Jansen RK (2008) Extensive rearrangements in the chloroplast genome of *Trachelium caeruleum*

are associated with repeats and tRNA genes. *Journal of Molecular Evolution*, **66**, 350–361.

Hong-Wa C (2016) *A taxonomic revision of the genus Noronhia Stadtm. ex Thouars (Oleaceae) in Madagascar and the Comoro Islands* (M Callmander, P Lowry, Eds,). Boissiera 70.

Jansen RK, Cai Z, Raubeson LA *et al.* (2007) Analysis of 81 genes from 64 plastid genomes resolves relationships in angiosperms and identifies genome-scale evolutionary patterns. *Proceedings of the National Academy of Sciences of the United States of America*, **104**, 19369–19374.

Jansen RK, Ruhlman TA (2012) Plastid genomes of seed plants. In: *Genomes of chloroplasts and mitochondria* (eds Bock R, Knoop V), pp. 103–126. Springer Science & Business Media.

Jensen PE, Leister D (2014) Chloroplast evolution, structure and functions. *F1000Prime Reports*, **6**, 1–14.

Kearse M, Moir R, Wilson A *et al.* (2012) Geneious Basic: an integrated and extendable desktop software platform for the organization and analysis of sequence data. *Bioinformatics*, **28**, 1647–1649.

Keeling PJ (2010) The endosymbiotic origin, diversification and fate of plastids. *Philosophical Transactions of the Royal Society B: Biological Sciences*, **365**, 729–748.

Khakhlova O, Bock R (2006) Elimination of deleterious mutations in plastid genomes by gene conversion. *Plant Journal*, **46**, 85–94.

Kode V, Mudd EA, Iamtham S, Day A (2005) The tobacco plastid *accD* gene is essential and is required for leaf development. *Plant Journal*, **44**, 237–244.

Kubo T, Kitazaki K, Matsunaga M, Kagami H, Mikami T (2011) Male sterility-inducing mitochondrial genomes: how do they differ? *Critical Reviews in Plant Sciences*, **30**, 378–400.

Laroche J, Li P, Maggia L, Bousquet J (1997) Molecular evolution of angiosperm mitochondrial introns and exons. *Proceedings of the National Academy of Sciences of the United States of America*, **94**, 5722–5727.

Lavin M, Doyle JJ (1990) Evolutionary significance of the loss of the chloroplast-DNA inverted repeat in the Leguminosae subfamily Papilionoideae. *Evolution*, **44**, 390–402.

Lee HL, Jansen RK, Chumley TW, Kim KJ (2007) Gene relocations within chloroplast genomes of Jasminum and Menodora (Oleaceae) are due to multiple, overlapping inversions. *Molecular Biology and Evolution*, **24**, 1161–1180.

Liu Y, Cui H, Zhang Q, Sodmergen (2004) Divergent potentials for cytoplasmic inheritance within the genus *Syringa*. A new trait associated with speciogenesis. *Plant Physiology*, **136**, 2762–2770.

Magee AM, Aspinall S, Rice DW *et al.* (2010) Localized hypermutation and associated gene losses in legume chloroplast genomes. *Genome Research*, **20**, 1700–1710.

Magnuson K, Jackowski S, Rock CO, Cronan JE (1993) Regulation of fatty acid synthesis. *Biochemical Society Transactions*, **57**, 522.

Martin W, Mentel M (2010) The origin of mitochondria. *Nature Education*, **3**, 58.

Maurizi MR, Clark WP, Kim SH, Gottesman S (1990) *ClpP* represents a unique family of serine proteases. *Journal of Biological Chemistry*, **265**, 12546–12552.

Mensous M, Van de Paer C, Manzi S *et al.* (2017) Diversity and evolution of plastomes in Saharan mimosoids: potential use for phylogenetic and population genetic studies. *Tree Genetics and Genomes*, **13**, 48.

Millen RS, Olmstead RG, Adams KL *et al.* (2001) Many parallel losses of *infA* from chloroplast DNA during angiosperm evolution with multiple independent transfers to the nucleus. *The Plant Cell*, **13**, 645–658.

Miller MA, Pfeiffer W, Schwartz T (2010) Creating the CIPRES Science Gateway for inference of large phylogenetic trees. *Proceedings of the Gateway Computing Environments Workshop (GCE)*, 1–8.

Mower JP, Sloan DB, Alverson AJ (2012) Plant mitochondrial genome diversity: the genomics revolution. In: *Plant Genome Diversity*, pp. 123–144. Springer Vienna.

Muller HJ (1964) The relation of recombination to mutational advance. *Mutation Research/Fundamental and Molecular Mechanisms of Mutagenesis*, **1**, 2–9.

Olinares PDB, Kim J, Van Wijk KJ (2011) The *Clp* protease system; a central component of the chloroplast protease network. *Biochimica et Biophysica Acta - Bioenergetics*, **1807**, 999–1011.

Palmer JD, Herbon LA (1988) Plant mitochondrial DNA evolves rapidly in structure, but slowly in sequence. *Journal of Molecular Evolution*, **28**, 87–97.

Palmer JD, Osorio B, Aldrich J, Thompson WF (1987) Chloroplast DNA evolution among legumes: loss of a large inverted repeat occurred prior to other sequence rearrangements. *Current Genetics*, **11**, 275–286.

Park S, Ruhlman TA, Weng M-L *et al.* (2017) Contrasting patterns of nucleotide substitution rates provide insight into dynamic evolution of plastid and mitochondrial genomes of *Geranium*. *Genome Biology and Evolution*, **9**, 1766–1780.

Parkinson CL, Mower JP, Qiu YL *et al.* (2005) Multiple major increases and decreases in mitochondrial substitution rates in the plant family Geraniaceae. *BMC Evolutionary Biology*, **5**, 73.

Ramundo S, Casero D, Muhlhaus T *et al.* (2014) Conditional depletion of the *Chlamydomonas* chloroplast *ClpP* protease activates nuclear genes involved in autophagy and plastid protein quality control. *The Plant Cell*, **26**, 2201–2222.

Revell LJ (2012) phytools: an R package for phylogenetic comparative biology (and other things). *Methods in Ecology and Evolution*, **3**, 217–223.

Rodrigue N, Lartillot N (2016) Detecting adaptation in protein-coding genes using a Bayesian site-heterogeneous mutation-selection codon substitution model. *Molecular Biology and Evolution*, **34**, 204–214.

Rousseau-Gueutin M, Huang X, Higginson E *et al.* (2013) Potential functional replacement of the plastidic acetyl-CoA carboxylase subunit (*accD*) gene by recent transfers to the nucleus in some angiosperm lineages. *Plant Physiology*, **161**, 1918–1929.

Sloan DB, Alverson AJ, Wu M, Palmer JD, Taylor DR (2012) Recent acceleration of plastid sequence and structural evolution coincides with extreme mitochondrial divergence in the angiosperm genus *Silene*. *Genome Biology and Evolution*, **4**, 294–306.

Sloan DB, Oxelman B, Rautenberg A, Taylor DR (2009) Phylogenetic analysis of mitochondrial substitution rate variation in the angiosperm tribe *Sileneae*. *BMC Evolutionary Biology*, **9**, 260.

Sloan DB, Wu Z (2014) History of plastid DNA insertions reveals weak deletion and AT mutation biases in angiosperm mitochondrial genomes. *Genome Biology and Evolution*, **6**, 3210–3221.

Smith S, Donoghue M (2008) Rates of molecular evolution are linked to life history in flowering plants. *Science*, **322**, 86–9.

Smith DR, Keeling PJ (2015) Mitochondrial and plastid genome architecture: reoccurring themes, but significant differences at the extremes. *Proceedings of the National Academy of Sciences of the United States of America*, **112**, 201422049.

Sodmergen, Bai HH, He JX *et al.* (1998) Potential for biparental cytoplasmic inheritance in *Jasminum officinale* and *Jasminum nudiflorum*. *Sexual Plant Reproduction*, **11**, 107–112.

Sollars ESA, Harper AL, Kelly LJ *et al.* (2017) Genome sequence and genetic diversity of European ash trees. *Nature*, **541**, 212–216.

Stamatakis A (2014) RAxML version 8: a tool for phylogenetic analysis and post-analysis of large phylogenies. *Bioinformatics*, **30**, 1312–1313.

Strathern JN, Shafer BK, McGill CB (1995) DNA synthesis errors associated with double-strand-break repair. *Genetics*, **140**, 965–972.

Ueda M, Fujimoto M, Arimura SI *et al.* (2007) Loss of the *rpl32* gene from the chloroplast genome and subsequent acquisition of a preexisting transit peptide within the nuclear gene in *Populus*. *Gene*, **402**, 51–56.

Van de Paer C, Bouchez O, Besnard G (2017) Prospects on the evolutionary mitogenomics of plants: a case study on the olive family (Oleaceae). *Molecular Ecology Resources*.

Wakil S, Stoops J, Joshi V (1983) Fatty acid synthesis and its regulation. *Annual Review of Biochemistry*, **52**, 537–579.

Wallander E, Albert V (2000) Phylogeny and classification of Oleaceae based on *rps16* and *trnL-F* sequence data. *American Journal of Botany*, **87**, 1827–1841.

Wang WC, Chen SY, Zhang XZ (2016) Chloroplast genome evolution in Actinidiaceae: *ClpP* loss, heterogenous divergence and phylogenomic practice. *PLoS ONE*, **11**, 1–17.

Wang YH, Qu XJ, Chen SY, Li DZ, Yi TS (2017) Plastomes of Mimosoideae: structural and size variation, sequence divergence, and phylogenetic implication. *Tree Genetics and Genomes*, **13**, 41.

Weng M, Blazier JC, Govindu M, Jansen RK (2013) Reconstruction of the ancestral plastid genome in Geraniaceae reveals a correlation between genome rearrangements, repeats and nucleotide substitution rates. *Molecular Biology and Evolution*, **31**, 645–659.

Wicke S, Schneeweiss GM, DePamphilis CW, Müller KF, Quandt D (2011) The evolution of the plastid chromosome in land plants: gene content, gene order, gene function. *Plant Molecular Biology*, **76**, 273–297.

Wolfe KH, Li WH, Sharp PM (1987) Rates of nucleotide substitution vary greatly among plant mitochondrial, chloroplast, and nuclear DNAs. *Proceedings of the National Academy of Sciences of the United States of America*, **84**, 9054–9058.

Yang Z (2007) PAML 4: phylogenetic analysis by maximum likelihood. *Molecular Biology and Evolution*, **24**, 1586–1591.

Yang Z, Nielsen R, Goldman N, Pedersen AMK (2000) Codon-substitution models for heterogeneous selection pressure at amino acid sites. *Genetics*, **155**, 431–449.

Yang Y, Sterling J, Storici F, Resnick MA, Gordenin DA (2008) Hypermutability of damaged single-strand DNA formed at double-strand breaks and uncapped telomeres in yeast *Saccharomyces cerevisiae*. *PLoS Genetics*, **4**, e1000264.

Zedane L, Hong-Wa C, Murienne J *et al.* (2016) Museomics illuminate the history of an extinct, paleoendemic plant lineage (*Hesperelaea*, Oleaceae) known from an 1875 collection from Guadalupe Island, Mexico. *Biological Journal of the Linnean Society*, **117**, 44–57.

Zhang Q, Liu Y, Sodmergen (2003) Examination of the cytoplasmic DNA in male reproductive cells

to determine the potential for cytoplasmic inheritance in 295 angiosperm species. *Plant and Cell Physiology*, **44**, 941–951.

Zhang J, Nielsen R, Yang Z (2005) Evaluation of an improved branch-site likelihood method for detecting positive selection at the molecular level. *Molecular Biology and Evolution*, **22**, 2472–2479.

Zhang Q, Sodmergen (2010) Why does biparental plastid inheritance revive in angiosperms? *Journal of Plant Research*, **123**, 201–206.

Zhu A, Guo W, Gupta S, Fan W, Mower JP (2016) Evolutionary dynamics of the plastid inverted repeat: the effects of expansion, contraction, and loss on substitution rates. *New Phytologist*, **209**, 1747–1756.

Zhu A, Guo W, Jain K, Mower JP (2014) Unprecedented heterogeneity in the synonymous substitution rate within a plant genome. *Molecular Biology and Evolution*, **31**, 1228–1236.

DISCUSSION ET PERSPECTIVES

Durant ce travail de thèse, j'ai pu aborder différents aspects sur la diversité et l'évolution des génomes cytoplasmiques. Tout d'abord, l'assemblage des tout premiers mitogénomes d'Oleaceae nous a permis d'étudier leur diversité structurelle mais aussi leur origine complexe, notamment en relation avec des transferts récurrents de régions plastidiques. Pour la première fois, des séquences mitochondriales et plastidiques ont été assemblées sur un large échantillonnage d'Oleaceae, ce qui nous permet désormais d'analyser l'histoire évolutive parallèle de ces génomes cytoplasmiques ouvrant de nombreuses perspectives de recherches, aussi bien appliquées que fondamentales. Nous pouvons spéculer que les approches développées dans ce travail vont fortement influencer les travaux sur la taxonomie et la biogéographie des Oleaceae, d'une part, et d'autre part, nos recherches devraient aussi stimuler de nouvelles recherches en biologie moléculaire et fonctionnelle des gènes cytoplasmiques chez les plantes en général.

I/ Mitogénome et phylogénomique des plantes

1. Reconstruction d'un mitogénome complet de plante à partir d'ADN ancien

Tout d'abord, la reconstruction du mitogénome complet du genre éteint *Hesperelaea* (CHAPITRE I) ouvre de nouvelles perspectives dans l'analyse d'ADN provenant d'échantillons muséologiques pour étudier l'histoire évolutive de groupes de plantes. A partir d'un spécimen d'herbier récolté en 1875, nous avons reconstruit le mitogénome d'*H. palmeri* (~658 kb) en utilisant des données de type "shotgun". L'utilisation de telles données génomiques peut permettre d'éclaircir le statut taxonomique d'un spécimen et de résoudre ses relations phylogénétiques avec d'autres taxa (Bakker *et al.* 2016 ; Zedane *et al.* 2016). Par ailleurs, l'intégration d'espèces ou genres rares ou éteints a une importance majeure dans les études biogéographiques de groupes de plantes. Afin d'expliquer la distribution géographique actuelle des espèces, un échantillonnage exhaustif des différents clades du groupe de plantes étudié est généralement nécessaire pour inférer des hypothèses de dispersion, de temps de divergence et d'aires ancestrales de distribution des espèces. L'intégration dans la phylogénie d'espèces rares issues d'herbiers peut également être utilisée dans l'étude de traits fonctionnels ou de traits d'histoire de vie. Par exemple, l'inclusion de spécimens d'herbier de *Sartidia* dans l'étude phylogénétique des Poaceae a permis de révéler deux transitions indépendantes de la photosynthèse C_3 vers C_4 chez la sous famille des Aristidoideae et même d'identifier quelques bases génétiques associées à ces transitions (Christin *et al.* 2008 ; Besnard *et al.* 2014). Aussi, l'utilisation d'ADN provenant d'échantillons d'herbier peut permettre d'analyser les variations de fréquences de certains allèles dans un passé récent. Par exemple, la substitution responsable de la résistance aux herbicides chez l'ACCase de certaines graminées a été

identifiée sur un spécimen d'herbier récolté en 1888, i.e. avant l'utilisation d'herbicides dans l'agriculture intensive moderne (Délye *et al.* 2013).

La reconstruction de mitogénomes ou de plastomes complets d'échantillons d'herbiers dépend en majeure partie de la qualité de l'ADN et de son degré de dégradation. Un facteur important pour limiter la dégradation de l'ADN est d'optimiser le conditionnement de l'échantillon en herbier. La prolifération de bactéries et de champignons peut fortement contaminer l'échantillon. Pour éviter de telles contaminations, il est souvent essentiel de soumettre l'échantillon à une exposition à la chaleur ou au froid au moment de sa fixation afin d'éliminer les champignons contaminants et les arthropodes (Bakker *et al.* 2016). Une fois fixés, les échantillons sont protégés des rayons ultra-violets et conservés à des températures modérées et sous de faibles niveaux d'humidité (Staats *et al.* 2011). Les échantillons subissent généralement des cycles de refroidissement à -20 °C tous les deux ans (Staats *et al.* 2011). Au moment de la fixation, les échantillons provenant de régions tropicales humides nécessitent en particulier une exposition à la chaleur plus élevée afin d'éviter le développement de moisissure (Bakker *et al.* 2016). Néanmoins, l'exposition d'échantillons à de fortes températures (60-70°C) est susceptible de provoquer des cassures double brin de l'ADN, résultant à la fragmentation de l'ADN (Bakker *et al.* 2016). Staats *et al.* (2011) ont montré que les niveaux de fragmentation de l'ADN sont équivalents entre l'ADN nucléaire et l'ADN des organites. En particulier, il n'y a pas de différences en termes de fragmentation entre l'ADN mitochondrial et l'ADN chloroplastique. Néanmoins, l'ADN chloroplastique d'herbier présente des niveaux élevés de transitions C→T/G→A comparé à l'ADN chloroplastique frais (Staats *et al.* 2011). Ces transitions semblent être en partie dues à la désamination hydrolytique de la cytosine en uracile, possiblement liée à l'humidité ambiante (Staats *et al.* 2011). A la différence de l'ADN chloroplastique, l'ADN mitochondrial n'est pas soumis à de telles transitions et pourrait ainsi constituer une ressource de plus grand intérêt dans les études d'ADN ancien, par exemple comme nous l'avons mentionné dans le chapitre II pour des analyses archéo-génétiques de restes d'olivier.

2. Assemblage de mitogénomes de plantes

Dans le CHAPITRE II, nous avons ensuite reconstruit le mitogénome complet de *Chionanthus rupicolu*s et de six accessions d'*Olea europaea*. Notre étude a confirmé que les mitogénomes de plantes sont très variables en termes de structure à l'échelle interspécifique, et surtout entre lignées maternelles d'une même espèce. Les mitogénomes des différentes lignées d'*O. europaea* présentent de larges réarrangements génomiques (déjà mis en évidence par la méthode RFLP ; Besnard *et al.* 2000), dus à la dynamique structurelle inhérente aux mitogénomes de plantes. Par ailleurs, les mitogénomes d'*O. europaea* sont variables en termes de contenu en régions non-codantes. En particulier, le contenu en régions répétées est largement variable entre les mitogénomes des différentes lignées d'olivier. Les répétitions contenues dans le mitogénome sont responsables de la dynamique structurelle des mitogénomes de plantes. Notre étude a permis d'élargir nos connaissances quant à la structure des mitogénomes, et en particulier au degré de variabilité structurelle des mitogénomes entre lignées maternelles d'une même espèce. De plus, les huit mitogénomes complets reconstruits

dans les CHAPITRES I et II constituent des ressources génomiques d'intérêt, pouvant servir de références pour reconstruire les mitogénomes ou des régions mitochondriales d'espèces proches. Enfin, la reconstruction des mitogénomes des différentes lignées d'*O. europaea* nous a permis d'identifier deux gènes chimériques potentiellement associés à un type de stérilité mâle cytoplasmique chez l'olivier.

Besnard *et al.* (2000) ont mis en évidence dans des croisements contrôlés une stérilité mâle cytoplasmique (CMS) liée à une lignée maternelle d'olivier (appelée E3, ou MCK). Néanmoins, aucun gène mitochondrial n'avait été jusqu'alors identifié comme étant associé à la CMS. Dans notre étude, nous avons identifié deux cadres de lecture ouverts (ORFs) potentiellement liés à cette CMS par comparaison de mitogénomes. Ces deux ORFs sont chimériques, l'une contenant une portion de *sdh3*, l'autre une portion de *rbcL* dérivé du plastome. Bien que l'identification de ces ORFs par comparaison de mitogénomes est insuffisante pour véritablement confirmer que ces gènes sont associés à la CMS chez la lignée E3, plusieurs éléments nous laissent à penser que l'ORF chimérique portant une partie de *sdh3* est un candidat intéressant pour la CMS. D'une part, parmi les gènes classiquement retrouvés dans la plupart des mitogénomes de Lamiales, *sdh3* est le seul gène qui soit absent, et en particulier dans le mitogénome de la lignée E3, ce qui laisse à penser que la CMS pourrait être liée au dysfonctionnement de ce gène. D'autre part, un pseudogène de *sdh3* est également suspecté comme un candidat impliqué dans une CMS chez *Nicotiana tabacum* (Zhang *et al.* 2013). Pour tester notre hypothèse, d'autres analyses sont bien évidemment nécessaires. On pourrait envisager d'effectuer une PCR quantitative pour mesurer le niveau d'expression des gènes chimériques de la lignée E3. D'autre part, on pourrait introduire un gène fonctionnel de *sdh3* dans le mitogénome afin de détecter d'éventuelles conséquences sur la production de pollen (e.g. "complémentation" de la mutation), ce qui nécessiterait de maîtriser des technologies de modification génétique chez l'olivier.

La reconstruction de mitogénomes de plantes présente donc un intérêt majeur pour mieux comprendre et caractériser la diversité structurelle des mitogénomes, mais aussi pour identifier des gènes cytoplasmiques associés à des traits fonctionnels tels que la CMS. Néanmoins, la reconstruction de mitogénomes de plantes reste laborieuse. Jusqu'à aujourd'hui, aucune méthode automatisée et standardisée n'a été développée pour assembler et résoudre les complexités structurelles des mitogénomes de plantes. Des logiciels spécialisés dans l'assemblage de génomes d'organites tels que l'assembleur développé par Coissac (2016) arrêtent l'assemblage du mitogénome au niveau des régions répétées et des zones dérivées du plastome. Pour chacun des mitogénomes reconstruits dans notre étude, nous avons dû assembler manuellement ces régions en comparant avec les séquences homologues du plastome. La résolution des régions dérivées du plastome (dites *mtpt*) se base sur une faible divergence entre la région *mtpt* et la région plastidique homologue. La résolution de ces régions *mtpt* s'effectue alors en identifiant les reads présentant les substitutions en plus faible couverture, le mitogénome étant généralement représenté plus faiblement que le plastome dans le pool de reads généré par "genome skimming". Aujourd'hui, la reconstruction de mitogénomes complets de plantes peut être envisagée dans des études se limitant à quelques accessions mais est inadaptée pour des études à plus large échantillonnage taxonomique.

Néanmoins, le développement de technologies de séquençage à haut débit générant des reads toujours plus longs pourrait faciliter l'assemblage des mitogénomes de plantes dans les années futures (sur matériel frais). Le séquençage PacBio présente déjà des performances intéressantes pour résoudre les problématiques rencontrées dans l'assemblage des mitogénomes. Le séquençage PacBio est considéré comme une méthode de séquençage en temps réel, qui permet de générer des reads beaucoup plus longs avec des runs plus rapides que les autres méthodes de séquençage (Rhoads & Au 2015). Cependant, le séquençage PacBio génère un taux d'erreurs important et le coût par base reste élevé. On pourrait néanmoins envisager de combiner le séquençage de type "genome skimming" et le séquençage PacBio (ou une autre méthodologie générant des longs reads) pour résoudre les difficultés d'assemblage du mitogénome des plantes.

3. Origines des régions non codantes dans les mitogénomes de plantes

Les mitogénomes reconstruits dans notre étude contiennent environ 90% de régions inter-géniques non codantes dont des séquences répétées (entre ~5 et 20% du mitogénome). Cependant, l'origine de ~70% du mitogénome reste inconnue. Dans le CHAPITRE II, nous avons mis en évidence des intégrations récurrentes de régions d'origine chloroplastique dans le mitogénome (*mtpt*), ce qui rejoint les résultats d'études précédentes (Wang *et al.* 2007; Sloan & Wu 2014). Afin d'identifier les régions *mtpt* dans les mitogénomes, nous avons fixé un seuil minimum de similarité avec le plastome de 85%. Cependant, ce seuil permet d'identifier uniquement des événements d'intégration relativement récents de régions *mtpt*. Il est donc possible qu'une partie du mitogénome soit issue d'événements d'intégration de régions chloroplastiques beaucoup plus anciens, qui seraient désormais plus difficiles à détecter du fait d'une plus grande divergence avec le plastome. De la même façon, on peut supposer qu'il y a eu des intégrations de séquences d'origine nucléaire ou des événements de transferts horizontaux (e.g. fusions de mitogénomes) avec d'autres plantes comme déjà révélés chez *Amborella trichopoda* (Bergthorsson *et al.* 2004; Rice *et al.* 2013). On peut également envisager qu'une partie des régions non codantes du mitogénome de plantes dérive de l'ADN de la bactérie endosymbiotique à l'origine de la mitochondrie. La longue divergence de ces séquences depuis ces événements d'endosymbiose rend désormais leur identification impossible. Enfin, les séquences répétées du mitogénome entraînent de nombreux événements de recombinaison au sein du mitogénome des plantes. Ces événements de recombinaison peuvent être à l'origine de duplications. Le mitogénome des plantes est donc potentiellement composé d'anciennes duplications ayant divergé au cours du temps.

4. Phylogénie à partir de données mitogénomiques.

Dans le CHAPITRE II, nous avons montré que l'utilisation des mitogénomes complets peut permettre de résoudre les relations phylogénétiques entre les lignées maternelles d'*O. europaea*, jusqu'alors non résolues à partir d'ADN chloroplastique. Bien que l'ADN mitochondrial évolue plus lentement que l'ADN chloroplastique, sa taille est beaucoup plus

grande (~700 kb vs. ~150 kb) et peut donc contenir plus de sites informatifs pour résoudre certaines relations phylogénétiques. Néanmoins, l'utilisation de mitogénomes complets dans des études phylogénétiques nécessite une analyse préliminaire pour vérifier l'orthologie des régions *mtpt* présentes dans les mitogénomes. Notre étude a montré que l'orthologie des régions *mtpt* est vérifiée entre les lignées d'*O. europaea*, mais plusieurs événements d'intégration d'ADN chloroplastique dans les mitogénomes semblent avoir eu lieu à l'échelle de la tribu des *Oleeae*. Si il y a eu plusieurs événements d'intégration d'ADN chloroplastique, les régions *mtpt* ne sont pas nécessairement orthologues. Dans ce cas, il convient d'éliminer les régions *mtpt* de l'alignement avant toute reconstruction phylogénétique. Notre étude a également montré que la reconstruction phylogénétique basée sur des régions mitochondriales codantes et non-codantes relativement conservées confirme la topologie des *Oleeae* obtenue à partir de plastomes complets. La comparaison des topologies de reconstructions phylogénétiques basées sur l'ADN chloroplastique et mitochondrial présente surtout un intérêt majeur dans le cas d'une évolution contrastée du plastome et du mitogénome, comme c'est le cas chez les Oleaceae.

II/ Conséquences génomiques d'un "paternal leakage" des organites

Dans le CHAPITRE III, nous avons discuté d'un lien potentiel entre le passage occasionnel de chloroplastes dans le pollen ("paternal leakage") et de larges réarrangements du plastome associés à une accélération du taux d'évolution chez certains gènes chloroplastiques au sein des *Jasmineae* et d'un clade de *Ligustrinae*. Un tel lien a par ailleurs été détecté chez d'autres groupes de plantes tels que des lignées d'acacias (Mensous *et al.* 2017 ; Wang *et al.* 2017), de *Silene* (Sloan *et al.* 2012), de Campanulaceae (Cosner *et al.* 1997, 2004), de Geraniaceae (Parkinson *et al.* 2005 ; Park *et al.* 2017) ou de *Plantago* (Cho *et al.* 2004). En dépit de ces observations similaires chez des groupes de plantes éloignés, la relation de cause à effet entre "paternal leakage" et réarrangements génomiques associés à une accélération du taux d'évolution reste mal comprise.

1. Lien entre "paternal leakage" et recombinaison

Bien que la transmission des organites est uniparentale et maternelle chez la plupart des plantes, la transmission occasionnelle d'organites par le pollen ("paternal leakage") a été observée indépendamment chez de nombreuses lignées de plantes (cf. Introduction). Une des raisons avancée pour expliquer le "paternal leakage" est l'évitement de l'accumulation de mutations délétères liée à l'absence de recombinaison (Greiner *et al.* 2014). En permettant la recombinaison (ou conversion génique) entre les copies maternelles et paternelles des organites, le "paternal leakage" pourrait contrecarrer le cliquet de Muller. Cependant, aucune preuve évidente de recombinaison entre les copies maternelles et paternelles des plastomes n'a été démontrée jusque maintenant. Le taux de GC ne varie pas entre les lignées à transmission maternelle des organites et celles présentant du "paternal leakage", suggérant peu de recombinaison [qui, dans le cas contraire, se traduirait par une augmentation du taux de GC; Marais (2003)]. Cependant, les événements de recombinaison suite à du "paternal leakage" sont vraisemblablement des événements rares.

2. Identification des modes de transmission des organites au sein des Oleaceae

Dans notre étude, il serait nécessaire de déterminer le mode de transmission des organites chez davantage d'espèces. Nos conclusions seraient d'autant plus robustes si nous démontrons une transmission maternelle stricte des organites dans les différents clades d'Oleaceae, mis à part les *Jasmineae* et un clade de *Ligustrinae* pour lesquels on observe une accélération du taux d'évolution du plastome. Par ailleurs, il serait nécessaire d'effectuer des croisements contrôlés pour confirmer que la descendance peut hériter véritablement de la copie paternelle du plastome. En effet, le chloroplaste paternel peut être transmis mais pas forcément hérité. Le mode de dégradation des organites est peu documenté chez les plantes. En revanche, il a été montré que les mécanismes de dégradation des mitochondries dans le monde animal sont variés et peuvent intervenir avant ou après fécondation. Chez les mammifères par exemple, la

copie paternelle du mitogénome entre dans le cytoplasme de l'oocyte après fécondation, avant d'être dégradée dans un second temps (Sato & Sato 2013).

3. Estimation de la fréquence du "paternal leakage" chez le troène en conditions semi-contrôlées

Bien que je n'ai pas présenté dans ce manuscrit cette expérimentation, nous avions envisagé de détecter du "paternal leakage" par croisements semi-contrôlés chez *Ligustrum vulgare*. Pierre Saumitou-Laprade et Philippe Vernet du laboratoire Evo-Eco-Paléo de Lille ont récolté des individus de *L. vulgare* provenant de huit populations naturelles relativement distantes en Europe. Cinq populations provenaient de localités situées dans le nord de la France (parfois distantes de plusieurs centaines de kilomètres), une population était située dans les Cévennes, une autre à Grenoble et une dernière était localisée en Allemagne. Ces différents individus ont été disposés en jardin expérimental et sans contrôle de la pollinisation. L'idée de notre expérimentation était d'identifier un individu ayant produit des graines dont le chlorotype était unique. Dans un tel cas, il serait possible de détecter dans les descendances un variant chloroplastique d'origine paternelle. Nous avons choisi 48 individus qui présentaient des génotypes différents (d'après des données microsatellites nucléaires), provenant des différentes populations. Nous avons séquencé deux fragments d'environ 600-700 bp d'ADN chloroplastique (un au niveau d'*accD* et un autre au niveau de *clpP*) présentant une grande variabilité interspécifique, et génotypé trois longs microsatellites [no 19, 22, 51 ; Besnard *et al.* (2011)]. Nous n'avons détecté aucune variabilité entre les individus de *L. vulgare*, ce qui ne nous a pas permis d'aller plus loin dans cette étude. En revanche, *L. ovalifolium* présentait de nombreux polymorphismes par rapport à *L. vulgare*. On suppose que ces espèces puissent s'hybrider. Une expérimentation qui serait à envisager serait donc de pulvériser du pollen de *L. vulgare* sur des individus de *L. ovalifolium* et de récolter les graines à l'automne. Si on détecte de l'hétéroplasmie chez les hybrides liée à la présence du chlorotype de *L. vulgare*, cela permettrait de confirmer sans ambiguïté qu'il existe du "paternal leakage".

4. Autres mécanismes que le "paternal leakage" pour limiter l'accumulation de mutations délétères

D'autres mécanismes que le "paternal leakage" pourraient permettre de limiter le fardeau de mutations délétères. Les travaux récents de Radzvilavicius *et al.* (2017) soulignent que l'hypothèse selon laquelle le "paternal leakage" permettrait de limiter l'accumulation de mutations délétères via de la recombinaison repose essentiellement sur des modèles théoriques de génétique des populations basés sur l'ADN nucléaire (Felsenstein & Yokoyama 1976 ; Charlesworth *et al.* 1993). Or, Radzvilavicius *et al.* (2017) remettent en cause l'application de telles hypothèses pour l'ADN mitochondrial. Radzvilavicius *et al.* (2017) précisent que, à la différence du génome nucléaire, plusieurs copies du mitogénome se situent dans un nucléoïde mitochondrial, que plusieurs nucléoïdes mitochondriaux se situent dans une mitochondrie et que plusieurs mitochondries se situent dans une cellule. Dans ce cas, la sélection contre les mutations délétères agit essentiellement sur les mutations qui ont un effet

délétère sur la fitness de la cellule, i.e. au niveau du groupe de copies de mitogénomes dans la cellule. Or, à chaque division cellulaire, le groupe de copies de mitogénomes dans une cellule change du fait de la ségrégation stochastique. Une cellule fille n'aura pas le même groupe de copies de mitogénomes que la cellule mère, ce qui a pour conséquence d'augmenter la variance inter-cellulaire et donc de diluer le fardeau de mutations délétères. Une autre explication du maintien du "paternal leakage" chez certaines lignées de plantes à fleurs pourrait être liée à de la sélection adaptative.

5. "Paternal leakage" et évolution adaptative

Une de nos hypothèses pour expliquer l'augmentation du nombre de substitutions non-synonymes dans des gènes plastidiques de *Jasmineae* et d'un clade de *Ligustrinae* est liée à une évolution adaptative. Comme exposé dans l'introduction du CHAPITRE III, la possibilité de recombinaison pourrait permettre à la sélection d'être plus efficace et en conséquence de relâcher la pression de sélection visant à préserver l'intégrité maximale de la séquence plastidique. Une augmentation du taux de mutation pourrait mener à une apparition plus fréquente de mutations délétères, mais en outre d'autres mutations qui les compenseraient pourraient aussi être positivement sélectionnées. Cette évolution accélérée avec la possibilité de mutations compensatoires pourrait offrir l'opportunité d'atteindre de nouveaux optimums adaptatifs. Afin de tester l'hypothèse selon laquelle le "paternal leakage" pourrait promouvoir l'évolution adaptative, nous envisageons, avec la collaboration d'Hervé Philippe, d'utiliser le modèle de mutation-sélection développé par Rodrigue & Lartillot (2016). Dans les modèles habituels tels que ceux utilisés par PAML, la sélection positive est détectée si $\omega > 1$. Or, l'évolution adaptative peut opérer et se traduire par une augmentation du nombre de substitutions non-synonymes, sans pour autant que ω soit supérieur à 1. Dans leur modèle, Rodrigue & Lartillot (2016) définissent un modèle neutre qui infère un profil d'acides aminés optimal pour chaque codon. Ils introduisent un nouveau paramètre, ω^*, qui absorbe toute déviation à ce modèle neutre. De l'évolution adaptative peut alors être détectée si $\omega^* > 1$.

CONCLUSION GENERALE

Ce projet de thèse a permis d'ouvrir de nouvelles perspectives dans l'utilisation des données mitogénomiques et chloroplastiques. Nous avons montré que la reconstruction de gènes ou de génomes d'organites à partir d'ADN ancien provenant de spécimens d'herbiers est possible, ce qui permettra d'inclure de nombreuses espèces rares dans les futurs travaux phylogénétiques des systématiciens et biogéographes. Par ailleurs, nous avons montré que l'utilisation des données mitogénomiques peut permettre de résoudre certaines relations phylogénétiques et ne reflète pas forcément la même histoire évolutive que l'ADN chloroplastique, signifiant que ces données pourraient être complémentaires dans certaines études. Enfin, nous avons mis en évidence un possible lien entre "paternal leakage" et réarrangements génomiques associés à une accélération du taux d'évolution. Néanmoins, la relation de cause à effet reste mal comprise. Ce travail de thèse pose une base pour de futures études sur les conséquences génomiques d'un "paternal leakage" et sur les interactions nucléo-cytoplasmiques.

Références

Bakker FT, Lei D, Yu J *et al.* (2016) Herbarium genomics: plastome sequence assembly from a range of herbarium specimens using an iterative organelle genome assembly pipeline. *Biological Journal of the Linnean Society*, **117**, 33–43.

Bergthorsson U, Richardson AO, Young GJ, Goertzen LR, Palmer JD (2004) Massive horizontal transfer of mitochondrial genes from diverse land plant donors to the basal angiosperm *Amborella*. *Proceedings of the National Academy of Sciences of the United States of America*, **101**, 17747–17752.

Besnard G, Christin PA, Malé PJG *et al.* (2014) From museums to genomics: old herbarium specimens shed light on a C_3 to C_4 transition. *Journal of Experimental Botany*, **65**, 6711–6721.

Besnard G, Hernández P, Khadari B, Dorado G, Savolainen V (2011) Genomic profiling of plastid DNA variation in the Mediterranean olive tree. *BMC Plant Biology*, **11**, 80.

Besnard G, Khadari B, Villemur P, Bervillé A (2000) Cytoplasmic male sterility in the olive (*Olea europaea* L.). *Theoretical and Applied Genetics*, **100**, 1018–1024.

Charlesworth D, Morgan MT, Charlesworth B (1993) Mutation accumulation in finite outbreeding and inbreeding populations. *Genetical Research*, **61**, 39–56.

Cho Y, Mower JP, Qiu YL, Palmer JD (2004) Mitochondrial substitution rates are extraordinarily elevated and variable in a genus of flowering plants. *Proceedings of the National Academy of Sciences of the United States of America*, **101**, 17741–17746.

Coissac E (2016) ORG.asm 0.2.04: A de novo assembler dedicated to organelle genome assembling. *See https://pypi.python.org/pypi/ORG.asm/0.2.04*.

Cosner ME, Jansen RK, Palmer JD, Downie SR (1997) The highly rearranged chloroplast genome of *Trachelium caeruleum* (Campanulaceae): multiple inversions, inverted repeat expansion and contraction, transposition, insertions/deletions, and several repeat families. *Current Genetics*, **31**, 419–429.

Cosner ME, Raubeson LA, Jansen RK (2004) Chloroplast DNA rearrangements in Campanulaceae: phylogenetic utility of highly rearranged genomes. *BMC Evolutionary Biology*, **4**, 27.

Délye C, Deulvot C, Chauvel B (2013) DNA analysis of herbarium specimens of the grass weed *Alopecurus myosuroides* reveals herbicide pesistance pre-dated herbicides. *PLoS ONE*, **8**, 1–8.

Felsenstein J, Yokoyama S (1976) The evolutionary advantage of recombination. *Genetics*, **83**, 845–859.

Greiner S, Sobanski J, Bock R (2014) Why are most organelle genomes transmitted maternally? *BioEssays*, **37**, 80–94.

Jeffroy O, Brinkmann H, Delsuc F, Philippe H (2006) Phylogenomics : the beginning of incongruence ? *Trends in Genetics*, **22**, 225–231.

Marais G (2003) Biased gene conversion: implications for genome and sex evolution. *Trends in Genetics*, **19**, 330–338.

Mensous M, Van de Paer C, Manzi S *et al.* (2017) Diversity and evolution of plastomes in Saharan mimosoids: potential use for phylogenetic and population genetic studies. *Tree Genetics and Genomes*, **13**, 48.

Park S, Ruhlman TA, Weng M-L *et al.* (2017) Contrasting patterns of nucleotide substitution rates provide insight into dynamic evolution of plastid and mitochondrial genomes of *Geranium*. *Genome Biology and Evolution*, **9**, 1766–1780.

Parkinson CL, Mower JP, Qiu YL *et al.* (2005) Multiple major increases and decreases in mitochondrial substitution rates in the plant family Geraniaceae. *BMC Evolutionary Biology*, **5**, 73.

Radzvilavicius A, Kokko H, Christie J (2017) Mitigating mitochondrial genome erosion without recombination. *Genetics*, doi.org/10.1534/genetics.117.300273.

Rhoads A, Au KF (2015) PacBio sequencing and its applications. *Genomics, Proteomics and Bioinformatics*, **13**, 278–289.

Rice DW, Alverson AJ, Richardson AO *et al.* (2013) Horizontal transfer of entire genomes via mitochondrial fusion in the angiosperm *Amborella*. *Science*, **342**, 1468–1473.

Rodrigue N, Lartillot N (2016) Detecting adaptation in protein-coding genes using a Bayesian site-heterogeneous mutation-selection codon substitution model. *Molecular Biology and Evolution*, **34**, 204–214.

Sato M, Sato K (2013) Maternal inheritance of mitochondrial DNA by diverse mechanisms to eliminate paternal mitochondrial DNA. *Biochimica et Biophysica Acta*, **1833**, 1979–1984.

Sloan DB, Alverson AJ, Wu M, Palmer JD, Taylor DR (2012) Recent acceleration of plastid sequence and structural evolution coincides with extreme mitochondrial divergence in the angiosperm genus *Silene*. *Genome Biology and Evolution*, **4**, 294–306.

Sloan DB, Wu Z (2014) History of plastid DNA insertions reveals weak deletion and AT mutation biases in angiosperm mitochondrial genomes. *Genome Biology and Evolution*, **6**, 3210–3221.

Staats M, Cuenca A, Richardson JE *et al.* (2011) DNA damage in plant herbarium tissue. *PLoS ONE*, **6**, e28448.

Wang YH, Qu XJ, Chen SY, Li DZ, Yi TS (2017) Plastomes of Mimosoideae: structural and size variation, sequence divergence, and phylogenetic implication. *Tree Genetics and Genomes*, **13**, 41.

Wang D, Wu YW, Shih ACC *et al.* (2007) Transfer of chloroplast genomic DNA to mitochondrial genome occurred at least 300 MYA. *Molecular Biology and Evolution*, **24**, 2040–2048.

Zedane L, Hong-Wa C, Murienne J *et al.* (2016) Museomics illuminate the history of an extinct, paleoendemic plant lineage (*Hesperelaea*, Oleaceae) known from an 1875 collection from Guadalupe Island, Mexico. *Biological Journal of the Linnean Society*, **117**, 44–57.

Zhang C, Li F, Jia X, Li Z, Feng Q (2013) Bioinformatics analysis of mitochondrial genes *sdh3* and *sdh4* of cytoplasmic male sterility in tobacco. *Chinese Tobacco Science*, **3**, 42–47.

ANNEXE I

Supplementary Material

Mitogenomics of *Hesperelaea*, an extinct genus of Oleaceae

VAN DE PAER C., HONG-WA C., JEZIORSKI C., BESNARD G.

Gene, 2016, **594**: 197-202

SI Appendix includes:

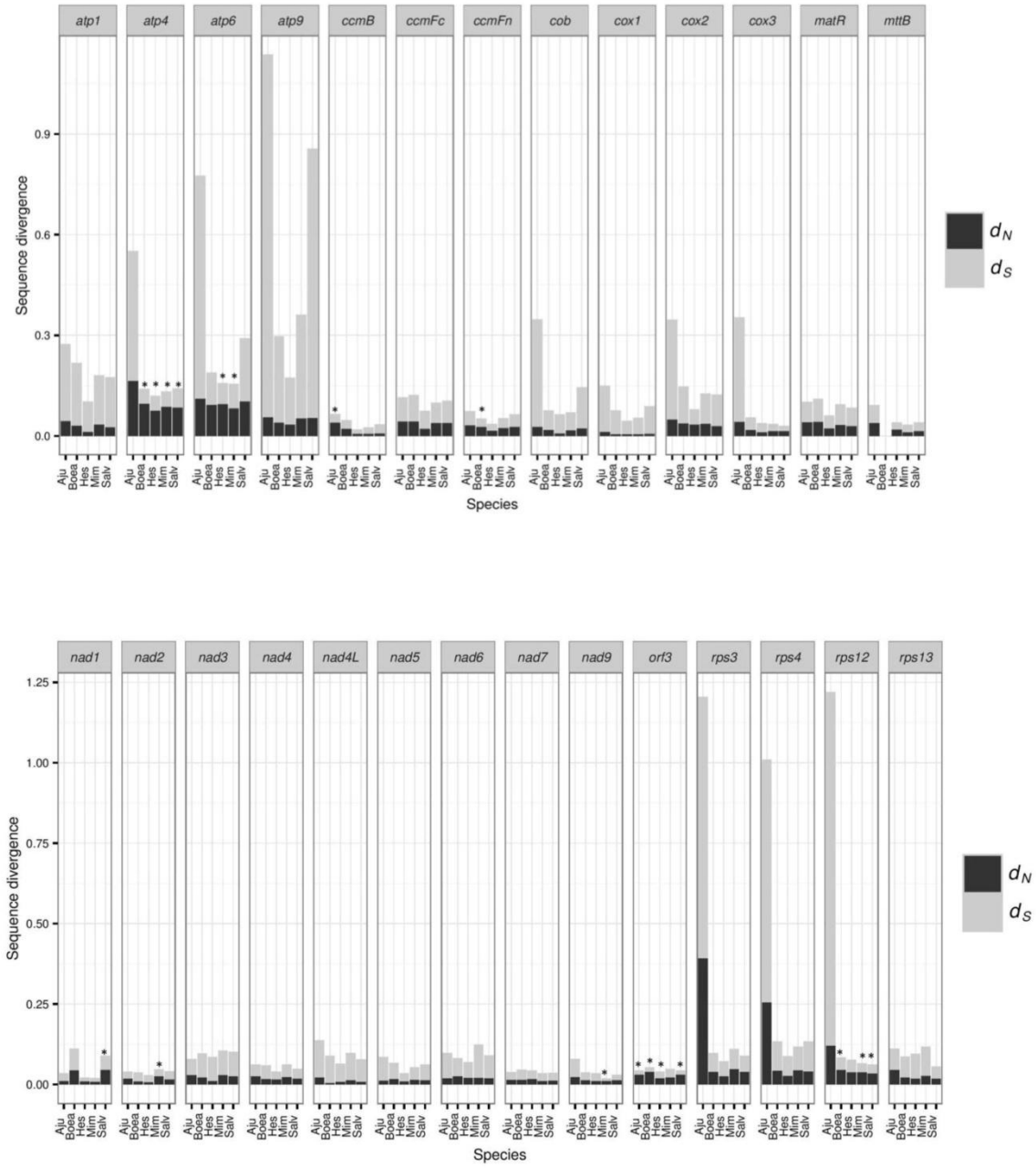


Figure S1. Substitution rates among 27 protein-coding genes in five Lamiales species. For each protein-coding gene, non-synonymous (d_N) and synonymous (d_S) divergence rates were calculated against *Nicotiana tabacum* using PAML (Yang, 2007). The star symbol (*) indicates d_N/d_S ratio (ω) superior to 1. (Aju: *Ajuga reptans*; Boea*: Boea hygrometrica*; Hes: *Hesperelaea palmeri*; Mim: *Mimulus guttatus*; Salv: *Salvia miltiorrhiza*)

Table S1. List of repeat pairs predicted using VMATCH (Kurtz, 2016) in the mitogenome of *Hesperelaea palmeri*. We retained repeats with a 40-bp minimum length and with at least 95% of identity. (D: direct orientation; P: palindromic orientation)

Repeat length (bp)	Start position	Match direction	Start position	Identity percentage (%)
7975	192316	P	435791	100.00
496	135775	D	201125	98.39
309	172567	D	610037	99.35
176	8692	D	262772	98.86
164	202770	P	267594	99.39
156	261094	P	643905	100.00
153	88691	P	98009	98.69
145	130405	D	513821	100.00
129	16246	D	472714	98.45
125	16561	D	473033	95.20
120	183482	D	272866	93.33
118	88726	P	98009	99.15
118	376013	D	482787	99.15
107	204776	P	500880	100.00
105	58304	D	85353	100.00
98	125629	P	459480	100.00
98	527237	D	654484	100.00
95	469137	P	654440	100.00
94	142048	P	508300	100.00
94	341217	P	615268	98.94
90	201688	D	459479	98.89
90	469149	P	571588	97.78
89	125638	P	201689	98.88
87	39758	D	633208	98.85
86	142156	P	508198	100.00
85	15892	D	472347	95.46
83	571595	D	654440	97.59
82	102791	D	391865	98.78
81	215553	D	428181	100.00
80	191683	P	521201	100.00
77	7434	P	28907	100.00
76	494294	D	540035	97.37
75	314298	D	606097	98.67
74	288508	D	344748	97.30
73	502372	D	513490	98.63

Table S1. Continued

Repeat length (bp)	Start position	Match direction	Start position	Identity percentage (%)
71	114004	P	201141	97.18
71	114004	P	135791	98.59
69	67020	D	555343	97.10
69	132572	P	650916	97.10
69	513019	P	552428	98.55
66	201778	D	459579	98.5
64	469131	D	504862	100.00
63	1	P	351561	100.00
62	16098	D	472557	95.2
62	308369	P	638443	96.77
62	313708	P	551989	98.39
61	125040	P	404562	96.72
59	19198	P	513083	98.31
59	88785	P	98009	100.00
59	126068	D	510693	98.31
59	238945	P	552360	96.61
58	504868	P	654477	100.00
57	393054	P	609882	98.25
56	45609	P	537758	98.21
56	251733	P	513082	98.21
55	6625	D	180011	98.18
55	19202	D	251733	100.00
55	113944	P	201435	96.36
55	113944	P	136085	98.18
54	19200	D	341527	100.00
54	288438	D	344694	96.30
54	341527	P	513086	98.15
53	14479	D	199336	96.23
53	14479	P	436693	96.23
53	41011	D	251732	100.00
53	52272	D	433928	96.23
53	282272	D	344623	98.11
53	342033	D	469057	96.23
52	19202	D	41012	100.00
52	36584	P	405527	96.15
52	41012	D	341529	100.00
52	41012	P	513086	98.08
52	251733	D	341529	100.00
52	385227	D	654545	96.15
51	129290	P	458242	98.04
51	314927	D	385227	98.04

Table S1. Continued

Repeat length (bp)	Start position	Match direction	Start position	Identity percentage (%)
51	349634	D	508263	98.04
51	469137	P	527237	100.00
51	504868	P	527237	100.00
50	130529	P	314519	96.00
50	168060	D	168096	98.00
50	365567	D	497112	98.00
50	365613	D	497172	100.00
50	494209	D	539960	96.00
49	31246	D	128836	100.00
49	124229	P	613841	97.96
49	313998	P	620051	100.00
49	314925	D	527296	95.92
49	347295	D	612152	100.00
48	16170	D	472638	95.8
48	29011	D	83560	97.92
48	64383	P	142124	100.00
47	320223	D	403438	95.74
47	385227	D	527298	97.87
47	454881	P	583085	95.74
47	494323	D	540064	97.87
47	494882	D	540610	95.74
46	8124	P	32702	100.00
46	45841	P	537696	95.65
46	504880	P	571632	97.83
45	23408	P	382354	100.00
45	41163	D	506842	100.00
45	135776	D	576939	95.56
45	308386	P	638443	100.00
45	342061	D	504832	97.78
45	382960	D	641035	95.56
45	494160	D	539915	95.56
44	8413	D	395045	97.73
44	114100	P	137242	95.45
44	129292	P	513004	95.45
44	321762	D	382414	97.73
44	458247	D	513004	97.73
43	16055	D	472503	95.3
43	41228	P	433845	95.35
43	125058	P	404562	100.00

Table S1. End

Repeat length (bp)	Start position	Match direction	Start position	Identity percentage (%)
43	129273	D	433033	100.00
43	136153	D	155433	95.35
43	136538	D	381742	100.00
43	313727	P	551989	100.00
43	314925	D	654543	97.67
42	45623	P	537758	100.00
42	54763	P	248929	95.24
42	116025	P	577381	95.24
42	450071	D	537324	95.24
42	505127	D	516942	95.24
41	5935	P	382085	95.12
41	6625	P	219574	100.00
41	29119	D	83644	95.12
41	68397	D	463470	97.56
41	83556	D	113089	97.56
41	84454	P	381746	95.12
41	137257	D	357334	95.12
41	148300	D	540879	95.12
41	180011	P	219574	100.00
41	216731	D	521486	100.00
41	382096	D	640153	95.12
41	425682	P	507295	95.12
41	527235	D	571637	95.12
40	3160	P	85352	95.00
40	16473	D	472942	95.00
40	28975	D	516746	95.00
40	29010	D	113092	97.50
40	52289	D	433068	95.00
40	65224	D	442676	95.00
40	65224	P	193366	95.00
40	98745	P	163695	97.50
40	99893	D	314877	95.00
40	111174	P	314296	95.00
40	113906	P	136133	95.00
40	138673	P	433879	95.00
40	331236	P	331238	95.00
40	494765	D	540490	95.00

Table S2. List of the plastid-derived regions, their location in the mitogenome of *Hesperelaea palmeri*, their identities and the percent of gaps compared to the homeologous region from the plastome [GenBank: NC_025787]. Only plastid-derived regions with a minimum length of 100 bp and with at least 80% of identity with the *H. palmeri* plastome were retained. The 15 fragments in bold show an identity with the plastome superior to 96%.

Fragment	Start position	Final position	Identities	Gaps
Fragment 1	**54**	**1526**	**1453/1484 (97%)**	**15/1484 (1%)**
Fragment 2	**1540**	**3191**	**1647/1652 (99%)**	**0/1652 (0%)**
Fragment 3	15816	16695	56/914 (82%)	58/914 (6%)
Fragment 4	**34035**	**37178**	**3103/3166 (98%)**	**22/3166 (0%)**
Fragment 5	**145126**	**148340**	**3209/3215 (99%)**	**0/3215 (0%)**
Fragment 6	**148300**	**152507**	**4178/4216 (99%)**	**8/4216 (0%)**
Fragment 7*	**194090**	**195799**	**1655/1722 (96%)**	**28/1722 (1%)**
Fragment 8	**206177**	**209684**	**3480/3542 (98%)**	**36/3542 (1%)**
Fragment 9	335093	335285	164/193 (84%)	4/193 (2%)
Fragment 10	335745	336707	846/1010 (83%)	54/1010 (5%)
Fragment 11	**379173**	**381554**	**2365/2433 (97%)**	**53/2433 (2%)**
Fragment 12	**381540**	**381688**	**148/149 (99%)**	**0/149 (0%)**
Fragment 13	405425	405578	126/157 (80%)	13/157 (8%)
Fragment 14	**406449**	**413450**	**6939/7054 (98%)**	**55/7054 (0%)**
Fragment 15	**413424**	**421212**	**7716/7826 (98%)**	**65/7826 (0%)**
Fragment 16*	**440282**	**441991**	**1655/1722 (96%)**	**28/1722 (1%)**
Fragment 17	**470279**	**480109**	**9665/9920 (97%)**	**114/9920 (1%)**
Fragment 18	**491195**	**492517**	**1303/1324 (98%)**	**8/1324 (0%)**
Fragment 19	540767	540873	101/119 (84%)	12/119 (10%)
Fragment 20	**638374**	**643913**	**5527/5571 (99%)**	**31/5571 (0%)**
Fragment 21	653303	653526	184/228 (80%)	4/228 (1%)

***** Fragments 7 and 16 are parts of the long duplicated zone of ca. 8 kb

Table S3. Comparison of gene content and number of repeats > 1 kb among the mitogenomes of Lamiales and *Nicotiana tabacum* (Solanales). The annotations of the mitogenomes were extracted from GenBank. (+: gene presence; -: gene absence; Ψ: pseudogene)

Gene	*Nicotiana tabacum*	*Hesperelaea palmeri*	*Boea hygrometrica*	*Mimulus guttatus*	*Salvia miltiorrhiza*	*Ajuga reptans*
atp1	+	+ (ATG)*	+	+	+	+
atp4	+	+ (ATG)	+	+	+	+
atp6	+	+ (ACG)	+	+	+	+
atp8	-	+ (ATG)	+	+	+	+
atp9	+	+ (ATG)	+	+	+	+
ccmB	+	+ (ATG)	+	+	+	+
ccmC	-	+ (ATG)	+	+	+	+
ccmFc	+	+ (ATG)	+	+	+	+
ccmFn	+	+ (ATG)	+	+	+	+
cox1	+	+ (ATG)	+	+	+	+
cox2	+	+ (ATG)	+	+	+	+
cox3	+	+ (ATG)	+	+	+	+
cytB	+	+ (ATG)	+	+	+	+
matR	+	+ (ATG)	+	+	+	+
mttB	+	+ (TTG)	-	+	+	+
nad1	+	+ (ATG)	+	+	+	+
nad2	+	+ (ATG)	+	+	+	+
nad3	+	+ (ATG)	+	+	+	+
nad4	+	+ (ATG)	+	+	+	+
nad4L	+	+ (ACG)	+	+	+	+
nad5	+	+ (ATG)	+	+	+	+
nad6	+	+ (ATG)	+	+	+	+
nad7	+	+ (ATG)	+	+	+	+
nad9	+	+ (ATG)	+	+	+	+
orf3	+	+ (ATG)	+	+	+	+
orf101	-	+ (ATG)	-	-	+	-
orf111a	-	+ (ATG)	-	-	+	-
orf117a	-	+ (ATG)	-	-	+	-
orf121	-	+ (ATG)	-	-	+	-
orf122c	-	+ (ATG)	-	-	+	-
orf182a	-	+ (ATG)	-	-	+	-
petG	-	-	-	-	+	-
petL	-	+ (ATG)	-	-	+	-
rpl2	-	+ (ATG)	-	Ψ	+	+
rpl5	+	+ (ATG)	-	+	+	-
rpl10	-	+ (ATG)	-	+	+	-
rpl16	+	+ (GTG)	Ψ	+	+	-
rpl23	-	+ (ATG)	-	-	+	+

***** Start codon is given in brackets for each protein-coding gene of *Hesperelaea*

Table S3. End

Gene	*Nicotiana tabacum*	*Hesperelaea palmeri*	*Boea hygrometrica*	*Mimulus guttatus*	*Salvia miltiorrhiza*	*Ajuga reptans*
rps3	+	+ (ATG)	+	+	+	+
rps4	+	+ (ACG)	+	+	+	Ψ
rps7	-	+ (ATG)	+	-	+	-
rps10	+	+ (ACG)	-	+	+	-
rps12	+	+ (ATG)	+	+	+	+
rps13	+	+ (ATG)	+	+	+	+
rps14	Ψ	+ (ATG)	-	+	+	-
rps19	+	-	-	-	-	-
sdh3	+	+ (ATG)	+	+	-	-
sdh4	-	+ (GTG)	-	+	+	-
rrn5	+	+	+	+	+	+
rrn18	+	+	+	+	+	+
rrn26	+	+	+	+	+	+
trnC-GCA	+	+	+	+	+	+
trnD-GTC	+	+	+	+	+	+
trnE-TTC	+	+	+	+	+	+
trnF-GAA	+	+	+	+	+	+
trnfM-CAT	+	-	-	+	+	+
trnG-GCC	+	+	+	+	+	-
trnH-GTG	+	+	+	+	+	-
trnI-CAT	-	+	-	+	+	+
trnK-TTT	+	+	+	+	+	-
trnL-CAA	-	-	+	+	+	-
trnL-GAG	-	-	-	+	-	-
trnL-TAA	-	-	+	+	-	-
trnM-CAT	+	+	+	+	+	+
trnN-GTT	+	+	+	+	+	+
trnP-TGG	+	+	+	+	+	+
trnQ-TTG	+	+	+	+	+	+
trnR-ACG	-	+	+	-	+	-
trnR-TCT	-	+	-	-	+	-
trnS-GCT	-	+	+	+	+	+
trnS-GGA	+	+	+	+	+	-
trnS-TGA	-	+	+	+	+	+
trnT-TGT	-	-	+	+	-	-
trnV-GAC	-	-	+	-	+	-
trnW-CCA	+	+	+	+	+	+
trnY-GTA	+	+	+	+	+	+
Total protein-coding genes	**32**	**46**	**30**	**36**	**46**	**30**
Total rRNA genes	**3**	**3**	**3**	**3**	**3**	**3**
Total tRNA genes	**15**	**19**	**21**	**22**	**22**	**14**
Number of repeats > 1kb	**5**	**1**	**1**	**3**	**3**	**0**

* Start codon is given in brackets for each protein-coding gene of *Hesperelaea*: Note that we reported 39 protein-coding genes with the standard ATG start codon. The ACG codon could be an alternative start codon in four protein-coding genes (*atp6*, *nad4L*, *rps4* and *rps10*), as reported by Giegé and Brennicke (1999), Picardi et al. (2010) and Mower et al. (2012b). The codon initiating translation in *sdh4* and *rpl16* genes is GTG, as already suggested in other plant mitogenomes (Bock et al., 1994). The ATG start codon is also not conserved in *mttB* and the TTG codon was an alternative start codon, as annotated in the mitogenome of *Salvia miltiorrhiza* (Mower et al., 2012a).

Table S4. List of ORFs detected with GENEIOUS v. 9.0.5 (Biomatters Ltd., 2015) in the mitogenome of *Hesperelaea palmeri*. We retained ORFs with a minimum length of 300 bp.

ORFs	Length (bp)	Start position	Final position	Orientation
orf1	300	109865	110164	reverse
orf2	300	87498	87797	reverse
orf4	300	141399	141698	forward
orf5	300	326463	326762	forward
orf6	300	440790	441089	forward
orf7	300	464017	464316	forward
orf8	300	630173	630472	reverse
orf9	300	519268	519567	reverse
orf10	300	518186	518485	reverse
orf11	300	521613	521912	forward
orf12	300	381349	381648	reverse
orf13	300	10460	10759	forward
orf14	300	194992	195291	reverse
orf15	303	273915	274217	forward
orf16	303	436778	437080	forward
orf17	303	279315	279617	reverse
orf18	303	199001	199303	reverse
orf19	306	52382	52687	reverse
orf20	306	331298	331603	forward
orf21	306	581600	581905	forward
orf22	306	91421	91726	forward
orf23	306	657960	658265	reverse
orf24	306	499339	499644	reverse
orf25	309	534818	535126	forward
orf26	309	506430	506738	forward
orf27	309	101161	101469	forward
orf28	309	497328	497636	reverse
orf29	312	219852	220163	forward
orf30	312	309468	309779	reverse
orf31	312	269059	269370	reverse
orf32	315	152180	152494	forward
orf33	315	343431	343745	forward
orf34	315	355310	355624	forward
orf35	315	549406	549720	reverse
orf36	318	388677	388994	forward
orf37	318	442363	442680	forward
orf38	318	624651	624968	forward
orf39	318	60539	60856	forward

Table S4. Continued

ORFs	Length	Start position	Final position	Orientation
orf40	318	435564	435881	reverse
orf41	318	193401	193718	reverse
orf42	321	98825	99145	forward
orf43	321	266670	266990	reverse
orf44	321	185364	185684	reverse
orf45	324	177748	178071	forward
orf46	324	252473	252796	reverse
orf47	327	567655	567981	reverse
orf48	330	252594	252923	forward
orf49	330	266195	266524	forward
orf50	330	301985	302314	forward
orf51	330	596928	597257	reverse
orf52	333	5083	5415	forward
orf53	333	21954	22286	forward
orf54	333	376316	376648	forward
orf55	333	520389	520721	forward
orf56	333	475430	475762	reverse
orf57	336	90115	90450	reverse
orf58	336	21306	21641	forward
orf59	336	339734	340069	forward
orf60	336	454078	454413	forward
orf61	336	239417	239752	reverse
orf62	339	621652	621990	forward
orf63	339	433425	433763	forward
orf64	339	500932	501270	reverse
orf65	339	526287	526625	forward
orf66	342	356961	357302	reverse
orf67	345	549356	549700	forward
orf68	348	76426	76773	reverse
orf69	348	11067	11414	reverse
orf70	348	100811	101158	forward
orf71	348	111218	111565	forward
orf72	348	459806	460153	reverse
orf73	348	524846	525193	forward
orf74	348	528254	528601	forward
orf75	351	139488	139838	reverse
orf76	351	101442	101792	reverse
orf77	351	10523	10873	reverse
orf78	351	321670	322020	forward
orf79	351	630834	631184	reverse
orf80	354	74481	74834	reverse
orf81	354	456675	457028	forward

Table S4. Continued

ORFs	Length	Start position	Final position	Orientation
orf82	354	508357	508710	forward
orf83	354	559239	559592	reverse
orf84	354	488007	488360	reverse
orf85	354	121219	121572	forward
orf86	357	344776	345132	forward
orf87	357	569464	569820	reverse
orf88	363	189762	190124	forward
orf89	366	86180	86545	reverse
orf90	366	119896	120261	forward
orf91	366	260060	260425	reverse
orf92	369	132637	133005	forward
orf93	372	271528	271899	reverse
orf94	375	397237	397611	forward
orf95	375	59584	59958	forward
orf96	375	386546	386920	reverse
orf97	378	30320	30697	reverse
orf98	378	21932	22309	reverse
orf99	378	43384	43761	forward
orf100	381	647801	648181	forward
orf102	381	83575	83955	forward
orf103	390	380080	380469	forward
orf104	393	184746	185138	forward
orf105	393	101488	101880	forward
orf106	396	33053	33448	forward
orf107	396	639715	640110	forward
orf108	399	174047	174445	forward
orf109	399	233604	234002	forward
orf110	399	36679	37077	forward
orf111	399	402022	402420	reverse
orf112	402	2251	2652	reverse
orf113	405	300822	301226	reverse
orf114	408	99588	99995	forward
orf115	411	176825	177235	reverse
orf116	414	89235	89648	reverse
orf117	417	134807	135223	forward
orf118	417	195194	195610	forward
orf119	417	440471	440887	reverse
orf120	420	165428	165847	forward
orf121	420	406602	407021	forward
orf122	423	253615	254037	forward
orf123	423	513960	514382	forward
orf124	426	342277	342702	forward

Table S4. Continued

ORFs	Length	Start position	Final position	Orientation
orf125	426	466575	467	Forward
orf126	435	161245	161679	reverse
orf127	435	272203	272637	reverse
orf128	447	208680	209126	forward
orf129	450	380719	381168	forward
orf130	456	78332	78787	forward
orf131	459	559640	560098	forward
orf132	462	574819	575280	reverse
orf133	465	555656	556120	reverse
orf134	465	417367	417831	reverse
orf135	471	226011	226481	forward
orf136	471	631262	631732	reverse
orf137	486	378625	379110	reverse
orf138	489	379	867	reverse
orf139	495	322409	322903	reverse
orf140	501	309193	309693	forward
orf141	519	121158	121676	reverse
orf142	528	508562	509089	forward
orf143	540	151856	152395	reverse
orf144	540	104629	105168	reverse
orf145	546	301463	302008	reverse
orf146	561	529819	530379	forward
orf147	564	555795	556358	reverse
orf148	567	415133	415699	reverse
orf149	570	149347	149916	reverse
orf150	570	441530	442099	forward
orf151	570	193982	194551	reverse
orf152	576	605529	606104	reverse
orf153	582	9493	10074	forward
orf154	594	136632	137225	reverse
orf155	594	269735	270328	reverse
orf156	600	379439	380038	forward
orf157	627	64465	65091	forward
orf158	633	141386	142018	forward
orf159	633	516801	517433	forward
orf160	642	476839	477480	reverse
orf161	663	207978	208640	forward
orf162	672	146002	146673	reverse
orf163	732	180372	181103	reverse
orf164	744	407244	407987	forward
orf165	765	445409	446173	reverse
orf166	774	387306	388079	forward

Table S4. End

ORFs	Length	Start position	Final position	Orientation
orf167	822	262017	262838	Reverse
orf168	834	46796	47629	reverse
orf169	870	113054	113923	forward
orf170	885	608880	609764	reverse
orf171	921	150179	151099	reverse
orf172	945	206273	207217	forward
orf173	963	474096	475058	reverse
orf174	1026	448145	449170	reverse
orf175	1062	154322	155383	reverse
orf176	1062	34661	35722	forward
orf177	1173	388076	389248	forward
orf178	1176	200288	201463	reverse
orf179	1224	410964	412187	forward
orf180	1242	545617	546858	forward
orf181	1386	418088	419473	reverse
orf182	1485	478610	480094	reverse
orf183	1605	585577	587181	reverse
orf184	1653	415689	417341	reverse
orf185	1953	510750	512702	forward
orf186	2283	389149	391431	forward
orf187	2757	177238	179994	reverse

References

Biomatters Ltd., 2015. GENEIOUS v. 9.0.5. Available at http://www.geneious.com/.

Bock, H., Brennicke, A., Schuster, W., 1994. *Rps3* and *rpl16* genes do not overlap in *Oenothera* mitochondria: GTG as a potential translation initiation codon in plant mitochondria? Plant Mol. Biol. 24, 811−818.

Giegé, P., Brennicke, A., 1999. RNA editing in *Arabidopsis* mitochondria effects 441 C to U changes in ORFs. Proc. Natl. Acad. Sci. USA 96, 15324–15329.

Kurtz, S., 2016. The VMATCH large scale sequence analysis software. Available at http://vmatch.de/.

Mower, J.P., Case, A.L., Floro, E.R., Willis, J.H., 2012a. Evidence against equimolarity of large repeat arrangements and a predominant master circle structure of the mitochondrial genome from a monkeyflower (*Mimulus guttatus*) lineage with cryptic CMS. Genome Biol. Evol. 4, 670−686.

Mower, J.P., Sloan, D.B., Alverson, A.J., 2012b. Plant mitochondrial genome diversity: the genomics revolution. In: Wendel, J.F., Greilhuber, J., Dolezel, J., & Leitch, I.J. (eds), Plant Genome Diversity plant genomes: Their Residents, and their Evolutionary Dynamics, Vol. 1, pp. 123−144. Springer, Vienna

Picardi, E., Horner, D.S., Chiara, M., Schiavon, R., Valle, G., Pesole, G., 2010. Large-scale detection and analysis of RNA editing in grape mtDNA by RNA deep-sequencing. Nucleic Acids Res., 38, 4755−4767.

Yang, Z., 2007. PAML 4: Phylogenetic analysis by maximum likelihood. Mol Biol Evol. 24, 1586–1591.

ANNEXE II

Supplementary Material

Prospects on the evolutionary mitogenomics of plants: a case study on the olive family (Oleaceae)

VAN DE PAER C, BOUCHEZ O, BESNARD G

Molecular Ecology Resources, 2017, 1-17

SI Appendix includes:

Table S1. Summary of plastome and mitogenome sequencing and assembly for each individual. (N: number; SD: Standard Deviation)

Tree accession	Sequencer (HiSeq)	N of paired-end reads	Read length	Mean insert size	N of merged reads	N of unmerged reads	Mean coverage of the plastome	Mean coverage of the Mitogenome[a]	*N50*[b] (Mitogenome)	Relative mean coverage: mitogenome vs. plastome (in %)
O. e. laperrinei E1-*l*1.1	3000	58,412,297	150	362	7,506,658	91,707,366	7,464.8× (*SD* = 966.6)	1,712.2× (*SD* = 174.8)	32463 (7)	23
O. e. europaea E1-*e*.4	2000	6,378,934	101	233	1,471,069	9,767,926	477.2× (*SD* =106.7)	54.7× (*SD* = 14.2)	32463 (7)	11
O. e. europaea E2.1	3000	27,916,604	150	419	1,757,602	50,895,394	2,013.6× (*SD* = 324.9)	412.5× (*SD* = 55.7)	29,426 (7)	20
O. e. guanchica M-*g*1.3	3000	6,064,147	150	315	1,674,271	8,107,320	326.1× (*SD* = 108.3)	33.5× (*SD* = 9.5)	40,477 (6)	10
O. e. europaea E3.1	3000	36,544,906	150	416	2,751,685	64,771,250	3,222.6× (*SD* = 459.8)	258.7× (*SD* = 32.9)	28,988 (8)	8
O. e. cuspidata C2.13	2000	5,606,536	101	379	628,669	9,941,264	218.7× (*SD* = 33.7)	22.5× (*SD* = 8.4)	53,990 (5)	10
Chionanthus parkinsonii	3000	14,886,027	150	198	9,570,299[c]	6,489,146	264.4× (*SD* = 53.2)	170.7× (*SD* = 51.4)	-	65
Chionanthus rupicolus	2500	19,149,492	125	202	12,539,371[c]	12,733,070	1,282.7× (*SD* = 232.7)	164.3× (*SD* = 52.2)	69,824 (4)	13
Fontanesia fortunei	2000	11,371,103	100	277	1,344,747	20,005,632	824.0× (*SD* = 137.3)	32.0× (*SD* = 14.2)	-	4
Forestiera isabelae	3000	8,243,092	150	164	6,299,465[c]	1,098,614	172.4× (*SD* = 31.0)	45.2× (*SD* = 18.8)	-	26
Forsythia × intermedia	2000	7,513,296	100	231	1,443,519	12,092,640	665.5× (*SD* = 129.4)	60.6× (*SD* = 28.7)	-	9
Fraxinus ornus	2000	9,238,000	100	265	958,119	16,517,852	744.1× (*SD* = 122.9)	47.8× (*SD* = 13.3)	-	6
Hesperelaea palmeri	2000	10,694,511	100	118	9,620,010	2,011,824	329.8× (*SD* = 144.9)	34.6× (*SD* = 11.0)	98,574 (3)	10
Nestegis apetala	2000	3,209,498	101	383	197,511	6,014,740	367.0× (*SD* = 61.0)	26.3× (*SD* = 8.2)	-	7
Noronhia lowryi	2000	3,803,214	101	230	1,422,249	4,731,472	224.4× (*SD* = 83.9)	33.0× (*SD* = 12.4)	-	15
Olea exasperata	2000	5,070,492	101	258	1,440,334	7,235,042	319.3× (*SD* = 39.0)	26.3× (*SD* = 7.6)	-	8
Schrebera arborea	3000	5,611,849	150	157	4,241,967[c]	700,056	64.0× (*SD* = 23.2)	17.1× (*SD* = 9.4)	-	27
Syringa vulgaris	2000	7,880,981	100	265	730,094	14,225,174	681.8× (*SD* = 142.7)	46.4× (*SD* = 15.3)	-	7

[a] Average coverage depth of mitogenomes was estimated excluding *mtpt* regions; [b] *N50* was calculated on single-copy mitochondrial contigs (excluding *mtpt* regions) after the initial step of the mtDNA assembly. This statistic defines the assembly quality in terms of contiguity. The *N50* length (in bp) is defined as the shortest sequence length at 50% of the genome. The number in brackets corresponds to the number of contigs superior or equal to *N50*; [c] Note that the percentage of merged reads is higher in the four herbarium accessions (>64% of paired-end reads) than in DNA extracted from fresh material (<30%). This is due to the more fragmented DNA usually extracted from herbarium specimens.

Table S2. List of individuals from GenBank used to reconstruct the phylogenetic tree in Fig. S5. Olive plastid DNA lineage and haplotypes are given (according to Besnard *et al.* 2007, 2018).

Species and origin of individuals	**GenBank no**	**Plastid lineage, and haplotype**
Olea europaea L. subsp. *europaea*		
- Manzanilla de Sevilla, Spain	FN996972	Cp-II, E1-*e*.1
- Haut Atlas, Morocco	FN997650	Cp-II, E2.9
- Gué de Constantine 20, Algeria	FN997651	Cp-II, E3.3
Olea europaea subsp. *maroccana* (Greuter & Burdet) P. Vargas *et al.*		
- Imouzzer S1, High Atlas, Morocco	FN998900	Cp-II, M-*m*.1
Olea europaea subsp. *cuspidata* (Wall. ex G. Don) Cif.		
- Maui 1, Hawaii	FN650747	Cp-I, A.1
- Almihwit 5.1, Yemen	FN996943	Cp-I, C2.1
- Guangzhou CH1, China	FN996944	Cp-I, C1.1
Olea woodiana Knobl. subsp. *woodiana*		
- A.Costa 02 [MPU], South Africa	FN998901	-

Table S3. Comparison of gene and pseudogene content among the mitogenomes of *Olea europaea* and *Chionanthus rupicolus*. Whole-gene copy numbers >1 are given. (+: gene presence; -: gene absence)

Product group	**Gene**	**E1-*e*.4** ***O. e. europaea***	**E1-*l*1.1** ***O. e. laperrinei***	**E2.1** ***O. e. europaea***	**E3.1** ***O. e. europaea***	**M-*g*1.3** ***O. e. guanchica***	**C2.13** ***O. e. cuspidata***	***Chionanthus rupicolus***
Complex I	*nad1*	+	+	2+[a]	+	+	2+[b]	+
	nad2	+	+	+	+	+	+	+
	nad3	+	+	+	+	+	+	+
	nad4	+	+	+	+	+	+	+
	nad4L	+	+	+	+	+	+	+
	nad5	+	+	+	+	+	+	+
	nad6	+	+	+	+	+	+	+
	nad7	+	+	+	+	+	+	+
	nad9	+	+	+	+	+	+	+
Complex II	*sdh3*	+	+	+	-	+	+	+
	sdh4	+	+	+	3+	+	+	+
Complex III	*cytB*	+	+	+	+	+	+	+
Complex IV	*cox1*	+	+	+	+	+	+	+
	cox2	+	+	+	+	+	+	+
	cox3	+	+	+	3+	+	+	+
Complex V	*atp1*	+	+	+	+	+	+	+
	atp4	+	+	+	+	+	+	+
	atp6	+	+	+	+	+	+	+
	atp8	+	+	+	+	+	+	+
	atp9	3+	3+	4+	2+	2+	2+	+
Cytochrome c biogenesis	*ccmB*	+	+	2+	+	+	2+	+
	ccmC	+	+	+	+	+	+	+
	ccmFc	+	+	+	+	+	+	+
	ccmFn	+	+	+	+	+	+	+
Ribosome	*rpl2*	+	+	+	+	+	+	+
	rpl5	3+	3+	3+	2+	+	2+	+
	rpl10	+	+	+	+	+	+	+
	rpl16	+	+	+	+	+	+	+
	*rpl23**	+	+	+	+	+	+	-
	rps3	+	+	+	+	+	+	+
	rps4	+	+	+	+	+	+	+
	rps7	+	+	+	+	+	2+	+
	rps10	+	+	+	+	+	+	+
	rps12	+	+	+	+	+	+	+
	rps13	+	+	+	+	+	+	+
	rps14	3+	3+	3+	2+	+	2+	+

Table S3. End

Product group	Gene	E1-*e*.4 *O. e. europaea*	E1-*l*1.1 *O. e. laperrinei*	E2.1 *O. e. europaea*	E3.1 *O. e. europaea*	M-*g*1.3 *O. e. guanchica*	C2.13 *O. e. cuspidata*	*Chionanthus rupicolus*
Other proteins	*matR*	+	+	2+	+	+	+	+
	mttB	+	+	+	2+	+	+	+
	*petL**	+	+	*ψ?*	+	+	+	-
	*petN**	+	+	+	+	+	+	+
	*psbD**	+	+	+	+	+	+	-
rRNA	*rrn5*	+	+	+	+	+	+	+
	rrn18	+	+	+	+	+	+	+
	rrn26	+	+	+	+	+	+	+
tRNA	*trnC-GCA*	+	+	+	+	+	+	+
	*trnC-GCA**	-	-	-	-	-	-	+
	*trnD-GTC**	2+	2+	2+	2+	2+	2+	2+
	trnE-TTC	+	+	+	+	+	+	+
	*trnE-TTC**	+	+	+	+	+	+	+
	trnF-GAA	+	+	+	+	+	+	+
	trnG-GCC	+	+	+	+	+	+	+
	*trnfM-CAT**	+	+	+	+	+	+	-
	trnH-GTG	+	+	+	+	+	+	+
	trnK-TTT	+	+	+	+	+	+	+
	trnM-CAT	3+	3+	3+	3+	3+	3+	3+
	*trnM-CAT**	+	+	+	+	+	+	+
	*trnN-GTT**	2+	2+	2+	2+	2+	2+	2+
	trnP-TGG	+	+	+	+	+	+	+
	trnQ-TTG	+	+	+	+	+	+	+
	trnS-GCT	+	+	+	+	+	+	+
	trnS-TGA	+	+	+	+	+	+	+
	*trnS- GGA**	+	+	+	+	+	+	+
	*trnT-GGT**	+	+	+	+	+	+	-
	*trnV-TAC**	+	+	+	+	+	+	-
	*trnW-CCA**	+	+	+	+	+	+	+
	trnY-GTA	+	+	+	+	+	+	+
	*trnY-GTA**	+	+	+	+	+	+	+
Pseudogenes	*ψ-ccmC*	-	-	2+	2+	2+	+	-
	orf_E3a	-	-	-	2+	-	-	-
	orf_E3b	-	-	-	+	-	-	-

[a] Two copies of exons 4 and 5; [b] Two copies of exon 4; * Chloroplast-derived protein-coding and tRNA genes

Table S4. Repeat content among the mitogenomes of *Olea europaea* and *Chionanthus rupicolus*

Repeat length (bp)	**E1-*e*.4** ***O. e. europaea***	**E1-*l*1.1** ***O. e. laperrinei***	**E2.1** ***O. e. europaea***	**E3.1** ***O. e. europaea***	**M-*g*1.3** ***O. e. guanchica***	**C2.13** ***O. e. cuspidata***	***Chionanthus rupicolus***
40-60	65	65	60	65	61	63	78
61-80	20	20	22	20	21	22	27
81-100	14	14	15	11	15	10	21
101-120	1	1	2	3	3	3	1
121-140	4	4	2	3	3	3	2
141-160	3	3	3	5	4	3	3
161-180	0	0	0	2	2	1	1
181-200	0	0	1	1	1	1	1
201-220	1	1	1	2	1	1	2
221-240	0	0	0	0	1	0	0
241-260	2	2	1	2	1	2	1
261-280	0	0	0	0	1	0	1
281-300	0	0	1	1	1	1	0
301-400	0	0	0	0	0	0	1
401-500	1	1	1	0	0	0	0
501-600	0	0	0	0	0	0	0
601-700	0	0	0	0	1	0	0
701-800	0	0	0	0	0	0	0
801-900	0	0	2	0	0	0	0
901-1,000	0	0	0	0	0	1	0
1,001-10,000	3	3	2	2	1	5	0
>10,000	1	1	2	3	0	1	0

Table S5. Estimates of the nucleotide diversity (π in %; Nei & Li 1979) among the olive complex for the nine *mtpt* regions (Fig. 2) and their ptDNA homologous regions. The standard error of π was estimated by bootstrapping on sequences (100 replicates).

Genomic region	Length (bp)	Parsimony-informative sites	mtDNA	ptDNA
mtpt1	497	12	0.0	0.0
mtpt2	3110	14	$2.8\times10^{-4} \pm 2.1\times10^{-4}$	$2.1\times10^{-4} \pm 1.6\times10^{-4}$
mtpt3	3154	52	$6.5\times10^{-4} \pm 2.6\times10^{-4}$	$1.1\times10^{-4} \pm 1.1\times10^{-4}$
mtpt4	1898	8	$4.6\times10^{-4} \pm 3.6\times10^{-4}$	$1.8\times10^{-4} \pm 1.6\times10^{-4}$
mtpt5	512	7	$1.0\times10^{-3} \pm 1.0\times10^{-3}$	0.0
mtpt6	1013	98	$7.0\times10^{-3} \pm 5.6\times10^{-4}$	$3.4\times10^{-4} \pm 3.3\times10^{-4}$
mtpt7	646	9	$3.8\times10^{-3} \pm 1.8\times10^{-3}$	$1.4\times10^{-3} \pm 9.4\times10^{-4}$
mtpt8	712	69	$1.4\times10^{-3} \pm 8.6\times10^{-4}$	$9.9\times10^{-4} \pm 6.4\times10^{-4}$
mtpt9	2183	23	$7.4\times10^{-4} \pm 4.4\times10^{-4}$	0.0

Table S6. GenBank numbers for each assembled mitochondrial gene region in ten Oleaceae species

Gene	*Chionanthus parkinsonii*	*Fontanesia fortunei*	*Forestiera isabelae*	*Forsythia× intermedia*	*Fraxinus ornus*	*Nestegis apetala*	*Noronhia lowryi*	*Olea exasperata*	*Schrebera arborea*	*Syringa vulgaris*
atp1	MG431454	MG431455	MG431456	MG431457	MG431458	MG431459	MG431460	MG431461	MG431462	MG431463
atp4	MG431464	MG431465	MG431466	MG431467	MG431468	MG431469	MG431470	MG431471	MG431472	MG431473
atp6	MG431474	MG431475	MG431476	MG431477	MG431478	MG431479	MG431480	MG431481	MG431482	MG431483
atp8	MG431484	MG431485	MG431486	MG431487	MG431488	MG431489	MG431490	MG431491	MG431492	MG431493
atp9	MG431494	MG431495	MG431496	MG431497	MG431498	MG431499	MG431500	MG431501	MG431502	MG431503
ccmB	MG431504	MG431505	MG431506	MG431507	MG431508	MG431509	MG431510	MG431511	MG431512	MG431513
ccmC	MG431514	MG431515	MG431516	MG431517	MG431518	MG431519	MG431520	MG431521	MG431522	MG431523
ccmFc	MG431714	MG431715	MG431716	MG431717	MG431718	MG431719	MG431720	MG431721	MG431722	MG431723
ccmFn	MG431524	MG431525	MG431526	MG431527	MG431528	MG431529	MG431530	MG431531	MG431532	MG431533
cox1	MG431534	MG431535	MG431536	MG431537	MG431538	MG431539	MG431540	MG431541	MG431542	MG431543
cox2	MG431794	MG431795	MG431796	MG431797	MG431798	MG431799	MG431800	MG431801	MG431802	MG431803
cox3	MG431544	MG431545	MG431546	MG431547	MG431548	MG431549	MG431550	MG431551	MG431552	MG431553
cytB	MG431554	MG431555	MG431556	MG431557	MG431558	MG431559	MG431560	MG431561	MG431562	MG431563
matR	MG431564	MG431565	MG431566	MG431567	MG431568	MG431569	MG431570	MG431571	MG431572	MG431573
mttB	MG431574	MG431575	MG431576	MG431577	MG431578	MG431579	MG431580	MG431581	MG431582	MG431583
*nad1**	MG431724	MG431725	MG431726	MG431727	MG431728	MG431729	MG431730	MG431731	MG431732	MG431733
*nad2**	MG431734	MG431735	MG431736	MG431737	MG431738	MG431739	MG431740	MG431741	MG431742	MG431743
nad3	MG431584	MG431585	MG431586	MG431587	MG431588	MG431589	MG431590	MG431591	MG431592	MG431593
*nad4**	MG431784	MG431785	MG431786	MG431787	MG431788	MG431789	MG431790	MG431791	MG431792	MG431793
nad4L	MG457284	MG457285	MG457286	MG457287	MG457288	MG457289	MG457290	MG457291	MG457292	MG457293
*nad5**	MG431744	MG431745	MG431746	MG431747	MG431748	MG431749	MG431750	MG431751	MG431752	MG431753
nad6	MG431594	MG431595	MG431596	MG431597	MG431598	MG431599	MG431600	MG431601	MG431602	MG431603
nad7	MG431764	MG431765	MG431766	MG431767	MG431768	MG431769	MG431770	MG431771	MG431772	MG431773
nad9	MG431604	MG431605	MG431606	MG431607	MG431608	MG431609	MG431610	MG431611	MG431612	MG431613
rpl5	MG431614	MG431615	MG431616	MG431617	MG431618	MG431619	MG431620	MG431621	MG431622	MG431623
rpl10	MG431624	MG431625	MG431626	MG431627	MG431628	MG431629	MG431630	MG431631	MG431632	MG431633
rpl16	MG431634	MG431635	MG431636	MG431637	MG431638	MG431639	MG431640	MG431641	MG431642	MG431643
rps3	MG431754	MG431755	MG431756	MG431757	MG431758	MG431759	MG431760	MG431761	MG431762	MG431763
rps4	MG431644	MG431645	MG431646	MG431647	MG431648	MG431649	MG431650	MG431651	MG431652	MG431653
rps7	MG431654	MG431655	MG431656	MG431657	MG431658	MG431659	MG431660	MG431661	MG431662	MG431663
rps10	MG431774	MG431775	MG431776	MG431777	MG431778	MG431779	MG431780	MG431781	MG431782	MG431783
rps12	MG431664	MG431665	MG431666	MG431667	MG431668	MG431669	MG431670	MG431671	MG431672	MG431673
rps13	MG431674	MG431675	MG431676	MG431677	MG431678	MG431679	MG431680	MG431681	MG431682	MG431683
rps14	MG431684	MG431685	MG431686	MG431687	MG431688	MG431689	MG431690	MG431691	MG431692	MG431693
sdh3	MG431694	MG431695	MG431696	MG431697	MG431698	MG431699	MG431700	MG431701	MG431702	MG431703
sdh4	MG431704	MG431705	MG431706	MG431707	MG431708	MG431709	MG431710	MG431711	MG431712	MG431713

***** One intron of *nad1*, *nad2, nad5* (all species) and *nad4* (*F. fortunei* and *S. arborea*) was not assembled (and indicated by 100 N in the submitted gene accession)

Table S7. Nucleotide polymorphisms at the 21 parsimony informative sites among the plastomes of the olive complex. Sites are numbered according to their position in the plastome of *O. e. laperrinei* E1-*l*1.1. All sites were located either in the LSC or SSC (Long or Short Single Copies, respectively). When the site is located in a coding sequence, the gene name, the position of the substitution in the codon and the amino-acid change for non-synonymous substitutions are indicated.

Accession	**Sites**																				
	2	**9082**	**13362**	**18642**	**45532**	**48090**	**51579**	**52165**	**56950**	**64704**	**67654**	**83304**	**85525**	**85589**	**112754**	**114452**	**115364**	**118316**	**122532**	**127059**	**129178**
	LSC	**LSC**	**LSC**	**LSC**	**LSC**	**LSC**	**LSC**	**LSC**	**LSC**	**LSC**	**LSC**	**LSC**	**LSC**	**LSC**	**SSC**	**SSC**	**SSC**	**SSC**	**SSC**	**SSC**	**SSC**
O. e. cuspidata C2.13	A	C	T	A	A	C	A	T	G	G	T	T	T	G	C	G	T	T	G	G	T
O. e. europaea E3.1	A	**A**	T	A	A	**A**	A	T	G	G	T	**G**	T	G	C	G	T	T	G	G	T
O. e. guanchica M-*g*1.3	A	C	T	A	A	**A**	**C**	T	G	G	**G**	**G**	T	G	C	G	T	T	**T**	G	T
O. e. europaea E2.1	A	**A**	T	A	A	**A**	**C**	T	G	G	**G**	T	T	G	**A**	G	T	T	G	G	T
O. e. europaea E1-*e*.4	**G**	**A**	**G**	**G**	**G**	C	**C**	**G**	**A**	**A**	**G**	T	**G**	**A**	**A**	**A**	**A**	**A**	**T**	**C**	**A**
O. e. laperrinei E1.*l*1.1	**G**	C	**G**	**G**	**G**	C	**C**	**G**	**A**	**A**	**G**	T	**G**	**A**	C	**A**	**A**	**A**	**T**	**C**	**A**
Gene location				*rpoC2*			*psbG*	*ndhC*		*petA*		*rpl14*	*rps3*	*rps3*	*ndhF*	*ndhF*		*ndhD*		*ycf1*	*ycf1*
Position in codon				2nd			3rd	3rd		3rd		3rd	1st	3rd	2nd	2nd		3rd		2nd	3rd
Non syn. changes								L/F				L/F	N/H	W/L		A/V				T/S	

Table S8. Nucleotide polymorphisms at the 156 parsimony informative sites among the mitogenomes of the olive complex. Sites are numbered according to their position in the mitogenome of *O. e. laperrinei* E1-*l*1.1. When the site is located in a coding sequence, the gene name, the position of the substitution in the codon and the amino-acid change for non-synonymous substitutions are given. We also indicate sites that were detected in *mtpt* regions.

Accession	Sites																													
	5100	**6444**	**17291**	**23100**	**26122**	**31568**	**39873**	**42361**	**43990**	**44938**	**46133**	**49983**	**49984**	**53846**	**67978**	**73599**	**74037**	**77942**	**81198**	**94185**	**96862**	**97419**	**97422**	**99292**	**99969**	**102429**	**106218**	**110323**	**112202**	**123269**
	mtpt																										***mtpt***			
O. e. cuspidata C2.13	G	G	T	C	-*	A	G	T	C	G	G	G	G	A	G	T	C	A	G	C	A	G	A	G	C	T	T	C	G	T
O. e. europaea E3.1	**T**	**T**	T	C	**A**	**C**	**T**	T	C	-	-	G	G	A	G	T	**A**	A	G	C	A	G	A	**A**	C	T	**A**	**A**	G	**A**
O. e. guanchica M-g1.3	G	G	T	C	**A**	**A**	**T**	T	**G**	G	G	G	G	**A**	G	T	C	A	G	**G**	A	G	A	**A**	**G**	T	T	C	**T**	**A**
O. e. europaea E2.1	**T**	**T**	T	**A**	**A**	**C**	G	**A**	**G**	G	**A**	-	-	**G**	G	T	**A**	A	**T**	**G**	**C**	**A**	**G**	**T**	C	T	**A**	**A**	**T**	T
O. e. europaea E1-*e*.4	G	**T**	**G**	**A**	**T**	**C**	G	**A**	C	**T**	**A**	**A**	**T**	**G**	**C**	**A**	C	**C**	**T**	C	**C**	**A**	**G**	**T**	**G**	**G**	**A**	C	**T**	**A**
O. e. laperrinei E1-*l*1.1	G	**T**	**G**	**A**	**T**	**C**	G	**A**	C	**T**	**A**	**A**	**T**	**G**	**C**	**A**	C	**C**	**T**	C	**C**	**A**	**G**	**T**	**G**	**G**	**A**	C	**T**	**A**
Gene location																									*cox3*					
Position in codon																									1st					
Non syn. changes																									L/I					

* "-" means that the region is missing

Table S8. Continued

Accession	Sites																													
	123712	**127747**	**133314**	**136867**	**137204**	**137490**	**137777**	**152560**	**157274**	**162261**	**162268**	**168739**	**169298**	**169700**	**169727**	**170796**	**189737**	**194408**	**206535**	**211367**	**211428**	**212669**	**222771**	**222882**	**223149**	**223194**	**223195**	**231189**	**248167**	**248735**
				mtpt	***mtpt***																										
O. e. cuspidata C2.13	C	A	G	T	A	T	C	A	A	C	C	A	A	T	A	T	T	T	-	T	A	C	T	G	G	A	T	C	C	G
O. e. europaea E3.1	**A**	A	G	**G**	**T**	**G**	C	**C**	A	**A**	C	A	A	T	A	T	T	T	**A**	T	A	C	T	G	**T**	A	T	**G**	C	**T**
O. e. guanchica M-*g*1.3	**A**	A	G	**G**	A	**G**	**A**	**C**	A	C	C	A	**C**	**C**	**G**	T	T	-	**A**	T	A	C	T	**T**	G	**C**	T	**G**	C	G
O. e. europaea E2.1	**A**	A	G	T	A	T	**A**	A	**G**	**A**	C	A	A	**C**	**G**	T	G	T	**A**	T	A	C	**G**	**T**	G	**C**	T	**G**	C	**T**
O. e. europaea E1-*e*.4	**A**	**C**	**T**	**G**	A	**G**	C	A	**G**	**A**	**A**	**C**	**C**	**C**	**G**	**A**	G	**C**	**C**	**G**	**C**	**A**	**G**	G	**T**	A	**G**	**G**	**A**	**T**
O. e. laperrinei E1-*l*1.1	C	**C**	**T**	**G**	**T**	**G**	C	A	**G**	**A**	**A**	**C**	A	**C**	**G**	**A**	G	**C**	**C**	**G**	**C**	**A**	**G**	G	G	A	**G**	C	**A**	**T**
Gene location																														
Position in codon																														
Non syn. changes																														

Table S8. Continued

Accession	Sites																													
	259497	**265110**	**276117**	**279456**	**281992**	**282455**	**282932**	**283164**	**295854**	**300051**	**301105**	**301753**	**302527**	**302609**	**303919**	**308204**	**311059**	**311775**	**326439**	**334081**	**334581**	**338510**	**339310**	**349116**	**351907**	**354564**	**356733**	**380955**	**383338**	**392471**
					mtpt																									
O. e. cuspidata C2.13	G	T	C	T	A	G	T	G	G	T	C	C	G	T	C	T	T	C	T	C	T	G	A	A	C	G	T	A	T	T
O. e. europaea E3.1	G	T	**A**	T	**C**	G	**G**	G	G	T	C	**A**	G	T	C	T	T	C	T	C	T	G	**C**	A	**A**	G	T	A	T	**G**
O. e. guanchica M-*g*1.3	G	T	**A**	T	A	G	T	**T**	**T**	T	C	**A**	G	T	C	T	T	**A**	T	**A**	**A**	G	**C**	A	C	G	T	**C**	T	T
O. e. europaea E2.1	G	T	**A**	T	A	**C**	**G**	-	**T**	T	C	C	G	T	C	T	-	-	T	**A**	**A**	G	**C**	**T**	**A**	**T**	T	**C**	T	**G**
O. e. europaea E1-*e*.4	**T**	**G**	**A**	**C**	**C**	**C**	**G**	**T**	**T**	**A**	**A**	**A**	**T**	**A**	**A**	**C**	**C**	**A**	**G**	C	**A**	**T**	A	**T**	**A**	**T**	**G**	**C**	G	**G**
O. e. laperrinei E1.*l*1.1	**T**	**G**	C	**C**	A	**C**	**G**	**T**	**T**	**A**	**A**	C	**T**	**A**	**A**	**C**	**C**	C	**G**	C	**A**	**T**	C	**T**	**A**	**T**	**G**	**C**	G	**G**
Gene location																														
Position in codon																														
Non syn. changes																														

Table S8. Continued

Accession	Sites																														
	399823	**413679**	**416549**	**416818**	**416913**	**417505**	**418532**	**418532**	**421440**	**421444**	**430394**	**433100**	**433102**	**435700**	**462965**	**465836**	**469087**	**473913**	**476879**	**482112**	**486679**	**496068**	**498087**	**504968**	**512223**	**522023**	**523270**	**530752**	**531126**	**535430**
																															mtpt
O. e. cuspidata C2.13	A	G	T	T	G	A	C	G	T	C	G	G	G	T	C	C	A	G	C	A	G	G	C	C	C	A	A	G	C	T
O. e. europaea E3.1	**C**	**C**	T	**A**	G	**C**	C	G	T	C	G	G	G	T	C	C	A	G	C	A	G	G	C	C	**A**	**C**	A	G	C	**G**
O. e. guanchica M-*g*1.3	**C**	G	T	**A**	**T**	**C**	**G**	**C**	-	-	G	G	G	**G**	**A**	C	**T**	**T**	C	A	G	G	C	**A**	C	**C**	**T**	**T**	**A**	**G**
O. e. europaea E2.1	A	**C**	T	T	**T**	**C**	C	G	**G**	C	**T**	**T**	G	**G**	**A**	**A**	**T**	**T**	C	A	**T**	**T**	**A**	**A**	**A**	A	**T**	**T**	**A**	T
O. e. europaea E1-*e*.4	A	**C**	**A**	T	**T**	A	C	G	**G**	**T**	G	**T**	**T**	**G**	C	**A**	**T**	**T**	**A**	**C**	**T**	**T**	**A**	**A**	**A**	**C**	**T**	**T**	**A**	**G**
O. e. laperrinei E1.*l*1.1	A	**C**	**A**	T	**T**	**C**	**G**	**C**	**G**	**T**	**T**	**T**	**T**	**G**	**A**	**A**	**T**	**T**	**A**	**C**	**T**	**T**	**A**	**A**	**A**	**C**	**T**	**T**	**A**	**G**
Gene location				*atp1*	*atp1*	*atp1*																*atp4*								
Position in codon				3rd	1st	3rd																1st								
Non syn. changes					L/I	F/L																Q/K								

Table S8. Continued

Accession	Sites																												
	537598	**537609**	**544032**	**546114**	**551289**	**558335**	**563412**	**563767**	**587413**	**587593**	**588311**	**595120**	**602501**	**605913**	**608724**	**609518**	**609658**	**609738**	**609878**	**611806**	**616230**	**619053**	**619060**	**622580**	**622941**	**625645**	**627248**	**637501**	**638174**
																												mtpt	
O. e. cuspidata C2.13	A	G	C	T	C	T	G	A	C	C	G	A	A	C	A	C	T	C	G	T	G	C	G	A	A	A	C	A	C
O. e. europaea E3.1	A	G	**A**	**G**	C	T	G	A	C	**A**	**T**	**T**	**C**	C	A	C	T	C	G	T	G	**A**	G	A	**C**	A	C	A	**G**
O. e. guanchica M-*g*1.3	A	G	C	**G**	**A**	T	**T**	**C**	C	**A**	G	A	A	C	A	C	T	C	G	**G**	**A**	C	G	A	**C**	A	**A**	-	-
O. e. europaea E2.1	A	G	**A**	T	**A**	T	G	A	**A**	C	G	A	A	C	A	C	T	**A**	G	T	**A**	C	G	**T**	A	**T**	**A**	A	C
O. e. europaea E1-*e*.4	**C**	**C**	**A**	T	**A**	**G**	**T**	**C**	**A**	C	**T**	**T**	**C**	**A**	**G**	**T**	**A**	**A**	**T**	T	**A**	T	**T**	**T**	A	**T**	**A**	**C**	**G**
O. e. laperrinei E1.*l*1.1	**C**	**C**	**A**	T	**A**	**G**	**T**	**C**	**A**	C	G	**T**	**C**	**A**	**G**	**T**	**A**	**A**	**T**	**G**	**A**	T	**T**	**T**	A	**T**	**A**	**C**	**G**
Gene location																													
Position in codon																													
Non syn. changes																													

Table S8. End

Accession	Sites						
	648972	**649951**	**652195**	**675733**	**692527**	**692716**	**698722**
O. e. cuspidata C2.13	T	C	T	G	-	-	C
O. e. europaea E3.1	T	**A**	T	G	T	T	-
O. e. guanchica M-*g*1.3	T	**A**	T	**T**	T	T	C
O. e. europaea E2.1	**C**	C	T	G	**C**	T	-
O. e. europaea E1-*e*.4	**C**	**A**	**A**	**T**	**C**	**G**	**A**
O. e. laperrinei E1./1.1	**C**	**A**	**A**	**T**	**C**	**G**	**A**
Gene location Position in codon Non syn. changes							

Fig. S1. Master circles of the seven reconstructed mitogenomes of *Olea europaea* (six accessions) and *Chionanthus rupicolus*. The outer circle represents the gene content, with protein-coding genes, rRNA and tRNA genes. The inner circle represents repeated sequences with a minimum length of 100 bp in the mitogenome. The large repeats >10 kb are represented in light blue color. The maps were performed with OGDRAW v1.2 (Lohse *et al.* 2007) and CIRCOS (Krzywinski *et al.* 2009).

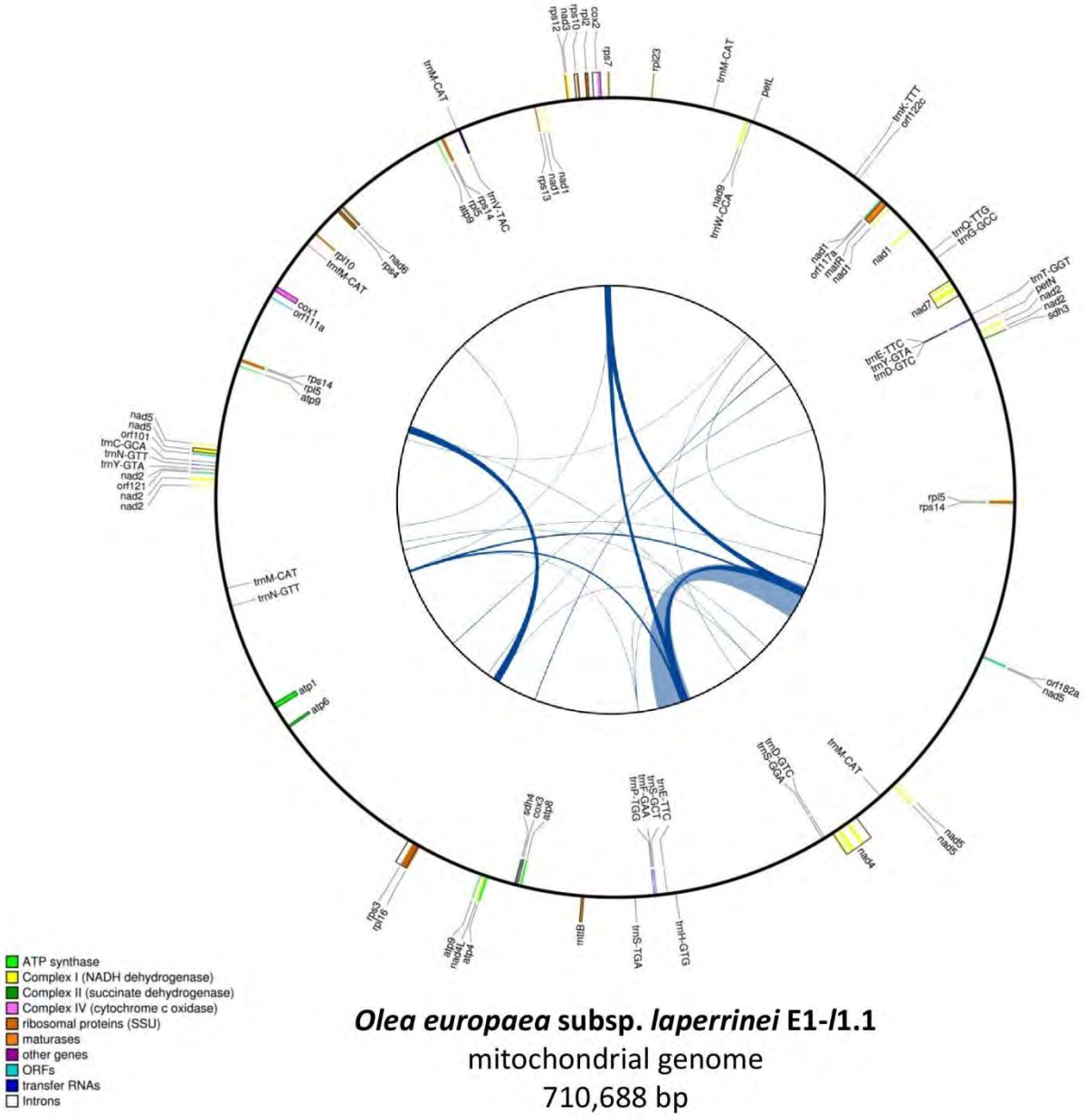


Fig. S1. Continued

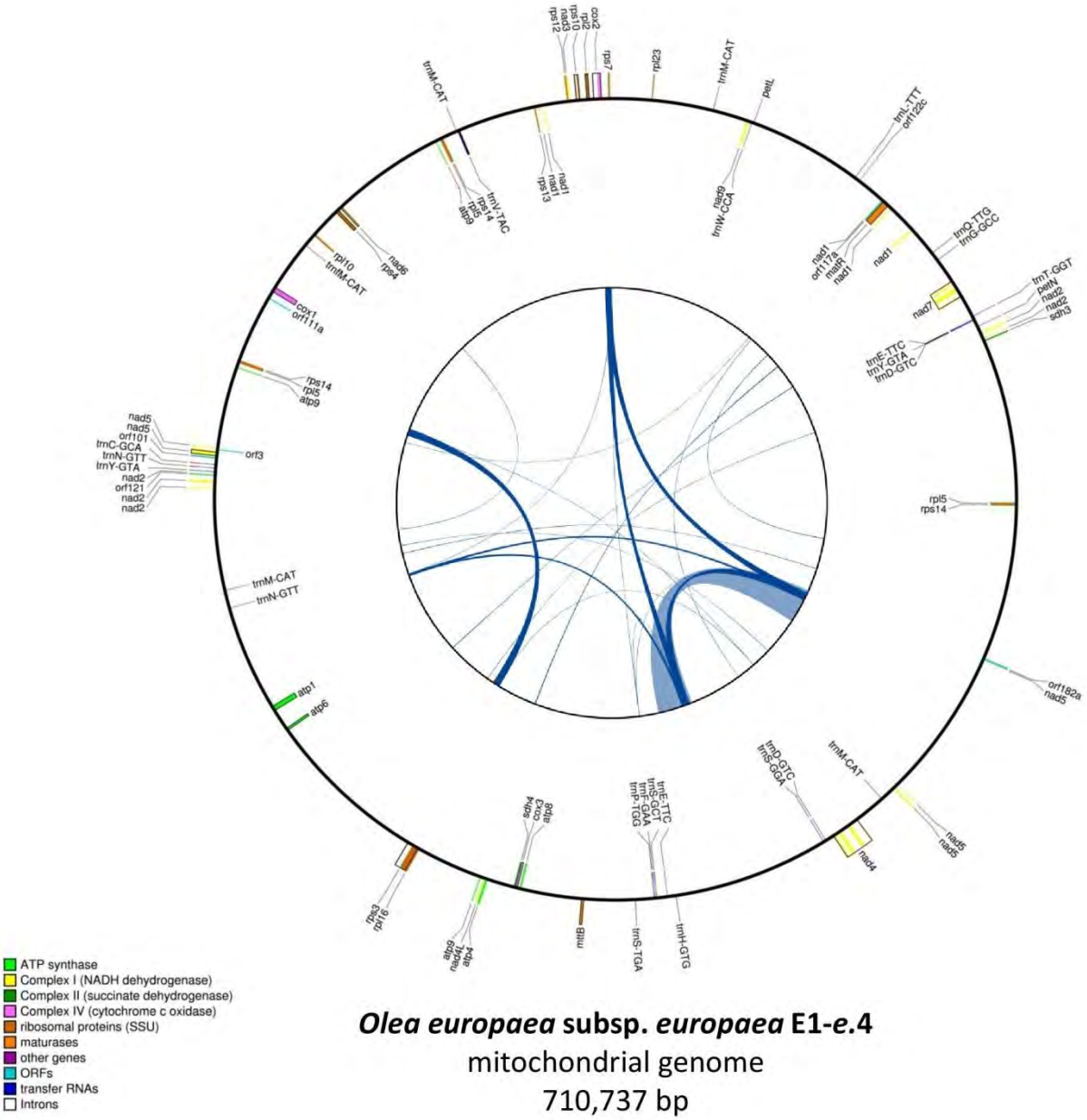


Fig. S1. Continued

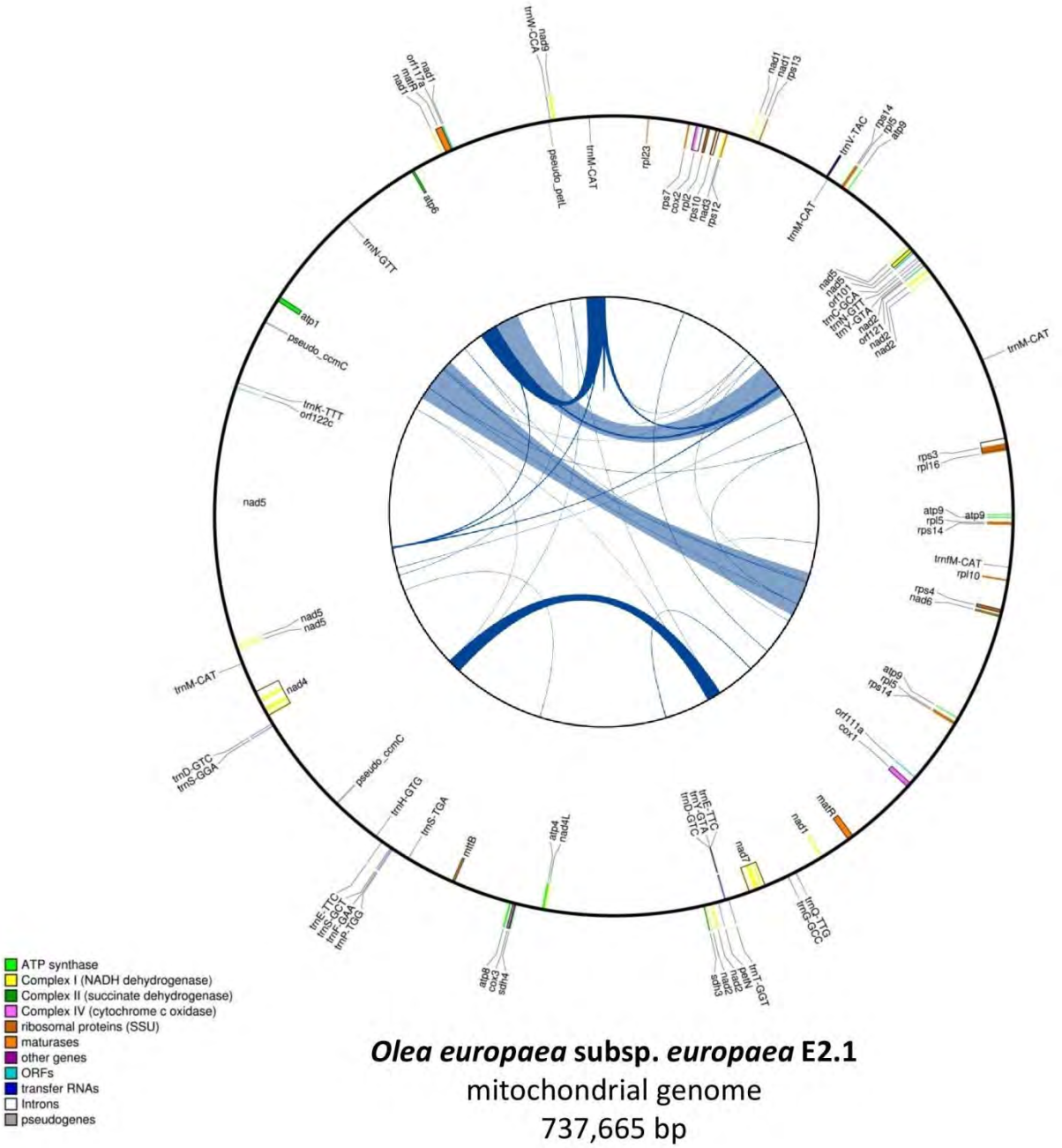


Fig. S1. Continued

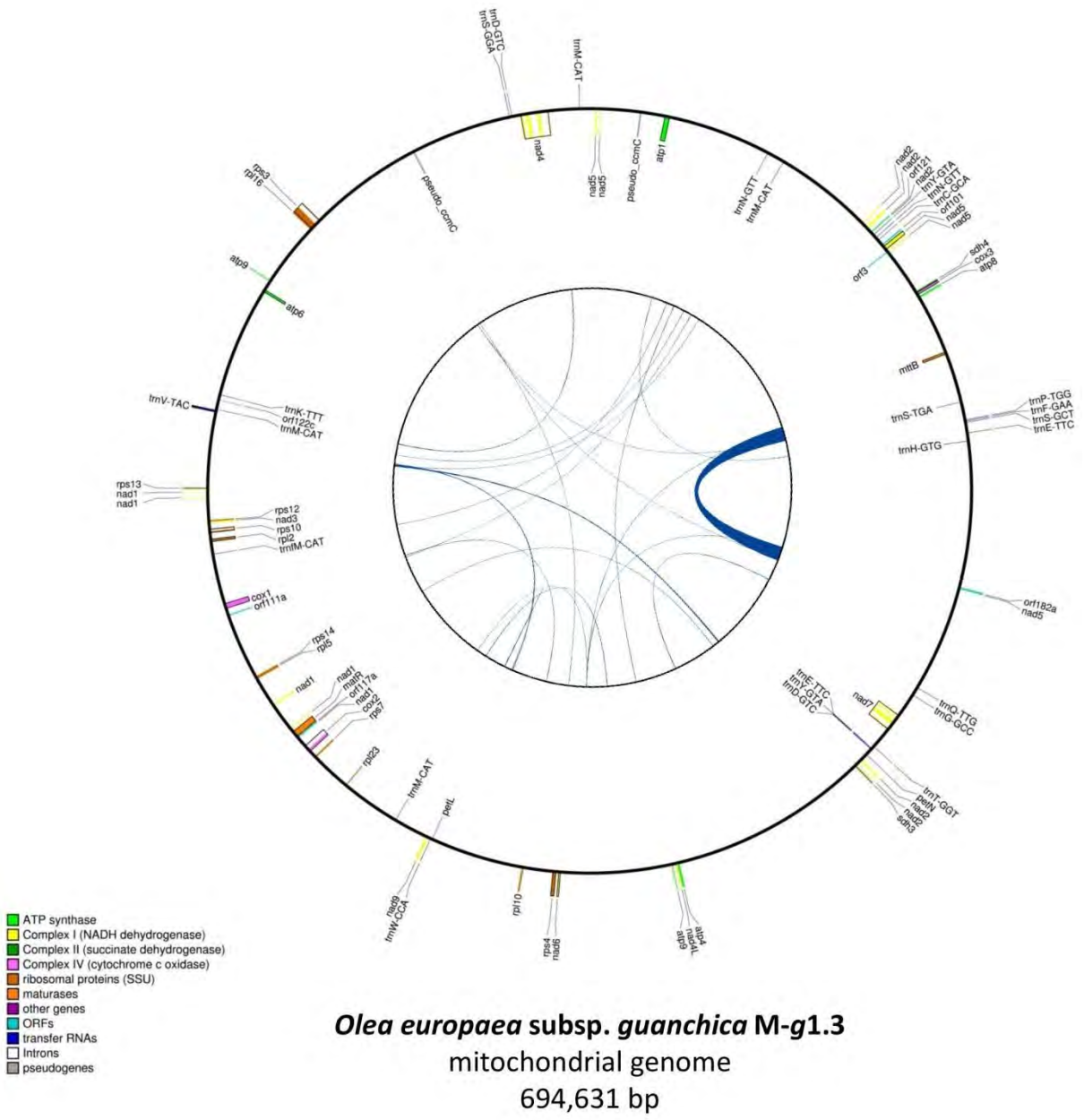


Fig. S1. Continued

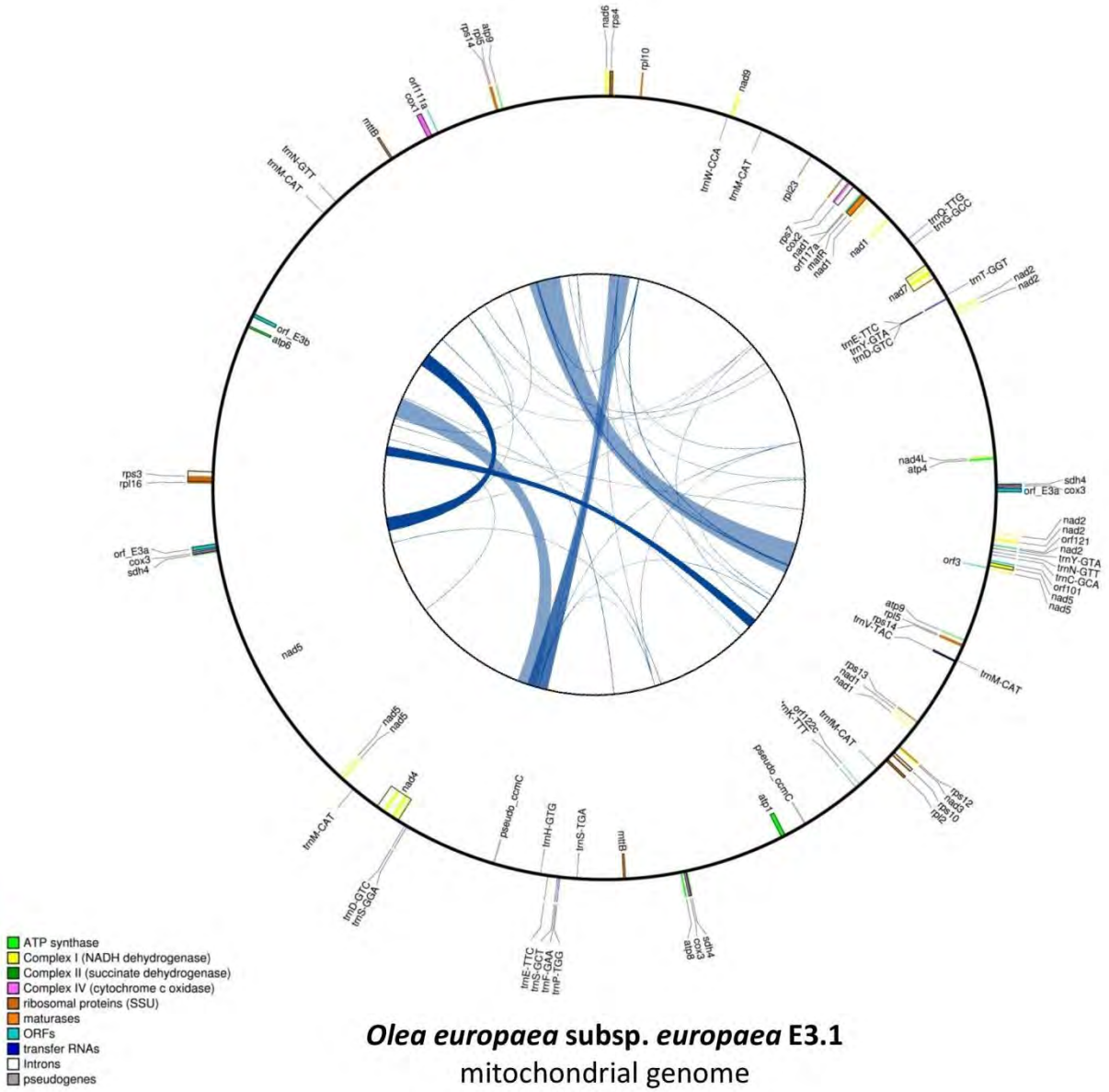


Fig. S1. Continued

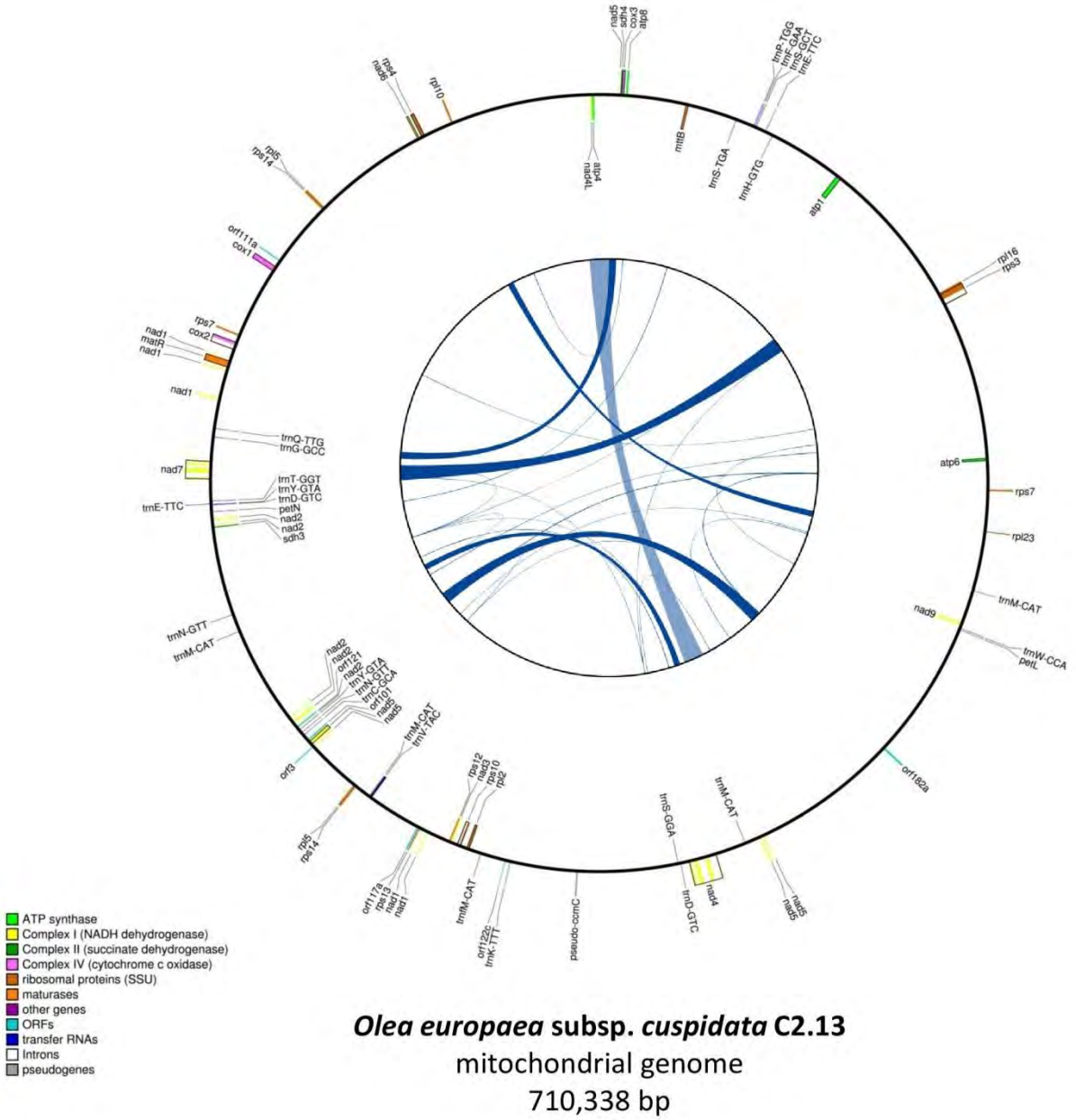
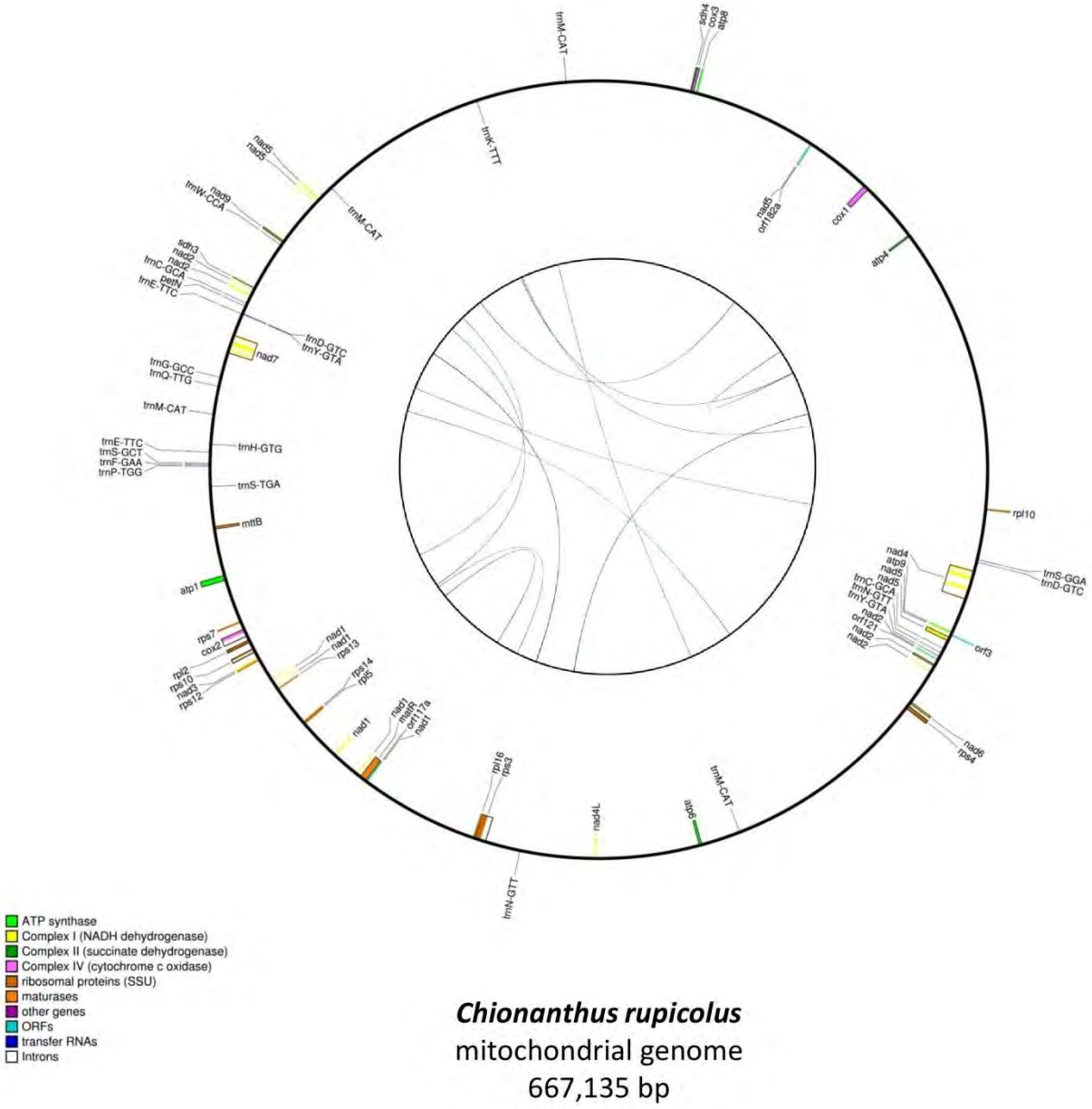


Fig. S1. End



Chionanthus rupicolus
mitochondrial genome
667,135 bp

Fig. S2. Alignment of the six *O. europaea* mitogenomes using PROGRESSIVEMAUVE (Darling *et al.* 2010). PROGRESSIVEMAUVE detects and aligns conserved regions (called locally collinear blocks, LCB) between genomes. Homologous regions between mitogenomes have the same color, and are linked by a line.

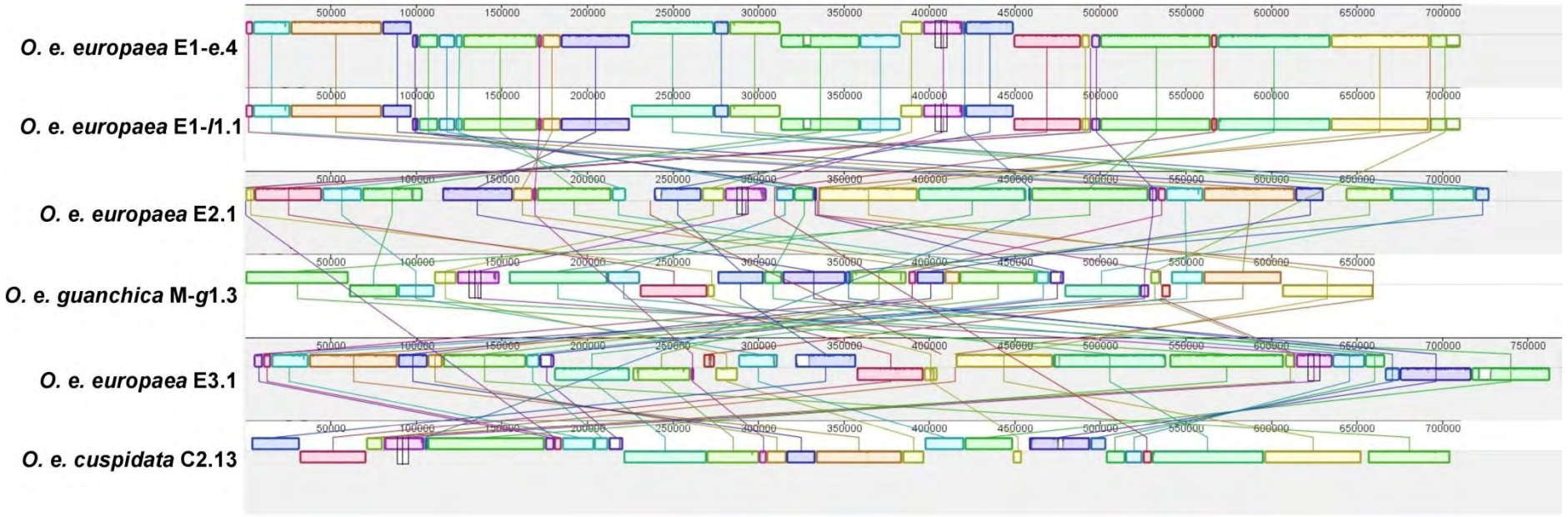


Fig. S3. Chimeric ORFs (*orf_E3a* and *orf_E3b*) specific to the E3.1 mitogenome. The number of bases of the known gene detected in the chimeric ORF is indicated. We represented the genes closest to each ORF, with the distance in bp between them.

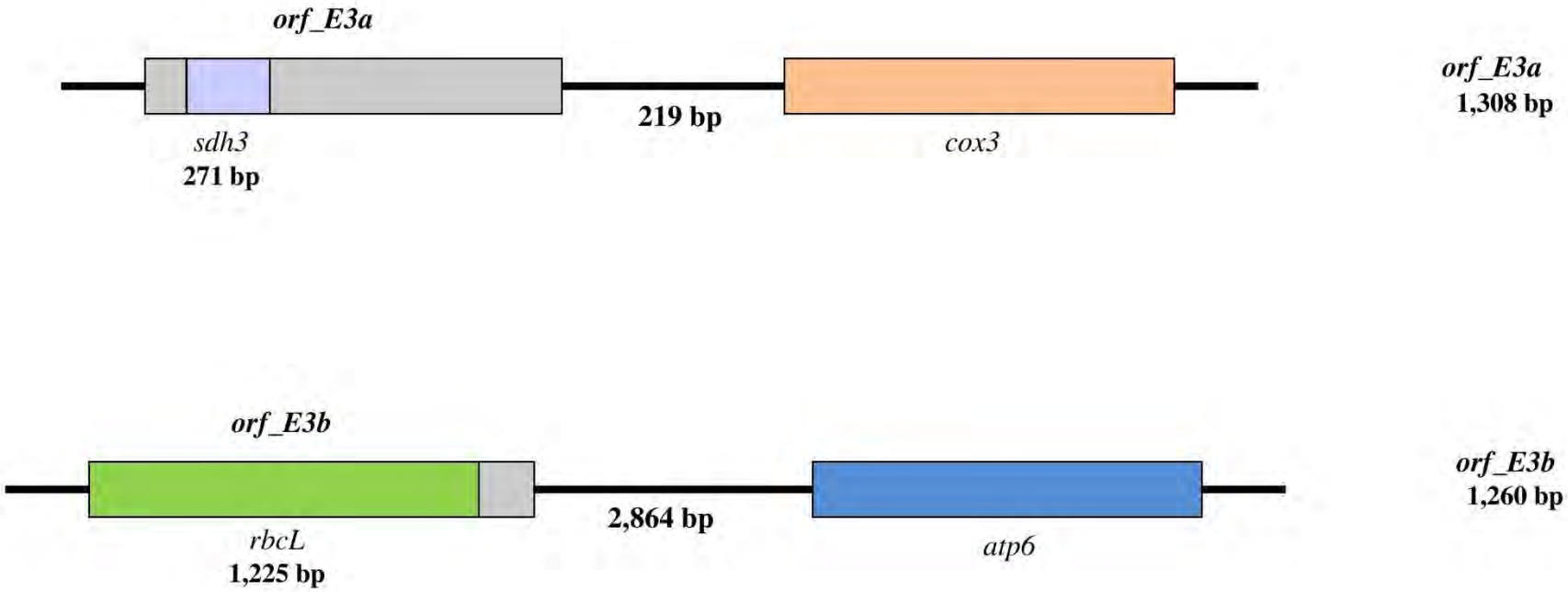


Fig. S4. ML phylogenetic trees reconstructed on nine *mtpt* segments and their homologous plastid regions extracted from plastomes of *Boea hygrometrica*, 16 accessions belonging to tribe *Oleeae*, and two accessions belonging to *Fontanesieae* and *Forsythieae*. The *mtpt* regions are represented in red and their corresponding homologous plastid sequences in green. Both phylogenetic trees were reconstructed with RAxML v.8 (Stamatakis 2014) using a GTR + G model, applying the rapid bootstrap algorithm with 1,000 iterations. We rooted the phylogenetic tree at midpoint.

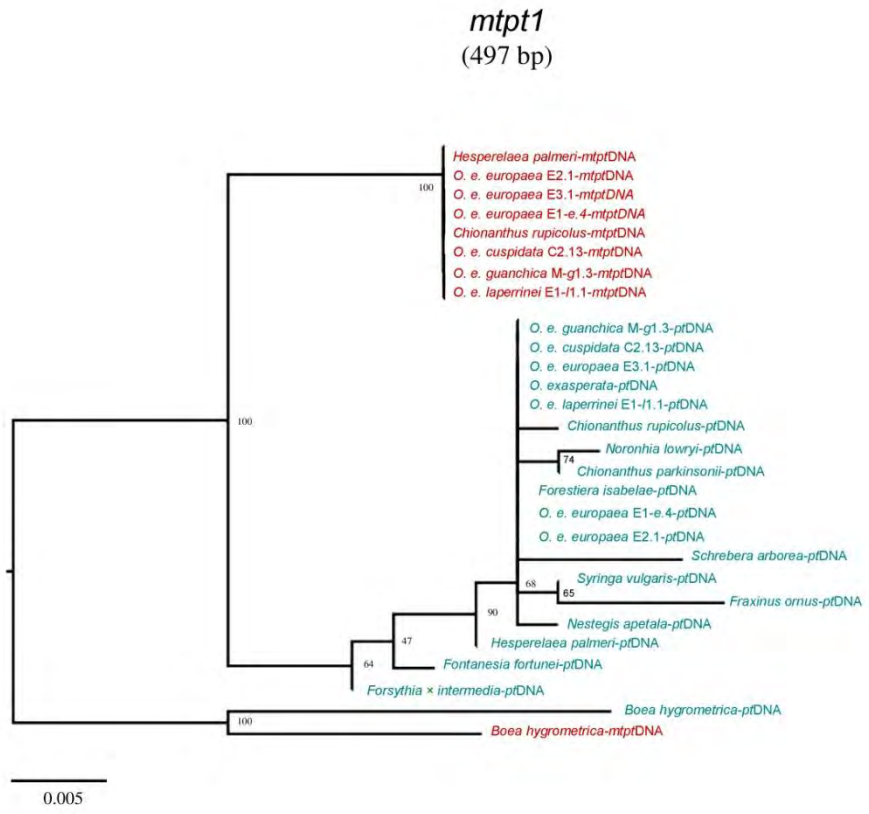
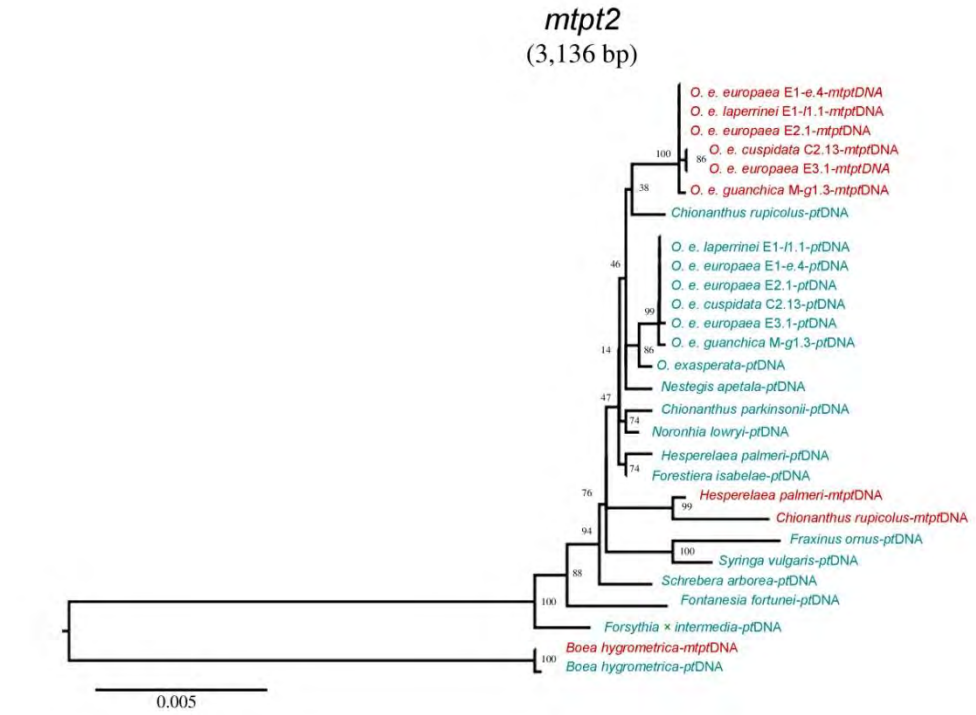


Fig. S4. Continued



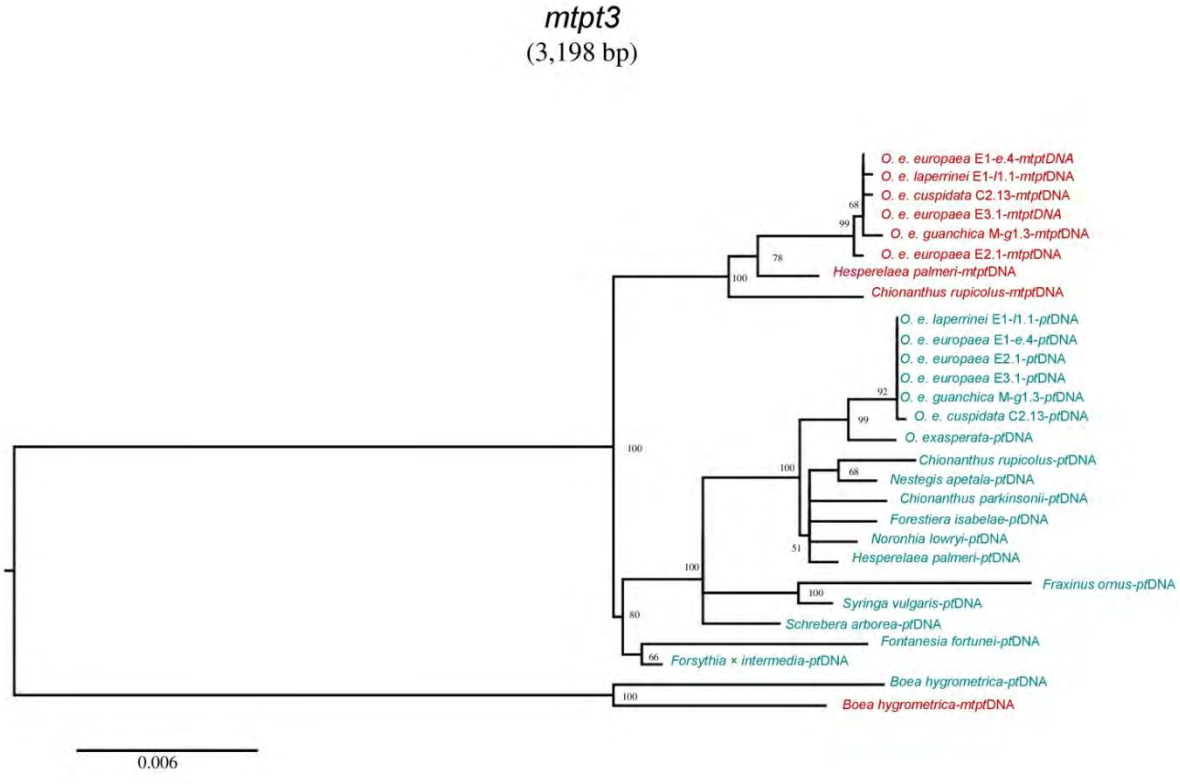


Fig. S4. Continued

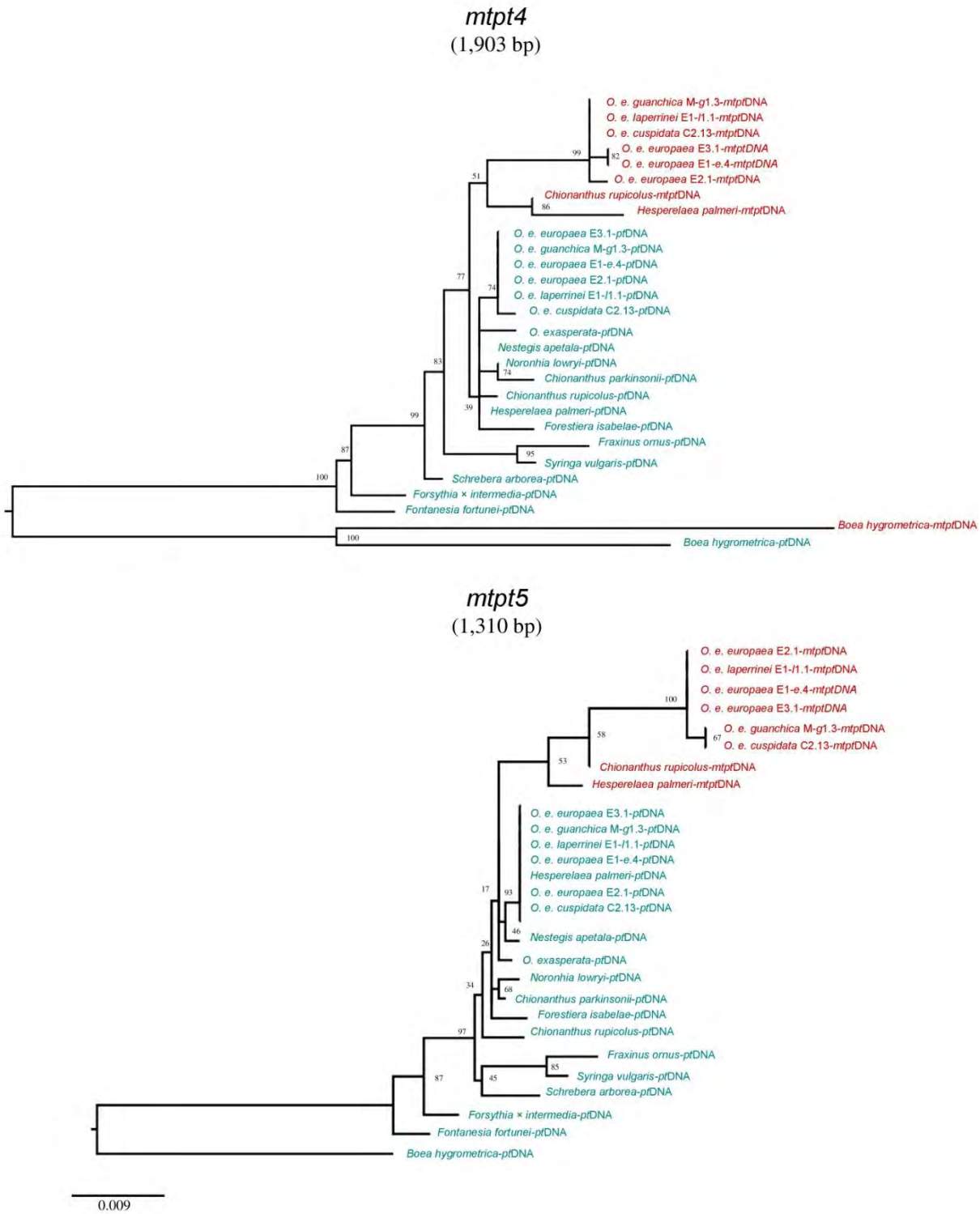


Fig. S4. Continued

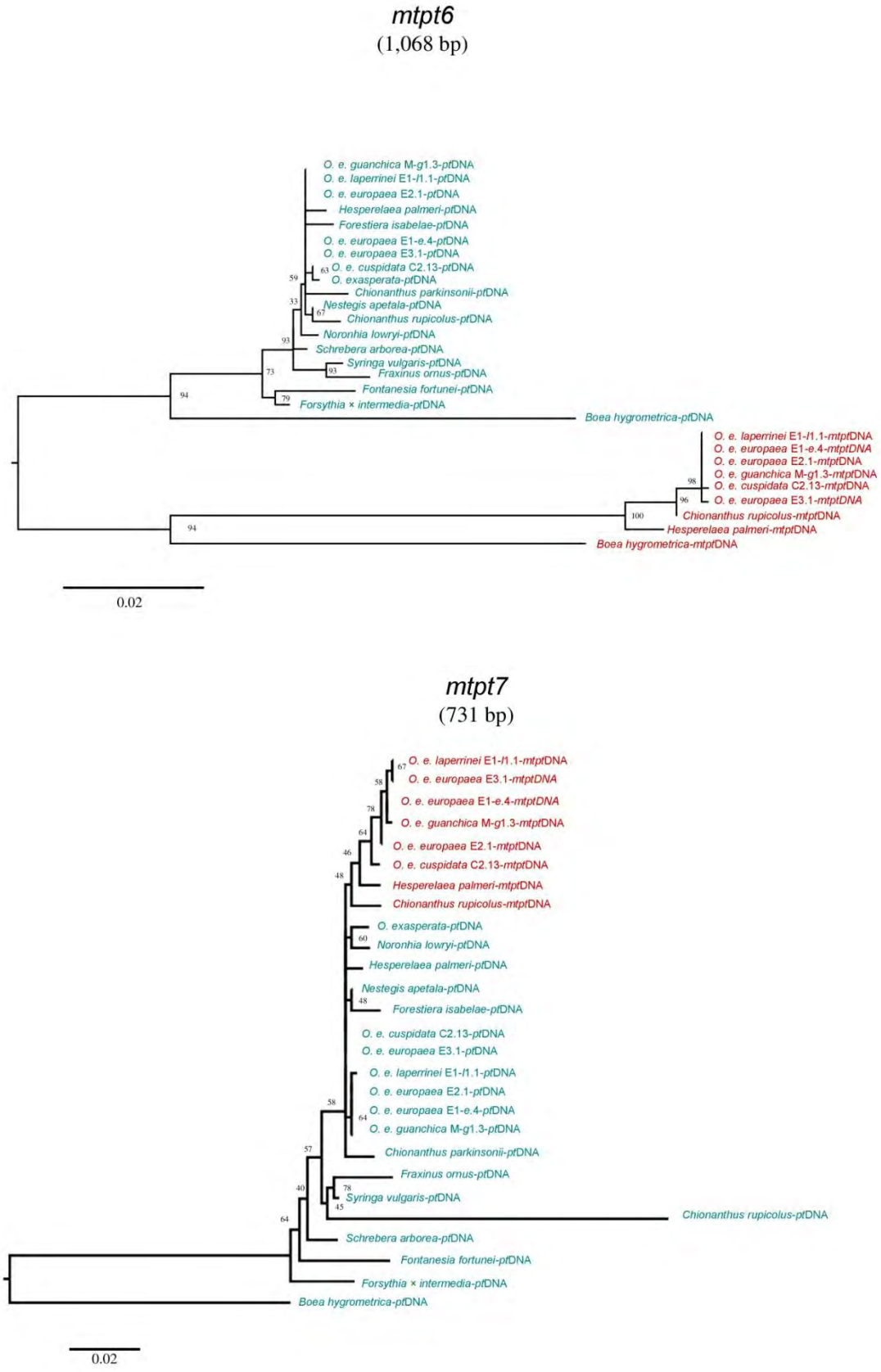


Fig. S4. End

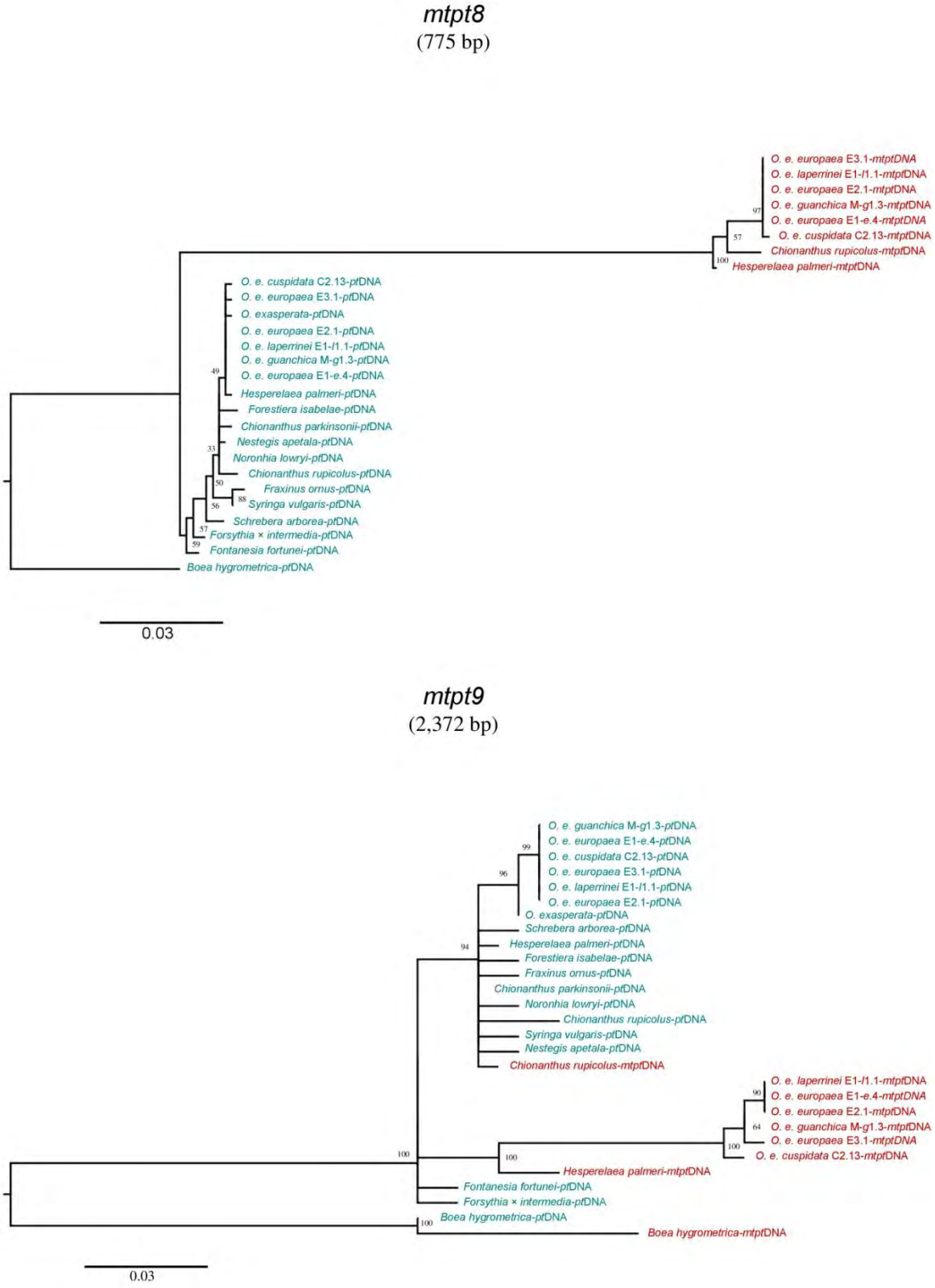
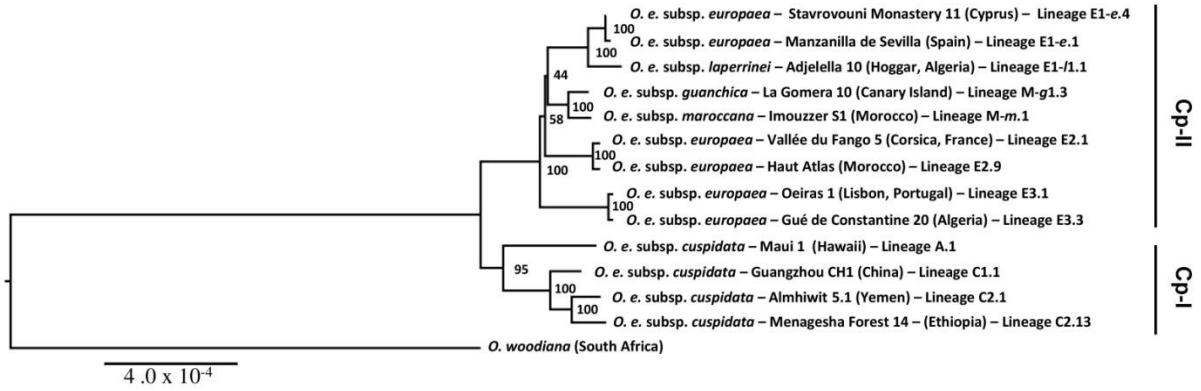

mtpt8
(775 bp)
O. e. europaea E3.1-mtptDNA
O. e. laperrinei E1-l1.1-mtptDNA
O. e. europaea E2.1-mtptDNA
O. e. guanchica M-g1.3-mtptDNA
O. e. europaea E1-e.4-mtptDNA
O. e. cuspidata C2.13-mtptDNA
Chionanthus rupicolus-mtptDNA
Hesperelaea palmeri-mtptDNA
O. e. cuspidata C2.13-ptDNA
O. e. europaea E3.1-ptDNA
O. exasperata-ptDNA
O. e. europaea E2.1-ptDNA
O. e. laperrinei E1-l1.1-ptDNA
O. e. guanchica M-g1.3-ptDNA
O. e. europaea E1-e.4-ptDNA
Hesperelaea palmeri-ptDNA
Forestiera isabelae-ptDNA
Chionanthus parkinsonii-ptDNA
Nestegis apetala-ptDNA
Noronhia lowryi-ptDNA
Chionanthus rupicolus-ptDNA
Fraxinus ornus-ptDNA
Syringa vulgaris-ptDNA
Schrebera arborea-ptDNA
Forsythia × intermedia-ptDNA
Fontanesia fortunei-ptDNA
Boea hygrometrica-ptDNA
0.03
mtpt9
(2,372 bp)
O. e. guanchica M-g1.3-ptDNA
O. e. europaea E1-e.4-ptDNA
O. e. cuspidata C2.13-ptDNA
O. e. europaea E3.1-ptDNA
O. e. laperrinei E1-l1.1-ptDNA
O. e. europaea E2.1-ptDNA
O. exasperata-ptDNA
Schrebera arborea-ptDNA
Hesperelaea palmeri-ptDNA
Forestiera isabelae-ptDNA
Fraxinus ornus-ptDNA
Chionanthus parkinsonii-ptDNA
Noronhia lowryi-ptDNA
Chionanthus rupicolus-ptDNA
Syringa vulgaris-ptDNA
Nestegis apetala-ptDNA
Chionanthus rupicolus-mtptDNA
O. e. laperrinei E1-l1.1-mtptDNA
O. e. europaea E1-e.4-mtptDNA
O. e. europaea E2.1-mtptDNA
O. e. guanchica M-g1.3-mtptDNA
O. e. europaea E3.1-mtptDNA
O. e. cuspidata C2.13-mtptDNA
Hesperelaea palmeri-mtptDNA
Fontanesia fortunei-ptDNA
Forsythia × intermedia-ptDNA
Boea hygrometrica-ptDNA
Boea hygrometrica-mtptDNA
0.03

Fig. S5. ML phylogenetic tree of the olive complex based on complete plastomes. We reconstructed a phylogenetic tree based on plastomes of the six newly sequenced accessions and eight *Olea* accessions from GenBank. *O. woodiana* is used as outgroup to root the tree. The phylogenetic tree was obtained using RAxML v.8 (Stamatakis 2014) with a GTR + G model. We applied the rapid bootstrap algorithm with 1,000 iterations.



References

Besnard G, Rubio de Casas R, Vargas P (2007) Plastid and nuclear DNA polymorphism reveals historical processes of isolation and reticulation in the olive tree complex (*Olea europaea* L.). *Journal of Biogeography*, **34**, 736–752.

Besnard G, Terral JF, Cornille A (2018) On the origins and domestication of the olive: a review and perspectives. *Annals of Botany*, **121**, doi:10.1093/aob/mcx145.

Darling AE, Mau B, Perna NT (2010) progressiveMauve: multiple genome alignment with gene gain, loss and rearrangement. *PLoS ONE*, **5**, e11147.

Krzywinski M et al, Schein J, Birol I *et al.* (2009) Circos: an information aesthetic for comparative genomics. *Genome Research*, **19**, 1639–1645.

Lohse M, Drechsel O, Bock R (2007) OrganellarGenomeDRAW (OGDRAW): a tool for the easy generation of high-quality custom graphical maps of plastid and mitochondrial genomes. *Current Genetics*, **52**, 267–274.

Nei M, Li WH (1979) Mathematical model for studying genetic variation in terms of restriction endonucleases. *Proceedings of the National Academy of Sciences of the United States of America*, **76**, 5269–5273.

Stamatakis A (2014) RAxML version 8: a tool for phylogenetic analysis and post-analysis of large phylogenies. *Bioinformatics*, **30**, 1312–1313.

ANNEXE III

Supplementary Material

Genomic consequences of occasional biparental transmission of organelles in two lineages of the olive family (Oleaceae)

VAN DE PAER C., BESNARD G.

In preparation

SI Appendix includes:

Table S1. List of species for which we excluded the mtDNA regions.

Table S2. Results of the strict molecular clock violation tests for cpDNA (Table S2.A) and mtDNA (Table S2.B) genes.

Table S3. Results of local molecular clock tests for cpDNA (Table S3.A) and mtDNA (Table S3.B) genes.

Table S4. Results of positive selection tests for cpDNA (Table S4.A) and mtDNA (Table S4.B) genes.

Fig. S1. Physical maps of the inverted repeats (IR1 and IR2) and the small single copy (SSC) of the plastomes

Fig. S2. Mean substitutions rates among cpDNA (Fig. S2.A) and mtDNA (Fig. S2.B) genes

Fig. S3. Mean substitutions rates among cpDNA (Fig. S3.A) and mtDNA (Fig. S3.B) functional categories

Table S1. List of species for which we excluded the mtDNA regions due to a too low depth coverage, which did not allow us to reconstructed the complete mtDNA sequence.

	atp1	*ccmFn*	*cox1*	*nad7*	*matR*	*rps3*	*rps4*	*rps10*	*rps12*	*sdh3*	*cox2 intron1*	*nad2 intron4*	*nad4 intron2*
Chionanthus macrobotrys	+	+	+	+	+	+	+	+	+	-	+	+	+
Fontanesia fortunei	+	+	+	+	+	+	+	+	+	+	+	+	-
Forestiera ligustrina	+	+	+	+	+	-	+	+	+	+	+	+	+
Forestiera pubescens var *parviflora*	+	+	+	+	+	+	+	+	-	+	+	+	+
Fraxinus excelsior	+	+	+	+	-	+	+	-	+	+	+	+	+
Haenianthus salicifolius	-	+	+	+	+	+	+	+	+	+	+	+	+
Olea capensis capensis	-	+	+	+	+	+	+	+	+	+	+	+	+
Olea europaea europaea E3	+	+	+	+	+	+	+	+	+	-	+	+	+
Olea javanica	+	-	+	+	+	+	-	+	+	+	+	+	+
Priogymnanthus hasslerianus	+	+	-	+	+	+	+	+	+	+	-	-	+
Schrebera arborea	+	+	+	-	+	+	+	+	+	+	+	+	-

Table S2. Results of the strict molecular clock violation tests for cpDNA (Table S2.A) and mtDNA (Table S2.B) genes. We performed these tests on Oleaceae accessions without clades *Jasmineae* and *LVP*, except for *Sy. vulgaris*. To determine the model which fitted better to data, we compared AIC between the model of strict molecular clock (noted *clock_1*) and the model of relaxed molecular clock (noted *clock_0*). We considered that models were not significantly different if the difference between their AIC was inferior to six (ΔAIC<6). For each gene, the model with the lowest AIC value was indicated in bold and highlighted in gray.

Table S2.A

gene	nb taxa	lnL clock0	lnL clock1	AIC clock0	AIC clock1	Interpretation
accD	91	-4150	-4206	8662	**8596**	strict molecular clock
atpA	91	-3042	-3098	6446	**6380**	strict molecular clock
atpB	91	-3059	-3122	6480	**6428**	strict molecular clock
atpE	91	-971	-1015	2304	**2214**	strict molecular clock
atpF	91	-1146	-1171	2654	**2526**	strict molecular clock
atpH	91	-433	-450	1228	**1084**	strict molecular clock
atpI	91	-1442	-1475	3246	**3134**	strict molecular clock
ccsA	91	-3524	-2954	7410	**6092**	strict molecular clock
cemA	91	-3524	-1855	7410	**3894**	strict molecular clock
clpP	91	-1271	-1304	2904	**2792**	strict molecular clock
infA	91	-484	-496	1330	**1176**	strict molecular clock
matK	91	-5058	-5138	10478	**10460**	strict molecular clock
ndhA	91	-2578	-2629	5518	**5442**	strict molecular clock
ndhB	91	-2249	-2263	4860	**4710**	strict molecular clock
ndhC	91	-813	-843	1988	**1870**	strict molecular clock
ndhD	91	-4300	-4354	8962	**8892**	strict molecular clock
ndhE	91	-712	-746	1786	**1676**	strict molecular clock
ndhF	91	-8250	-8330	16862	**16844**	strict molecular clock
ndhG	91	-1292	-1341	2946	**2866**	strict molecular clock
ndhH	91	-2904	-2959	6170	**6102**	strict molecular clock
ndhI	91	-1315	-1351	2992	**2886**	strict molecular clock
ndhJ	91	-929	-957	2220	**2098**	strict molecular clock
petA	91	-2180	-2221	4722	**4626**	strict molecular clock
petB	91	-85	-91	532	**366**	strict molecular clock
petD	91	-938	-963	2238	**2110**	strict molecular clock
petG	91	-194	-200	750	**584**	strict molecular clock
petL	91	-227	-244	816	**672**	strict molecular clock
petN	91	-158	-167	678	**518**	strict molecular clock
psaA	91	-4286	-4340	8934	**8864**	strict molecular clock
psaB	91	-4133	-4183	8628	**8550**	strict molecular clock
psaC	91	-485	-502	1332	**1188**	strict molecular clock
psaI	91	-211	-222	784	**628**	strict molecular clock
psaJ	91	-270	-283	902	**750**	strict molecular clock
psbA	91	-2039	-2077	4440	**4338**	strict molecular clock
psbB	91	-3071	-3117	6504	**6418**	strict molecular clock
psbC	91	-2741	-2783	5844	**5750**	strict molecular clock
psbD	91	-1942	-1980	4246	**4144**	strict molecular clock
psbE	91	-411	-424	1184	**1032**	strict molecular clock
psbF	91	-247	-265	856	**714**	strict molecular clock
psbG	91	-1446	-1488	3254	**3160**	strict molecular clock
psbH	91	-414	-432	1190	**1048**	strict molecular clock
psbI	91	-199	-207	760	**598**	strict molecular clock
psbJ	91	-232	-221	826	**626**	strict molecular clock
psbK	91	-426	-430	1214	**1044**	strict molecular clock
psbL	91	-205	-215	772	**614**	strict molecular clock
psbM	91	-192	-208	746	**600**	strict molecular clock
psbN	91	-245	-240	852	**664**	strict molecular clock
psbT	91	-238	-255	838	**694**	strict molecular clock
psbZ	91	-295	-303	952	**790**	strict molecular clock
rbcL	91	-4132	-4191	8626	**8566**	strict molecular clock
rpl2	91	-1246	-1260	2854	**2704**	strict molecular clock
rpl14	91	-973	-993	2308	**2170**	strict molecular clock
rpl16	91	-984	-1019	2330	**2222**	strict molecular clock
rpl20	91	-897	-928	2156	**2040**	strict molecular clock
rpl22	91	-1591	-1648	3544	**3480**	strict molecular clock
rpl23	91	-408	-414	1178	**1012**	strict molecular clock
rpl32	91	-770	-833	1902	**1850**	strict molecular clock
rpl33	91	-473	-499	1308	**1182**	strict molecular clock
rpl36	91	-219	-231	800	**646**	strict molecular clock
rpoA	91	-2787	-2853	5936	**5890**	strict molecular clock
rpoB	91	-6902	-6960	14166	**14104**	strict molecular clock
rpoC1	91	-4461	-4512	9284	**9208**	strict molecular clock
rpoC2	91	-10781	-10876	**21924**	21936	relaxed molecular clock
rps2	91	-1553	-1597	3468	**3378**	strict molecular clock
rps3	91	-1743	-1785	3848	**3754**	strict molecular clock
rps4	91	-1224	-1247	2810	**2678**	strict molecular clock
rps7	91	-827	-741	2016	**1666**	strict molecular clock
rps8	91	-950	-986	2262	**2156**	strict molecular clock
rps11	91	-959	-992	2280	**2168**	strict molecular clock
rps12	91	-710	-726	1782	**1636**	strict molecular clock
rps14	91	-645	-675	1652	**1534**	strict molecular clock
rps15	91	-697	-736	1756	**1656**	strict molecular clock
rps16	91	-583	-613	1528	**1410**	strict molecular clock
rps18	91	-606	-629	1574	**1442**	strict molecular clock
rps19	91	-724	-758	1810	**1700**	strict molecular clock
ycf1	91	-24917	-25074	**50196**	50332	relaxed molecular clock
ycf2	91	-10784	-10841	21930	**21866**	strict molecular clock
ycf3	91	-894	-917	2150	**2018**	strict molecular clock
ycf4	91	-1177	-1219	2716	**2622**	strict molecular clock
ycf15	91	-283	-289	928	**762**	strict molecular clock

Table S2.B

gene	nb taxa	lnL clock0	lnL clock1	AIC clock0	AIC clock1	Interpretation
atp1	89	-3219	-3335	**6792**	6850	relaxed molecular clock
atp4	91	-1114	-1154	2590	**2490**	strict molecular clock
atp6	91	-2356	-2568	**5074**	5318	relaxed molecular clock
atp8	91	-1032	-1070	2426	**2322**	strict molecular clock
atp9	91	-486	-477	1334	**1136**	strict molecular clock
ccmB	91	-943	-959	2248	**2100**	strict molecular clock
ccmC	91	-1192	-1218	2746	**2618**	strict molecular clock
ccmFc	91	-2499	-2586	5360	**5354**	strict molecular clock
ccmFn	90	-2808	-2841	5974	**5864**	strict molecular clock
cox1	90	-2463	-2484	5284	**5150**	strict molecular clock
cox2	91	-1503	-1549	3368	**3280**	strict molecular clock
cox3	91	-1313	-1353	2988	**2888**	strict molecular clock
cytB	91	-2158	-2198	4678	**4578**	strict molecular clock
matR	90	-3376	-3414	7110	**7010**	strict molecular clock
mttB	91	-1523	-1597	3408	**3376**	strict molecular clock
nad1	91	-1511	-1534	3384	**3250**	strict molecular clock
nad2	91	-2187	-2217	4736	**4616**	strict molecular clock
nad3	91	-722	-770	1806	**1722**	strict molecular clock
nad4	91	-2345	-2389	5052	**4960**	strict molecular clock
nad4L	91	-522	-490	1406	**1162**	strict molecular clock
nad5	91	-3284	-3351	6930	**6884**	strict molecular clock
nad6	91	-1141	-1178	2644	**2538**	strict molecular clock
nad7	90	-2045	-1918	4448	**4018**	strict molecular clock
nad9	91	-1073	-1109	2508	**2400**	strict molecular clock
rpl5	91	-936	-960	2234	**2102**	strict molecular clock
rpl10	91	-834	-861	2030	**1904**	strict molecular clock
rpl16	91	-819	-841	2000	**1864**	strict molecular clock
rps3	90	-3091	-3159	6540	**6500**	strict molecular clock
rps4	90	-2082	-2132	4522	**4446**	strict molecular clock
rps7	91	-753	-775	1868	**1732**	strict molecular clock
rps10	90	-969	-1004	2296	**2190**	strict molecular clock
rps12	90	-719	-740	1796	**1662**	strict molecular clock
rps13	91	-776	-827	1914	**1836**	strict molecular clock
rps14	91	-480	-494	1322	**1170**	strict molecular clock
sdh3	89	-642	-667	1638	**1514**	strict molecular clock
sdh4	91	-1557	-1641	3476	**3464**	strict molecular clock

Table S3. Results of local molecular clock tests for cpDNA (Table S3.A) and mtDNA (Table S3.B) genes. We performed these tests on all 100 Oleaceae accessions. To determine the model which fitted better to data, we compared AIC between the model of strict molecular clock (noted *clock_1*), the model of relaxed molecular clock (noted *clock_0*) and the three models of local molecular clock which assume a distinct evolutionary rate (i) in both clades *Jasmineae* and *LVP* (*clock2_Jasmineae_ LVP*), (ii) in clade *Jasmineae* only (*clock2_Jasmineae*), (iii) in clade *LVP* only (*clock2_ LVP*). For each gene, models with the lowest AIC value was indicated in bold and highlighted in gray. We considered that models were not significantly different if the difference between their AIC was inferior to six (ΔAIC<6).

Table S3.A

gene	lnL clock0	lnL clock1	lnL *clock2_Jasmineae_LVP*	lnL *clock2_Jasmineae*	lnL *clock2_LVP*	AIC *clock0*	AIC *clock1*	AIC *clock2_Jasmineae_LVP*	AIC *clock2_Jasmineae*	AIC *clock2_LVP*	Interpretation
accD	-12143	-12635	-	-	-12219	24668	25464	-	-	**24634**	*local molecular clock (LVP)*
atpA	-3628	-3733	-3699	-3706	-3727	7654	7668	**7604**	7616	7658	*local molecular clock (Jasmineae & LVP)*
atpB	-3518	-3606	-3588	-3593	-3601	7434	7414	**7382**	7390	7406	*local molecular clock (Jasmineae & LVP)*
atpE	-1104	-1161	-1149	-1150	-1160	2606	2524	**2504**	**2504**	2524	local molecular clock (at least: *Jasmineae*)
atpF	-1344	-1387	-1378	-1375	-1382	3086	2976	2962	**2954**	2968	local molecular clock (*Jasmineae*)
atpH	-464	-486	-486	-486	-486	1326	**1174**	**1178**	**1176**	**1176**	*n.s.*
atpI	-1618	-1668	-1663	-1664	-1667	3634	3538	**3532**	**3532**	3538	local molecular clock (at least: *Jasmineae*)
ccsA	-3524	-3628	-3591	-3598	-3624	7446	7458	**7388**	7400	7452	*local molecular clock (Jasmineae & LVP)*
cemA	-2257	-2364	-2335	-2394	-2340	4912	4930	**4876**	4992	4884	*local molecular clock (Jasmineae & LVP)*
clpP	-4435	-4985	-4526	-4770	-4859	9268	10172	**9258**	9744	9922	*local molecular clock (Jasmineae & LVP)*
infA	-1002	-1078	-1026	-1032	-1074	2402	2358	**2258**	2270	2352	*local molecular clock (Jasmineae & LVP)*
matK	-5058	-6051	-6015	-6020	-6048	**10514**	12304	12236	12244	12300	relaxed molecular clock
ndhA	-3010	-3081	-3068	-3069	-3094	6418	6364	**6342**	**6342**	6392	local molecular clock (at least: *Jasmineae*)
ndhB	-2439	-2471	-2458	-3069	-2471	5276	5144	**5122**	6342	5146	*local molecular clock (Jasmineae & LVP)*
ndhC	-946	-994	-984	-986	-993	2290	2190	**2174**	**2176**	2190	local molecular clock (at least: *Jasmineae*)
ndhD	-5086	-5188	-5149	-5180	-5178	10570	10578	**10504**	10564	10560	*local molecular clock (Jasmineae & LVP)*
ndhE	-874	-871	-864	-864	-870	2146	1944	**1934**	**1932**	1944	local molecular clock (at least: *Jasmineae*)
ndhF	-10119	-10314	-10235	-10258	-10298	**20636**	20830	20676	20720	20800	relaxed molecular clock
ndhG	-1509	-1573	-1564	-1566	-1577	3416	3348	**3334**	**3336**	3358	local molecular clock (at least: *Jasmineae*)
ndhH	-3239	-3313	-3302	-3298	-3308	6876	6828	6810	**6800**	6820	local molecular clock (*Jasmineae*)
ndhI	-1468	-1508	-1509	-1506	-1508	3334	**3218**	3224	**3216**	3220	*n.s.*
ndhJ	-1046	-1092	-1084	-1085	-1091	2490	2386	**2374**	**2374**	2386	local molecular clock (at least: *Jasmineae*)
petA	-2476	-2536	-2524	-2534	-2534	5350	5274	**5254**	5272	5272	*local molecular clock (Jasmineae & LVP)*
petB	-84	-91	-91	-91	-91	566	**384**	**388**	**386**	**386**	*n.s.*
petD	-1009	-1039	-1038	-1038	-1039	2416	**2280**	**2282**	**2280**	**2282**	*n.s.*
petG	-207	-215	-213	-213	-215	812	**632**	**632**	**630**	**634**	*n.s.*
petL	-237	-255	-260	-255	-255	872	**712**	726	**714**	**714**	*n.s.*
petN	-175	-187	-187	-187	-189	748	**576**	**580**	**578**	582	*n.s.*
psaA	-4678	-4745	-4735	-4738	-4743	9754	9692	**9676**	**9680**	9690	local molecular clock (at least: *Jasmineae*)
psaB	-4592	-4661	-4646	-4647	-4660	9582	9524	**9498**	**9498**	9524	local molecular clock (at least: *Jasmineae*)
psaC	-536	-559	-557	-557	-559	1470	**1320**	**1320**	**1318**	**1322**	*n.s.*
psaI	-233	-247	-246	-246	-247	864	**696**	**698**	**696**	**698**	*n.s.*
psaJ	-312	-332	-326	-326	-331	1022	866	**858**	**856**	866	local molecular clock (at least: *Jasmineae*)
psbA	-2303	-2357	-2353	-2355	-2354	5004	**4916**	**4912**	**4914**	**4912**	*n.s.*
psbB	-3417	-3490	-3474	-3475	-3489	7232	7182	**7154**	**7154**	7182	local molecular clock (at least: *Jasmineae*)
psbC	-2741	-3075	-3069	-3069	-3075	**5880**	6352	6344	6342	6354	relaxed molecular clock
psbD	-2095	-2140	-2137	-2140	-2138	4588	**4482**	**4480**	**4484**	**4480**	*n.s.*
psbE	-448	-467	-465	-466	-466	1294	**1136**	**1136**	**1136**	**1136**	*n.s.*
psbF	-251	-270	-270	-270	-270	900	**742**	**746**	**744**	**744**	*n.s.*
psbG	-1560	-1613	-1605	-1609	-1611	3518	3428	**3416**	3422	3426	*local molecular clock (Jasmineae & LVP)*
psbH	-476	-500	-493	-493	-500	1350	1202	**1192**	**1190**	1204	local molecular clock (at least: *Jasmineae*)
psbI	-208	-217	-217	-217	-217	814	**636**	**640**	**638**	**638**	*n.s.*
psbJ	-221	-234	-233	-233	-234	840	**670**	**672**	**670**	**672**	*n.s.*
psbK	-488	-502	-498	-499	-502	1374	**1206**	**1202**	**1202**	**1208**	*n.s.*
psbL	-231	-228	-227	-227	-228	860	**658**	**660**	**658**	**660**	*n.s.*
psbM	-223	-226	-225	-225	-226	844	**654**	**656**	**654**	**656**	*n.s.*
psbN	-261	-255	-255	-255	-255	920	**712**	**716**	**714**	**714**	*n.s.*
psbT	-256	-275	-275	-275	-275	910	**752**	**756**	**754**	**754**	*n.s.*
psbZ	-336	-351	-349	-350	-352	1070	**904**	**904**	**904**	**908**	*n.s.*
rbcL	-4678	-4769	-4748	-4750	-4766	9754	9740	**9702**	**9704**	9736	local molecular clock (at least: *Jasmineae*)
rpl2	-1479	-1528	-1500	-1506	-1524	3356	3258	**3206**	3216	3252	*local molecular clock (Jasmineae & LVP)*
rpl14	-1179	-1220	-1202	-1204	-1219	2756	2642	**2610**	**2612**	2642	local molecular clock (at least: *Jasmineae*)
rpl16	-1180	-1228	-1218	-1218	-1235	2758	2658	**2642**	**2640**	2674	local molecular clock (at least: *Jasmineae*)
rpl20	-1454	-1538	-1487	-1501	-1529	3306	3278	**3180**	3206	3262	*local molecular clock (Jasmineae & LVP)*
rpl22	-1860	-1932	-1912	-1913	-1931	4118	4066	**4030**	**4030**	4066	local molecular clock (at least: *Jasmineae*)
rpl23	-525	-539	-534	-535	-538	1448	1280	**1274**	**1274**	1280	local molecular clock (at least: *Jasmineae*)
rpl32	-1053	-1122	-1115	-1117	-1121	2504	2446	**2436**	**2438**	2446	local molecular clock (at least: *Jasmineae*)
rpl33	-610	-645	-635	-635	-645	1618	1492	**1476**	**1474**	1494	local molecular clock (at least: *Jasmineae*)
rpl36	-283	-307	-299	-300	-306	964	816	**804**	**804**	816	local molecular clock (at least: *Jasmineae*)
rpoA	-3333	-3425	-3396	-3403	-3420	7064	7052	**6998**	7010	7044	*local molecular clock (Jasmineae & LVP)*
rpoB	-8248	-8393	-8310	-8340	-8381	16894	16988	**16826**	16884	16966	*local molecular clock (Jasmineae & LVP)*
rpoC1	-5527	-5663	-5582	-5590	-5657	11452	11528	**11370**	11384	11518	*local molecular clock (Jasmineae & LVP)*
rpoC2	-12991	-13264	-13081	-13102	-13176	26380	26730	**26368**	26408	26556	*local molecular clock (Jasmineae & LVP)*
rps2	-2009	-2112	-2055	-2064	-2106	4416	4426	**4316**	4332	4416	*local molecular clock (Jasmineae & LVP)*
rps3	-2418	-2520	-2469	-2478	-2515	5234	5242	**5144**	5160	5234	*local molecular clock (Jasmineae & LVP)*
rps4	-1503	-1559	-1535	-1537	-1566	3404	3320	**3276**	**3278**	3336	local molecular clock (at least: *Jasmineae*)
rps7	-943	-1002	-968	-975	-1000	2284	2206	**2142**	2154	2204	*local molecular clock (Jasmineae & LVP)*
rps8	-1225	-1296	-1265	-1267	-1296	2848	2794	**2736**	**2738**	2796	local molecular clock (at least: *Jasmineae*)
rps11	-1436	-1543	-1497	-1483	-1541	3270	3288	3200	**3170**	3286	local molecular clock (*Jasmineae*)
rps12	-868	-911	-889	-891	-910	2134	2024	**1984**	**1988**	2024	local molecular clock (at least: *Jasmineae*)
rps14	-713	-767	-746	-756	-767	1824	**1736**	1698	1716	**1738**	*n.s.*
rps15	-918	-979	-962	-962	-989	2234	2160	**2130**	**2128**	2182	local molecular clock (at least: *Jasmineae*)
rps16	-684	-727	-715	-716	-726	1766	1656	**1636**	**1636**	1656	local molecular clock (at least: *Jasmineae*)
rps18	-1454	-1551	-1480	-1484	-1547	3306	3304	**3166**	3172	3298	*local molecular clock (Jasmineae & LVP)*
rps19	-1222	-1308	-1262	-1266	-1305	2842	2818	**2730**	2736	2814	*local molecular clock (Jasmineae & LVP)*
ycf1	-38150	-39353	-38346	-38653	-39143	76698	78908	**76898**	77510	78490	*local molecular clock (Jasmineae & LVP)*
ycf2	-18174	-19132	-18294	-18568	-18880	**36746**	38466	36794	37340	37964	relaxed molecular clock
ycf3	-988	-1019	-1014	-1014	-1019	2374	2240	**2234**	**2232**	2242	local molecular clock (at least: *Jasmineae*)
ycf4	-1796	-1969	-1878	-1933	-1894	3990	4140	**3962**	4070	3992	*local molecular clock (Jasmineae & LVP)*
ycf15	-542	-557	-551	-558	-553	1482	1316	**1308**	1320	**1310**	*local molecular clock (at least: LVP)*

Table S3.B

gene	lnL *clock0*	lnL *clock1*	lnL *clock2_Jasmineae_LVP*	lnL *clock2_Jasmineae*	lnL *clock2_LVP*	AIC *clock0*	AIC *clock1*	AIC *clock2_Jasmineae_LVP*	AIC *clock2_Jasmineae*	AIC *clock2_LVP*	Interpretation
atp1	-3436	-3583	-3576	-3583	-3576	**7262**	7364	7354	7366	7352	relaxed molecular clock
atp4	-1174	-1206	-1204	-1206	-1204	2746	**2614**	**2614**	**2616**	**2612**	*n.s.*
atp6	-2426	-2588	-2584	-2588	-2584	**5250**	5378	5374	5380	5372	relaxed molecular clock
atp8	-1139	-1123	-1121	-1121	-1123	2676	**2448**	**2448**	**2446**	**2450**	*n.s.*
atp9	-465	-478	-478	-478	-478	1328	**1158**	**1162**	**1160**	**1160**	*n.s.*
ccmB	-943	-956	-954	-955	-955	2284	**2114**	**2114**	**2114**	**2114**	*n.s.*
ccmC	-1200	-1218	-1218	-1218	-1218	2798	**2638**	**2642**	**2640**	**2640**	*n.s.*
ccmFc	-2623	-2695	-2694	-2694	-2695	5644	**5592**	**5594**	**5592**	**5594**	*n.s.*
ccmFn	-2874	-2908	-2907	-2907	-2907	6142	**6016**	**6018**	**6016**	**6016**	*n.s.*
cox1	-3942	-3978	-3965	-3967	-3977	8278	8156	**8134**	**8136**	8156	local molecular clock (at least: *Jasmineae*)
cox2	-1586	-1632	-1630	-1630	-1631	3570	**3466**	**3466**	**3464**	**3466**	*n.s.*
cox3	-1314	-1345	-1344	-1344	-1345	3026	**2892**	**2894**	**2892**	**2894**	*n.s.*
cytB	-2287	-2329	-2323	-2324	-2328	4972	4860	**4852**	**4852**	4860	local molecular clock (at least: *Jasmineae*)
matR	-3461	-3502	-3501	-3501	-3502	7316	**7204**	**7206**	**7204**	**7206**	*n.s.*
mttB	-1752	-1814	-1788	-1792	-1812	3902	3830	**3782**	3788	3828	*local molecular clock (Jasmineae & LVP)*
nad1	-1534	-1555	-1555	-1555	-1555	3466	**3312**	**3316**	**3314**	**3314**	*n.s.*
nad2	-2435	-2250	-2250	-2250	-2250	5268	**4702**	**4706**	**4704**	**4704**	*n.s.*
nad3	-748	-786	-786	-786	-786	1894	**1774**	**1778**	**1776**	**1776**	*n.s.*
nad4	-2363	-2404	-2403	-2403	-2404	5124	**5010**	**5012**	**5010**	**5012**	*n.s.*
nad4L	-479	-492	-492	-492	-492	1356	**1186**	**1190**	**1188**	**1188**	*n.s.*
nad5	-3402	-3465	-3462	-3462	-3465	7202	**7132**	**7130**	**7128**	7134	*n.s.*
nad6	-1162	-1197	-1196	-1196	-1196	2722	**2596**	**2598**	**2596**	**2596**	*n.s.*
nad7	-1942	-1972	-1969	-1969	-1971	4278	**4144**	**4142**	**4140**	**4144**	*n.s.*
nad9	-1088	-1119	-1119	-1119	-1119	2574	**2440**	**2444**	**2442**	**2442**	*n.s.*
rpl5	-983	-1012	-1010	-1010	-1012	2364	**2226**	**2226**	**2224**	**2228**	*n.s.*
rpl10	-868	-886	-885	-886	-885	2134	**1974**	**1976**	**1976**	**1974**	*n.s.*
rpl16	-934	-847	-847	-847	-847	2266	**1896**	**1900**	**1898**	**1898**	*n.s.*
rps3	-3239	-3316	-3309	-3309	-3315	6872	6832	**6822**	**6820**	6832	local molecular clock (at least: *Jasmineae*)
rps4	-2142	-2200	-2200	-2200	-2200	4678	**4600**	**4604**	**4602**	**4602**	*n.s.*
rps7	-761	-778	-778	-778	-778	1920	**1758**	**1762**	**1760**	**1760**	*n.s.*
rps10	-987	-1025	-1025	-1025	-1025	2368	**2250**	**2254**	**2252**	**2252**	*n.s.*
rps12	-749	-773	-772	-772	-772	1892	**1746**	**1748**	**1746**	**1746**	*n.s.*
rps13	-838	-885	-882	-882	-885	2074	**1972**	**1970**	**1968**	1974	*n.s.*
rps14	-555	-501	-501	-501	-501	1508	**1204**	**1208**	**1206**	**1206**	*n.s.*
sdh3	-676	-704	-702	-704	-702	1742	**1606**	**1606**	**1608**	**1604**	*n.s.*
sdh4	-1600	-1650	-1649	-1650	-1649	3598	**3502**	**3504**	**3504**	**3502**	*n.s.*

Table S4. Results of positive selection tests for cpDNA (Table S4.A) and mtDNA (Table S4.B) genes, using PAML v.4 (Yang 2007). For each clade independently, we compared AIC values of the model M1a with both models A and A' under the hypothesis H1 (assuming a potential positive selection on the internal branch of the clade) and the hypothesis H2 (assuming a potential positive selection on all branches of the clade). For each gene, models with the lowest AIC value is indicated in bold. We considered that models were non-significantly different if the difference between their AIC was inferior to six (ΔAIC<6). We highlighted in blue for *Jasmineae* and in gray for *LVP* the models which were not significantly different from the model with the lowest AIC value.

Table S4.A

gene	clade	hypothesis	model	npar	lnL	AIC
accD	*LVP*	H1	A	194	-11693	23774
accD	*LVP*	H1	A'	193	-11693	23772
accD	*LVP*	H2	A	194	-11693	23774
accD	*LVP*	H2	A'	193	-11693	23772
accD	-	-	M1a	192	-11693	**23770**
atpA	*Jasmineae*	H1	A	192	-3272	6928
atpA	*Jasmineae*	H1	A'	191	-3272	6926
atpA	*Jasmineae*	H2	A	192	-3268	**6920**
atpA	*Jasmineae*	H2	A'	191	-3269	**6920**
atpA	*LVP*	H1	A	195	-3165	6720
atpA	*LVP*	H1	A'	194	-3165	**6718**
atpA	*LVP*	H2	A	195	-3165	6720
atpA	*LVP*	H2	A'	194	-3165	**6718**
atpA	-	-	M1a	200	-3515	7430
atpB	*Jasmineae*	H1	A	192	-3208	6800
atpB	*Jasmineae*	H1	A'	191	-3210	6802
atpB	*Jasmineae*	H2	A	192	-3208	6800
atpB	*Jasmineae*	H2	A'	191	-3208	**6798**
atpB	*LVP*	H1	A	195	-3149	6688
atpB	*LVP*	H1	A'	194	-3149	**6686**
atpB	*LVP*	H2	A	195	-3149	6688
atpB	*LVP*	H2	A'	194	-3149	**6686**
atpB	-	-	M1a	200	-3408	7216
atpE	*Jasmineae*	H1	A	192	-1022	2428
atpE	*Jasmineae*	H1	A'	191	-1022	**2426**
atpE	*Jasmineae*	H2	A	192	-1022	2428
atpE	*Jasmineae*	H2	A'	191	-1022	**2426**
atpE	*LVP*	H1	A	195	-1003	2396
atpE	*LVP*	H1	A'	194	-1003	**2394**
atpE	*LVP*	H2	A	195	-1003	2396
atpE	*LVP*	H2	A'	194	-1003	**2394**
atpE	-	-	M1a	200	-1080	2560
atpF	*Jasmineae*	H1	A	192	-1241	**2866**
atpF	*Jasmineae*	H1	A'	191	-1246	2874
atpF	*Jasmineae*	H2	A	192	-1246	2876
atpF	*Jasmineae*	H2	A'	191	-1246	2874
atpF	*LVP*	H1	A	195	-1214	2818
atpF	*LVP*	H1	A'	194	-1214	**2816**
atpF	*LVP*	H2	A	195	-1214	2818
atpF	*LVP*	H2	A'	194	-1214	**2816**
atpF	-	-	M1a	200	-1326	3052
atpH	*Jasmineae*	H1	A	192	-409	1202
atpH	*Jasmineae*	H1	A'	191	-409	1200
atpH	*Jasmineae*	H2	A	192	-408	1200
atpH	*Jasmineae*	H2	A'	191	-408	**1198**
atpH	*LVP*	H1	A	195	-404	1198
atpH	*LVP*	H1	A'	194	-404	**1196**
atpH	*LVP*	H2	A	195	-404	1198
atpH	*LVP*	H2	A'	194	-404	**1196**
atpH	-	-	M1a	200	-423	1246
atpI	*Jasmineae*	H1	A	192	-1500	3384
atpI	*Jasmineae*	H1	A'	191	-1500	**3382**
atpI	*Jasmineae*	H2	A	192	-1500	3384
atpI	*Jasmineae*	H2	A'	191	-1500	**3382**
atpI	*LVP*	H1	A	195	-1461	3312
atpI	*LVP*	H1	A'	194	-1461	**3310**
atpI	*LVP*	H2	A	195	-1461	3312
atpI	*LVP*	H2	A'	194	-1461	**3310**
atpI	-	-	M1a	200	-1569	3538
ccsA	*Jasmineae*	H1	A	192	-3293	6970
ccsA	*Jasmineae*	H1	A'	191	-3293	**6968**
ccsA	*Jasmineae*	H2	A	192	-3293	6970
ccsA	*Jasmineae*	H2	A'	191	-3293	**6968**
ccsA	*LVP*	H1	A	195	-3094	6578
ccsA	*LVP*	H1	A'	194	-3094	**6576**
ccsA	*LVP*	H2	A	195	-3094	6578
ccsA	*LVP*	H2	A'	194	-3094	**6576**
ccsA	-	-	M1a	200	-3523	7446

Table S4.A Continued

gene	clade	hypothesis	model	npar	lnL	AIC
cemA	*Jasmineae*	H1	A	192	-1906	**4196**
cemA	*Jasmineae*	H1	A'	191	-1910	4202
cemA	*Jasmineae*	H2	A	192	-1907	4198
cemA	*Jasmineae*	H2	A'	191	-1908	4198
cemA	*LVP*	H1	A	195	-2106	4602
cemA	*LVP*	H1	A'	194	-2106	**4600**
cemA	*LVP*	H2	A	195	-2106	4602
cemA	*LVP*	H2	A'	194	-2106	**4600**
cemA	-	-	M1a	200	-2235	4870
clpP	*Jasmineae*	H1	A	192	-2518	5420
clpP	*Jasmineae*	H1	A'	191	-2545	5472
clpP	*Jasmineae*	H2	A	192	-2487	**5358**
clpP	*Jasmineae*	H2	A'	191	-2522	5426
clpP	*LVP*	H1	A	195	-3106	6602
clpP	*LVP*	H1	A'	194	-3106	**6600**
clpP	*LVP*	H2	A	195	-3106	6602
clpP	*LVP*	H2	A'	194	-3106	**6600**
clpP	-	-	M1a	200	-4384	9168
infA	*Jasmineae*	H1	A	192	-884	2152
infA	*Jasmineae*	H1	A'	191	-887	2156
infA	*Jasmineae*	H2	A	192	-873	**2130**
infA	*Jasmineae*	H2	A'	191	-874	**2130**
infA	*LVP*	H1	A	195	-582	1554
infA	*LVP*	H1	A'	194	-582	**1552**
infA	*LVP*	H2	A	195	-582	1554
infA	*LVP*	H2	A'	194	-582	**1552**
infA	-	-	M1a	200	-1007	2414
matK	*Jasmineae*	H1	A	192	-5500	11384
matK	*Jasmineae*	H1	A'	191	-5500	11382
matK	*Jasmineae*	H2	A	192	-5495	11374
matK	*Jasmineae*	H2	A'	191	-5495	**11372**
matK	*LVP*	H1	A	195	-5258	**10906**
matK	*LVP*	H1	A'	194	-5268	10924
matK	*LVP*	H2	A	195	-5268	10926
matK	*LVP*	H2	A'	194	-5268	10924
matK	-	-	M1a	200	-5830	12060
ndhA	*Jasmineae*	H1	A	192	-2798	5980
ndhA	*Jasmineae*	H1	A'	191	-2799	5980
ndhA	*Jasmineae*	H2	A	192	-2795	5974
ndhA	*Jasmineae*	H2	A'	191	-2795	**5972**
ndhA	*LVP*	H1	A	195	-2633	5656
ndhA	*LVP*	H1	A'	194	-2633	**5654**
ndhA	*LVP*	H2	A	195	-2633	5656
ndhA	*LVP*	H2	A'	194	-2633	**5654**
ndhA	-	-	M1a	200	-2934	6268
ndhB	*Jasmineae*	H1	A	192	-2332	5048
ndhB	*Jasmineae*	H1	A'	191	-2332	**5046**
ndhB	*Jasmineae*	H2	A	192	-2333	5050
ndhB	*Jasmineae*	H2	A'	191	-2333	5048
ndhB	*LVP*	H1	A	195	-2253	4896
ndhB	*LVP*	H1	A'	194	-2253	**4894**
ndhB	*LVP*	H2	A	195	-2253	4896
ndhB	*LVP*	H2	A'	194	-2253	**4894**
ndhB	-	-	M1a	200	-2388	5176
ndhC	*Jasmineae*	H1	A	192	-886	2156
ndhC	*Jasmineae*	H1	A'	191	-886	**2154**
ndhC	*Jasmineae*	H2	A	192	-886	2156
ndhC	*Jasmineae*	H2	A'	191	-886	**2154**
ndhC	*LVP*	H1	A	195	-845	2080
ndhC	*LVP*	H1	A'	194	-845	**2078**
ndhC	*LVP*	H2	A	195	-845	2080
ndhC	*LVP*	H2	A'	194	-845	**2078**
ndhC	-	-	M1a	200	-932	2264
ndhD	*Jasmineae*	H1	A	192	-4767	9918
ndhD	*Jasmineae*	H1	A'	191	-4767	**9916**
ndhD	*Jasmineae*	H2	A	192	-4767	9918
ndhD	*Jasmineae*	H2	A'	191	-4767	**9916**
ndhD	*LVP*	H1	A	195	-4533	9456
ndhD	*LVP*	H1	A'	194	-4533	**9454**
ndhD	*LVP*	H2	A	195	-4533	9456
ndhD	*LVP*	H2	A'	194	-4533	**9454**
ndhD	-	-	M1a	200	-5044	10488

Table S4.A Continued

gene	clade	hypothesis	model	npar	lnL	AIC
ndhE	*Jasmineae*	H1	A	192	-753	1890
ndhE	*Jasmineae*	H1	A'	191	-753	**1888**
ndhE	*Jasmineae*	H2	A	192	-753	1890
ndhE	*Jasmineae*	H2	A'	191	-753	**1888**
ndhE	*LVP*	H1	A	195	-705	1800
ndhE	*LVP*	H1	A'	194	-705	**1798**
ndhE	*LVP*	H2	A	195	-705	1800
ndhE	*LVP*	H2	A'	194	-705	**1798**
ndhE	-	-	M1a	200	-783	1966
ndhF	*Jasmineae*	H1	A	192	-9279	18942
ndhF	*Jasmineae*	H1	A'	191	-9279	18940
ndhF	*Jasmineae*	H2	A	192	-9275	18934
ndhF	*Jasmineae*	H2	A'	191	-9275	**18932**
ndhF	*LVP*	H1	A	195	-8952	18294
ndhF	*LVP*	H1	A'	194	-8952	**18292**
ndhF	*LVP*	H2	A	195	-8952	18294
ndhF	*LVP*	H2	A'	194	-8952	**18292**
ndhF	-	-	M1a	200	-10028	20456
ndhG	*Jasmineae*	H1	A	192	-1433	**3250**
ndhG	*Jasmineae*	H1	A'	191	-1436	3254
ndhG	*Jasmineae*	H2	A	192	-1436	3256
ndhG	*Jasmineae*	H2	A'	191	-1436	3254
ndhG	*LVP*	H1	A	195	-1333	3056
ndhG	*LVP*	H1	A'	194	-1333	**3054**
ndhG	*LVP*	H2	A	195	-1333	3056
ndhG	*LVP*	H2	A'	194	-1333	**3054**
ndhG	-	-	M1a	200	-1496	3392
ndhH	*Jasmineae*	H1	A	192	-3054	6492
ndhH	*Jasmineae*	H1	A'	191	-3054	**6490**
ndhH	*Jasmineae*	H2	A	192	-3055	6494
ndhH	*Jasmineae*	H2	A'	191	-3055	6492
ndhH	*LVP*	H1	A	195	-2893	6176
ndhH	*LVP*	H1	A'	194	-2893	**6174**
ndhH	*LVP*	H2	A	195	-2893	6176
ndhH	*LVP*	H2	A'	194	-2893	**6174**
ndhH	-	-	M1a	200	-3142	6684
ndhI	*Jasmineae*	H1	A	192	-1372	3128
ndhI	*Jasmineae*	H1	A'	191	-1372	**3126**
ndhI	*Jasmineae*	H2	A	192	-1372	3128
ndhI	*Jasmineae*	H2	A'	191	-1372	**3126**
ndhI	*LVP*	H1	A	195	-1311	3012
ndhI	*LVP*	H1	A'	194	-1311	**3010**
ndhI	*LVP*	H2	A	195	-1311	3012
ndhI	*LVP*	H2	A'	194	-1311	**3010**
ndhI	-	-	M1a	200	-1417	3234
ndhJ	*Jasmineae*	H1	A	192	-997	2378
ndhJ	*Jasmineae*	H1	A'	191	-997	**2376**
ndhJ	*Jasmineae*	H2	A	192	-997	2378
ndhJ	*Jasmineae*	H2	A'	191	-997	**2376**
ndhJ	*LVP*	H1	A	195	-947	2284
ndhJ	*LVP*	H1	A'	194	-947	**2282**
ndhJ	*LVP*	H2	A	195	-947	2284
ndhJ	*LVP*	H2	A'	194	-947	**2282**
ndhJ	-	-	M1a	200	-1031	2462
petA	*Jasmineae*	H1	A	192	-2306	4996
petA	*Jasmineae*	H1	A'	191	-2306	**4994**
petA	*Jasmineae*	H2	A	192	-2307	4998
petA	*Jasmineae*	H2	A'	191	-2307	4996
petA	*LVP*	H1	A	195	-2226	4842
petA	*LVP*	H1	A'	194	-2226	**4840**
petA	*LVP*	H2	A	195	-2226	4842
petA	*LVP*	H2	A'	194	-2226	**4840**
petA	-	-	M1a	200	-2414	5228
petB	*Jasmineae*	H1	A	192	-74	532
petB	*Jasmineae*	H1	A'	191	-74	**530**
petB	*Jasmineae*	H2	A	192	-74	532
petB	*Jasmineae*	H2	A'	191	-74	**530**
petB	*LVP*	H1	A	195	-74	538
petB	*LVP*	H1	A'	194	-74	**536**
petB	*LVP*	H2	A	195	-74	538
petB	*LVP*	H2	A'	194	-74	**536**
petB	-	-	M1a	200	-75	550

Table S4.A Continued

gene	clade	hypothesis	model	npar	lnL	AIC
petD	*Jasmineae*	H1	A	192	-938	2260
petD	*Jasmineae*	H1	A'	191	-938	**2258**
petD	*Jasmineae*	H2	A	192	-938	2260
petD	*Jasmineae*	H2	A'	191	-938	**2258**
petD	*LVP*	H1	A	195	-918	2226
petD	*LVP*	H1	A'	194	-918	**2224**
petD	*LVP*	H2	A	195	-918	2226
petD	*LVP*	H2	A'	194	-918	**2224**
petD	-	-	M1a	200	-963	2326
petG	*Jasmineae*	H1	A	192	-196	776
petG	*Jasmineae*	H1	A'	191	-196	**774**
petG	*Jasmineae*	H2	A	192	-196	776
petG	*Jasmineae*	H2	A'	191	-196	**774**
petG	*LVP*	H1	A	195	-183	756
petG	*LVP*	H1	A'	194	-183	**754**
petG	*LVP*	H2	A	195	-183	756
petG	*LVP*	H2	A'	194	-183	**754**
petG	-	-	M1a	200	-196	792
petL	*Jasmineae*	H1	A	192	-215	814
petL	*Jasmineae*	H1	A'	191	-215	812
petL	*Jasmineae*	H2	A	192	-214	812
petL	*Jasmineae*	H2	A'	191	-214	**810**
petL	*LVP*	H1	A	195	-203	796
petL	*LVP*	H1	A'	194	-203	**794**
petL	*LVP*	H2	A	195	-203	796
petL	*LVP*	H2	A'	194	-203	**794**
petL	-	-	M1a	200	-215	830
petN	*Jasmineae*	H1	A	192	-157	698
petN	*Jasmineae*	H1	A'	191	-157	**696**
petN	*Jasmineae*	H2	A	192	-157	698
petN	*Jasmineae*	H2	A'	191	-157	**696**
petN	*LVP*	H1	A	195	-144	678
petN	*LVP*	H1	A'	194	-144	**676**
petN	*LVP*	H2	A	195	-144	678
petN	*LVP*	H2	A'	194	-144	**676**
petN	-	-	M1a	200	-157	714
psaA	*Jasmineae*	H1	A	192	-4342	9068
psaA	*Jasmineae*	H1	A'	191	-4342	**9066**
psaA	*Jasmineae*	H2	A	192	-4342	9068
psaA	*Jasmineae*	H2	A'	191	-4342	**9066**
psaA	*LVP*	H1	A	195	-4266	8922
psaA	*LVP*	H1	A'	194	-4266	**8920**
psaA	*LVP*	H2	A	195	-4266	8922
psaA	*LVP*	H2	A'	194	-4266	**8920**
psaA	-	-	M1a	200	-4488	9376
psaB	*Jasmineae*	H1	A	192	-4303	8990
psaB	*Jasmineae*	H1	A'	191	-4301	**8984**
psaB	*Jasmineae*	H2	A	192	-4302	8988
psaB	*Jasmineae*	H2	A'	191	-4302	8986
psaB	*LVP*	H1	A	195	-4086	8562
psaB	*LVP*	H1	A'	194	-4086	**8560**
psaB	*LVP*	H2	A	195	-4086	8562
psaB	*LVP*	H2	A'	194	-4086	**8560**
psaB	-	-	M1a	200	-4405	9210
psaC	*Jasmineae*	H1	A	192	-492	1368
psaC	*Jasmineae*	H1	A'	191	-492	**1366**
psaC	*Jasmineae*	H2	A	192	-492	1368
psaC	*Jasmineae*	H2	A'	191	-492	**1366**
psaC	*LVP*	H1	A	195	-456	1302
psaC	*LVP*	H1	A'	194	-456	**1300**
psaC	*LVP*	H2	A	195	-456	1302
psaC	*LVP*	H2	A'	194	-456	**1300**
psaC	-	-	M1a	200	-499	1398
psaI	*Jasmineae*	H1	A	192	-217	818
psaI	*Jasmineae*	H1	A'	191	-217	**816**
psaI	*Jasmineae*	H2	A	192	-217	818
psaI	*Jasmineae*	H2	A'	191	-217	**816**
psaI	*LVP*	H1	A	195	-196	782
psaI	*LVP*	H1	A'	194	-196	**780**
psaI	*LVP*	H2	A	195	-196	782
psaI	*LVP*	H2	A'	194	-196	**780**
psaI	-	-	M1a	200	-217	834

Table S4.A Continued

gene	clade	hypothesis	model	npar	lnL	AIC
psaJ	*Jasmineae*	H1	A	192	-294	972
psaJ	*Jasmineae*	H1	A'	191	-294	**970**
psaJ	*Jasmineae*	H2	A	192	-294	972
psaJ	*Jasmineae*	H2	A'	191	-294	**970**
psaJ	*LVP*	H1	A	195	-257	904
psaJ	*LVP*	H1	A'	194	-257	**902**
psaJ	*LVP*	H2	A	195	-257	904
psaJ	*LVP*	H2	A'	194	-257	**902**
psaJ	-	-	M1a	200	-294	988
psbA	*Jasmineae*	H1	A	192	-2075	4534
psbA	*Jasmineae*	H1	A'	191	-2075	**4532**
psbA	*Jasmineae*	H2	A	192	-2075	4534
psbA	*Jasmineae*	H2	A'	191	-2075	**4532**
psbA	*LVP*	H1	A	195	-2093	4576
psbA	*LVP*	H1	A'	194	-2093	**4574**
psbA	*LVP*	H2	A	195	-2093	4576
psbA	*LVP*	H2	A'	194	-2093	**4574**
psbA	-	-	M1a	200	-2206	4812
psbB	*Jasmineae*	H1	A	192	-3160	6704
psbB	*Jasmineae*	H1	A'	191	-3160	**6702**
psbB	*Jasmineae*	H2	A	192	-3160	6704
psbB	*Jasmineae*	H2	A'	191	-3160	**6702**
psbB	*LVP*	H1	A	195	-3045	6480
psbB	*LVP*	H1	A'	194	-3045	**6478**
psbB	*LVP*	H2	A	195	-3045	6480
psbB	*LVP*	H2	A'	194	-3045	**6478**
psbB	-	-	M1a	200	-3276	6952
psbC	*Jasmineae*	H1	A	192	-2792	5968
psbC	*Jasmineae*	H1	A'	191	-2792	**5966**
psbC	*Jasmineae*	H2	A	192	-2792	5968
psbC	*Jasmineae*	H2	A'	191	-2792	**5966**
psbC	*LVP*	H1	A	195	-2728	5846
psbC	*LVP*	H1	A'	194	-2728	**5844**
psbC	*LVP*	H2	A	195	-2728	5846
psbC	*LVP*	H2	A'	194	-2728	**5844**
psbC	-	-	M1a	200	-2893	6186
psbD	*Jasmineae*	H1	A	192	-1951	4286
psbD	*Jasmineae*	H1	A'	191	-1951	**4284**
psbD	*Jasmineae*	H2	A	192	-1952	4288
psbD	*Jasmineae*	H2	A'	191	-1952	4286
psbD	*LVP*	H1	A	195	-1905	**4200**
psbD	*LVP*	H1	A'	194	-1906	**4200**
psbD	*LVP*	H2	A	195	-1905	**4200**
psbD	*LVP*	H2	A'	194	-1906	**4200**
psbD	-	-	M1a	200	-1997	4394
psbE	*Jasmineae*	H1	A	192	-408	1200
psbE	*Jasmineae*	H1	A'	191	-408	**1198**
psbE	*Jasmineae*	H2	A	192	-408	1200
psbE	*Jasmineae*	H2	A'	191	-408	**1198**
psbE	*LVP*	H1	A	195	-409	1208
psbE	*LVP*	H1	A'	194	-409	**1206**
psbE	*LVP*	H2	A	195	-409	1208
psbE	*LVP*	H2	A'	194	-409	**1206**
psbE	-	-	M1a	200	-426	1252
psbF	*Jasmineae*	H1	A	192	-230	844
psbF	*Jasmineae*	H1	A'	191	-230	**842**
psbF	*Jasmineae*	H2	A	192	-230	844
psbF	*Jasmineae*	H2	A'	191	-230	**842**
psbF	*LVP*	H1	A	195	-224	838
psbF	*LVP*	H1	A'	194	-224	**836**
psbF	*LVP*	H2	A	195	-224	838
psbF	*LVP*	H2	A'	194	-224	**836**
psbF	-	-	M1a	200	-230	860
psbG	*Jasmineae*	H1	A	192	-1469	3322
psbG	*Jasmineae*	H1	A'	191	-1469	**3320**
psbG	*Jasmineae*	H2	A	192	-1469	3322
psbG	*Jasmineae*	H2	A'	191	-1469	**3320**
psbG	*LVP*	H1	A	195	-1432	3254
psbG	*LVP*	H1	A'	194	-1432	**3252**
psbG	*LVP*	H2	A	195	-1432	3254
psbG	*LVP*	H2	A'	194	-1432	**3252**
psbG	-	-	M1a	200	-1508	3416

Table S4.A Continued

gene	clade	hypothesis	model	npar	lnL	AIC
psbH	*Jasmineae*	H1	A	192	-432	1248
psbH	*Jasmineae*	H1	A'	191	-432	**1246**
psbH	*Jasmineae*	H2	A	192	-432	1248
psbH	*Jasmineae*	H2	A'	191	-432	**1246**
psbH	*LVP*	H1	A	195	-403	1196
psbH	*LVP*	H1	A'	194	-403	**1194**
psbH	*LVP*	H2	A	195	-403	1196
psbH	*LVP*	H2	A'	194	-403	**1194**
psbH	-	-	M1a	200	-448	1296
psbI	*Jasmineae*	H1	A	192	-192	768
psbI	*Jasmineae*	H1	A'	191	-192	**766**
psbI	*Jasmineae*	H2	A	192	-192	768
psbI	*Jasmineae*	H2	A'	191	-192	**766**
psbI	*LVP*	H1	A	195	-182	754
psbI	*LVP*	H1	A'	194	-182	**752**
psbI	*LVP*	H2	A	195	-182	754
psbI	*LVP*	H2	A'	194	-182	**752**
psbI	-	-	M1a	200	-192	784
psbJ	*Jasmineae*	H1	A	192	-200	**784**
psbJ	*Jasmineae*	H1	A'	191	-203	788
psbJ	*Jasmineae*	H2	A	192	-200	**784**
psbJ	*Jasmineae*	H2	A'	191	-203	788
psbJ	*LVP*	H1	A	195	-192	774
psbJ	*LVP*	H1	A'	194	-192	**772**
psbJ	*LVP*	H2	A	195	-192	774
psbJ	*LVP*	H2	A'	194	-192	**772**
psbJ	-	-	M1a	200	-204	808
psbK	*Jasmineae*	H1	A	192	-466	1316
psbK	*Jasmineae*	H1	A'	191	-466	**1314**
psbK	*Jasmineae*	H2	A	192	-465	**1314**
psbK	*Jasmineae*	H2	A'	191	-466	**1314**
psbK	*LVP*	H1	A	195	-428	1246
psbK	*LVP*	H1	A'	194	-427	**1242**
psbK	*LVP*	H2	A	195	-428	1246
psbK	*LVP*	H2	A'	194	-427	**1242**
psbK	-	-	M1a	200	-481	1362
psbL	*Jasmineae*	H1	A	192	-209	802
psbL	*Jasmineae*	H1	A'	191	-209	**800**
psbL	*Jasmineae*	H2	A	192	-209	802
psbL	*Jasmineae*	H2	A'	191	-209	**800**
psbL	*LVP*	H1	A	195	-197	784
psbL	*LVP*	H1	A'	194	-197	**782**
psbL	*LVP*	H2	A	195	-197	784
psbL	*LVP*	H2	A'	194	-197	**782**
psbL	-	-	M1a	200	-209	818
psbM	*Jasmineae*	H1	A	192	-190	764
psbM	*Jasmineae*	H1	A'	191	-190	**762**
psbM	*Jasmineae*	H2	A	192	-190	764
psbM	*Jasmineae*	H2	A'	191	-190	**762**
psbM	*LVP*	H1	A	195	-173	736
psbM	*LVP*	H1	A'	194	-173	**734**
psbM	*LVP*	H2	A	195	-173	736
psbM	*LVP*	H2	A'	194	-173	**734**
psbM	-	-	M1a	200	-190	780
psbN	*Jasmineae*	H1	A	192	-220	824
psbN	*Jasmineae*	H1	A'	191	-220	**822**
psbN	*Jasmineae*	H2	A	192	-220	824
psbN	*Jasmineae*	H2	A'	191	-220	**822**
psbN	*LVP*	H1	A	195	-219	828
psbN	*LVP*	H1	A'	194	-219	**826**
psbN	*LVP*	H2	A	195	-219	828
psbN	*LVP*	H2	A'	194	-219	**826**
psbN	-	-	M1a	200	-225	850
psbT	*Jasmineae*	H1	A	192	-222	828
psbT	*Jasmineae*	H1	A'	191	-222	**826**
psbT	*Jasmineae*	H2	A	192	-222	828
psbT	*Jasmineae*	H2	A'	191	-222	**826**
psbT	*LVP*	H1	A	195	-226	842
psbT	*LVP*	H1	A'	194	-226	**840**
psbT	*LVP*	H2	A	195	-226	842
psbT	*LVP*	H2	A'	194	-226	**840**
psbT	-	-	M1a	200	-232	864

Table S4.A Continued

gene	clade	hypothesis	model	npar	lnL	AIC
psbZ	*Jasmineae*	H1	A	192	-308	1000
psbZ	*Jasmineae*	H1	A'	191	-308	**998**
psbZ	*Jasmineae*	H2	A	192	-308	1000
psbZ	*Jasmineae*	H2	A'	191	-308	**998**
psbZ	*LVP*	H1	A	195	-282	954
psbZ	*LVP*	H1	A'	194	-282	**952**
psbZ	*LVP*	H2	A	195	-282	954
psbZ	*LVP*	H2	A'	194	-282	**952**
psbZ	-	-	M1a	200	-316	1032
rbcL	*Jasmineae*	H1	A	192	-4479	9342
rbcL	*Jasmineae*	H1	A'	191	-4480	9342
rbcL	*Jasmineae*	H2	A	192	-4474	**9332**
rbcL	*Jasmineae*	H2	A'	191	-4479	9340
rbcL	*LVP*	H1	A	195	-4213	8816
rbcL	*LVP*	H1	A'	194	-4213	**8814**
rbcL	*LVP*	H2	A	195	-4213	8816
rbcL	*LVP*	H2	A'	194	-4213	**8814**
rbcL	-	-	M1a	200	-4602	9604
rpl2	*Jasmineae*	H1	A	192	-1381	3146
rpl2	*Jasmineae*	H1	A'	191	-1381	3144
rpl2	*Jasmineae*	H2	A	192	-1375	**3134**
rpl2	*Jasmineae*	H2	A'	191	-1376	**3134**
rpl2	*LVP*	H1	A	195	-1263	2916
rpl2	*LVP*	H1	A'	194	-1263	**2914**
rpl2	*LVP*	H2	A	195	-1263	2916
rpl2	*LVP*	H2	A'	194	-1263	**2914**
rpl2	-	-	M1a	200	-1441	3282
rpl14	*Jasmineae*	H1	A	192	-1069	**2522**
rpl14	*Jasmineae*	H1	A'	191	-1076	2534
rpl14	*Jasmineae*	H2	A	192	-1071	2526
rpl14	*Jasmineae*	H2	A'	191	-1071	2524
rpl14	*LVP*	H1	A	195	-1009	2408
rpl14	*LVP*	H1	A'	194	-1009	**2406**
rpl14	*LVP*	H2	A	195	-1009	2408
rpl14	*LVP*	H2	A'	194	-1009	**2406**
rpl14	-	-	M1a	200	-1146	2692
rpl16	*Jasmineae*	H1	A	192	-1102	2588
rpl16	*Jasmineae*	H1	A'	191	-1103	2588
rpl16	*Jasmineae*	H2	A	192	-1101	**2586**
rpl16	*Jasmineae*	H2	A'	191	-1103	2588
rpl16	*LVP*	H1	A	195	-975	2340
rpl16	*LVP*	H1	A'	194	-975	**2338**
rpl16	*LVP*	H2	A	195	-975	2340
rpl16	*LVP*	H2	A'	194	-975	**2338**
rpl16	-	-	M1a	200	-1140	2680
rpl20	*Jasmineae*	H1	A	192	-1103	2590
rpl20	*Jasmineae*	H1	A'	191	-1104	2590
rpl20	*Jasmineae*	H2	A	192	-1098	**2580**
rpl20	*Jasmineae*	H2	A'	191	-1100	2582
rpl20	*LVP*	H1	A	195	-1192	2774
rpl20	*LVP*	H1	A'	194	-1192	**2772**
rpl20	*LVP*	H2	A	195	-1192	2774
rpl20	*LVP*	H2	A'	194	-1192	**2772**
rpl20	-	-	M1a	200	-1428	3256
rpl22	*Jasmineae*	H1	A	192	-1756	**3896**
rpl22	*Jasmineae*	H1	A'	191	-1757	**3896**
rpl22	*Jasmineae*	H2	A	192	-1759	3902
rpl22	*Jasmineae*	H2	A'	191	-1759	3900
rpl22	*LVP*	H1	A	195	-1642	3674
rpl22	*LVP*	H1	A'	194	-1642	**3672**
rpl22	*LVP*	H2	A	195	-1642	3674
rpl22	*LVP*	H2	A'	194	-1642	**3672**
rpl22	-	-	M1a	200	-1834	4068
rpl23	*Jasmineae*	H1	A	192	-423	1230
rpl23	*Jasmineae*	H1	A'	191	-423	**1228**
rpl23	*Jasmineae*	H2	A	192	-423	1230
rpl23	*Jasmineae*	H2	A'	191	-423	**1228**
rpl23	*LVP*	H1	A	195	-471	1332
rpl23	*LVP*	H1	A'	194	-471	**1330**
rpl23	*LVP*	H2	A	195	-471	1332
rpl23	*LVP*	H2	A'	194	-471	**1330**
rpl23	-	-	M1a	200	-501	1402

Table S4.A Continued

gene	clade	hypothesis	model	npar	lnL	AIC
rpl32	*Jasmineae*	H1	A	192	-987	2358
rpl32	*Jasmineae*	H1	A'	191	-989	2360
rpl32	*Jasmineae*	H2	A	192	-988	2360
rpl32	*Jasmineae*	H2	A'	191	-986	**2354**
rpl32	*LVP*	H1	A	195	-842	2074
rpl32	*LVP*	H1	A'	194	-842	**2072**
rpl32	*LVP*	H2	A	195	-842	2074
rpl32	*LVP*	H2	A'	194	-842	**2072**
rpl32	-	-	M1a	200	-1055	2510
rpl33	*Jasmineae*	H1	A	192	-537	1458
rpl33	*Jasmineae*	H1	A'	191	-537	**1456**
rpl33	*Jasmineae*	H2	A	192	-537	1458
rpl33	*Jasmineae*	H2	A'	191	-537	**1456**
rpl33	*LVP*	H1	A	195	-515	1420
rpl33	*LVP*	H1	A'	194	-515	**1418**
rpl33	*LVP*	H2	A	195	-515	1420
rpl33	*LVP*	H2	A'	194	-515	**1418**
rpl33	-	-	M1a	200	-591	1582
rpl36	*Jasmineae*	H1	A	192	-243	**870**
rpl36	*Jasmineae*	H1	A'	191	-245	872
rpl36	*Jasmineae*	H2	A	192	-244	872
rpl36	*Jasmineae*	H2	A'	191	-244	**870**
rpl36	*LVP*	H1	A	195	-231	852
rpl36	*LVP*	H1	A'	194	-231	**850**
rpl36	*LVP*	H2	A	195	-231	852
rpl36	*LVP*	H2	A'	194	-231	**850**
rpl36	-	-	M1a	200	-274	948
rpoA	*Jasmineae*	H1	A	192	-3096	6576
rpoA	*Jasmineae*	H1	A'	191	-3096	6574
rpoA	*Jasmineae*	H2	A	192	-3091	**6566**
rpoA	*Jasmineae*	H2	A'	191	-3093	6568
rpoA	*LVP*	H1	A	195	-2962	6314
rpoA	*LVP*	H1	A'	194	-2962	**6312**
rpoA	*LVP*	H2	A	195	-2962	6314
rpoA	*LVP*	H2	A'	194	-2962	**6312**
rpoA	-	-	M1a	200	-3297	6994
rpoB	*Jasmineae*	H1	A	192	-7672	15728
rpoB	*Jasmineae*	H1	A'	191	-7674	15730
rpoB	*Jasmineae*	H2	A	192	-7650	**15684**
rpoB	*Jasmineae*	H2	A'	191	-7663	15708
rpoB	*LVP*	H1	A	195	-7180	14750
rpoB	*LVP*	H1	A'	194	-7180	**14748**
rpoB	*LVP*	H2	A	195	-7180	14750
rpoB	*LVP*	H2	A'	194	-7180	**14748**
rpoB	-	-	M1a	200	-8111	16622
rpoC1	*Jasmineae*	H1	A	192	-5137	10658
rpoC1	*Jasmineae*	H1	A'	191	-5139	10660
rpoC1	*Jasmineae*	H2	A	192	-5126	**10636**
rpoC1	*Jasmineae*	H2	A'	191	-5131	10644
rpoC1	*LVP*	H1	A	195	-4639	**9668**
rpoC1	*LVP*	H1	A'	194	-4640	**9668**
rpoC1	*LVP*	H2	A	195	-4639	**9668**
rpoC1	*LVP*	H2	A'	194	-4640	**9668**
rpoC1	-	-	M1a	200	-5422	11244
rpoC2	*Jasmineae*	H1	A	192	-12184	24752
rpoC2	*Jasmineae*	H1	A'	191	-12184	24750
rpoC2	*Jasmineae*	H2	A	192	-12180	24744
rpoC2	*Jasmineae*	H2	A'	191	-12180	**24742**
rpoC2	*LVP*	H1	A	195	-11364	23118
rpoC2	*LVP*	H1	A'	194	-11364	**23116**
rpoC2	*LVP*	H2	A	195	-11364	23118
rpoC2	*LVP*	H2	A'	194	-11364	**23116**
rpoC2	-	-	M1a	200	-12881	26162
rps2	*Jasmineae*	H1	A	192	-1852	4088
rps2	*Jasmineae*	H1	A'	191	-1857	4096
rps2	*Jasmineae*	H2	A	192	-1843	**4070**
rps2	*Jasmineae*	H2	A'	191	-1852	4086
rps2	*LVP*	H1	A	195	-1650	3690
rps2	*LVP*	H1	A'	194	-1650	**3688**
rps2	*LVP*	H2	A	195	-1650	3690
rps2	*LVP*	H2	A'	194	-1650	**3688**
rps2	-	-	M1a	200	-1988	4376

Table S4.A Continued

gene	clade	hypothesis	model	npar	lnL	AIC
rps3	*Jasmineae*	H1	A	192	-2148	4680
rps3	*Jasmineae*	H1	A'	191	-2149	4680
rps3	*Jasmineae*	H2	A	192	-2146	4676
rps3	*Jasmineae*	H2	A'	191	-2146	**4674**
rps3	*LVP*	H1	A	195	-1932	4254
rps3	*LVP*	H1	A'	194	-1932	**4252**
rps3	*LVP*	H2	A	195	-1932	4254
rps3	*LVP*	H2	A'	194	-1932	**4252**
rps3	-	-	M1a	200	-2381	5162
rps4	*Jasmineae*	H1	A	192	-1413	3210
rps4	*Jasmineae*	H1	A'	191	-1420	3222
rps4	*Jasmineae*	H2	A	192	-1412	**3208**
rps4	*Jasmineae*	H2	A'	191	-1414	3210
rps4	*LVP*	H1	A	195	-1257	2904
rps4	*LVP*	H1	A'	194	-1257	**2902**
rps4	*LVP*	H2	A	195	-1257	2904
rps4	*LVP*	H2	A'	194	-1257	**2902**
rps4	-	-	M1a	200	-1484	3368
rps7	*Jasmineae*	H1	A	192	-842	2068
rps7	*Jasmineae*	H1	A'	191	-845	2072
rps7	*Jasmineae*	H2	A	192	-838	**2060**
rps7	*Jasmineae*	H2	A'	191	-843	2068
rps7	*LVP*	H1	A	195	-772	1934
rps7	*LVP*	H1	A'	194	-772	**1932**
rps7	*LVP*	H2	A	195	-772	1934
rps7	*LVP*	H2	A'	194	-772	**1932**
rps7	-	-	M1a	200	-922	2244
rps8	*Jasmineae*	H1	A	192	-1144	2672
rps8	*Jasmineae*	H1	A'	191	-1144	2670
rps8	*Jasmineae*	H2	A	192	-1140	2664
rps8	*Jasmineae*	H2	A'	191	-1140	**2662**
rps8	*LVP*	H1	A	195	-1020	2430
rps8	*LVP*	H1	A'	194	-1020	**2428**
rps8	*LVP*	H2	A	195	-1020	2430
rps8	*LVP*	H2	A'	194	-1020	**2428**
rps8	-	-	M1a	200	-1226	2852
rps11	*Jasmineae*	H1	A	192	-1266	2916
rps11	*Jasmineae*	H1	A'	191	-1271	2924
rps11	*Jasmineae*	H2	A	192	-1262	**2908**
rps11	*Jasmineae*	H2	A'	191	-1267	2916
rps11	*LVP*	H1	A	195	-1060	2510
rps11	*LVP*	H1	A'	194	-1060	**2508**
rps11	*LVP*	H2	A	195	-1060	2510
rps11	*LVP*	H2	A'	194	-1060	**2508**
rps11	-	-	M1a	200	-1410	3220
rps12	*Jasmineae*	H1	A	192	-823	2030
rps12	*Jasmineae*	H1	A'	191	-826	2034
rps12	*Jasmineae*	H2	A	192	-818	**2020**
rps12	*Jasmineae*	H2	A'	191	-828	2038
rps12	*LVP*	H1	A	195	-730	1850
rps12	*LVP*	H1	A'	194	-730	**1848**
rps12	*LVP*	H2	A	195	-730	1850
rps12	*LVP*	H2	A'	194	-730	**1848**
rps12	-	-	M1a	200	-861	2122
rps14	*Jasmineae*	H1	A	192	-692	1768
rps14	*Jasmineae*	H1	A'	191	-692	**1766**
rps14	*Jasmineae*	H2	A	192	-693	1770
rps14	*Jasmineae*	H2	A'	191	-693	1768
rps14	*LVP*	H1	A	195	-625	1640
rps14	*LVP*	H1	A'	194	-625	**1638**
rps14	*LVP*	H2	A	195	-625	1640
rps14	*LVP*	H2	A'	194	-625	**1638**
rps14	-	-	M1a	200	-695	1790
rps15	*Jasmineae*	H1	A	192	-856	2096
rps15	*Jasmineae*	H1	A'	191	-856	2094
rps15	*Jasmineae*	H2	A	192	-855	2094
rps15	*Jasmineae*	H2	A'	191	-855	**2092**
rps15	*LVP*	H1	A	195	-719	1828
rps15	*LVP*	H1	A'	194	-719	**1826**
rps15	*LVP*	H2	A	195	-719	1828
rps15	*LVP*	H2	A'	194	-719	**1826**
rps15	-	-	M1a	200	-899	2198

Table S4.A End

gene	clade	hypothesis	model	npar	lnL	AIC
rps16	*Jasmineae*	H1	A	192	-641	1666
rps16	*Jasmineae*	H1	A'	191	-641	**1664**
rps16	*Jasmineae*	H2	A	192	-643	1670
rps16	*Jasmineae*	H2	A'	191	-643	1668
rps16	*LVP*	H1	A	195	-582	1554
rps16	*LVP*	H1	A'	194	-582	**1552**
rps16	*LVP*	H2	A	195	-582	1554
rps16	*LVP*	H2	A'	194	-582	**1552**
rps16	-	-	M1a	200	-663	1726
rps18	*Jasmineae*	H1	A	192	-1125	2634
rps18	*Jasmineae*	H1	A'	191	-1129	2640
rps18	*Jasmineae*	H2	A	192	-1118	**2620**
rps18	*Jasmineae*	H2	A'	191	-1123	2628
rps18	*LVP*	H1	A	195	-936	2262
rps18	*LVP*	H1	A'	194	-936	**2260**
rps18	*LVP*	H2	A	195	-936	2262
rps18	*LVP*	H2	A'	194	-936	**2260**
rps18	-	-	M1a	200	-1429	3258
rps19	*Jasmineae*	H1	A	192	-1119	2622
rps19	*Jasmineae*	H1	A'	191	-1120	2622
rps19	*Jasmineae*	H2	A	192	-1120	2624
rps19	*Jasmineae*	H2	A'	191	-1119	**2620**
rps19	*LVP*	H1	A	195	-794	1978
rps19	*LVP*	H1	A'	194	-794	**1976**
rps19	*LVP*	H2	A	195	-794	1978
rps19	*LVP*	H2	A'	194	-794	**1976**
rps19	-	-	M1a	200	-1212	2824
ycf1	*Jasmineae*	H1	A	192	-29618	59620
ycf1	*Jasmineae*	H1	A'	191	-29680	59742
ycf1	*Jasmineae*	H2	A	192	-29493	**59370**
ycf1	*Jasmineae*	H2	A'	191	-29592	59566
ycf1	*LVP*	H1	A	195	-29372	59134
ycf1	*LVP*	H1	A'	194	-29372	**59132**
ycf1	*LVP*	H2	A	195	-29372	59134
ycf1	*LVP*	H2	A'	194	-29372	**59132**
ycf1	-	-	M1a	200	-34230	68860
ycf2	*Jasmineae*	H1	A	192	-13093	26570
ycf2	*Jasmineae*	H1	A'	191	-13096	26574
ycf2	*Jasmineae*	H2	A	192	-13087	**26558**
ycf2	*Jasmineae*	H2	A'	191	-13096	26574
ycf2	*LVP*	H1	A	195	-12516	25422
ycf2	*LVP*	H1	A'	194	-12516	**25420**
ycf2	*LVP*	H2	A	195	NA	NA
ycf2	*LVP*	H2	A'	194	-12516	**25420**
ycf2	-	-	M1a	200	-15893	32186
ycf3	*Jasmineae*	H1	A	192	-944	2272
ycf3	*Jasmineae*	H1	A'	191	-944	**2270**
ycf3	*Jasmineae*	H2	A	192	-944	2272
ycf3	*Jasmineae*	H2	A'	191	-944	**2270**
ycf3	*LVP*	H1	A	195	-859	2108
ycf3	*LVP*	H1	A'	194	-859	**2106**
ycf3	*LVP*	H2	A	195	-859	2108
ycf3	*LVP*	H2	A'	194	-859	**2106**
ycf3	-	-	M1a	200	-944	2288
ycf4	*Jasmineae*	H1	A	192	-1308	3000
ycf4	*Jasmineae*	H1	A'	191	-1308	**2998**
ycf4	*Jasmineae*	H2	A	192	-1308	3000
ycf4	*Jasmineae*	H2	A'	191	-1308	**2998**
ycf4	*LVP*	H1	A	195	-1626	3642
ycf4	*LVP*	H1	A'	194	-1626	**3640**
ycf4	*LVP*	H2	A	195	-1626	3642
ycf4	*LVP*	H2	A'	194	-1626	**3640**
ycf4	-	-	M1a	200	-1782	3964
ycf15	*Jasmineae*	H1	A	192	-277	**938**
ycf15	*Jasmineae*	H1	A'	191	-281	944
ycf15	*Jasmineae*	H2	A	192	-278	940
ycf15	*Jasmineae*	H2	A'	191	-281	944
ycf15	*LVP*	H1	A	195	-496	1382
ycf15	*LVP*	H1	A'	194	-496	**1380**
ycf15	*LVP*	H2	A	195	-496	1382
ycf15	*LVP*	H2	A'	194	-496	**1380**
ycf15	-	-	M1a	200	-516	1432

Table S4.B

gene	clade	hypothesis	model	npar	lnL	AIC
atp1	*Jasmineae*	H1	A	189	-3220	6818
atp1	*Jasmineae*	H1	A'	188	-3221	6818
atp1	*Jasmineae*	H2	A	189	-3222	6822
atp1	*Jasmineae*	H2	A'	188	-3220	**6816**
atp1	*LVP*	H1	A	191	-3260	6902
atp1	*LVP*	H1	A'	190	-3259	**6898**
atp1	*LVP*	H2	A	191	-3258	**6898**
atp1	*LVP*	H2	A'	190	-3259	**6898**
atp1	-	-	M1a	196	-3360	7112
atp4	*Jasmineae*	H1	A	193	-1128	2642
atp4	*Jasmineae*	H1	A'	192	-1128	**2640**
atp4	*Jasmineae*	H2	A	193	-1128	2642
atp4	*Jasmineae*	H2	A'	192	-1128	**2640**
atp4	*LVP*	H1	A	195	-1131	2652
atp4	*LVP*	H1	A'	194	-1131	2650
atp4	*LVP*	H2	A	195	-1128	**2646**
atp4	*LVP*	H2	A'	194	-1133	2654
atp4	-	-	M1a	200	-1160	2720
atp6	*Jasmineae*	H1	A	193	-2407	**5200**
atp6	*Jasmineae*	H1	A'	192	-2408	**5200**
atp6	*Jasmineae*	H2	A	193	-2408	5202
atp6	*Jasmineae*	H2	A'	192	-2408	**5200**
atp6	*LVP*	H1	A	195	-2414	5218
atp6	*LVP*	H1	A'	194	-2415	5218
atp6	*LVP*	H2	A	195	-2416	5222
atp6	*LVP*	H2	A'	194	-2415	5218
atp6	-	-	M1a	200	-2444	5288
atp8	*Jasmineae*	H1	A	193	-1047	2480
atp8	*Jasmineae*	H1	A'	192	-1047	**2478**
atp8	*Jasmineae*	H2	A	193	-1047	2480
atp8	*Jasmineae*	H2	A'	192	-1047	**2478**
atp8	*LVP*	H1	A	195	-1020	2430
atp8	*LVP*	H1	A'	194	-1020	**2428**
atp8	*LVP*	H2	A	195	-1020	2430
atp8	*LVP*	H2	A'	194	-1020	**2428**
atp8	-	-	M1a	200	-1060	2520
atp9	*Jasmineae*	H1	A	193	-419	1224
atp9	*Jasmineae*	H1	A'	192	-419	**1222**
atp9	*Jasmineae*	H2	A	193	-419	1224
atp9	*Jasmineae*	H2	A'	192	-419	**1222**
atp9	*LVP*	H1	A	195	-412	1214
atp9	*LVP*	H1	A'	194	-412	**1212**
atp9	*LVP*	H2	A	195	-412	1214
atp9	*LVP*	H2	A'	194	-412	**1212**
atp9	-	-	M1a	200	-419	1238
ccmB	*Jasmineae*	H1	A	193	-928	2242
ccmB	*Jasmineae*	H1	A'	192	-929	2242
ccmB	*Jasmineae*	H2	A	193	-928	2242
ccmB	*Jasmineae*	H2	A'	192	-928	**2240**
ccmB	*LVP*	H1	A	195	-936	2262
ccmB	*LVP*	H1	A'	194	-936	2260
ccmB	*LVP*	H2	A	195	-936	2262
ccmB	*LVP*	H2	A'	194	-936	2260
ccmB	-	-	M1a	200	-928	**2256**
ccmC	*Jasmineae*	H1	A	193	-1180	2746
ccmC	*Jasmineae*	H1	A'	192	-1180	**2744**
ccmC	*Jasmineae*	H2	A	193	-1180	2746
ccmC	*Jasmineae*	H2	A'	192	-1180	**2744**
ccmC	*LVP*	H1	A	195	-1171	2732
ccmC	*LVP*	H1	A'	194	-1171	**2730**
ccmC	*LVP*	H2	A	195	-1171	2732
ccmC	*LVP*	H2	A'	194	-1171	**2730**
ccmC	-	-	M1a	200	-1186	2772
ccmFc	*Jasmineae*	H1	A	193	-2577	**5540**
ccmFc	*Jasmineae*	H1	A'	192	-2587	5558
ccmFc	*Jasmineae*	H2	A	193	-2581	5548
ccmFc	*Jasmineae*	H2	A'	192	-2587	5558
ccmFc	*LVP*	H1	A	195	-2538	5466
ccmFc	*LVP*	H1	A'	194	-2538	**5464**
ccmFc	*LVP*	H2	A	195	-2538	5466
ccmFc	*LVP*	H2	A'	194	-2538	**5464**
ccmFc	-	-	M1a	200	-2623	5646

Table S4.B Continued

gene	clade	hypothesis	model	npar	lnL	AIC
ccmFn	*Jasmineae*	H1	A	191	-2811	6004
ccmFn	*Jasmineae*	H1	A'	190	-2811	**6002**
ccmFn	*Jasmineae*	H2	A	191	-2811	6004
ccmFn	*Jasmineae*	H2	A'	190	-2811	**6002**
ccmFn	*LVP*	H1	A	193	-2803	5992
ccmFn	*LVP*	H1	A'	192	-2803	**5990**
ccmFn	*LVP*	H2	A	193	-2803	5992
ccmFn	*LVP*	H2	A'	192	-2803	**5990**
ccmFn	-	-	M1a	198	-2840	6076
cox1	*Jasmineae*	H1	A	191	-3759	7900
cox1	*Jasmineae*	H1	A'	190	-3759	7898
cox1	*Jasmineae*	H2	A	191	-3752	**7886**
cox1	*Jasmineae*	H2	A'	190	-3753	**7886**
cox1	*LVP*	H1	A	193	-3671	7728
cox1	*LVP*	H1	A'	192	-3670	**7724**
cox1	*LVP*	H2	A	193	-3670	7726
cox1	*LVP*	H2	A'	192	-3671	7726
cox1	-	-	M1a	198	-3854	8104
cox2	*Jasmineae*	H1	A	193	-1556	3498
cox2	*Jasmineae*	H1	A'	192	-1556	3496
cox2	*Jasmineae*	H2	A	193	-1553	**3492**
cox2	*Jasmineae*	H2	A'	192	-1556	3496
cox2	*LVP*	H1	A	195	-1484	3358
cox2	*LVP*	H1	A'	194	-1484	**3356**
cox2	*LVP*	H2	A	195	-1485	3360
cox2	*LVP*	H2	A'	194	-1486	3360
cox2	-	-	M1a	200	-1568	3536
cox3	*Jasmineae*	H1	A	193	-1290	2966
cox3	*Jasmineae*	H1	A'	192	-1290	**2964**
cox3	*Jasmineae*	H2	A	193	-1290	2966
cox3	*Jasmineae*	H2	A'	192	-1290	**2964**
cox3	*LVP*	H1	A	195	-1297	2984
cox3	*LVP*	H1	A'	194	-1297	2982
cox3	*LVP*	H2	A	195	-1297	2984
cox3	*LVP*	H2	A'	194	-1297	2982
cox3	-	-	M1a	200	-1290	**2980**
cytB	*Jasmineae*	H1	A	193	-2223	4832
cytB	*Jasmineae*	H1	A'	192	-2223	4830
cytB	*Jasmineae*	H2	A	193	-2206	**4798**
cytB	*Jasmineae*	H2	A'	192	-2223	4830
cytB	*LVP*	H1	A	195	-2133	4656
cytB	*LVP*	H1	A'	194	-2134	4656
cytB	*LVP*	H2	A	195	-2133	4656
cytB	*LVP*	H2	A'	194	-2133	**4654**
cytB	-	-	M1a	200	-2237	4874
matR	*Jasmineae*	H1	A	191	-3410	7202
matR	*Jasmineae*	H1	A'	190	-3410	**7200**
matR	*Jasmineae*	H2	A	191	-3410	7202
matR	*Jasmineae*	H2	A'	190	-3410	**7200**
matR	*LVP*	H1	A	193	-3363	7112
matR	*LVP*	H1	A'	192	-3363	**7110**
matR	*LVP*	H2	A	193	-3363	7112
matR	*LVP*	H2	A'	192	-3363	**7110**
matR	-	-	M1a	198	-3429	7254
mttB	*Jasmineae*	H1	A	193	-1681	3748
mttB	*Jasmineae*	H1	A'	192	-1682	3748
mttB	*Jasmineae*	H2	A	193	-1681	3748
mttB	*Jasmineae*	H2	A'	192	-1681	**3746**
mttB	*LVP*	H1	A	195	-1522	3434
mttB	*LVP*	H1	A'	194	-1522	**3432**
mttB	*LVP*	H2	A	195	-1522	3434
mttB	*LVP*	H2	A'	194	-1522	**3432**
mttB	-	-	M1a	200	-1709	3818
nad1	*Jasmineae*	H1	A	193	-1486	3358
nad1	*Jasmineae*	H1	A'	192	-1479	**3342**
nad1	*Jasmineae*	H2	A	193	-1479	3344
nad1	*Jasmineae*	H2	A'	192	-1479	**3342**
nad1	*LVP*	H1	A	195	-1472	3334
nad1	*LVP*	H1	A'	194	-1472	**3332**
nad1	*LVP*	H2	A	195	-1472	3334
nad1	*LVP*	H2	A'	194	-1472	**3332**
nad1	-	-	M1a	200	-1495	3390

Table S4.B Continued

gene	clade	hypothesis	model	npar	lnL	AIC
nad2	*Jasmineae*	H1	A	193	-2142	4670
nad2	*Jasmineae*	H1	A'	192	-2142	**4668**
nad2	*Jasmineae*	H2	A	193	-2142	4670
nad2	*Jasmineae*	H2	A'	192	-2142	**4668**
nad2	*LVP*	H1	A	195	-2152	4694
nad2	*LVP*	H1	A'	194	-2152	4692
nad2	*LVP*	H2	A	195	-2149	4688
nad2	*LVP*	H2	A'	194	-2149	**4686**
nad2	-	-	M1a	200	-2163	4726
nad3	*Jasmineae*	H1	A	193	-731	1848
nad3	*Jasmineae*	H1	A'	192	-731	**1846**
nad3	*Jasmineae*	H2	A	193	-731	1848
nad3	*Jasmineae*	H2	A'	192	-731	**1846**
nad3	*LVP*	H1	A	195	-745	1880
nad3	*LVP*	H1	A'	194	-745	**1878**
nad3	*LVP*	H2	A	195	-745	1880
nad3	*LVP*	H2	A'	194	-745	**1878**
nad3	-	-	M1a	200	-753	1906
nad4	*Jasmineae*	H1	A	193	-2306	4998
nad4	*Jasmineae*	H1	A'	192	-2306	**4996**
nad4	*Jasmineae*	H2	A	193	-2306	4998
nad4	*Jasmineae*	H2	A'	192	-2306	**4996**
nad4	*LVP*	H1	A	195	-2324	5038
nad4	*LVP*	H1	A'	194	-2324	**5036**
nad4	*LVP*	H2	A	195	-2324	5038
nad4	*LVP*	H2	A'	194	-2324	**5036**
nad4	-	-	M1a	200	-2324	5048
nad4L	*Jasmineae*	H1	A	193	-463	1312
nad4L	*Jasmineae*	H1	A'	192	-463	**1310**
nad4L	*Jasmineae*	H2	A	193	-463	1312
nad4L	*Jasmineae*	H2	A'	192	-463	**1310**
nad4L	*LVP*	H1	A	195	-453	1296
nad4L	*LVP*	H1	A'	194	-453	**1294**
nad4L	*LVP*	H2	A	195	-453	1296
nad4L	*LVP*	H2	A'	194	-453	**1294**
nad4L	-	-	M1a	200	-464	1328
nad5	*Jasmineae*	H1	A	193	-3324	7034
nad5	*Jasmineae*	H1	A'	192	-3324	7032
nad5	*Jasmineae*	H2	A	193	-3317	**7020**
nad5	*Jasmineae*	H2	A'	192	-3324	7032
nad5	*LVP*	H1	A	195	-3265	6920
nad5	*LVP*	H1	A'	194	-3264	**6916**
nad5	*LVP*	H2	A	195	-3264	6918
nad5	*LVP*	H2	A'	194	-3264	**6916**
nad5	-	-	M1a	200	-3350	7100
nad6	*Jasmineae*	H1	A	193	-1145	2676
nad6	*Jasmineae*	H1	A'	192	-1147	2678
nad6	*Jasmineae*	H2	A	193	-1146	2678
nad6	*Jasmineae*	H2	A'	192	-1145	**2674**
nad6	*LVP*	H1	A	195	-1126	2642
nad6	*LVP*	H1	A'	194	-1126	**2640**
nad6	*LVP*	H2	A	195	-1126	2642
nad6	*LVP*	H2	A'	194	-1127	2642
nad6	-	-	M1a	200	-1147	2694
nad7	*Jasmineae*	H1	A	191	-1875	4132
nad7	*Jasmineae*	H1	A'	190	-1875	**4130**
nad7	*Jasmineae*	H2	A	191	-1875	4132
nad7	*Jasmineae*	H2	A'	190	-1875	**4130**
nad7	*LVP*	H1	A	193	-1846	4078
nad7	*LVP*	H1	A'	192	-1846	**4076**
nad7	*LVP*	H2	A	193	-1846	4078
nad7	*LVP*	H2	A'	192	-1846	**4076**
nad7	-	-	M1a	198	-1884	4164
nad9	*Jasmineae*	H1	A	193	-1049	2484
nad9	*Jasmineae*	H1	A'	192	-1049	**2482**
nad9	*Jasmineae*	H2	A	193	-1049	2484
nad9	*Jasmineae*	H2	A'	192	-1049	**2482**
nad9	*LVP*	H1	A	195	-1038	2466
nad9	*LVP*	H1	A'	194	-1038	**2464**
nad9	*LVP*	H2	A	195	-1038	2466
nad9	*LVP*	H2	A'	194	-1038	**2464**
nad9	-	-	M1a	200	-1049	2498

Table S4.B Continued

gene	clade	hypothesis	model	npar	lnL	AIC
rpl5	*Jasmineae*	H1	A	193	-958	2302
rpl5	*Jasmineae*	H1	A'	192	-958	**2300**
rpl5	*Jasmineae*	H2	A	193	-958	2302
rpl5	*Jasmineae*	H2	A'	192	-958	**2300**
rpl5	*LVP*	H1	A	195	-928	2246
rpl5	*LVP*	H1	A'	194	-928	**2244**
rpl5	*LVP*	H2	A	195	-928	2246
rpl5	*LVP*	H2	A'	194	-928	**2244**
rpl5	-	-	M1a	200	-966	2332
rpl10	*Jasmineae*	H1	A	193	-846	2078
rpl10	*Jasmineae*	H1	A'	192	-846	**2076**
rpl10	*Jasmineae*	H2	A	193	-846	2078
rpl10	*Jasmineae*	H2	A'	192	-847	2078
rpl10	*LVP*	H1	A	195	-826	2042
rpl10	*LVP*	H1	A'	194	-826	**2040**
rpl10	*LVP*	H2	A	195	-826	2042
rpl10	*LVP*	H2	A'	194	-826	**2040**
rpl10	-	-	M1a	200	-853	2106
rpl16	*Jasmineae*	H1	A	193	-796	1978
rpl16	*Jasmineae*	H1	A'	192	-796	**1976**
rpl16	*Jasmineae*	H2	A	193	-797	1980
rpl16	*Jasmineae*	H2	A'	192	-796	**1976**
rpl16	*LVP*	H1	A	195	-788	1966
rpl16	*LVP*	H1	A'	194	-788	**1964**
rpl16	*LVP*	H2	A	195	-788	1966
rpl16	*LVP*	H2	A'	194	-788	**1964**
rpl16	-	-	M1a	200	-796	1992
rps3	*Jasmineae*	H1	A	191	-3177	**6736**
rps3	*Jasmineae*	H1	A'	190	-3182	6744
rps3	*Jasmineae*	H2	A	191	-3179	6740
rps3	*Jasmineae*	H2	A'	190	-3182	6744
rps3	*LVP*	H1	A	193	-3077	6540
rps3	*LVP*	H1	A'	192	-3078	6540
rps3	*LVP*	H2	A	193	-3077	6540
rps3	*LVP*	H2	A'	192	-3077	**6538**
rps3	-	-	M1a	198	-3202	6800
rps4	*Jasmineae*	H1	A	191	-2097	4576
rps4	*Jasmineae*	H1	A'	190	-2097	4574
rps4	*Jasmineae*	H2	A	191	-2097	4576
rps4	*Jasmineae*	H2	A'	190	-2096	**4572**
rps4	*LVP*	H1	A	193	-2096	4578
rps4	*LVP*	H1	A'	192	-2097	4578
rps4	*LVP*	H2	A	193	-2095	**4576**
rps4	*LVP*	H2	A'	192	-2097	4578
rps4	-	-	M1a	198	-2119	4634
rps7	*Jasmineae*	H1	A	193	-741	**1868**
rps7	*Jasmineae*	H1	A'	192	-742	**1868**
rps7	*Jasmineae*	H2	A	193	-741	**1868**
rps7	*Jasmineae*	H2	A'	192	-742	**1868**
rps7	*LVP*	H1	A	195	-740	1870
rps7	*LVP*	H1	A'	194	-740	**1868**
rps7	*LVP*	H2	A	195	-740	1870
rps7	*LVP*	H2	A'	194	-740	**1868**
rps7	-	-	M1a	200	-742	1884
rps10	*Jasmineae*	H1	A	191	-966	2314
rps10	*Jasmineae*	H1	A'	190	-966	**2312**
rps10	*Jasmineae*	H2	A	191	-966	2314
rps10	*Jasmineae*	H2	A'	190	-966	**2312**
rps10	*LVP*	H1	A	193	-946	2278
rps10	*LVP*	H1	A'	192	-946	**2276**
rps10	*LVP*	H2	A	193	-946	2278
rps10	*LVP*	H2	A'	192	-946	**2276**
rps10	-	-	M1a	198	-966	2328
rps12	*Jasmineae*	H1	A	191	-711	1804
rps12	*Jasmineae*	H1	A'	190	-711	**1802**
rps12	*Jasmineae*	H2	A	191	-711	1804
rps12	*Jasmineae*	H2	A'	190	-711	**1802**
rps12	*LVP*	H1	A	193	-729	1844
rps12	*LVP*	H1	A'	192	-728	**1840**
rps12	*LVP*	H2	A	193	-728	1842
rps12	*LVP*	H2	A'	192	-728	**1840**
rps12	-	-	M1a	198	-736	1868

Table S4.B End

gene	clade	hypothesis	model	npar	lnL	AIC
rps13	*Jasmineae*	H1	A	193	-808	2002
rps13	*Jasmineae*	H1	A'	192	-808	**2000**
rps13	*Jasmineae*	H2	A	193	-808	2002
rps13	*Jasmineae*	H2	A'	192	-808	**2000**
rps13	*LVP*	H1	A	195	-772	1934
rps13	*LVP*	H1	A'	194	-774	1936
rps13	*LVP*	H2	A	195	-772	1934
rps13	*LVP*	H2	A'	194	-772	**1932**
rps13	-	-	M1a	200	-822	2044
rps14	*Jasmineae*	H1	A	193	-478	1342
rps14	*Jasmineae*	H1	A'	192	-478	**1340**
rps14	*Jasmineae*	H2	A	193	-478	1342
rps14	*Jasmineae*	H2	A'	192	-478	**1340**
rps14	*LVP*	H1	A	195	-470	1330
rps14	*LVP*	H1	A'	194	-470	**1328**
rps14	*LVP*	H2	A	195	-470	1330
rps14	*LVP*	H2	A'	194	-470	**1328**
rps14	-	-	M1a	200	-478	1356
sdh3	*Jasmineae*	H1	A	189	-633	1644
sdh3	*Jasmineae*	H1	A'	188	-633	**1642**
sdh3	*Jasmineae*	H2	A	189	-633	1644
sdh3	*Jasmineae*	H2	A'	188	-633	**1642**
sdh3	*LVP*	H1	A	NA	-651	1684
sdh3	*LVP*	H1	A'	NA	-651	1682
sdh3	*LVP*	H2	A	NA	-646	**1674**
sdh3	*LVP*	H2	A'	NA	-651	1682
sdh3	-	-	M1a	196	-6559	13510
sdh4	*Jasmineae*	H1	A	193	-1487	3360
sdh4	*Jasmineae*	H1	A'	192	-1487	**3358**
sdh4	*Jasmineae*	H2	A	193	-1488	3362
sdh4	*Jasmineae*	H2	A'	192	-1487	**3358**
sdh4	*LVP*	H1	A	195	-1504	3398
sdh4	*LVP*	H1	A'	194	-1509	3406
sdh4	*LVP*	H2	A	195	-1504	3398
sdh4	*LVP*	H2	A'	194	-1504	**3396**
sdh4	-	-	M1a	200	-1519	3438

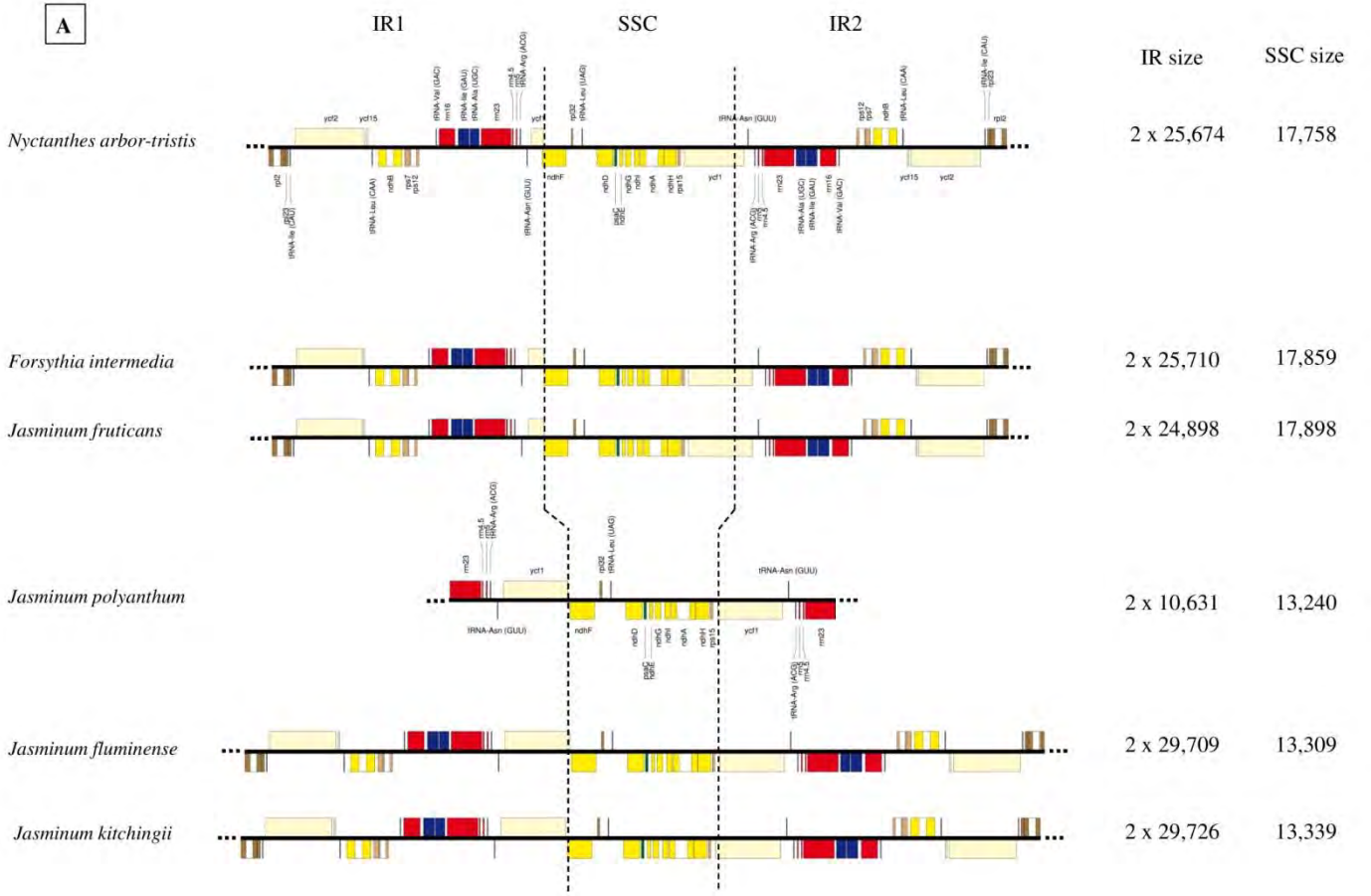


Fig. S1. Physical maps of the inverted repeats (IR1 and IR2) and the small single copy (SSC) of the plastomes of *Nyctanthes arbor-tristis*, *Forsythia intermedia* and species belonging to clades *Jasmineae* (Figure S1.A) and *Ligustrinae* (Figure S1.B). These maps were drawn with OGDRAW (Lohse *et al.* 2007). On the right, IR and SSC sizes are given in bp for each species.

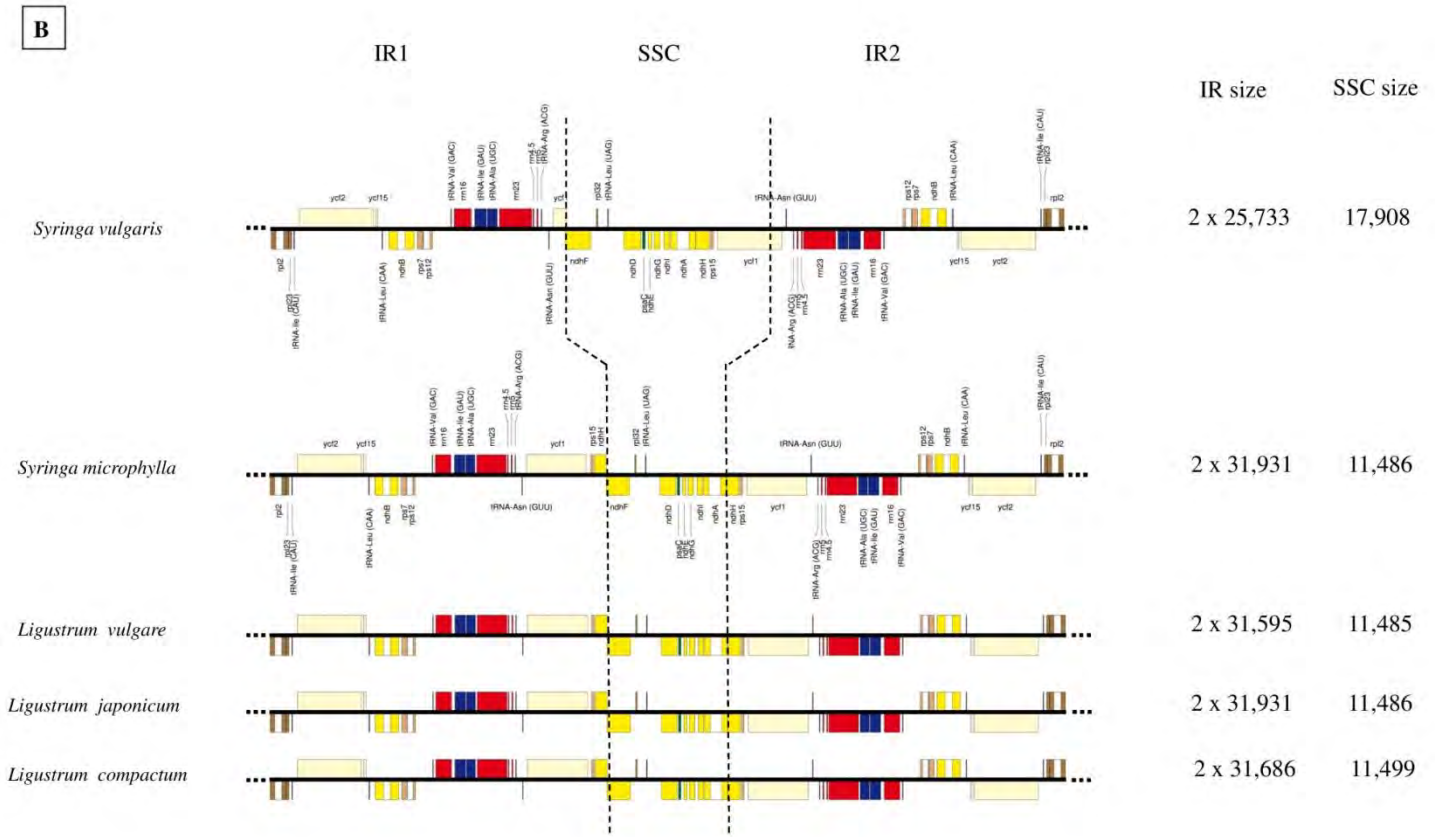

B
IR1
SSC
IR2
IR size
SSC size
Syringa vulgaris
2 x 25,733
17,908
Syringa microphylla
2 x 31,931
11,486
Ligustrum vulgare
2 x 31,595
11,485
Ligustrum japonicum
2 x 31,931
11,486
Ligustrum compactum
2 x 31,686
11,499

Fig. S1. End.

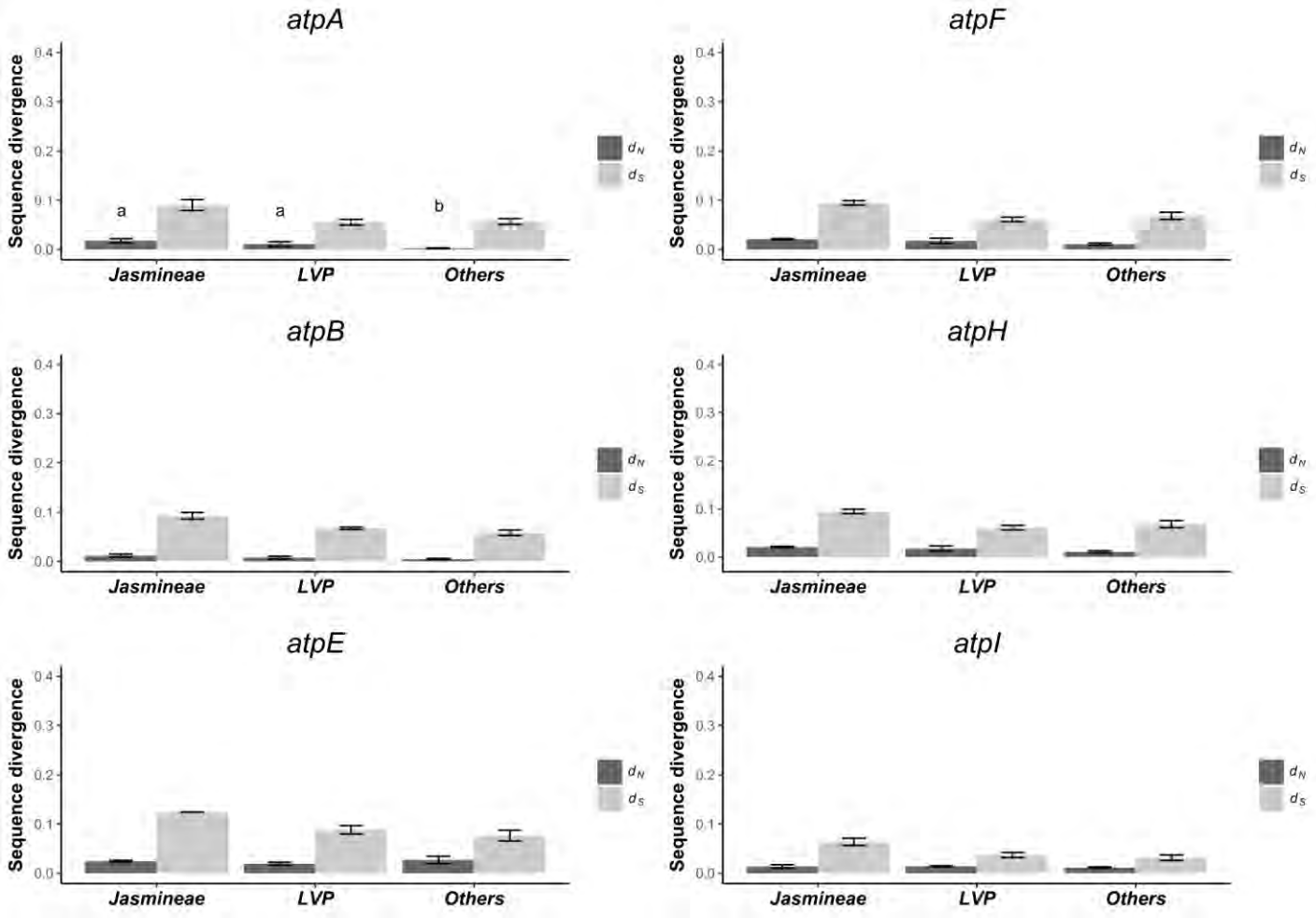


Fig. S2.A Mean substitutions rates among the 80 cpDNA genes for clade *Jasmineae* (noted '*Jasmineae*'), clade (noted '*LVP*') and for Oleaceae species without clades *Jasmineae* and *LVP* (noted 'Others'). For each gene, non-synonymous (d_N) and synonymous (d_S) were calculated against the outgroup species (*Nyctanthes arbor-tristis*) using PAML v.4 (Yang 2007). The significant differences between groups are indicated by Latin letters for d_N and by Greek letters for d_S.

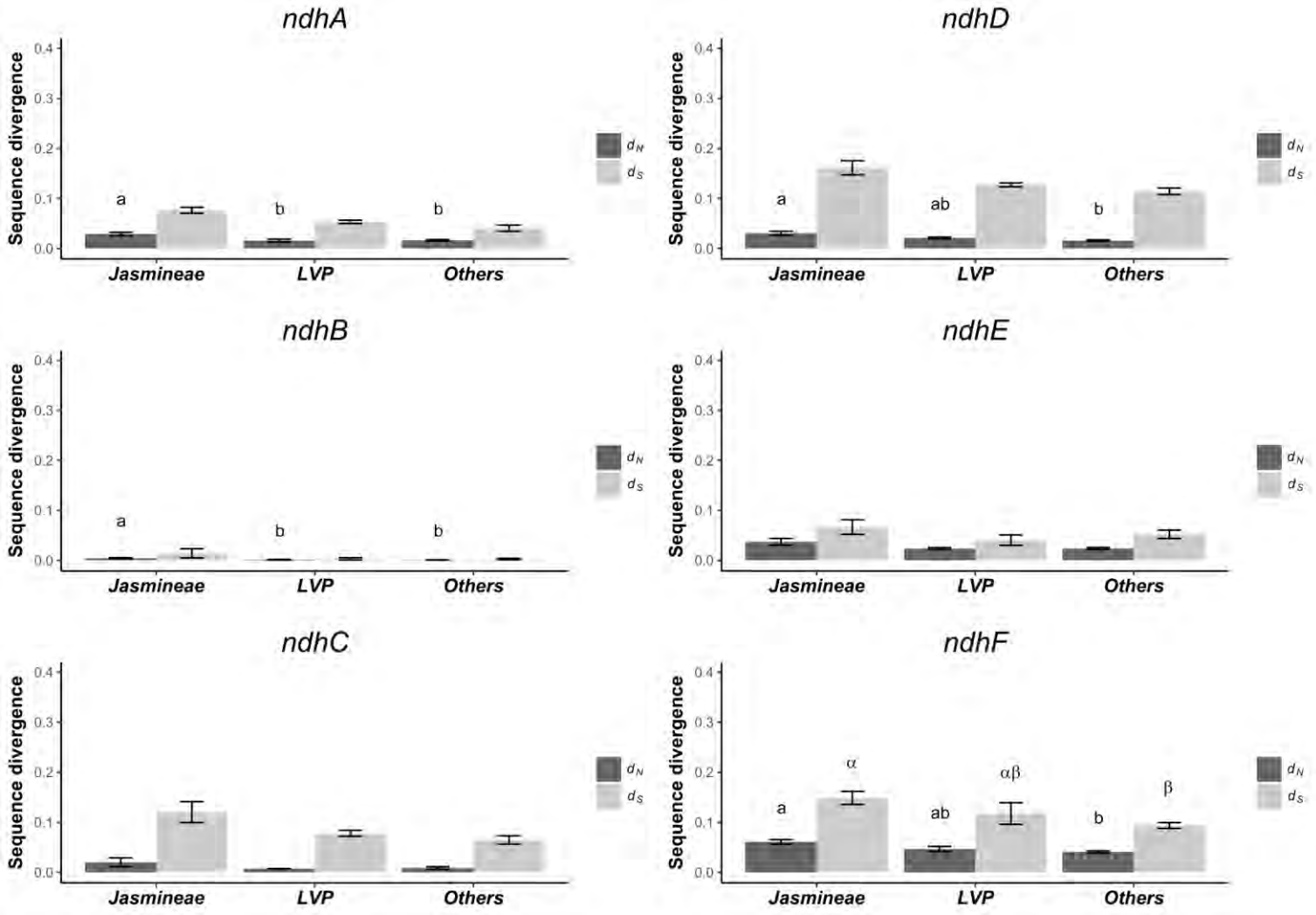


Fig. S2.A Continued

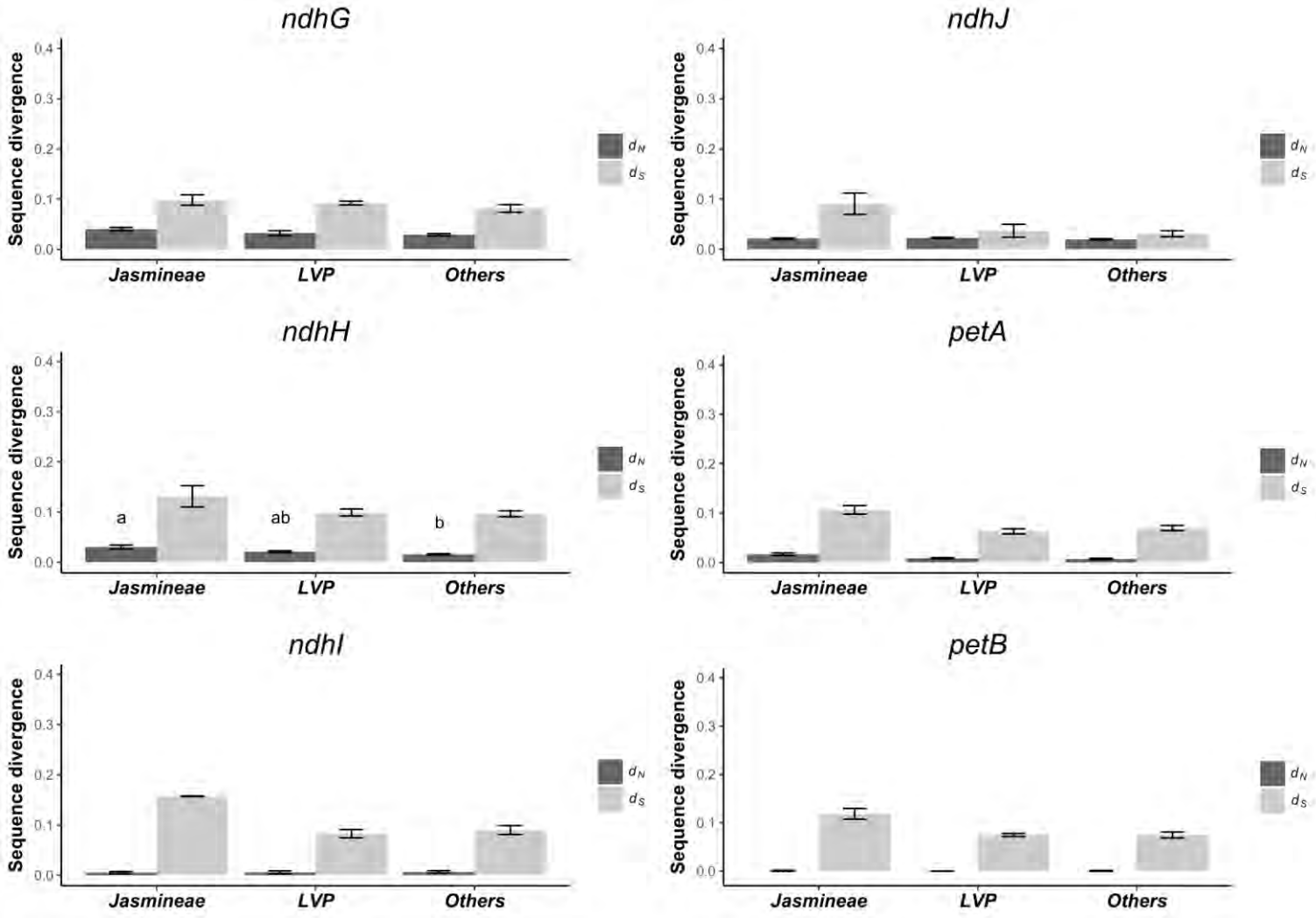


Fig. S2.A Continued

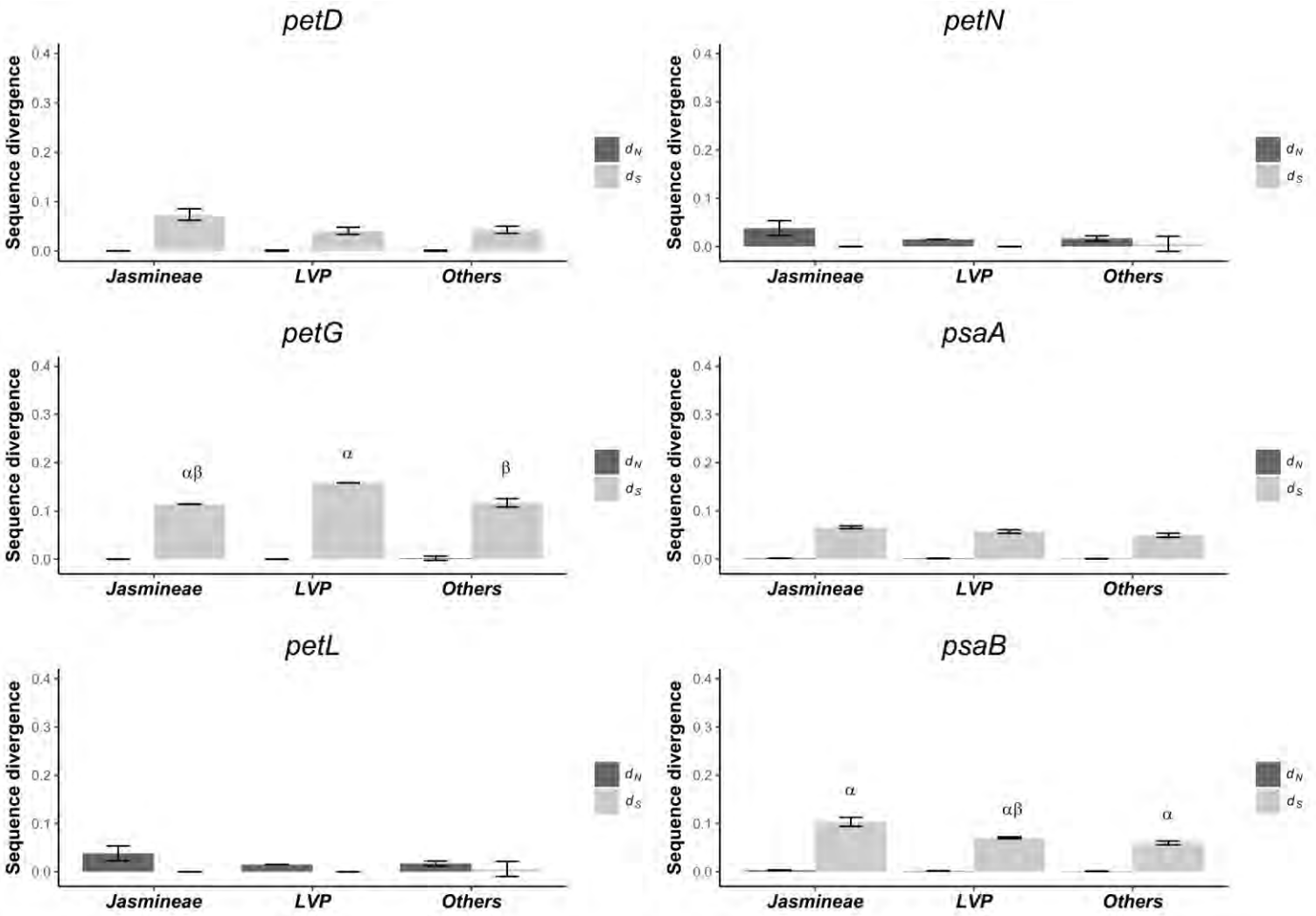


Fig. S2.A Continued

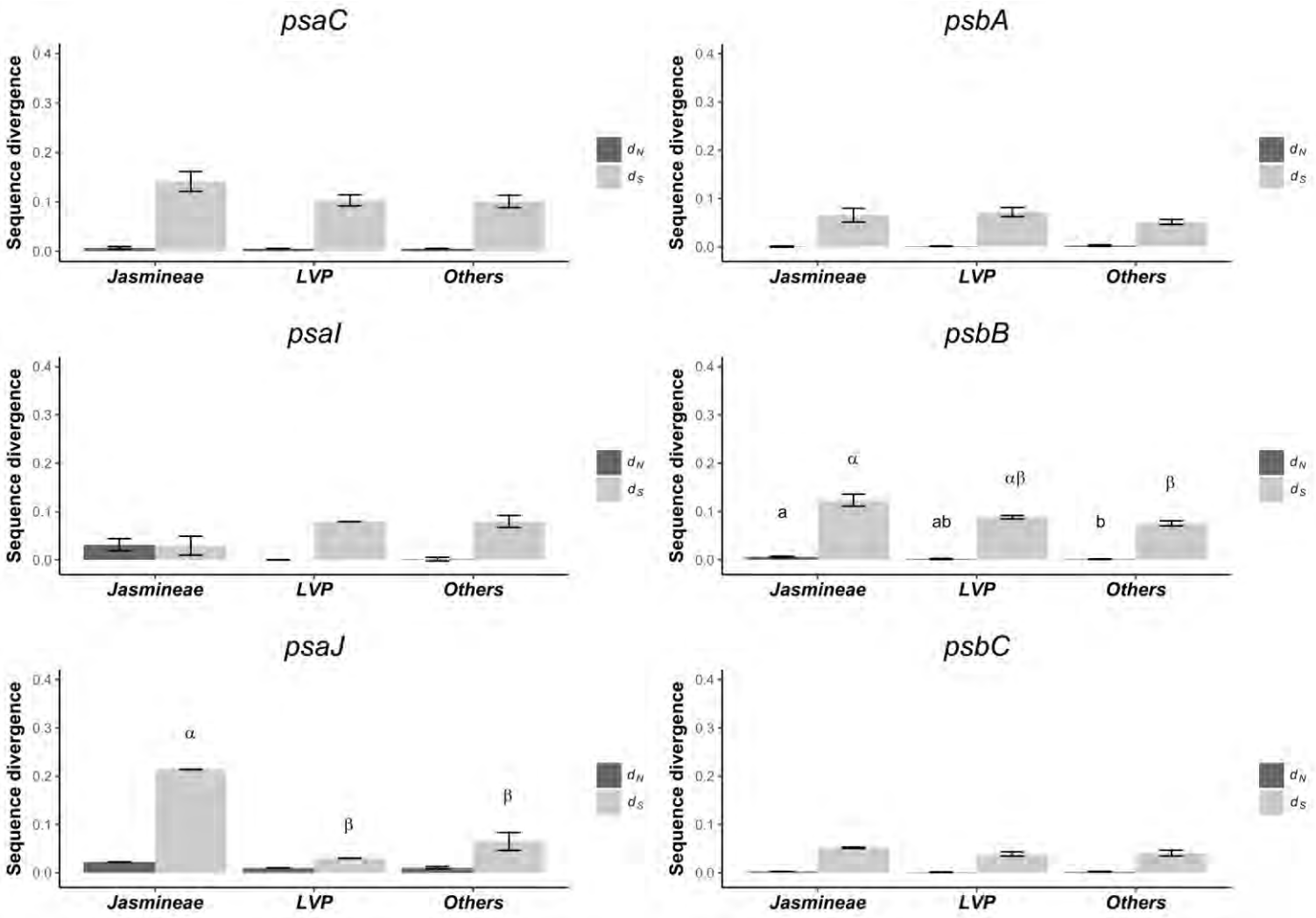


Fig. S2.A Continued

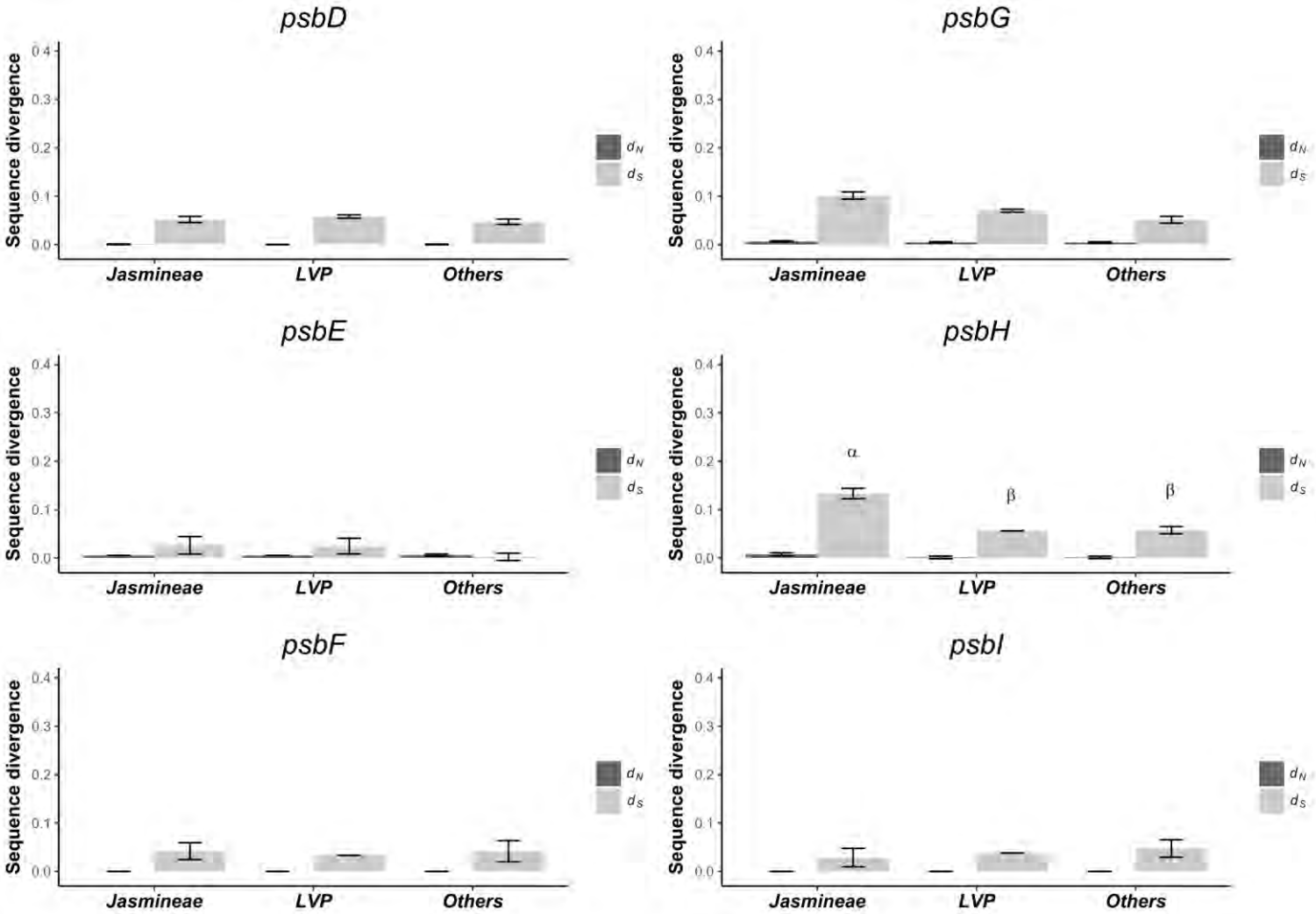


Fig. S2.A Continued

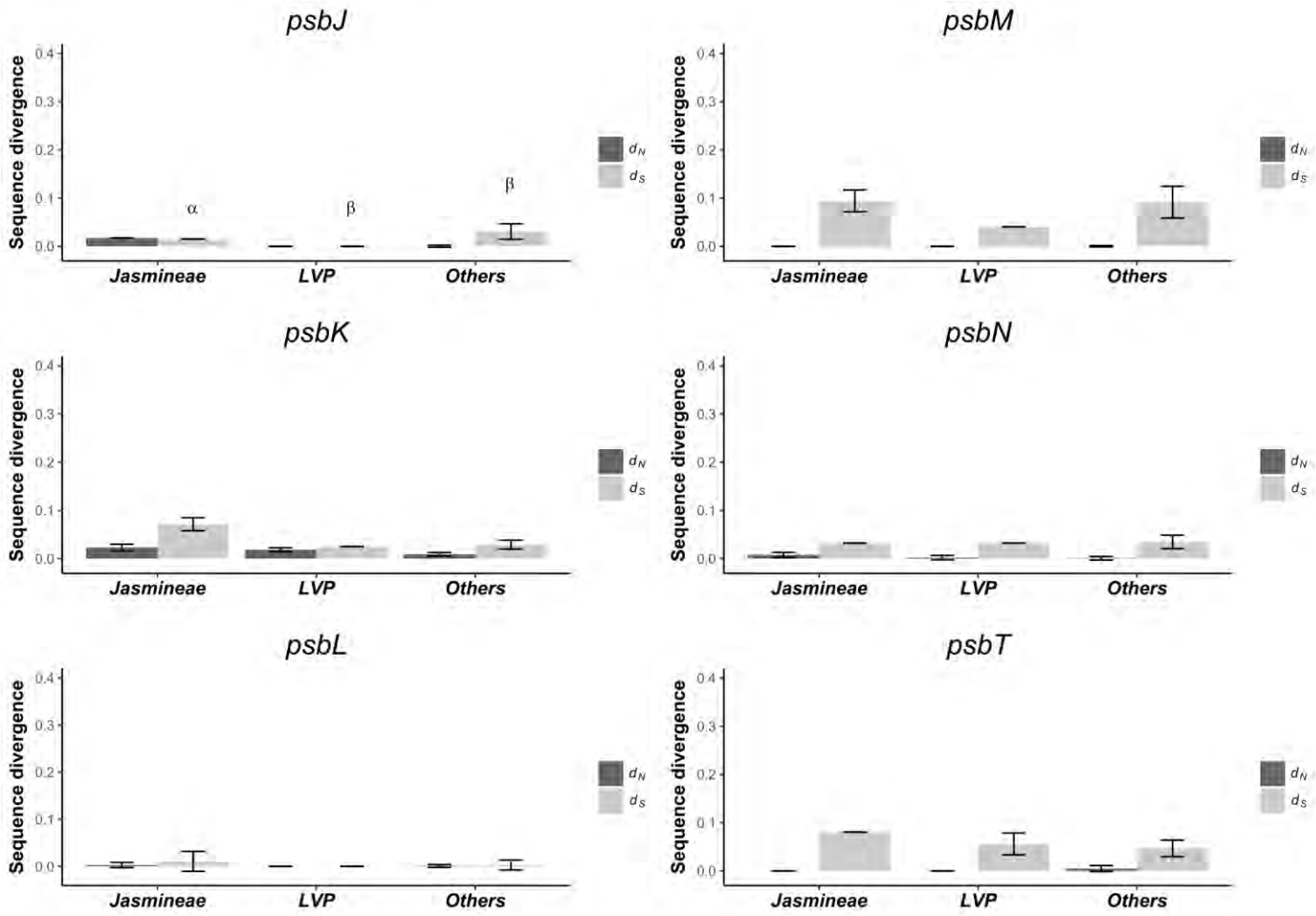


Fig. S2.A Continued

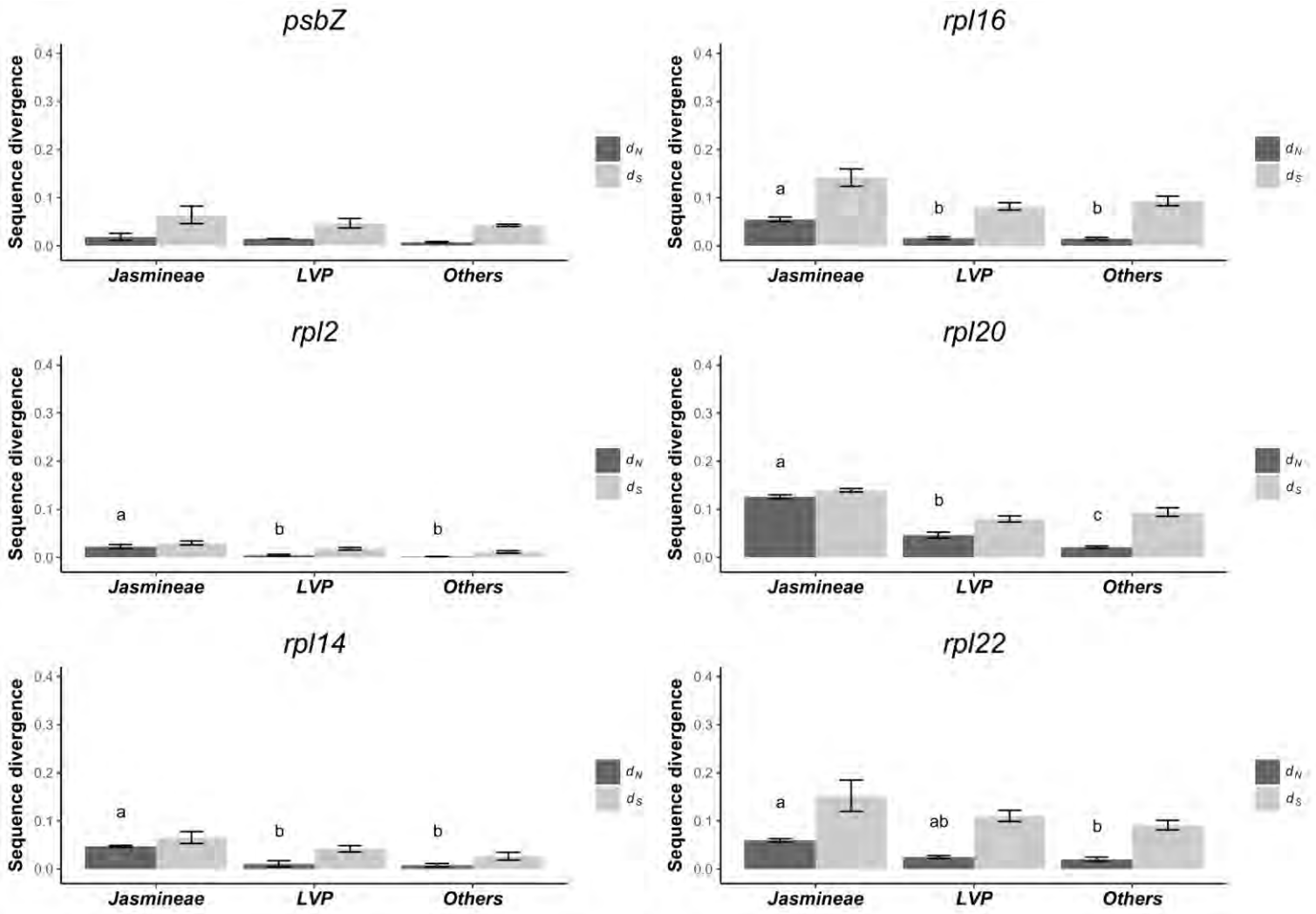


Fig. S2.A Continued

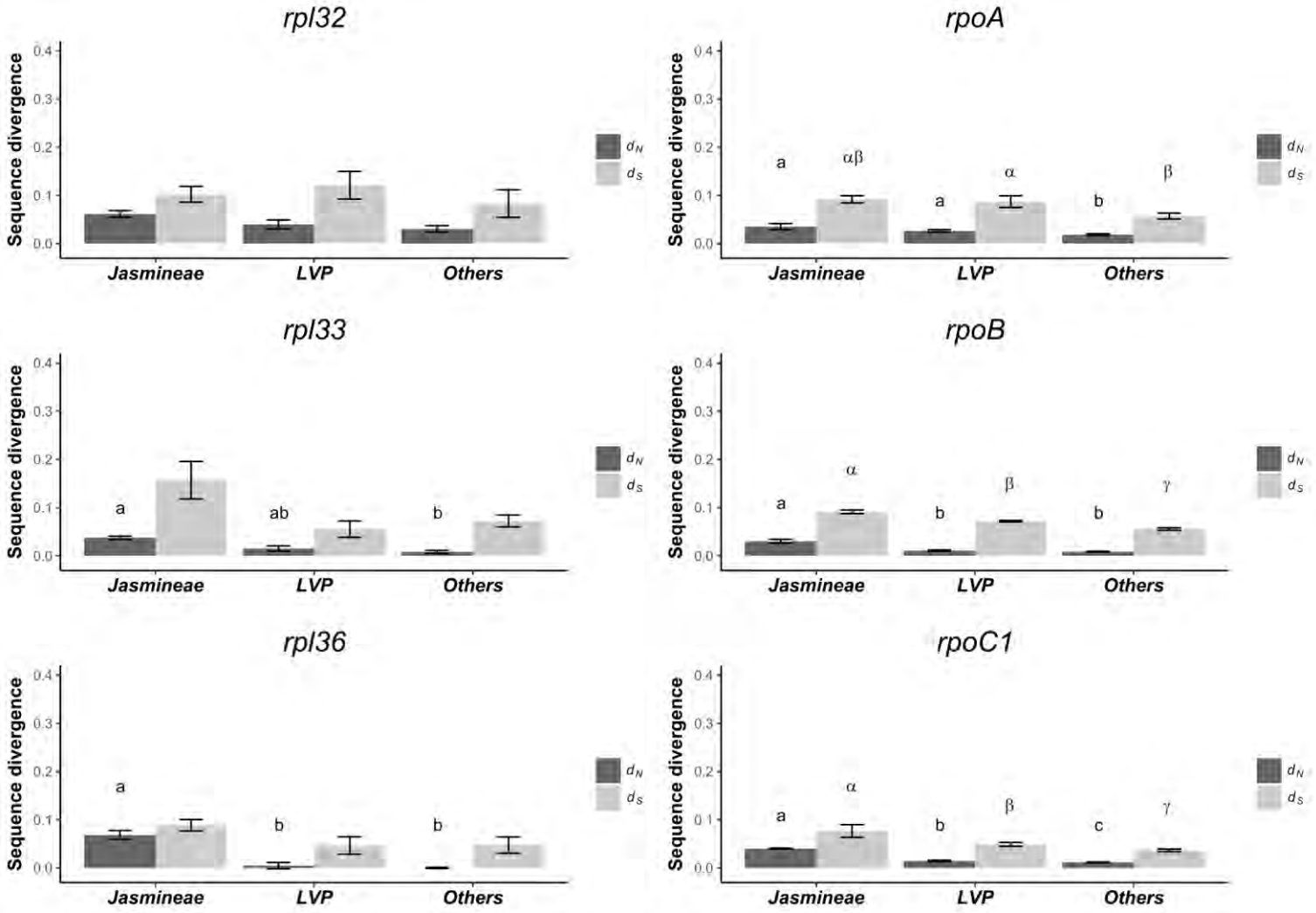


Fig. S2.A Continued

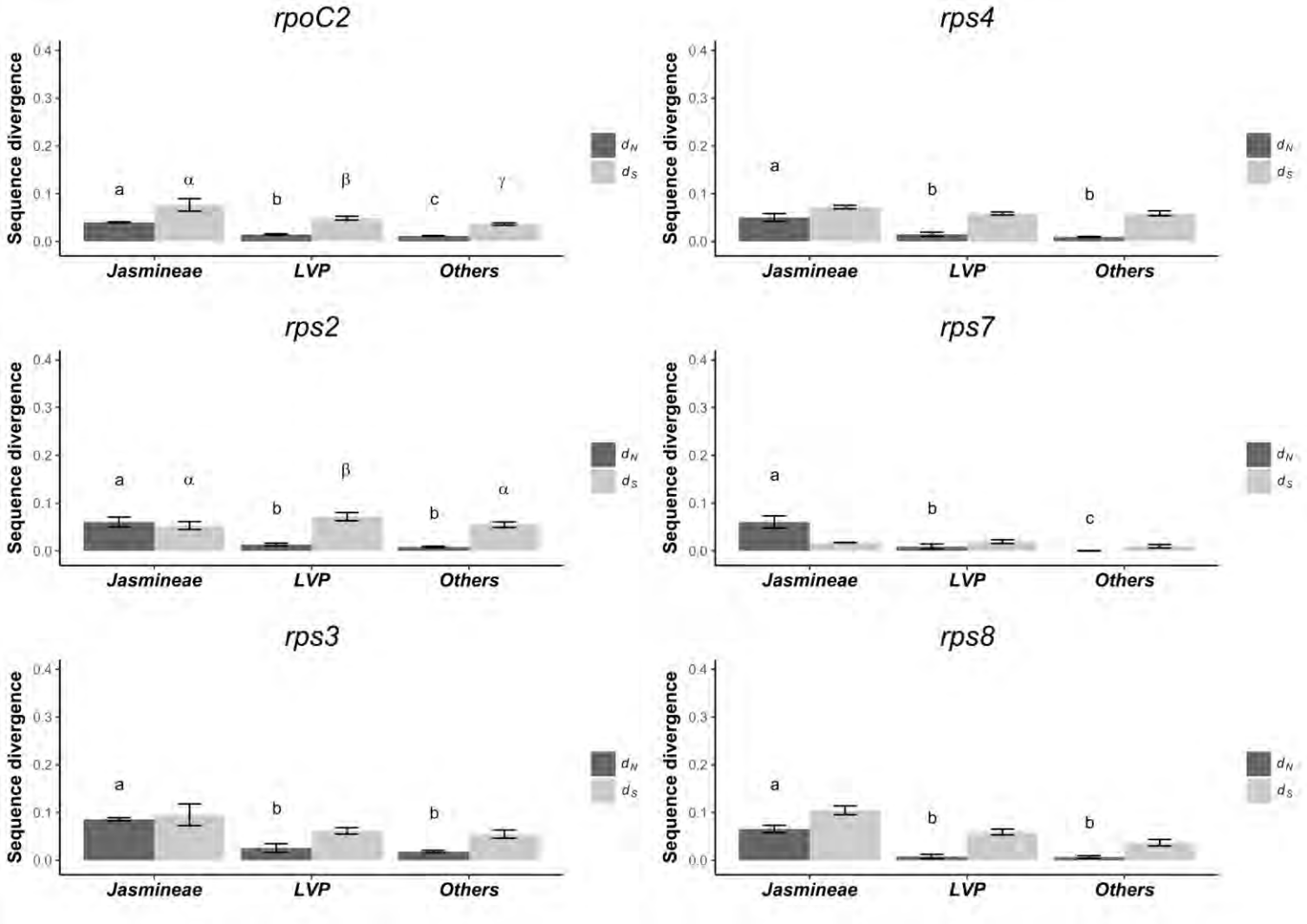


Fig. S2.A Continued

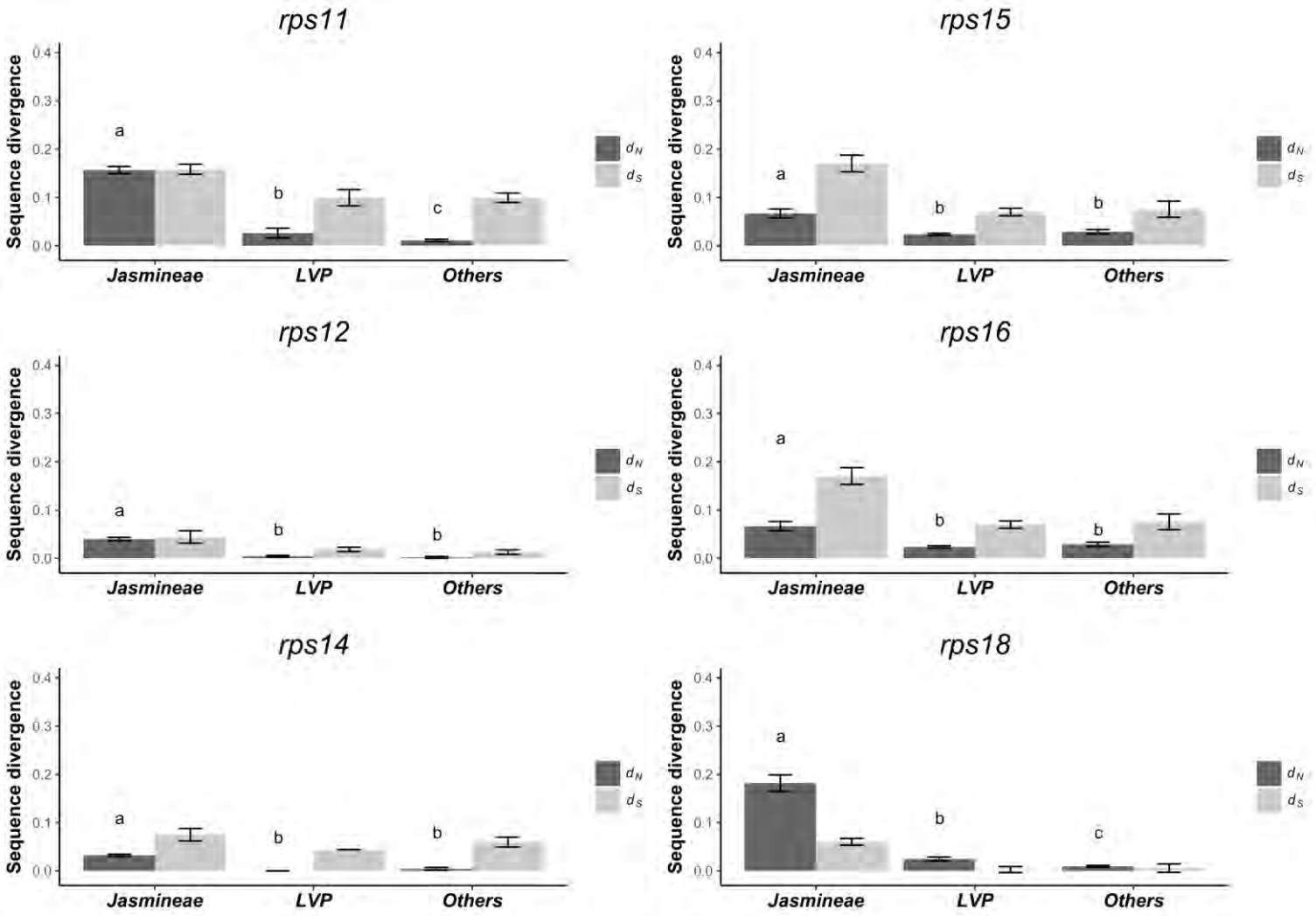


Fig. S2.A Continued

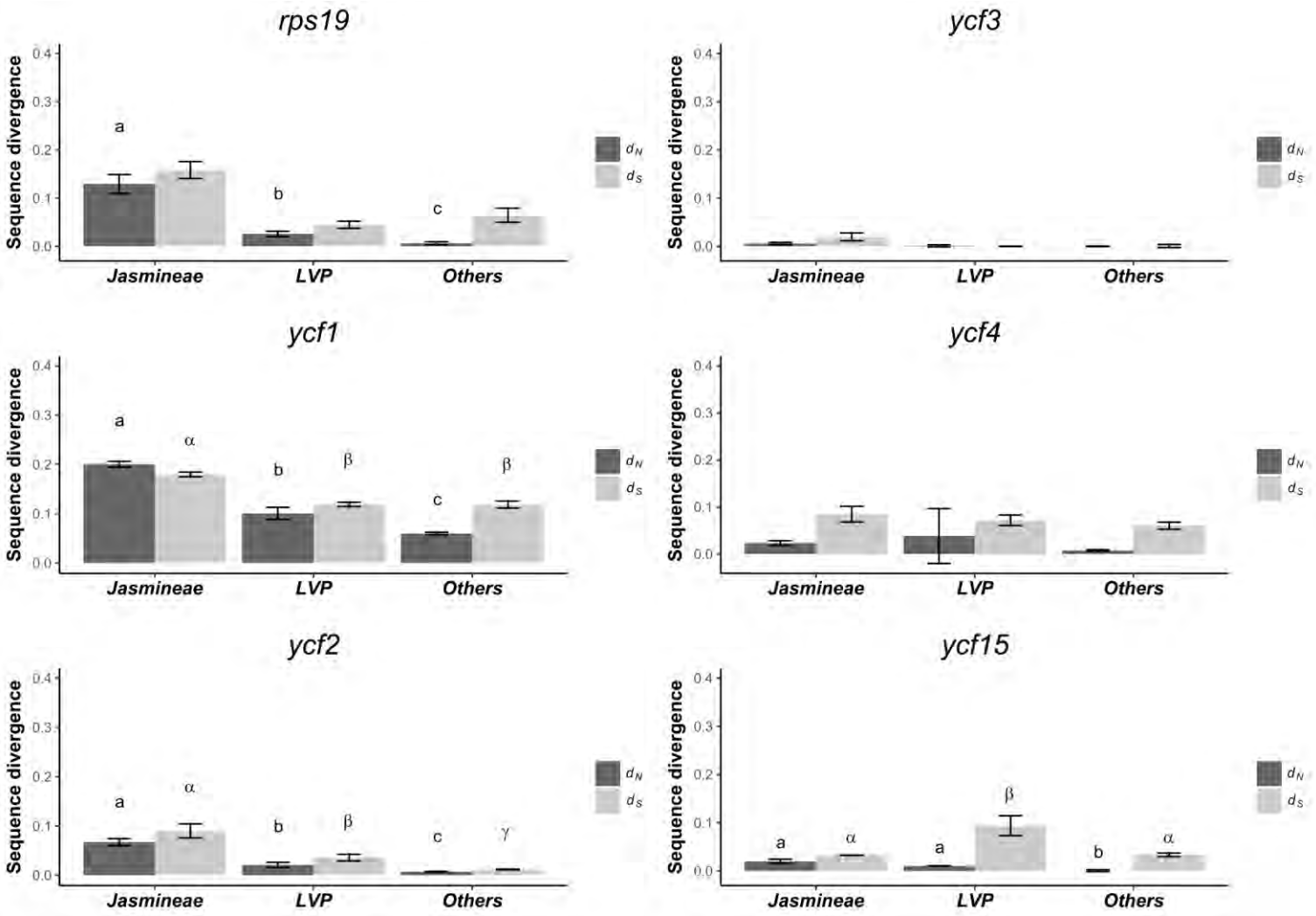


Fig. S2.A Continued

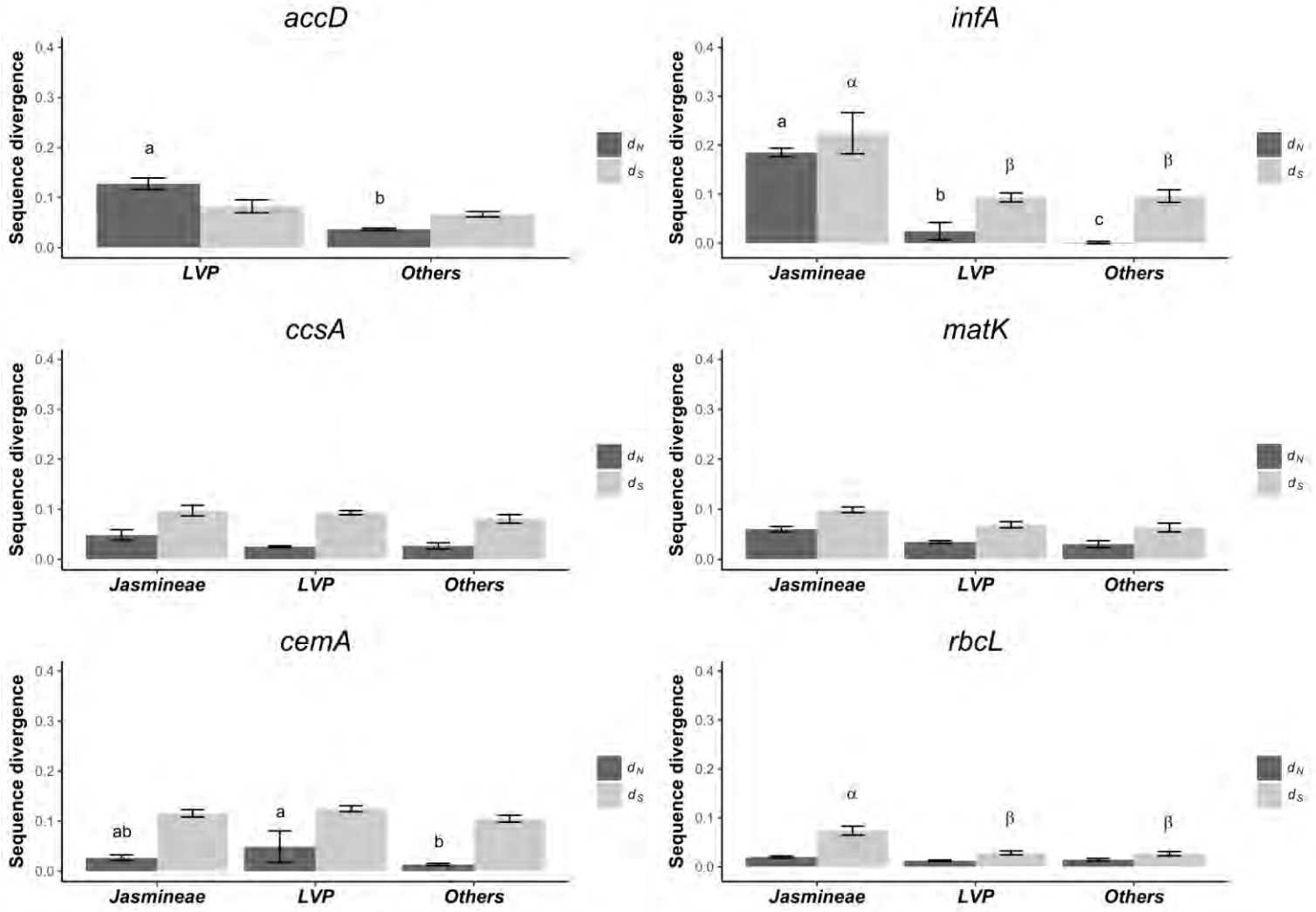


Fig. S2.A Continued

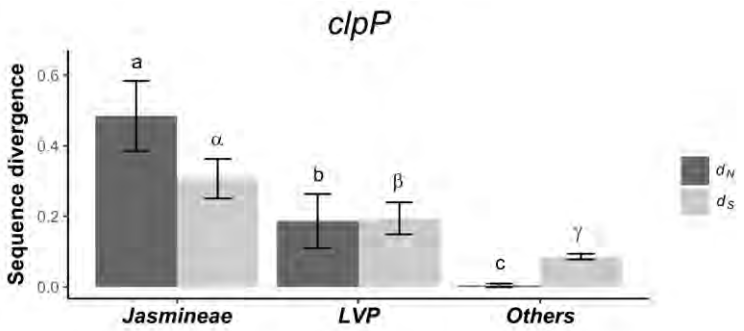

clpP
Sequence divergence
0.6
0.4
0.2
0.0
a
α
b
β
c
γ
Jasmineae
LVP
Others
d_N
d_S

Fig. S2.A End

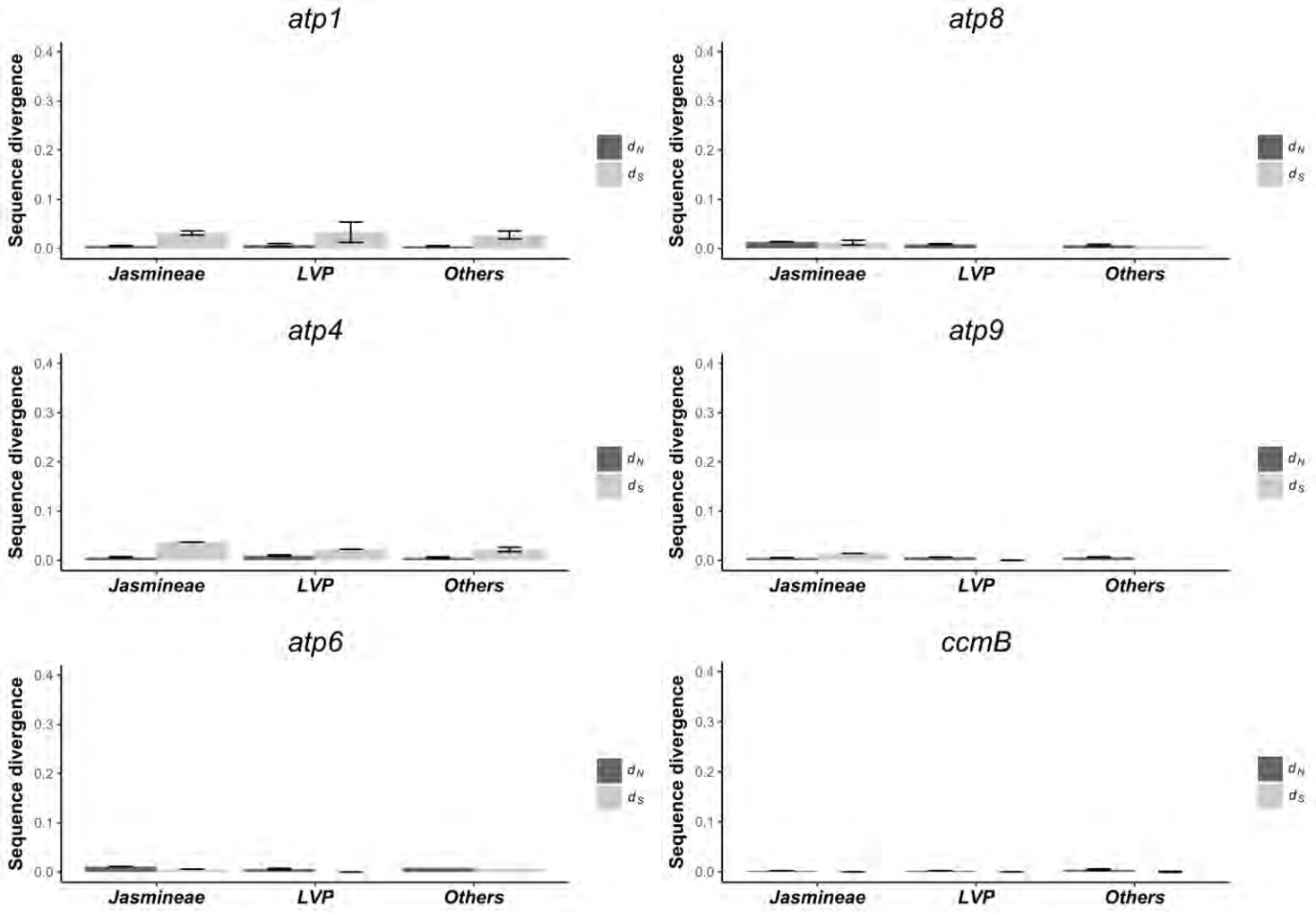


Fig. S2.B Mean substitutions rates among the 36 mtDNA genes for clade *Jasmineae* (noted '*Jasmineae*'), clade (noted '*LVP*') and for Oleaceae species without clades *Jasmineae* and *LVP* (noted 'Others'). For each gene, non-synonymous (d_N) and synonymous (d_S) were calculated against the outgroup species (*Nyctanthes arbor-tristis*) using PAML v.4 (Yang 2007). The significant differences between groups are indicated by Latin letters for d_N and by Greek letters for d_S.

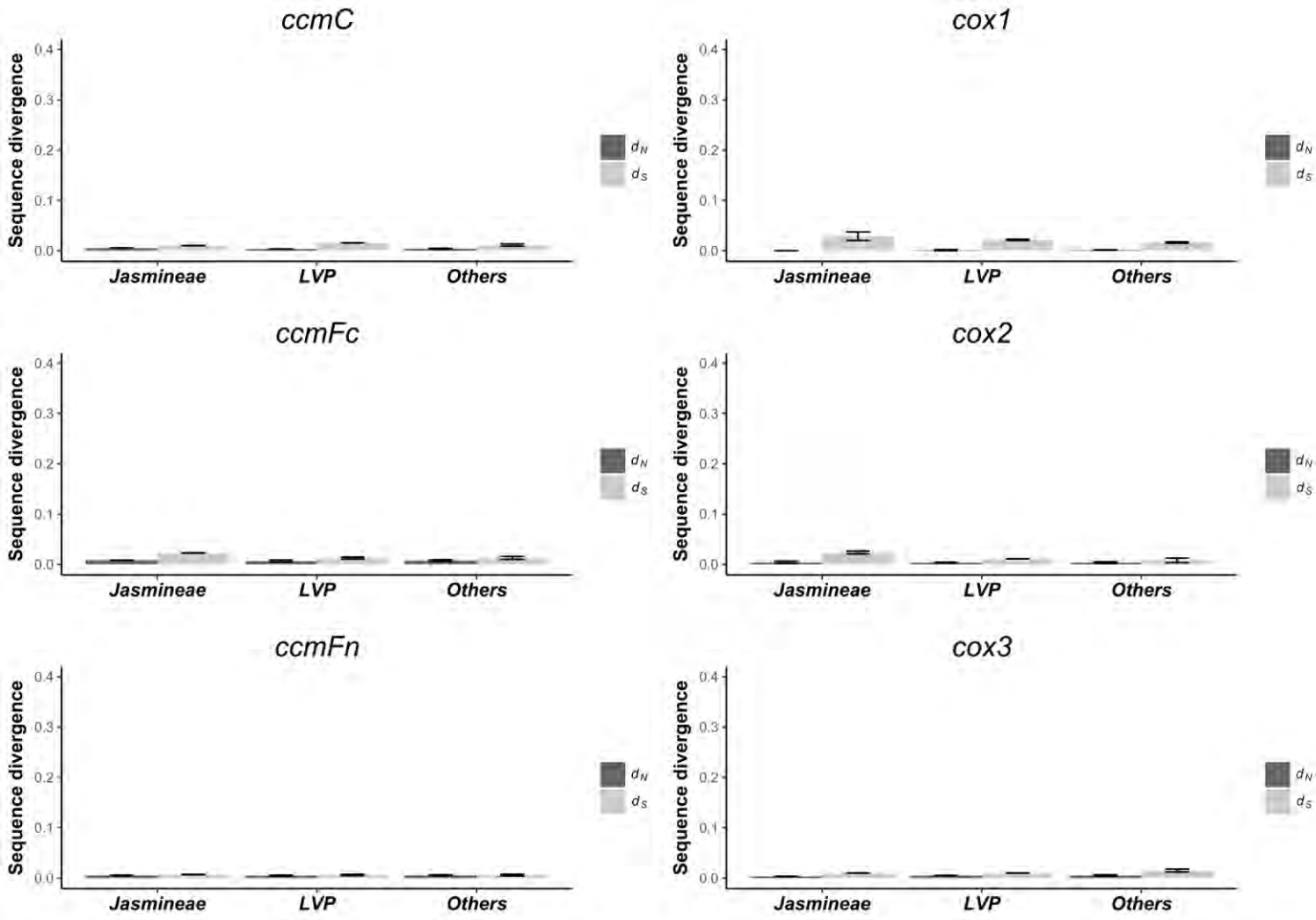


Fig. S2.B Continued

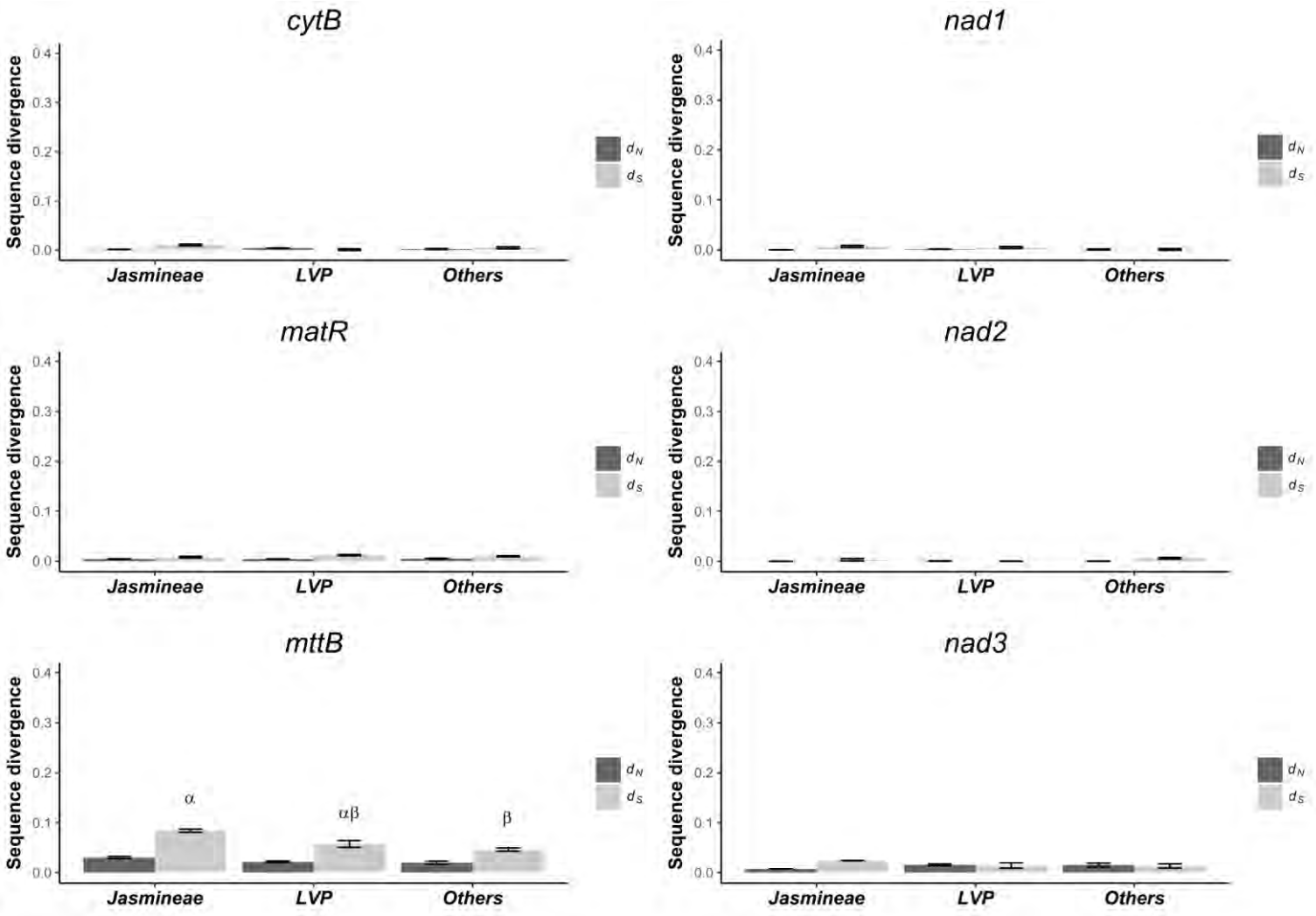


Fig. S2.B Continued

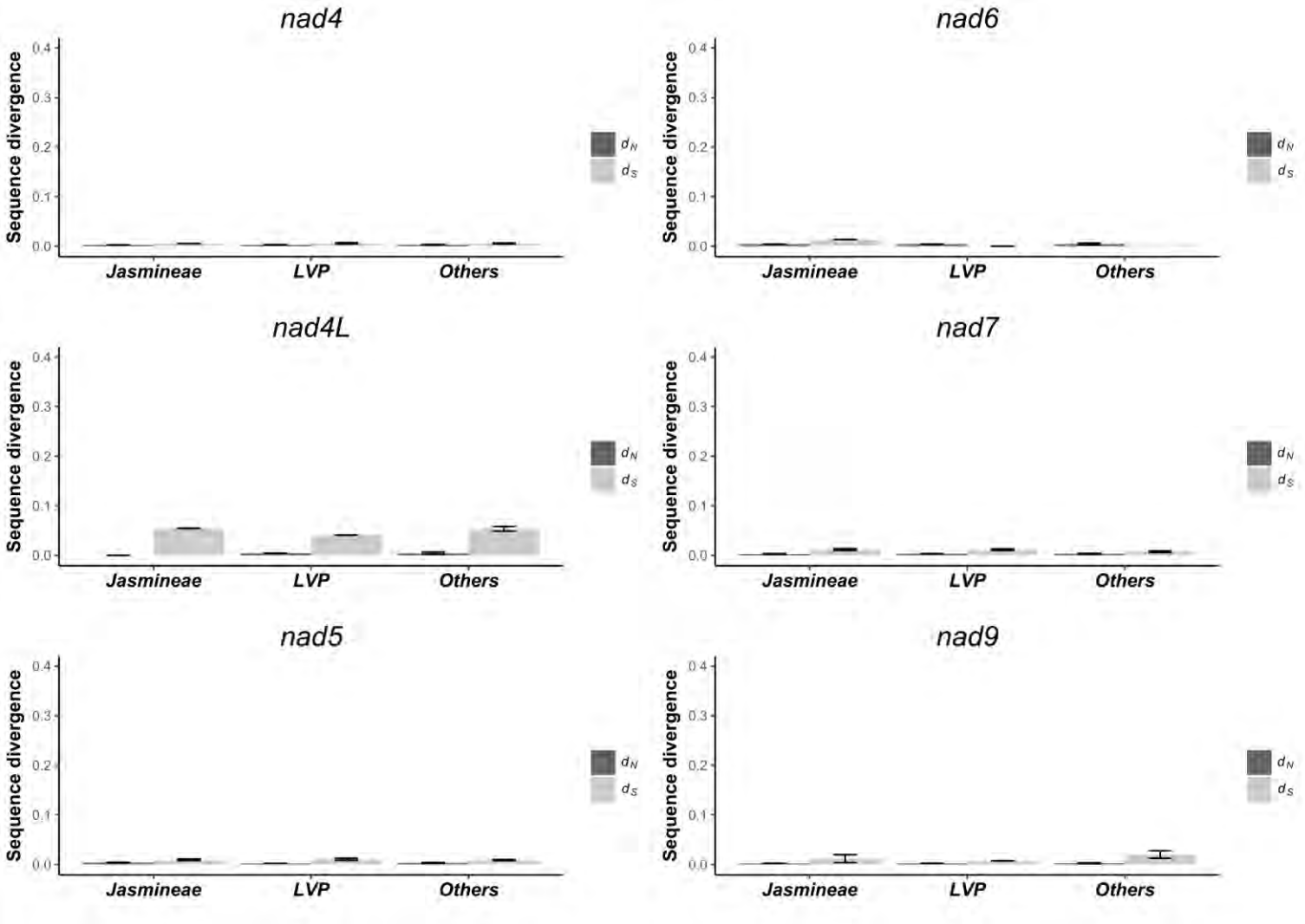


Fig. S2.B Continued

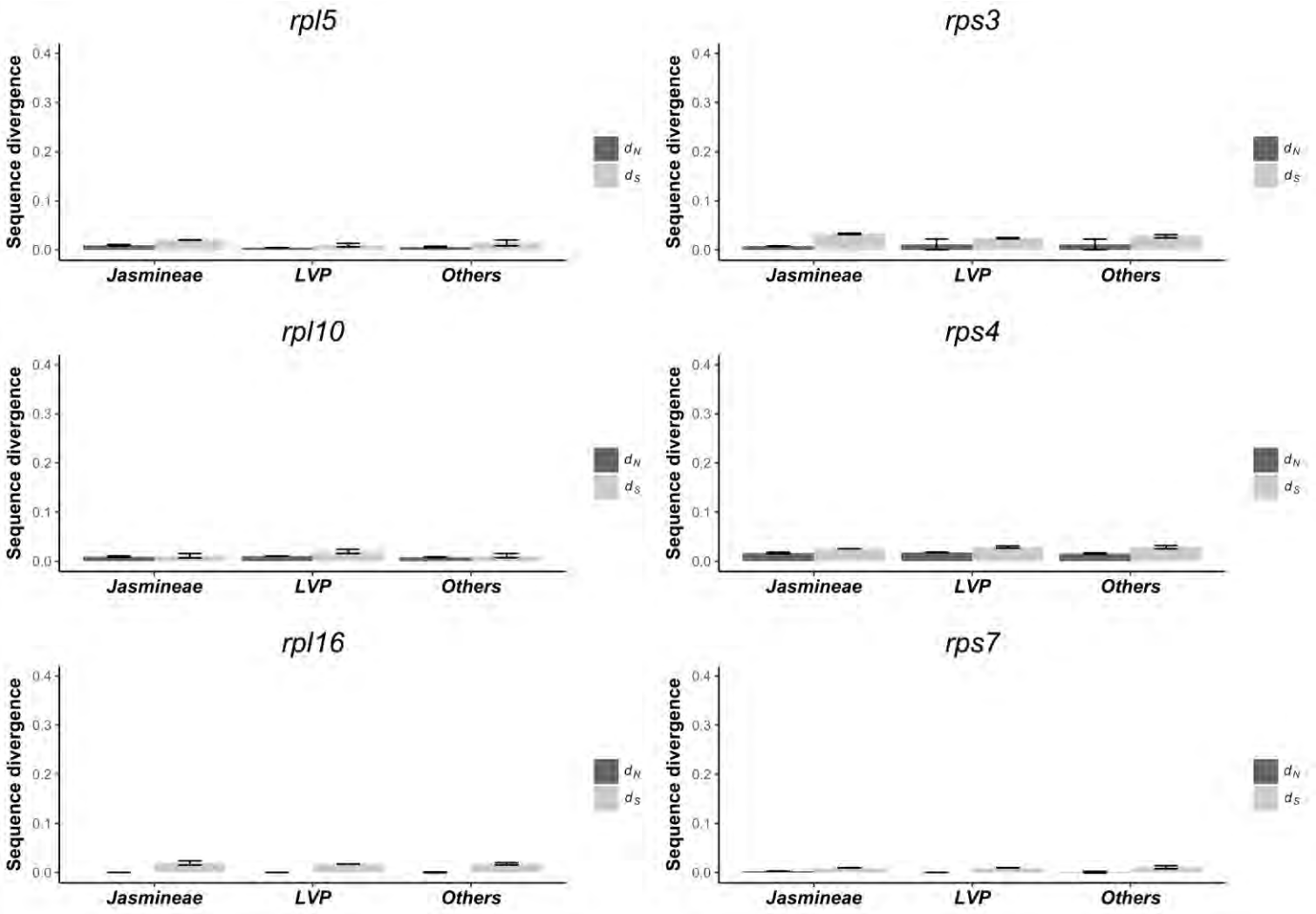


Fig. S2.B Continued

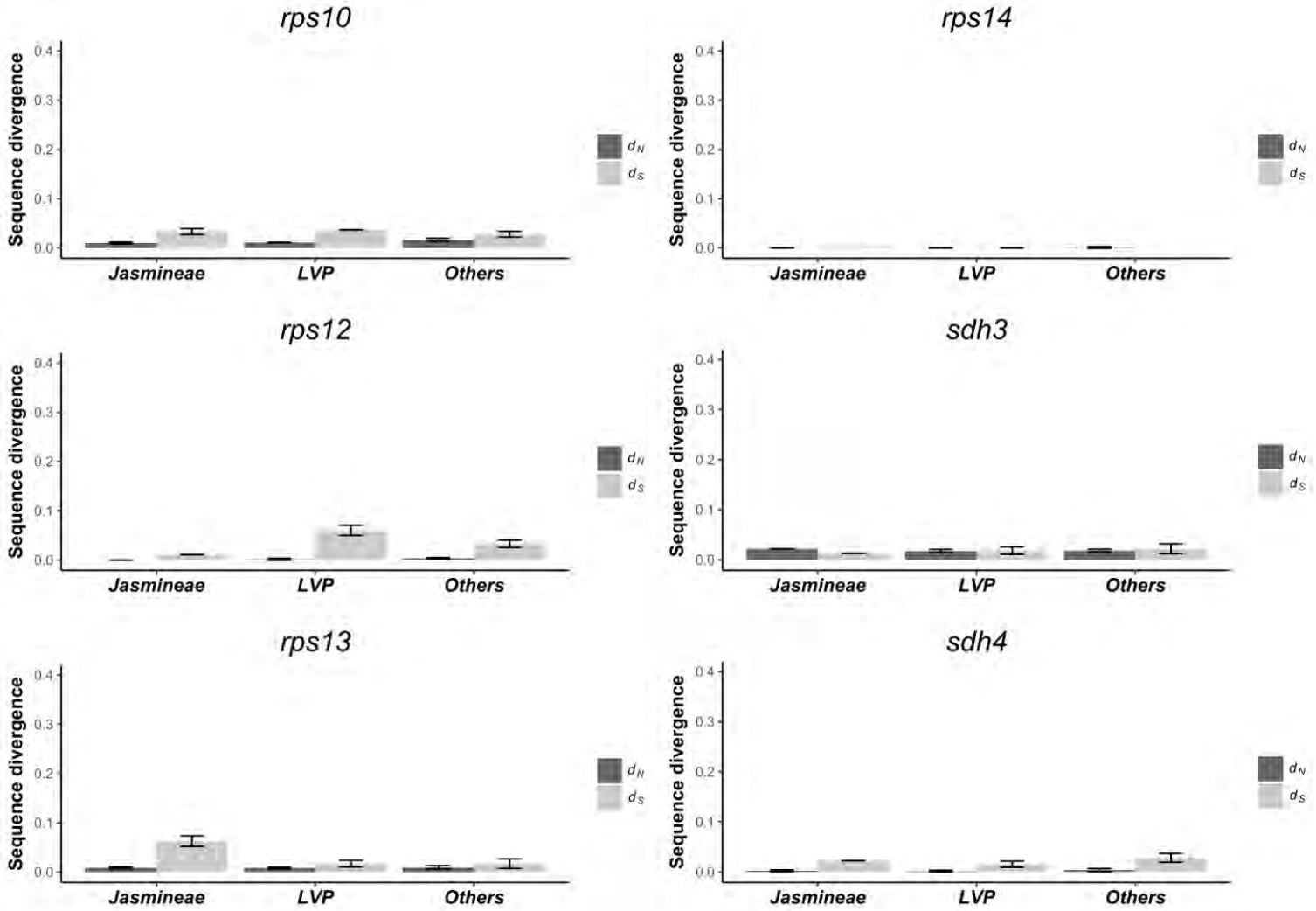

rps10
rps14
rps12
sdh3
rps13
sdh4
Sequence divergence
0.0
0.1
0.2
0.3
0.4
Jasmineae
LVP
Others
d_N
d_S

Fig. S2.B End

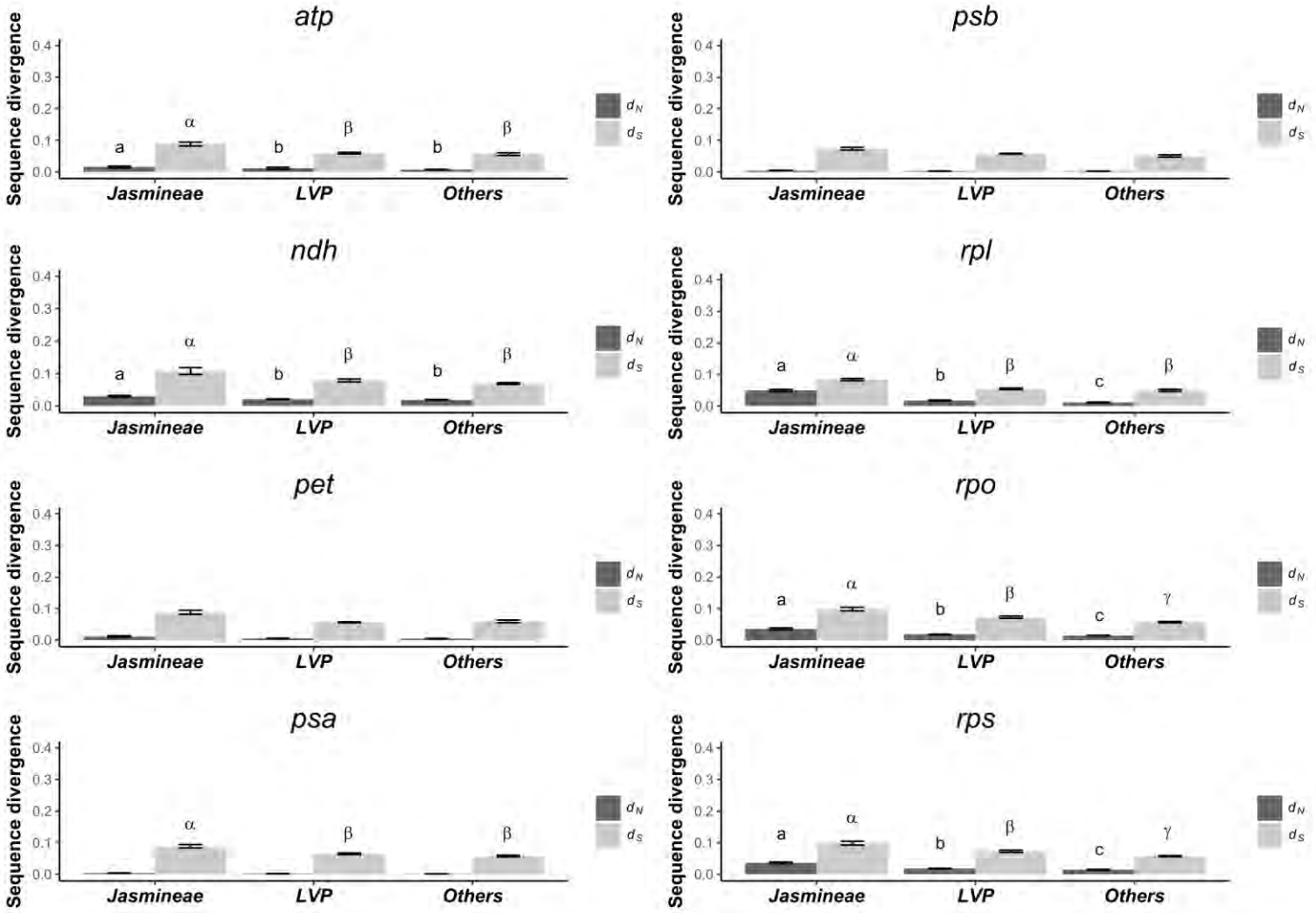


Fig. S3.A Mean substitutions rates among the eight cpDNA functional categories for clade *Jasmineae* (noted '*Jasmineae*'), clade *LVP* and for Oleaceae species without clades *Jasmineae* and *LVP* (noted 'Other'). For each functional category, non-synonymous (d_N) and synonymous (d_S) were calculated against the outgroup species (*Nyctanthes arbor-tristis*) using PAML v.4 (Yang 2007). The significant differences between groups are indicated by Latin letters for d_N and by Greek letters for d_S.

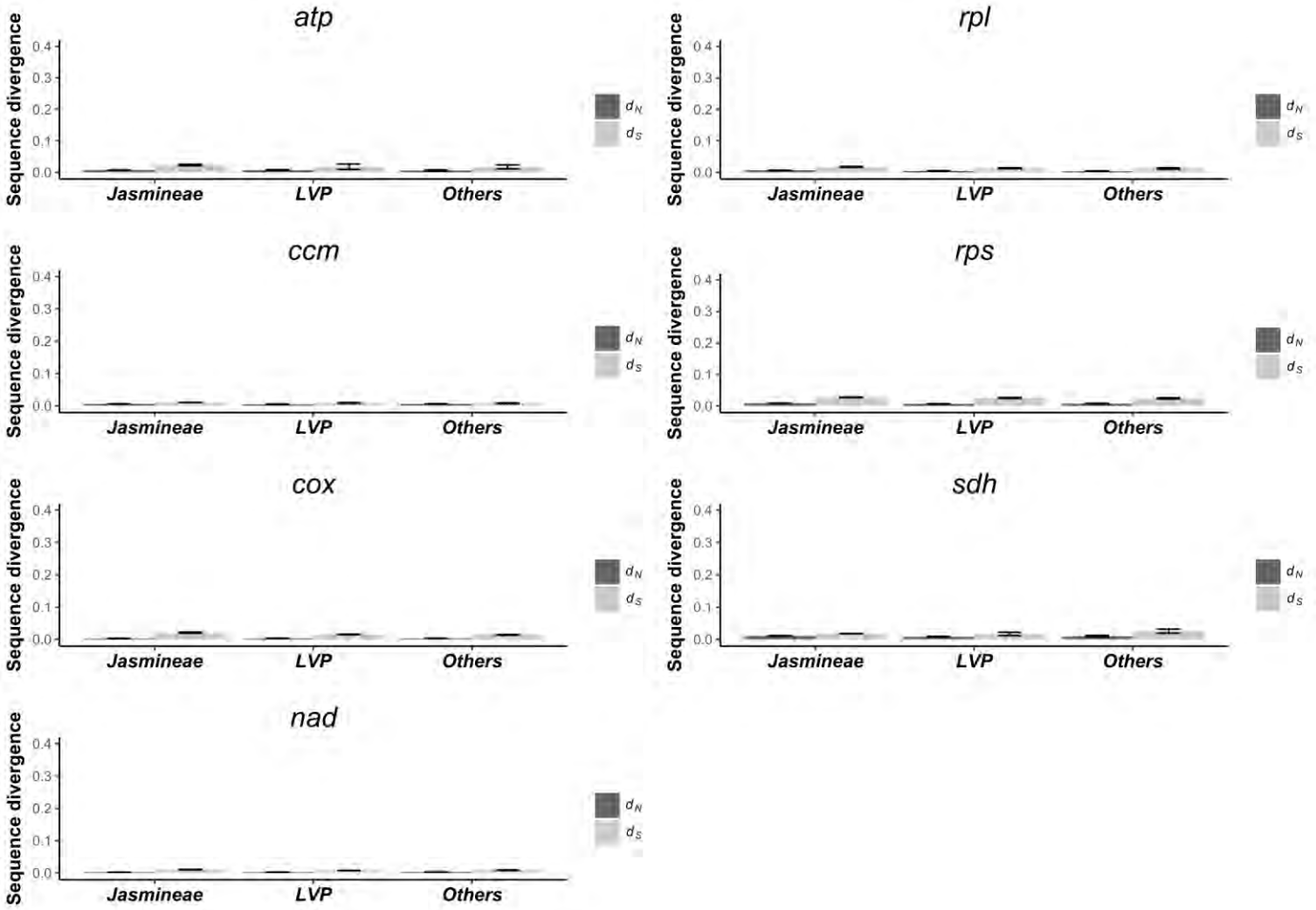


Fig. S3.B Mean substitutions rates among the seven mtDNA functional categories for clade *Jasmineae* (noted '*Jasmineae*'), clade *LVP* and for Oleaceae species without clades *Jasmineae* and *LVP* (noted 'Other'). For each functional category, non-synonymous (d_N) and synonymous (d_S) were calculated against the outgroup species (*Nyctanthes arbor-tristis*) using PAML v.4 (Yang 2007). The significant differences between groups are indicated by Latin letters for d_N and by Greek letters for d_S.

References

Lohse M, Drechsel O, Bock R (2007) OrganellarGenomeDRAW (OGDRAW): A tool for the easy generation of high-quality custom graphical maps of plastid and mitochondrial genomes. *Current Genetics*, **52**, 267–274.

Yang Z (2007) PAML 4: phylogenetic analysis by maximum likelihood. *Molecular Biology and Evolution*, **24**, 1586–1591.

ANNEXE IV

Diversity and evolution of plastomes in Saharan mimosoids: potential use for phylogenetic and population studies

MENSOUS M, VAN DE PAER C, MANZI S, BOUCHEZ O, BAALI-CHERIF D, BESNARD G

Tree Genetics & Genome, 2017, **13**: 48

Tree Genetics & Genomes (2017) 13:48
DOI 10.1007/s11295-017-1131-2

ORIGINAL ARTICLE

Diversity and evolution of plastomes in Saharan mimosoids: potential use for phylogenetic and population genetic studies

Mohamed Mensous[1] · Céline Van de Paer[2] · Sophie Manzi[2] · Olivier Bouchez[3,4] · Djamel Baâli-Cherif[5] · Guillaume Besnard[2]

Received: 22 December 2016 / Revised: 1 March 2017 / Accepted: 6 March 2017

Abstract Acacias (Mimosoideae) represent a major woody group in arid and subarid habitats of all tropical and subtropical regions. The genetic diversity and population dynamic of African species are still poorly investigated, in particular due to ploidy variation among and within species. Here, we aim to investigate the diversity of the plastid genome (or plastome) of Central Saharan mimosoids, in order to assess its potential utility for phylogenetic and population genetic analyses. We first used a genome skimming strategy to assemble the complete plastome plus the nuclear ribosomal DNA cluster of six species belonging to three genera (*Vachellia*, *Senegalia*, and *Faidherbia*). Phylogenetic relationships based on these data confirm the existence of three main evolutionary lineages in the Hoggar range (southern Algeria). An analysis of the plastome structure reveals an extension of the inverted repeat (IR) in *Faidherbia albida* as recently reported in two other genera of the same lineage (*Inga* and *Acacia s.s.*). Higher substitution rates are detected in this lineage, and our species sampling allows revealing genes (particularly *accD*, *clpP*, *rps2*, *rps3*, *ycf1*, *ycf2*, and *ycf4*) under positive selection following the IR extension. The reasons for this evolutionary transition need to be unraveled. We then develop 21 plastid microsatellites to be used on a large panel of mimosoid species. At a local scale, 18 of these loci reveal intra-specific polymorphism in at least one species. These markers may be useful to assess the genetic diversity of the plastome for comparative phylogeographies or population genetic studies.

Keywords Acacia · Chloroplast genome · Hoggar · Microsatellite · Nuclear ribosomal DNA · Phylogeny · Phylogeography · Population genetics · Positive selection

Mohamed Mensous and Céline Van de Paer contributed equally to this work.

Communicated by G. G. Vendramin

Electronic supplementary material The online version of this article (doi:10.1007/s11295-017-1131-2) contains supplementary material, which is available to authorized users.

✉ Mohamed Mensous
mmensous@ymail.com

✉ Guillaume Besnard
guillaume.besnard@univ-tlse3.fr

[1] Laboratoire de Recherche sur la Phoeniciculture, Faculté des Sciences de la Nature et de la Vie, Université Kasdi Merbah—Ouargla, 30000 Route de Ghardaia, Ouargla, Algeria

[2] UMR 5174 EDB (Laboratoire Evolution et Diversité Biologique), CNRS-UPS-ENSFEA-IRD, Université Paul Sabatier, Bât. 4R1, 31062 Toulouse Cedex 9, France

[3] INRA, GeT-PlaGe, Genotoul, 31326 Castanet-Tolosan, France

[4] GenPhySE, Université de Toulouse, INRA, INPT, ENVT, 31326 Castanet-Tolosan, France

[5] Laboratoire de Recherche sur les Zones Arides, USTHB/ENSA, BP44, Alger, Algeria

Introduction

Acacias (Fabaceae, Mimosoideae) represent a major group of woody legumes in arid and subarid habitats of tropical and subtropical regions, particularly in the Southern hemisphere (Lewis et al. 2005). They are highly diversified with ca. 1350–1450 species around the world and ca. 130–150 species in Africa (Ross 1981; Maslin et al. 2003; Lewis et al. 2005). Acacias are an important source of wood (for fuel and lumber), fibers (to produce ropes), or medicines for human populations and of forage for wild and domestic animals (Le Floc'h and Grouzis 2003; Lewis et al. 2005; Hobbs et al. 2014). In addition, they can be used to fight against the desertification

Published online: 28 March 2017

and soil erosion, for instance in the Sahelo-Saharan region where *Vachellia* and *Senegalia* spp. have been planted on a large scale as part of the "Great Green Wall" initiative (FAO 2014). Their ability to fix nitrogen, thanks to rhizobia in root nodules, can also favor the growth of other plants, especially for cereal crops in traditional agrosystems (Payne et al. 1998). In semi-desert conditions, acacias often dominate tree communities and several species can occur in sympatry, but taxon identification based on morphological criteria can be challenging, especially on saplings (Brenan 1983; Doran et al. 1983). A few species are widespread in the Central Saharan mountains. While some taxa are frequent such as *Vachellia tortilis* subsp. *raddiana* and *Vachellia flava*, others (as *Vachellia nilotica*, *Faidherbia albida*, and *Senegalia laeta*) have a very fragmented distribution (Table S1). In addition, the various uses of these woody species by autochthonous people strongly threaten the long persistence of some populations that may be locally endangered (Quézel 1965, 1997; Sahki et al. 2004; CSFD 2004; Chenoune 2005; Baâli-Cherif 2007). The inventory of such populations and the characterization of their genetic diversity are necessary to assist their conservation (FAO 1996). Knowledge on the population dynamic (e.g., turnover, life span of individuals and importance of clonal growth, seed- versus pollen-mediated gene flow) is also essential for defining optimal strategies of management (Borgel et al. 1992).

To date, population genetic studies have been performed only on a few African acacia species [e.g., *F. albida* (Joly et al. 1992; Harris et al. 1997), *Senegalia mellifera* (Guajardo et al. 2010), *Senegalia senegal* (Odee et al. 2012; Assoumane et al. 2013)]. Such studies also allowed unraveling phylogeographic patterns (Guajardo et al. 2010; Odee et al. 2012) that can bring important insights for land managers by depicting the recent history of species, identifying barriers to dispersal and distinguishing main gene pools with putative hotspots of diversity (e.g., Besnard et al. 2007; Migliore et al. 2012). A prerequisite step to such studies is the development of molecular markers based on genomic sequences. Nuclear microsatellite markers have been developed on a few African acacia species, including *V. tortilis* (Winters et al. 2013; Osmondi et al. 2015) and *S. senegal* (Assoumane et al. 2009). However, one major limitation to these approaches is related to the existence of variable ploidy levels among and within species (Bukhari 1997; Chevalier and Borgel 1998; Assoumane et al. 2013; Odee et al. 2012, 2015; Table S1). Consequently, phylogeographic patterns were also investigated using the plastid DNA variation that is haploid whatever the ploidy level (e.g., in *Senegalia* spp.; Guajardo et al. 2010; Odee et al. 2012). In these studies, microsatellite length variation (i.e., variable number of repeats in mono-nucleotide stretches) was a source of valuable information, as such loci are usually considered as the fastest evolving regions of the plastid genome (hereafter plastome) (Powell et al. 1995; Weising and Gardner 1999). Full plastomes of mimosoid legumes have also been recently released, mainly for Australian species (Dugas et al. 2015; Williams et al. 2015, 2016) and represent a valuable resource to develop new markers.

Plastids (e.g., chloroplasts) are usually maternally inherited in angiosperms (Corriveau and Coleman 1988). In mimosoids, maternal inheritance (with potential rare paternal leakage) has been reported for two species (i.e., *Acacia decurrens* and *Acacia mearnsii*) based on cytological observations of pollen grains (Reboud and Zeyl 1994). Polymorphism of the plastome can thus be useful to follow seed dispersal. The plastome is also more prone to stochastic events (i.e., genetic drift) because its effective population size is half that of diploid genomes, while its dispersal is limited due to seeds generally dispersed at shorter distance than pollen (Petit et al. 1993). For these two reasons, polymorphism of plastomes allows revealing pronounced geographic structure that is suitable for phylogeographic analyses (Petit et al. 1993). Furthermore, both a high number of copies per cell and the haploid nature of the plastome (due to mono-parental inheritance) facilitate its sequencing and analysis, even on poorly preserved DNAs (e.g., herbarium and forensic analyses; Besnard and Bervillé 2002; Pérez-Jiménez et al. 2013). Although plastid DNA markers offer many advantages for population genetic and phylogenetic analyses, they may, however, show some limitations because phylogenetic patterns are not reflecting the whole history of a plant lineage, especially when hybridizations and cytoplasmic captures have occurred (Harrison and Larson 2014). In addition, structural rearrangements [inverted repeat (IR) extension] and a faster evolution of some genes (in particular *clpP*) have been revealed in the plastomes of *Acacia ligulata* and *Inga leiocalycina* (Dugas et al. 2015; Williams et al. 2015). The reasons for this phenomenon still remain unclear, and additional investigations are necessary to understand the contrasted evolutionary trajectories of plastid genes in mimosoid plastomes.

In the present study, we investigated the plastome diversity of six mimosoid species from the Central Sahara (i.e., *Vachellia* spp., *S. laeta*, and *F. albida*) in order to evaluate its utility for phylogenetic and population genetic analyses. We first sequenced complete plastomes (and nuclear ribosomal DNA clusters) using a genome skimming strategy (Straub et al. 2012). Based on these data, phylogenies were reconstructed. Structural and evolutionary changes were detected in plastomes of a mono-phyletic lineage (tribe Ingeae + *Acacia* sensu stricto), and several fast-evolving genes under positive selection were identified. Finally, plastid microsatellite markers were developed and tested on populations of the six South Algerian species. Perspectives on the genetic characterization of mimosoid species from Central Africa are finally proposed.

Material and methods

Plant sampling

The six species of Mimosoideae occurring in the Hoggar range were sampled in the Tamanrasset district (Fig. S1). According to taxonomical and phylogenetic studies, these taxa belong to three different lineages (e.g., Lewis et al. 2005; Maslin 2008; Murphy 2008; Seigler and Ebinger 2009; Bouchenak-Khelladi et al. 2010; Kyalangalilwa et al. 2013). First, four species belong to the *Vachellia* lineage: *Vachellia tortilis* (Forssk.) Galasso & Banfi subsp. *raddiana* (Savi) Kyal. & Boatwr. (hereafter referred to *V. tortilis*), *Vachellia flava* (Forssk.) Kyal. & Boatwr., *Vachellia nilotica* (L.) P. J. H. Hurter & Mabb. subsp. *tomentosa* (Benth.) Kyal. & Boatwr. (hereafter referred to *V. nilotica*), and *Vachellia seyal* (Delile) P. J. H. Hurter; second, *Faidherbia albida* (Delile) A. Chev. belongs to tribe Ingeae; and third, *Senegalia laeta* (R. Br. ex Benth.) Seigler & Ebinger belongs to the *Senegalia* lineage. These species show a large distribution range toward the Sahelian and southern Africa, as well as on the Asian continent (Table S1). Variable ploidy levels were also reported, especially in *V. tortilis* and *V. nilotica* [i.e., $2n = 26$ ($2x$) to 104 ($8x$); Table S1]. Note that *S. laeta* was reported as triploid ($3x = 39$), and possibly issued from an interspecific hybridization (Giffard 1966; El Amin 1976; Ross 1979; Chevalier and Borgel 1998).

We sampled between 14 (*S. laeta*) and 42 mature individuals (*V. seyal*) per species (Table 1). For *Vachellia* spp. and *S. laeta*, only apparently unconnected ramets distant at a minimum of 10 m from each other were sampled to avoid the analysis of clones. While *V. tortilis* and *V. flava* are relatively frequent in the study area, the other species are present at a few locations (Fig. S1). Individuals of each species were collected on different sites (in wadis), except for *V. nilotica* that was found at only one site ("Amezgin") in the southern part of the study area. For *F. albida*, we collected all ramets (18) that we found in the study area. Four isolated trees were sampled for this taxon. At "In Zbib," some of the 10 collected ramets were relatively close from each other (<5 m), and clones are possibly present at this locality.

Plastome and nuclear ribosomal DNA sequencing

For all samples, DNA extractions were performed from 20 mg of leaves (dried in silica gel) with the Biosprint method (Qiagen Inc.). To generate complete plastomes (chloroplast DNA (cpDNA)) and nearly complete nuclear ribosomal DNA units (nrDNA) of each species, we used the genome skimming approach (i.e., shotgun; Straub et al. 2012). Two individuals of *V. tortilis* were analyzed (haplotypes 302 and 303; see the "Results" section), while one tree accession was used for the five other species. Five hundred nanograms of double-stranded DNA (according to the NanoDrop quantification) were used for shotgun sequencing with the Illumina technology (HiSeq 2000, Illumina Inc.). The library was constructed using the Illumina TruSeq DNA Sample Prep v.2 Kit following the instructions of the supplier. Purified fragments were A-tailed and ligated to sequencing indexed adapters. The seven libraries were multiplexed with 17 other libraries (generated in other projects). The pool of libraries was then hybridized to the HiSeq 2000 flow cell using the Illumina TruSeq PE Cluster Kit v.3. Bridge amplification was performed to generate clusters, and paired-end reads of 100 nucleotides were collected on the HiSeq 2000 sequencer using the Illumina TruSeq SBS Kit v.3 (200 cycles). Complete chloroplast sequences and nuclear ribosomal units were assembled using the approach described by Besnard et al. (2013). Finally, reads were mapped onto the different sequences obtained with Geneious v.9.1.2 (Kearse et al. 2012) to assess the depth of sequencing of all investigated regions. Gene annotation of plastomes was performed after their alignment (see below) using Geneious by transferring the annotations of the *Leucaena trichandra* plastome (NC_028733.1). Physical maps of plastomes were drawn with OGDRAW (Lohse et al. 2013).

Phylogenetic reconstructions and analysis of evolutionary rates among protein-coding genes

A phylogenetic tree based on the complete plastomes was reconstructed (for phylogenetic analyses based on the nrDNA cluster, see Supplementary Methods). Three plastome sequences were downloaded from GenBank. First, a cpDNA sequence of *L. trichandra* (NC_028733.1; Dugas et al. 2015) was used as outgroup to root the tree, since this taxon belongs to tribe Mimoseae that is sister to a clade encompassing all Mimosoideae species from the Hoggar mountains (Bouchenak-Khelladi et al. 2010). Then, the plastome of *Acacia ligulata* (NC_026134; Williams et al. 2015) and *Inga leiocalycina* (NC_028732; Dugas et al. 2015) were also added to our alignment as two additional members of a lineage that encompasses tribe Ingeae and *Acacia s.s.* (Bouchenak-Khelladi et al. 2010). Sequences were aligned with the application Mega v.7 (Kumar et al. 2016). The alignment was manually refined, and the second inverted repeat (IR) was removed. Tree inference was made under a maximum likelihood criterion, using RAxML v.8 (Stamatakis 2014) in the CIPRES platform. The alignment was divided into six partitions: (i) protein-coding genes of the large and the single-copy regions (LSC and SSC, respectively), (ii) non-coding regions of LSC and SSC, (iii) protein-coding genes of the IR, (iv) non-coding regions of the IR, (v) tRNA genes, and (vi) rRNA genes. We applied the GTR + G model for each partition and we used the rapid bootstrap algorithm with 1000 iterations. The

Table 1 Description of acacia populations analyzed in the present study, including the taxon, the location (with GPS coordinates; Lat., latitude; long., longitude), the location altitude (Alt.; in m), and the number of individuals analyzed per location (number)

Taxon	Location	GPS coordinates		Alt.	Number	Individuals used for GS
		Lat.	Long.			
Vachellia flava	Adriane	22.47° N	05.35° E	1483	14	"Adriane 1"
	Aglala	22.38° N	05.36° E	1485	13	–
Vachellia nilotica	Amezgin	22.33° N	05.23° E	1255	18	"Amezgin 20"
Vachellia tortilis	Nezmet	22.44° N	05.49° E	1338	13	"Nezmet 19 and 23"
	Adriane	22.47° N	05.35° E	1296	10	–
	Tassena	22.48° N	05.36° E	1451	8	–
	Amezgin	22.33° N	05.23° E	1255	3	–
	Mazoliet	22.58° N	05.40° E	1828	1	–
Vachellia seyal	Mazoliet	22.58° N	05.40° E	1828	23	"Mazoliet 36"
	Akarakar	23.02° N	05.42° E	1866	19	–
Faidherbia albida	In Zbib	22.47° N	05.37° E	1423	10	"In Zbib 5"
	Ouarsedakfis	22.49° N	05.58° E	1426	4	–
	Ihghi	23.03° N	05.12° E	1084	1	–
	Tassenna	22.48° N	05.36° E	1451	1	–
	Inemrus	23.08° N	05.59° E	1713	1	–
	Inarjiwen	23.11° N	06.02° E	1975	1	–
Senegalia laeta	In Zbib	22.47° N	05.37° E	1423	12	"In Zbib 8"
	Tassena	22.48° N	05.36° E	1451	2	–

The seven individuals analyzed with a genome skimming (GS) strategy are also indicated

phylogenetic tree was visualized using FigTree v1.4.2 (Rambaut 2014).

The evolutionary rates of each plastid protein-coding genes were then estimated and compared between species. We performed the analysis with one individual per species, using accession "Nez19" to represent *V. tortilis*. Sequence alignment of each gene was individually done using Muscle (Edgar 2004) and refined manually taking into account their amino acid translation. Non-synonymous and synonymous substitution rates (d_N and d_S) were calculated on all concatenated protein-coding genes and independently for each gene using the software *codeml* implemented in PAML v.4.8 (Yang 2007). We estimated d_N and d_S for each species against the outgroup species *L. trichandra.*

An IR extension in tribe Ingeae + *Acacia s.s.* was associated to an accelerated evolutionary rate, in particular on *clpP* (Dugas et al. 2015; Williams et al. 2015; see the "Results" section). We analyzed changes of selective pressures in tribe Ingeae + *Acacia s.s* for genes in which we previously detected an acceleration of substitution rates in *I. leiocalycina*, *F. albida*, and *A. ligulata* compared to other species. This analysis was performed across branches of the phylogenetic tree using *codeml* in PAML. The phylogenetic topology based on complete plastomes was used. On each gene, we tested three codon models, which used the non-synonymous versus synonymous substitution rate ratio (d_N/d_S, denoted ω). The ω ratio is used to evaluate selective pressures at the protein level, with $\omega < 1$, =1, or >1, indicating purifying selection, neutral selection, or positive selection, respectively. The first model, M1a, is a site model which allows ω to vary among sites. This model allows sites to be either under neutral selection ($\omega = 1$) or under purifying selection ($\omega < 1$; Yang et al. 2000). The second model, A, is a branch-site model that allows ω to vary among both sites and branches of the phylogenetic tree (Zhang et al. 2005). Branch-site models allow different ω for branches and need to assign foreground branches on which positive selection can potentially operate. In this model, sites can be attributed to four types of selection: (i) purifying selection on the whole tree ($\omega < 1$), (ii) neutral selection on the whole tree ($\omega = 1$), (iii) purifying selection on background branches ($\omega_0 < 1$) and positive or neutral selection on foreground branches ($\omega_2 \geq 1$), and (iv) neutral selection on background branches ($\omega_1 = 1$) and positive or neutral selection on foreground branches ($\omega_2 \geq 1$). The last model, A′, is the same as model A, except that positive selection on foreground branches is replaced by neutral selection ($\omega_2 = 1$; Zhang et al. 2005). In this study, we tested two hypotheses (Fig. S2): (i) positive selection occurred on the internal branch ancestral to *I. leiocalycina*, *A. ligulata*, and *F. albida*, on which the IR extension has occurred (hypothesis H1) and (ii) positive selection occurred on all branches of the clade including *I. leiocalycina*, *A. ligulata*, and *F. albida*, after the IR

extension (hypothesis H2). For each hypothesis, we performed two likelihood ratio tests (LRTs) to compare models. As model A is nested into both models M1a and A′, we can compare model A versus model M1a and model A versus model A′. To choose the hypothesis (H1 or H2) that fits the data best, we compared the best model of each hypothesis using the Akaike Information Criterion (AIC). In the case of non-significantly different models under a given hypothesis, we compared the AIC of the three models M1a, A, and A′ with that of the best model of the other hypothesis. Given the small number of species in our analysis, we also performed simulations to evaluate the frequency of the type-I error (i.e., the incorrect rejection of a true null hypothesis H0). For each gene, we generated 1000 simulated data under the null hypothesis, i.e., assuming two classes of sites [either under neutral selection ($\omega = 1$) or under purifying selection ($\omega < 1$)], using the software *evolverNSsites* in PAML. We used the value of ω and the proportion of sites attributed to each class estimated under model M1a with the real dataset. We then tested the three models M1a, A, and A′ under hypotheses H1 and H2 for each simulated data. In order to estimate the proportion of false positives, we estimated LRT relative to χ^2 for each simulated data to compare (i) models A with M1a and (ii) models A with A′.

Characterization of cpDNA microsatellite polymorphisms in mimosoid populations

Based on six complete plastome sequences of acacias from the Hoggar (one for each studied species), we targeted regions that show microsatellite motif(s) (i.e., mono-nucleotide stretches) in the six investigated species. We identified 73 regions with at least one mono-nucleotide stretch (min. 8 repeats) in all species. After preliminary tests, 21 loci (all located in intergenic spacers) were finally selected for their polymorphism and reliable amplification (Tables 2 and S2). These loci were hereafter referred to chloroplast simple sequence repeats (cpSSRs). The PCR primers were designed in flanking regions to specifically amplify short segments (inferior to 265 bp; Table 2). For locus multiplexing, the annealing temperature of all these primers needed to be similar, while the size of PCR products of each locus should be as different as possible. Finally, these primers were also designed to allow amplification of short DNA segments for the characterization of degraded DNAs. All primer pairs and specific characteristics of generated fragments are given in Tables 2 and S2. To reduce time and costs of the PCR characterization, we used the method described by Schuelke (2000). For each locus, an 18-bp tail of M13 was added on the forward primer (Table 2). Additionally, the 5′ end of the reverse primer was also tagged with the sequence GTGTCTT to minimize band stuttering (Table 2). Three or four loci were simultaneously amplified in the same reaction (PCR multiplexes defined in Table 2). Each PCR reaction (25 μl) contained 10 ng DNA template, 1× reaction buffer, 5 mM $MgCl_2$, 0.2 mM dNTPs, 0.2 μmol of one universal fluorescent-labeled M13(−21) primer (5′-TGTAAAACGACGGCCAGT-3′, labeled with one of the three following fluorochromes: YAK, 6-FAM or AT550), 0.04 μmol of the reverse primer, 0.01 μmol of the forward primer, and 0.5 U of *Taq* DNA polymerase (Promega). The reaction mixtures were incubated in a T1 thermocycler for 2 min at 95 °C, followed by 28 cycles of 30 s at 95 °C, 30 s at 57 °C, and 1 min at 72 °C, and then by 8 cycles of 30 s at 95 °C, 30 s at 51.5 °C, and 1 min at 72 °C. The last cycle was followed by a 20-min extension at 72 °C. The PCR products labeled with a fluorochrome were mixed together with GeneScan-600 LIZ as internal standard to run all loci at the same time. They were separated on an ABI Prism 3730 DNA Analyzer (Applied Biosystems), and the size of amplified fragments was determined with Geneious.

Based on fragment genotyping (cpSSRs), we defined a cpDNA haplotype for each individual (i.e., combination of alleles at the 21 loci). The probability that two individuals taken at random display a different haplotype was computed as $D = 1 - \Sigma p_i^2$, where p_i is the frequency of the haplotype i. Haplotype richness was computed with FSTAT (Goudet 1995). The relationships among cpDNA haplotypes were then visualized by constructing a reduced median network implemented in the application Network v.5 (Bandelt et al. 1999). cpSSRs are fast-evolving parts of the plastome, and thus, generally not informative when comparing too distant lineages. Such markers have proven to be valuable tools at the population level, and eventually at the species complex level (e.g., *Olea europaea*; Besnard et al. 2011). Consequently, we decided to reconstruct haplotype networks separately on *Vachellia* spp. and on *F. albida* (+ *S. laeta* that was used as outgroup). Multistate microsatellites were treated as ordered alleles and coded by the number of repeated motifs for each allele (e.g., number of nucleotides in mono-nucleotide stretches; see also Besnard et al. 2011). Basically, this coding strategy assumes that variation at cpSSRs is mainly due to single-step mutations (e.g., Besnard et al. 2011). For haplotypes whose plastome is not available, we checked length polymorphisms when superior to 2 bp compared to related haplotypes. These fragments were sequenced with the Sanger method as described in Besnard et al. (2011). For loci combining microsatellite motifs and non-SSR indels either in *Vachellia* spp. or in *F. albida* (i.e., loci 3 and 14), we separately coded the two types of characters based on available sequences. The presence or absence of such indels was coded as 1 and 0, respectively. In addition, size homoplasy is frequent on plastid microsatellites (Powell et al. 1995; Navascués and Emerson 2005). To downweight the impact of the most variable (and thus the most homoplasic) markers in the network reconstruction, we arbitrarily assigned four classes of weight to character changes according to the number of alleles at a locus (Bandelt et al.

Table 2 Primers used to amplify by PCR the 21 cpSSR loci with their allele size range (in bp) on the six studied acacia species, and M13-primer fluorochrome (Flu) used for each locus (Y = YAK, F = 6-FAM, or A = AT550)

No. of locus	Primers		Allele size range	Flu	Multiplex
	Forward[a]	Reverse[b]			
1	tgtaaaacgacggccagtATAATGAAACATTCCTTCGGAAAG	gtgtcttGCAATTTTGAATACTCGAACGG	101–110	Y	a
2	tgtaaaacgacggccagtCTGACCAAACAAATAATTGTCAG	gtgtcttAATGCGAAATCGATGACCTATG	113–120	Y	a
3	tgtaaaacgacggccagtGAGATAAAATTGGCGAATAAGAAAC	gtgtcttAAGTGGAATGAAGGGTATGATC	132–144	Y	a
4	tgtaaaacgacggccagtGGAAACTCAATGGAATTCATAACT	gtgtcttYGTGCTTATCCTAATTGTTGGT	148–151	Y	a
5	tgtaaaacgacggccagtGAATATTTTTAGAGAATGCTCGG	gtgtcttTAATAACGACTTGTATTTGTACGC	158–170	Y	b
9	tgtaaaacgacggccagtCCCTTAGAACCGTACTTGAG	gtgtcttCTTCTGAAGGTAGGAGAAAAG	228–234	Y	b
10	tgtaaaacgacggccagtTTCTTGCGTGTACCAATTGAAG	gtgtcttGCAATTGCCCCTTTCGATGA	241–265	Y	b
11	tgtaaaacgacggccagtTTTTAGTGAACGTGTCACAGTC	gtgtcttAATMGAATTTGTTTCTTCGTCTTTAC	93–99	F	c
12	tgtaaaacgacggccagtTGTGCTAAAGATTCAAAACCCG	gtgtcttTGCATAAAATGGAAATCCATTGG	104–119	F	c
13	tgtaaaacgacggccagtCATTAAGATTTATCCAGTCGAAG	gtgtcttCGAAGTCTTTCTGATAGTTKGAT	124–136	F	c
14	tgtaaaacgacggccagtATTGGAAGAGAATAGGGTTTCC	gtgtcttAGGGTTATAGACTCCGGAAC	144–155	F	c
15	tgtaaaacgacggccagtCTCACGCTCAATTACTTATGG	gtgtcttGAAGAGTCTTCCGAATATTCTG	162–166	F	d
16	tgtaaaacgacggccagtTTGATACCAATAATGGCAGC	gtgtcttAAGACAATGCGTATGTGCATC	173–181	F	d
17	tgtaaaacgacggccagtGGATTTGCACCAACGGAAAC	gtgtcttCATGGACTAAGACAAAACAGAG	191–204	F	d
20	tgtaaaacgacggccagtGGTTCAAATCATTTTATCGACATG	gtgtcttAAGAACCATAGCATTTCGTGATTC	247–262	F	d
21	tgtaaaacgacggccagtYCCCGRATCAAACACCTTAC	gtgtcttCAAGAGATTTTTACTCCCTTCG	106–110	A	e
23	tgtaaaacgacggccagtTCAKTCCAATCCCCACTC	gtgtcttCTCGCCTTTTCTTTTCYGGC	131–135	A	e
24	tgtaaaacgacggccagtAGAAATAGAGAACGAAGTAACTAG	gtgtcttTCTGTCTGAATCTGTCTGAATG	145–153	A	e
25	tgtaaaacgacggccagtGGGATAAGCTATAAGAGATTATCA	gtgtcttATCAGAATGGTTSGTTATTAAACC	158–161	A	f
29	tgtaaaacgacggccagtTTTAAGTCCCTTTTTGTTTTGCAC	gtgtcttACCCGTCGTTTTCGATTACG	216–229	A	f
30	tgtaaaacgacggccagtGCTTTAGCAGGTCTATTGATAG	gtgtcttCGAATTCCGAGTTTCACTGAG	238–255	A	f

Three or four loci were simultaneously amplified in six multiplexes (a to f, considering non-overlapping allele size range for a given fluorochrome). Additional loci features (microsatellite motifs and their genome position) are given in Table S2

[a] M13-tail on each forward primer: tgtaaaacgacggccagt

[b] Reverse-primer tailing: gtgtctt

1999): weight of 10 for loci with two alleles, 5 for loci with three alleles, 3 for loci with four alleles, and 1 for loci with five alleles and more.

Results

Assembly of plastomes and nuclear ribosomal DNA clusters

We first generated between 7,298,994 and 41,826,997 paired-end reads for the seven tree accessions analyzed by shotgun (Table S3). These data allowed us to assemble the plastome and nuclear ribosomal cluster for at least one individual per species. The sequencing depth of these genomic regions (Table S3) was relatively high, from 194 to 4257× for the plastome and from 235 to 10,064× for the ribosomal cluster, allowing their perfect assembly.

The plastome length varies from 162,754 bp (*S. laeta*) to 175,675 bp (*F. albida*; Table S3). The two *V. tortilis* plastomes are very similar, and differ by only three nucleotidic substitutions and one nucleotide indel in a poly-T stretch (corresponding to cpSSR locus 10; see below). The plastome of *Faidherbia* is longer than this of other acacia species by ca. 10–13 kb. This difference in the plastome length is mainly due

to an extension of the IR (Fig. S3) that notably encompasses the whole coding sequence of genes that are usually located in the SSC (*ycf1*, *rps15*, *ndhH*, *ndhA*, *ndhI*, *ndhG*, *ndhE*, *psaC* and *ndhD*; Xu et al. 2015), as similarly reported in *I. leiocalycina* and *A. ligulata* (Dugas et al. 2015; Williams et al. 2015). As expected, plastomes are AT-rich regions (GC content = 35.3 to 35.8%; Table S3). In contrast, the nrDNA clusters are GC-rich regions (GC content = 57.1 to 59.1%; Table S3). The assembled nrDNA regions are 6401 to 8104 bp long and comprise 5′ and 3′ external transcribed sequences (ETS), the 18S gene, the internal transcribed spacer 1 (ITS-1), the 5.8S gene, the internal transcribed spacer 2 (ITS-2), and the 26S gene. The internal gene spacer (IGS), which includes ETS, was not completely assembled due to the presence of repeated and inverted regions that render its assembly very challenging. Interestingly, two divergent units (i.e., 98.3% of homology on the whole sequence) were separated in *S. laeta*, the major and minor units representing approximately 65 and 35% of nrDNA reads, respectively (Table S3).

Phylogenetic topologies and molecular evolution of plastid protein coding genes

Phylogenetic relationships based on plastomes and nrDNA are mostly congruent (Figs. 1 and S4). In both phylogenetic trees, relationships between *Vachellia* spp. are fully congruent. This genus is sister to a clade including *Senegalia*, *Inga*, and *Acacia s.s*. While the placement of *S. laeta* as sister to a paraphyletic tribe Ingeae and *Acacia s.s.* (*F. albida*, *I. leiocalycina*, and *A. ligulata*) is strongly supported, the placement of *L. trichandra* differs in the two phylogenies (Figs. 1 and S4). In the plastome phylogenetic tree, branches of *F. albida*, *I. leiocalycina*, and *A. ligulata* are longer than those of other species, suggesting a faster evolution of the plastome for these taxa (Fig. 1).

Non-synonymous and synonymous substitution rates (d_N and d_S) were estimated on 78 plastid protein-coding genes (Figs. 2 and S5). These rates were computed for each species against the outgroup species *L. trichandra*. Both non-synonymous and synonymous substitution rates are globally higher in *F. albida*, *I. leiocalycina*, and *A. ligulata* ($0.0137 > d_N > 0.0315$; $0.0300 > d_S > 0.0455$) than in *S. laeta* and *Vachellia* spp. ($0.0073 > d_N > 0.0077$; $0.0258 > d_S > 0.0272$; Fig. S5). As shown by Dugas et al. (2015), *clpP* has quickly evolved in *A. ligulata* and *I. leiocalycina* (Fig. 2). A similar pattern is here detected in *F. albida.* For these three species that belong to a monophyletic lineage, we found an accelerated evolutionary rate for 24 genes compared with the other species (Fig. S5; Tables 3 and S4). Interestingly, an IR extension is observed in plastomes of *F. albida*, *I. leiocalycina*, and *A. ligulata* (see above). Changes of selective pressure during or after this IR extension were then tested with codon models on the 24 genes mentioned above. Under hypothesis H1, assuming positive selection on the internal branch ancestral to *F. albida*, *I. leiocalycina*, and *A. ligulata*, model A is significantly better than both null models for *clpP*, *ycf1*, and *ycf2* (models A and M1a are marginally different for *ycf1*, and *ycf2*; Table 3), while the three models are not significantly different for the other genes (Table S4). Under hypothesis H2, assuming positive selection on all branches of the clade including *F. albida*, *I. leiocalycina*, and *A. ligulata*, model A is significantly better than both null models for *accD*, *clpP*, *rps2*, *rps3*, *ycf1*, *ycf2*, and *ycf4* (models A and M1a are marginally different for *rps2*; Table 3), while the three models are not significantly different for the other genes (Table S4). Using sequence simulations, we did not detect any inflation of false positives for *accD*, *clpP*, *rps2*, *rps3*, *ycf1*, *ycf2*, and *ycf4* under H1 and H2 (Table S5) and type-I errors may thus not affect our analyses. Based on AIC comparison, hypothesis H2 fits to our data

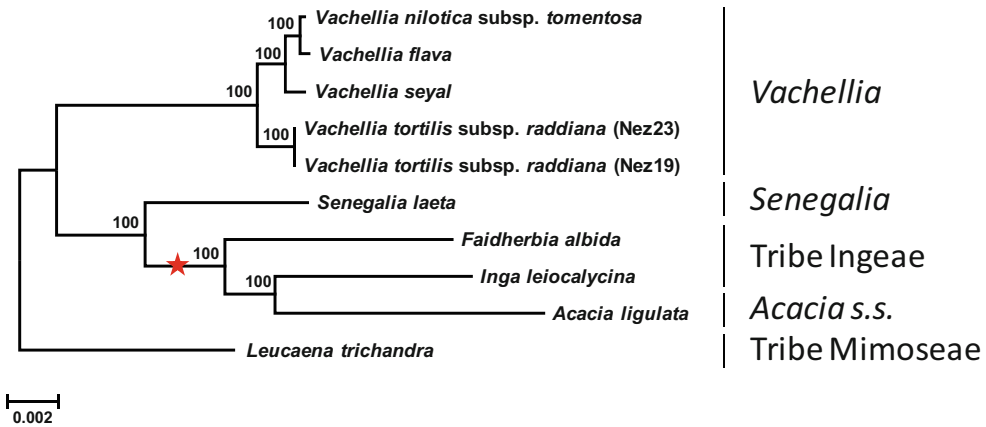


Fig. 1 Phylogenetic reconstruction of Mimosoideae based on complete plastomes. Tree inference was made under a maximum likelihood criterion, using RAxML v.8 (Stamatakis 2014). Main lineages (i.e., tribes and genera) usually recognized are indicated (see Lewis et al. 2005; Maslin 2008; Murphy 2008; Seigler and Ebinger 2009; Bouchenak-Khelladi et al. 2010; Kyalangalilwa et al. 2013). The *red star* indicates the internal branch ancestral to *I. leiocalycina*, *A. ligulata*, and *F. albida*, on which the IR extension has occurred (Fig. S3)

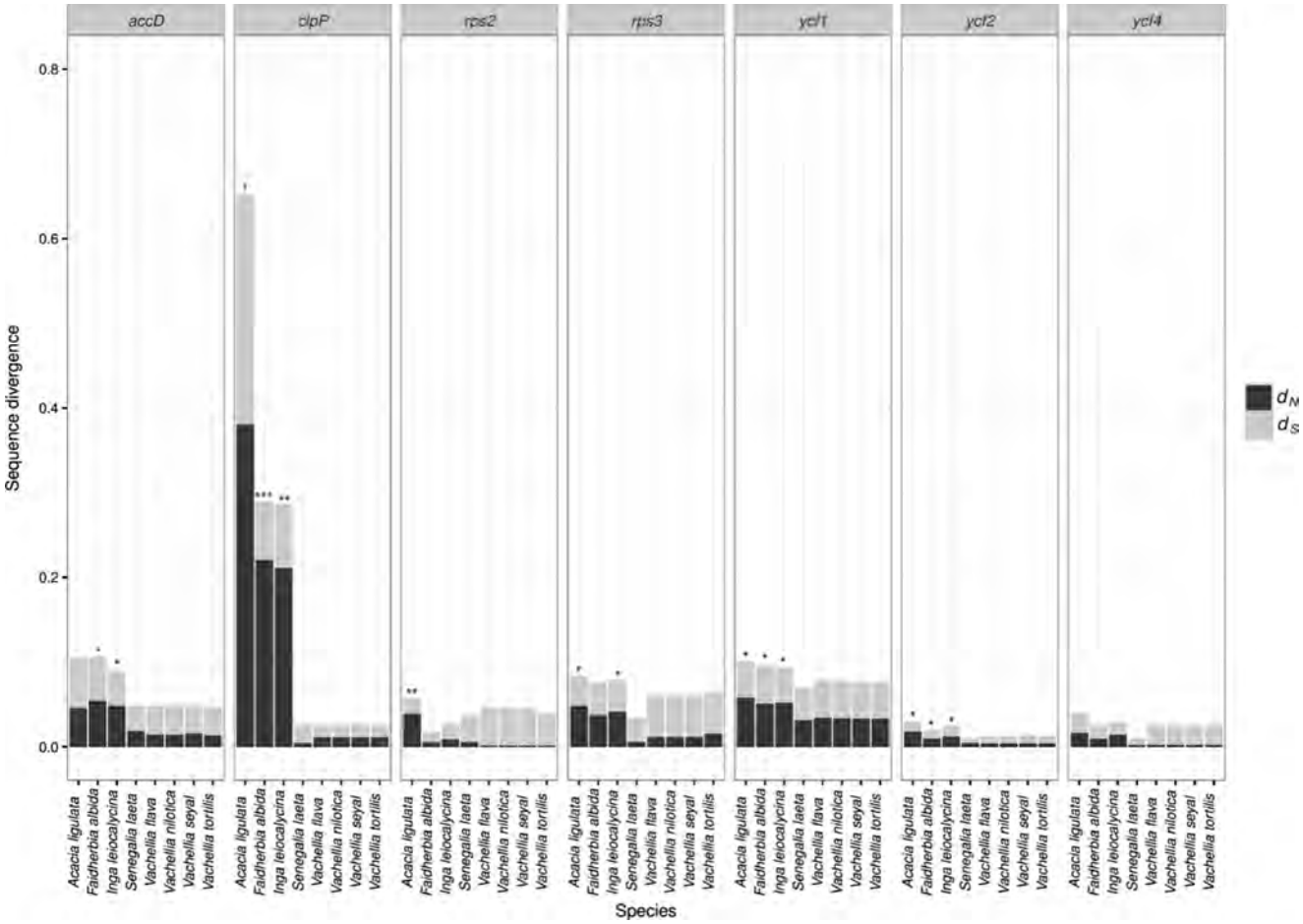


Fig. 2 Substitution rates among genes *accD*, *clpP*, *rps2*, *rps3*, *ycf1*, *ycf2*, and *ycf4* of the eight acacia species belonging to Ingeae, *Vachellia*, *Senegalia*, and *Acacia s.s*. For each protein-coding gene, non-synonymous (d_N) and synonymous (d_S) rates were calculated against the outgroup species (*Leucaena trichandra*) using PAML (Yang 2007). *ω (d_N/d_S ratio) > 1; **ω > 2; ***ω > 3. For substitution rates among other protein-coding genes, see Fig. S5

significantly better than hypothesis H1 for *clpP*, *ycf1* and *ycf2* (Table 4). Because models M1a, A, and A′ are not significantly different under hypothesis H1 for *accD*, *rps2*, *rps3*, and *ycf4*, we compared the AIC of these three models with that of the best model of hypothesis H2. We also found that hypothesis H2 fits to our data significantly better than hypothesis H1. Hypothesis H2, assuming positive selection on all branches of the clade, is thus the best-fit hypothesis for

Table 3 Summary of *p* values estimated by comparing model A versus model A′ and model A versus model A′ using a likelihood ratio test (LRT)

Hypothesis	Models	df	Gene						
			accD	*clpP*	*rps2*	*rps3*	*ycf1*	*ycf2*	*ycf4*
H1	A versus M1a	2	0.6393	0.0022***	1	0.9337	0.0584*	0.0595*	0.6311
	A versus A′	1	0.3442	0.0011***	1	0.7111	0.0172**	0.0126**	0.3388
H2	A versus M1a	2	#	#	#	#	#	#	0.0384**
	A versus A′	1	#	#	0.0563*	#	#	#	0.0444**

We performed these tests under two hypotheses assuming a potential positive selection either on the internal branch ancestral to *Faidherbia albida*, *Inga leiocalycina*, and *Acacia ligulata* (hypothesis H1) or on all branches of the clade including *F. albida*, *I. leiocalycina*, and *A. ligulata* (hypothesis H2). Results on other genes are given in Table S4

df degree(s) of freedom

Significance of tests: *$p \leq 0.1$; **$p \leq 0.05$; ***$p \leq 0.01$; #$p \leq 0.001$

Table 4 Summary of the Akaike Information Criterion (AIC) used to determine the best-fit hypothesis

Hypothesis	Npar	*accD*		*clpP*		*rps2*		*rps3*		*ycf1*		*ycf2*		*ycf4*	
		lnL	AIC	lnL	AIC	lnL	AIC	lnL	AIC	lnL	AIC	lnL	AIC	lnL	AIC
H1															
Model M1a	19	−3922.2	7882.3	–	–	−1190.0	2418.0	−1314.3	2666.7	–	–	–	–	−893.2	1824.4
Model A	21	−3921.7	7885.4	−2032.4	4106.9	−1190.0	2422.0	−1314.3	2670.5	−12,043.7	24,129.4	−10,639.8	21,321.5	−892.8	1827.5
Model A′	20	−3922.2	7884.3	–	–	−1190.0	2420.0	−1314.3	2668.7	–	–	–	–	−893.2	1826.4
H2															
Model A	21	−3895.5	*7833.0*	−2006.4	*4054.7*	−1180.2	*2402.4*	−1298.9	*2639.9*	−11,984.0	*24,010.1*	−10,623.0	*21,288.0*	−890.0	*1821.9*

For each gene, we compared the best-fit model of both hypotheses. For each gene, the AIC value indicating the best-fit hypothesis is in italics

lnL log-likelihood, *Npar* number of parameters in the model

accD, *clpP*, *rps2*, *rps3*, *ycf1*, *ycf2*, and *ycf4*. Under this hypothesis, model A detected between 0.4 and 10.6% of sites under positive selection for *ycf2* and *clpP*, respectively (posterior probability >0.95; Table S6).

Development of plastid microsatellites and inter- versus intra-specific variation

All 21 cpSSR loci rendered length polymorphism at the inter-specific level, while 18 were variable at the intra-specific level in at least one species (Tables 5 and 6). The combination of alleles of these 21 loci allows us to define a haplotype for each individual (Tables 6 and S7). Each haplotype is unique to a species. No variation was found in three taxa: *V. flava*, *V. seyal*, and *S. laeta*. In contrast, we detected diversity (D = 0.51 to 0.78) within the three other species, with three haplotypes in populations of *V. tortilis* and *F. albida* and five haplotypes in the population of *V. nilotica*. The haplotype network for *Vachellia* species reveals two small clusters of related haplotypes for *V. tortilis* and *V. nilotica* (Fig. 3). All haplotypes of *V. nilotica* are, however, not closely related, since one divergent haplotype (no 201) is detected in three individuals. The distribution of haplotypes in *V. tortilis* does not show obvious geographic structure in our study area since the three haplotypes are detected in all locations with more than two samples (Fig. 3). Although our *F. albida* sampling is limited, the distribution of haplotypes seems not random in this species and no diversity is found within populations (Fig. S6).

Discussion

On the phylogenetic relationships between Central Saharan mimosoids

In our study, we generated new genomic resources for African mimosoid species, in particular the first complete plastomes and long sequences of the nrDNA cluster. Such data are potentially useful to resolve relationships in the Mimosoideae (Dugas et al. 2015; Williams et al. 2015, 2016), for which taxon sampling in phylogenetic works is not yet comprehensive compared to other legumes of the subfamily Papilionoideae (Schwarz et al. 2015). Based on our complete plastomes, and as expected from Luckow et al. (2003), Bouchenak-Khelladi et al. (2010), and Kyalangalilwa et al. (2013), *S. laeta* is sister to a paraphyletic tribe Ingeae (*F. albida* and *I. leiocalycina*) and *Acacia s.s.* (*A. ligulata*), while *Vachellia* spp. is sister to this clade. The topology based on nrDNA is mostly congruent, but *Leucaena* shows an unexpected position (as sister to *Acacia s.s.*), possibly due to a too rapid and/or heterogeneous evolution of the ITS marker (e.g., saturated phylogenetic signal and long branch attraction

when comparing distantly related species; see Supplementary Methods).

Phylogenies reconstructed from plastomes and nrDNA units show fully congruent relationships between the four *Vachellia* species. A close relationship is inferred for *V. nilotica* and *V. flava* based on both cpDNA and nrDNA, and their non-distinction based on the nrDNA cluster suggests that these taxa may have diverged in a very recent past or belong to a species complex (e.g., Ali and Qaiser 1980, 1992). In addition, we assembled two different nrDNA types in *S. laeta*, and our phylogenetic analysis shows that these two sequences are not resolved in a mono-phyletic lineage within *Senegalia* but are intermingled with those of *S. senegal* (Supplementary Methods), supporting that *S. laeta* belongs to the *S. senegal* complex (Chevalier et al. 1994). Our observation also likely reflects the hybrid status of *S. laeta* as has been suggested by some authors (Giffard 1966; El Amin 1976; Ross 1979). The reported triploidy of this species (Ross 1979) suggests that *S. laeta* may be allopolyploid (or segmental autopolyploid) as frequently observed in the *S. senegal* complex (Chevalier and Borgel 1998; Assoumane et al. 2013; Odee et al. 2012, 2015; Diallo et al. 2015).

On contrasted evolutionary rates in plastomes of mimosoids

Our phylogenetic analyses based on plastomes confirm that the evolutionary rates are heterogeneous among mimosoids (Dugas et al. 2015; Williams et al. 2015). Compared to previous studies, adding members of *Faidherbia* (Ingeae), *Vachellia*, and *Senegalia* in the mimosoid plastome phylogeny allows delimiting a clade (tribe Ingeae + *Acacia s.s.*) with longer branches resulting from accelerated global substitution rates (Fig. S5). Interestingly, the three analyzed species of this clade also share an extension of the IR (Fig. S3) as already noticed by Dugas et al. (2015) and Williams et al. (2015), respectively, on *Acacia* and *Inga*. Our analyses confirm that several genes [(encoding proteins involved in different functions; caseinolytic protease (*clpP*), ribosomal protein (*rps2*, *rps3*), acetyl-coA carboxylase (*accD*), probable ATPase (*ycf2*), photosystem-I assembly protein (*ycf4*)] show accelerated synonymous and non-synonymous substitution rates in tribe Ingeae + *Acacia s.s.*, with evidence for positive selection (Fig. 2; Table 3). These analyses also suggest that positive selection was not episodic (in relation to the IR extension) but seems widespread on all branches of the clade following the IR extension (i.e., hypothesis H2; Table 4), especially on *clpP*, *accD*, *ycf1*, and *ycf2* as similarly observed in several lineages of Sileneae (Erixon and Oxelman 2008; Sloan et al. 2014; Rockenbach et al. 2017). An evolutionary transition related to the IR extension could be associated to changes of some selective pressures offering new evolutionary trajectories to some genes. IR extension or reduction coupled to an accelerated evolutionary rate have been also observed in plastomes of Papilionoideae legumes (e.g., Guo et al. 2007; Magee et al. 2010; Dugas et al. 2015) and other plant families [e.g., Geraniaceae (Chumley et al. 2006; Guisinger et al. 2011; Blazier et al. 2016), Oleaceae (Lee et al. 2007; C. Van de Paer et al., in prep.), Apocynaceae (Ku et al. 2013), Caryophyllaceae (Sloan et al. 2014), Apiaceae (Downie and Jansen 2015), Eucommiaceae (Wang et al. 2016)]. The reasons for such recurrent switches need to be unraveled. In tribe Ingeae and *Acacia s.s.*, this could be an indirect consequence of a non-strict maternal inheritance of chloroplasts (Reboud and Zeyl 1994) potentially leading to heteroplasmy than in turn could modify selective pressures. We suggest that the plastid inheritance mode should be better investigated in mimosoids to test for this hypothesis. Finally, contrasted

Table 5 Number of alleles and haplotypes per locus within and among species

Taxon	Locus no.																					N_h	R_S	D
	1	2	3	4	5	9	10	11	12	13	14	15	16	17	20	21	23	24	25	29	30			
Vachellia flava	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1.00	0.000
Vachellia nilotica	2	3	2	1	2	2	2	1	2	3	3	1	3	2	2	2	1	2	2	2	1	5	5.00	0.778
Vachellia tortilis	1	1	2	1	1	1	2	1	1	1	1	1	1	1	2	1	1	1	1	1	1	3	3.00	0.630
Vachellia seyal	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1.00	0.000
Faidherbia albida	1	1	3	1	1	1	2	1	2	2	1	2	1	2	2	1	2	2	1	1	1	3	2.96	0.512
Senegalia laeta	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1.00	0.000
All species	5	6	7	3	5	6	7	3	6	7	6	4	5	5	6	4	5	7	3	7	4	14	–	–

N_h number of haplotypes (combination of polymorphism from the 21 cpSSR loci) per species and among species, R_S haplotype richness (calculated for 14 individuals), D probability that two individuals show a different allele within a species

Table 6 Plastid haplotype profiles based on length variation at the 21 cpSSR loci

Hapl. no.	Locus no.																					Taxon
	1	2	3	4	5	9	10	11	12	13	14	15	16	17	20	21	23	24	25	29	30	
101	108	119	133	150	159	229	262	93	104	129	155	163	179	192	247	108	131	149	159	227	238	*Vachellia flava*
201	108	117	132	150	160	228	262	93	105	132[a]	146[a]	162	177	191	254[a]	106	132	146	159	228	238	*V. nilotica*
202	109	119	133	150	159	229	264	93	104	129	155	162	179	192	247	109	132	148	160	229	238	*V. nilotica*
203	109	118	133	150	159	229	264	93	104	129	155	162	179	192	247	109	132	148	160	229	238	*V. nilotica*
204	109	119	133	150	159	229	264	93	104	130	154	162	179	192	247	109	132	148	160	228	238	*V. nilotica*
205	109	119	133	150	159	229	264	93	104	129	154	162	178	192	247	109	132	148	160	229	238	*V. nilotica*
301	107	118	134	151	158	229	264	93	106	129	151	162	178	192	254[a]	110	133	147	159	226	238	*V. tortilis*
302	107	118	135	151	158	229	264	93	106	129	151	162	178	192	248	110	133	147	159	226	238	*V. tortilis*
303	107	118	135	151	158	229	265	93	106	129	151	162	178	192	248	110	133	147	159	226	238	*V. tortilis*
401	110	120	133	150	160	230	263	93	104	128	153	162	181	193	247	108	134	145	160	225	239	*V. seyal*
501	101	115	134	148	161	234	241	99	117	124	144	162	173	204	256	110	135	147	158	216	255	*Senegalia laeta*
601	101	113	143	148	170	231	246	94	119	134	146	164	173	202	261	106	133	152	159	218	241	*Faidherbia albida*
602	101	113	144	148	170	231	246	94	119	134	146	164	173	202	261	106	133	152	159	218	241	*F. albida*
603	101	113	138[a]	148	170	232	248	94	118	136	146	166	173	204	262	106	134	153	161	218	241	*F. albida*

For each locus, the length of the allele is given in base pair. Based on our sampling in the Hoggar mountains, each haplotype is unique to a taxon

Hapl. no. Haplotype number

[a] Alleles sequenced with the Sanger method to check the cause of length polymorphism

evolutionary rates also need to be taken into consideration in phylogenetic analyses of mimosoids, especially on genes under positive selection.

Utility of cpSSRs for phylogeographic reconstructions and population genetic analyses

Although our sampling is on a limited geographic area around Tamanrasset, most of our cpSSRs (18/21) revealed intraspecific polymorphism in at least one species. In addition, haplotypes (multiloci profiles) are found to be unique to a taxon, demonstrating the potential utility of our markers to identify species at this local scale. We also suggest our markers to be used on a larger sampling of populations (and species) to reconstruct phylogeographies at the continental scale. Such genetic analyses may also be valuable to identify gene pools and assist taxonomical revisions (e.g., Taylor and Barker 2012; Binks et al. 2015). Knowledge on the distribution of plastid haplotypes is potentially useful to develop tests of provenances and to trace the spread of lineages (for instance in invasive acacia species; Thompson et al. 2015). Based on released plastomes, other cpSSR loci could also be developed but many of them are not conserved among mimosoid genera and can not be used for comparative analyses at large taxonomical scale.

Our set of cpSSR markers, which is expected universal for the African acacias, may allow direct comparisons of diversity patterns on a mostly maternally inherited haploid genome, even on species complexes with distinct ploidy levels [e.g., *Vachellia* spp. (Bukhari 1997; El Ferchichi et al. 2009); *Senegalia* spp. (Assoumane et al. 2013; Odee et al. 2015)]. In our study, only three species harbor several haplotypes, and the highest diversity is detected in *V. nilotica* ($D = 0.78$) for which only one population was sampled (Fig. 3; Table 5). In contrast, its most closely species, *V. flava*, does not show any variation among two populations. The reason for such a difference remains unclear but could be due to distinct colonization histories and/or demographic processes. The occurrence of phylogenetically divergent chlorotypes in the *V. nilotica* population also suggests a possible event of hybridization/admixture between differentiated taxa/populations. The combined use of our cpSSR markers with bi-parentally inherited markers may help separating the contribution of pollen- and seed-mediated gene flows and unraveling such admixture events in species complexes [*Vachellia* spp. (Ali and Qaiser 1980, 1992); *Senegalia* spp. (Odee et al. 2012)]. While the nuclear genome analysis is limited by variable ploidy levels among and within species, the use of nuclear markers also remains essential to investigate the reproductive strategies of mimosoid species by assessing the importance of asexual

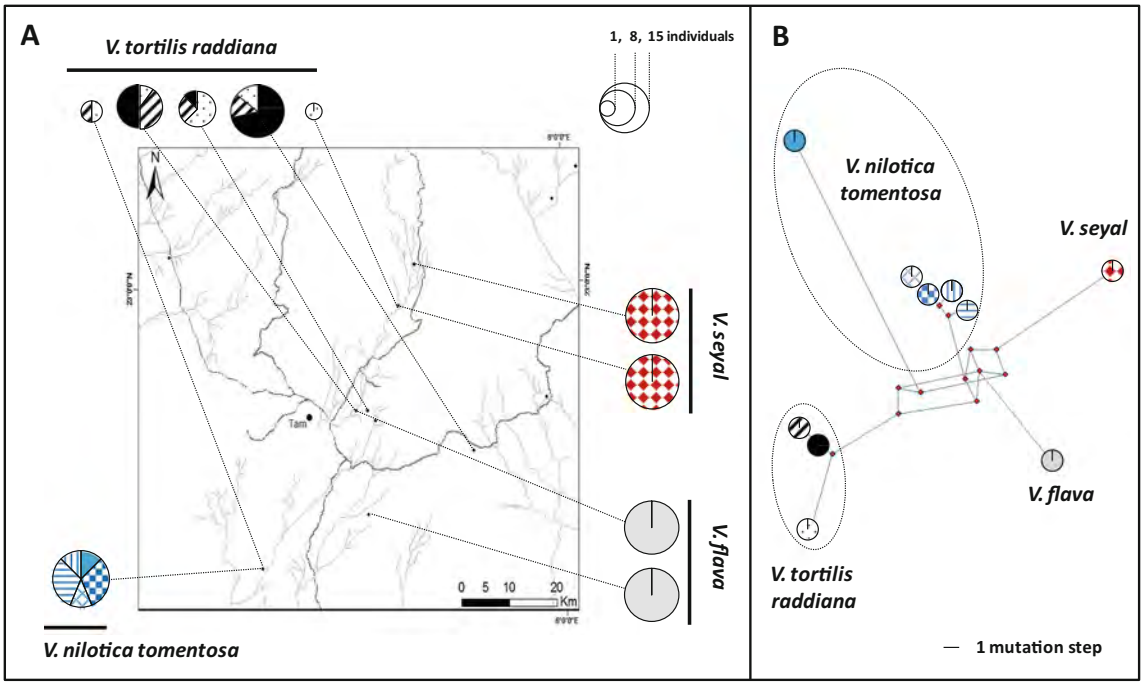


Fig. 3 Plastid diversity assessed with cpSSRs in *Vachellia* spp. **a** Distribution of haplotypes in locations sampled around Tamanrasset. **b** Phylogenetic relationships between plastid haplotypes of *Vachellia* spp. assessed with a reduced median network (Bandelt et al. 1999). The data matrix used for the network analysis is given in Table S8. The network without downweigthing character changes according to the number of alleles per locus is given in Fig. S7

(e.g., clonal growth or apomixis) and sexual reproduction in semi-desert conditions (e.g., Pichot et al. 2001; Andrew et al. 2003; Abdoun et al. 2005; Baali-Cherif and Besnard 2005; Migliore et al. 2013; Roberts et al. 2016). In particular, the population dynamic of *S. laeta*, which is reported to be a triploid hybrid (Ross 1979), deserves special attention.

Acknowledgments This work has been conducted in the Laboratoire Evolution & Diversité Biologique (EDB) part of the LABEX entitled TULIP managed by Agence Nationale de la Recherche (ANR-10-LABX-0041). CVDP and GB were funded by the Regional Council Midi-Pyrenees (AAP 13053637, 2014-EDB-UT3-DOCT). This work was performed in collaboration with the GeT core facility, Toulouse, France (http://get.genotoul.fr), and was supported by France Génomique National infrastructure, funded as part of "Investissement d'avenir" program managed by Agence Nationale pour la Recherche (ANR-10-INBS-09). We are grateful to Pascal-Antoine Christin for helpful advices on the use of positive selection tests and Salah Abdelaoui who helped for the plant sampling.

Author's contribution statement MM, DBC, and GB conceived the initial project; MM and DBC conducted field expeditions and sampling; MM, SM, OB, and GB performed laboratory works; CVDP and GB analyzed the data; MM, CVDP, and GB wrote the paper; and all authors commented and approved the final version of the manuscript.

Compliance with ethical standards

Conflict of interest The authors declare that they have no conflict of interest.

References

Abdoun F, Jull AJT, Guibal F, Thinon M (2005) Radial growth of the Sahara's oldest trees: *Cupressus dupreziana A. Camus*. Tree Struct Funct 19:661–670

Ali SI, Qaiser M (1980) Hybridisation in *Acacia nilotica* (Mimosoideae) complex. Bot J Linn Soc 80:69–77

Ali SI, Qaiser M (1992) Hybridization between *Acacia nilotica subsp. indica* and subsp. *cupressiformis*. Pak J Bot 24:88–94

Andrew RA, Miller JT, Peakall R, Crisp MD, Bayer RJ (2003) Genetic, cytogenetic and morphological patterns in a mixed mulga population: evidence for apomixis. Aust Syst Bot 16:69–80

Assoumane A, Vaillant A, Mayaki AZ, Verhaegen D (2009) Isolation and characterization of microsatellite markers for *Acacia senegal* (L.) Willd., a multipurpose arid and semi-arid tree. Mol Ecol Resour 9: 1380–1383

Assoumane A, Zoubeirou AM, Rodier-Goud M, Favreau B, Bezancon G, Verhaegen D (2013) Highlighting the occurrence of tetraploidy in *Acacia senegal* (L.) Willd. and genetic variation patterns in its natural range revealed by DNA microsatellite markers. Tree Genet Genomes 9:93–106

Baâli-Cherif D (2007) Etude des populations d'olivier de Laperrine (*Olea europaea* ssp. *laperrinei*) du Sahara Central Algérien (Hoggar et Tassili): aspects biologiques et caractérisation moléculaire. Thèse de Doctorat d'Etat, INA-Alger, 106 p

Baali-Cherif D, Besnard G (2005) High genetic diversity and clonal growth in relict populations of *Olea europaea* subsp. *laperrinei* (Oleaceae) from Hoggar, Algeria. Ann Bot 96:823–830

Bandelt HJ, Forster P, Röhl A (1999) Median-joining networks for inferring intraspecific phylogenies. Mol Biol Evol 16:37–48

Besnard G, Bervillé A (2002) On chloroplast DNA variations in the olive (*Olea europaea* L.) complex: comparison of RFLP and PCR polymorphisms. Theor Appl Genet 104:1157–1163

Besnard G, Christin PA, Baali-Cherif D, Bouguedoura N, Anthelme F (2007) Spatial genetic structure in the Laperrine's olive (*Olea europaea* subsp. *laperrinei*), a long-living tree from central-Saharan mountains. Heredity 99:649–657

Besnard G, Hernández P, Khadari B, Dorado G, Savolainen V (2011) Genomic profiling of plastid DNA variation in the Mediterranean olive tree. BMC Plant Biol 11:80

Besnard G, Christin PA, Malé PJG, Coissac E, Ralimanana H, Vorontsova MS (2013) Phylogenomics and taxonomy of Lecomtelleae (Poaceae), an isolated panicoid lineage from Madagascar. Ann Bot 112:1057–1066

Binks RM, O'Brien M, MacDonald B, Maslin B, Byrne M (2015) Genetic entities and hybridisation within the *Acacia microbotrya* species complex in Western Australia. Tree Genet Genomes 11:65

Blazier JC, Jansen RK, Mower JP, Govindu M, Zhang J, Weng ML, Ruhlman TA (2016) Variable presence of the inverted repeat and plastome stability in *Erodium*. Ann Bot 117:1209–1220

Borgel A, Cardoso C, Chevalier MH, Danthu P, Leblanc JM (1992) Diversité génétique des acacias sahéliens: exploitation par les voies clonales et sexuée. In: Wolf JN (ed) Interactions plantes microorganismes. Fondation Internationale pour la Sciences, Stockholm, pp 199–200

Bouchenak-Khelladi Y, Maurin O, Hurter J, van der Bank M (2010) The evolutionary history and biogeography of Mimosoideae (Leguminosae): an emphasis in African acacias. Mol Phylogenet Evol 57:495–508

Brenan JPM (1983) The present state of taxonomy of four species of Acacia (*A. albida*, *A. senegal*, *A. nilotica*, *A. tortilis*). FAO UN, Rome

Bukhari YM (1997) Cytoevolution of taxa in *Acacia* and *Prosopis* (Mimosaceae). Aust J Bot 45:879–889

Chenoune K (2005) La flore et la végétation du Hoggar. Bois For Trop 284:79–84

Chevalier MH, Borgel A (1998) Diversité génétique des acacias. In: Campa C, Grignon C, Gueye M & Hamon S (Eds), L'acacia au Sénégal. ORSTOM, ISRA. Paris: ORSTOM, pp. 287–308

Chevalier MH, Brizard JP, Diallo I, Leblanc JM (1994) La diversité génétique dans le complexe *Acacia senegal*. Bois For Trop 240: 5–12

Chumley TW, Palmer JD, Mower JP, Fourcade HM, Calie PJ, Boore JL, Jansen RK (2006) The complete chloroplast genome sequence of *Pelargonium* × *hortorum*: organization and evolution of the largest and most highly rearranged chloroplast genome of land plants. Mol Biol Evol 23:2175–2190

Corriveau JL, Coleman AW (1988) Rapid screening method to detect potential biparental inheritance of plastid DNA and results over 200 angiosperm species. Am J Bot 75:1443–1458

CSFD (2004) Impact des pratiques humaines sur la conservation et la gestion in situ des ressources génétiques forestières: cas d' *Acacia tortilis raddiana* et de *Balanites aegyptiaca*. Projet CSFD 57, 68 p

Diallo AM, Nielsen LR, Hansen JK, Ræbild A, Kjær ED (2015) Study of quantitative genetics of gum arabic production complicated by variability in ploidy level of *Acacia senegal* (L.) Willd. Tree Genet Genomes 11:80

Doran JC, Boland DJ, Turnbull JW, Gunn BV (1983) Guide des semences d'acacias des zones sèches: Récolte, extraction, nettoyage, conservation et traitement des graines d'acacias des zones sèches. FAO, Rome

Downie SR, Jansen RK (2015) A comparative analysis of whole plastid genomes from the Apiales: expansion and contraction of the inverted repeat, mitochondrial to plastid transfer of DNA, and identification of highly divergent noncoding regions. Syst Bot 40:336–351

Dugas DV, Hernandez D, Koenen EJM, Schwarz E, Straub S, Hughes CE, Jansen RK, Nageswara-Rao M, Staats M, Trujillo JT, Hajrah NH, Alharbi NS, Al-Malki AL, Sabir JSM, Bailey CD (2015) Mimosoid legume plastome evolution: IR expansion, tandem repeat expansions, and accelerated rate of evolution in *clpP*. Sci Rep 5: 16958

Edgar RC (2004) Muscle: a multiple sequence alignment method with reduced time and space complexity. BMC Bioinform 5:113

El Amin HM (1976) *Acacia laeta* (R. Br.) ex. Benth., considered as a species of hybrid origin. Sudan Silva 3:14–23

El Ferchichi Ouarda H, Walker DJ, Khouja ML, Correal E (2009) Diversity analysis of *Acacia tortilis* (Forsk.) Hayne ssp. *raddiana* (Savi) Brenan (Mimosaceae) using phenotypic traits, chromosome counting and DNA content approaches. Genet Resour Crop Evol 56: 1001–1010

Erixon P, Oxelman B (2008) Whole-gene positive selection, elevated synonymous substitution rates, duplication, and indel evolution of the chloroplast *clpP1* gene. PLoS One 3:e1386

FAO (1996) Role of *Acacia* species in the rural economy of dry Africa and the Near East. Conservation Guide 27, Rome, 137 p

FAO (2014) Action against desertification. http://www.fao.org/in-action/action-against-desertification/activities/africa/en/

Giffard PL (1966) Les gommiers: *Acacia senegal* Willd., *Acacia laeta* R. Br. Bois For Trop 105:21–32

Goudet J (1995) FSTAT (version 1.2): a computer program to calculate *F*-statistics. J Hered 86:485–486

Guajardo JCR, Schnabel A, Ennos R, Preuss S, Otero-Arnaiz A, Stone G (2010) Landscape genetics of the key African acacia species *Senegalia mellifera* (Vahl)—the importance of the Kenyan Rift Valley. Mol Ecol 19:5126–5139

Guisinger MM, Kuehl JV, Boore JL, Jansen RK (2011) Extreme reconfiguration of plastid genomes in the angiosperm family Geraniaceae: rearrangements, repeats, and codon usage. Mol Biol Evol 28:583–600

Guo X, Castillo-Ramírez S, González V, Bustos P, Fernández-Vázquez JL, Santamaría RI, Arellano J, Cevallos MA, Dávila G (2007) Rapid evolutionary change of common bean (*Phaseolus vulgaris* L.) plastome, and the genomic diversification of legume chloroplasts. BMC Genomics 8:228

Harris SA, Fagg CW, Barnes RD (1997) Isozyme variation in *Faidherbia albida* (Leguminosae, Mimosoideae). Plant Syst Evol 207:119–132

Harrison RG, Larson EL (2014) Hybridization, introgression, and the nature of species boundaries. J Hered 105:795–809

Hobbs JJ, Krzywinski K, Andersen GL, Talib M, Pierce RH, Saadallah AEM (2014) Acacia trees on the cultural landscapes of the Red Sea Hills. Biodivers Conserv 23:2923–2943

Joly HI, Zehnlo M, Danthu P, Aygalent C (1992) Population-genetics of an African acacia, *Acacia albida*: 1. Genetic diversity of populations from West Africa. Aust J Bot 40:59–73

Kearse M, Moir R, Wilson A, Stones-Havas S, Sturrock S, Buxton S, Cooper A, Markowitz S, Duran C, Thierer T, Ashton B, Meintjes P, Drummond A (2012) Geneious Basic: an integrated and extendable desktop software platform for the organization and analysis of sequence data. Bioinformatics 28:1647–1649

Ku C, Chung WC, Chen LL, Kuo CH (2013) The complete plastid genome sequence of Madagascar periwinkle *Catharanthus roseus* (L.) G. Don: plastid genome evolution, molecular marker identification, and phylogenetic implications in asterids. PLoS One 18:e68518

Kumar S, Stecher G, Tamura K (2016) MEGA7: Molecular Evolutionary Genetics Analysis version 7.0 for bigger datasets. Mol Biol Evol 33: 1870–1874

Kyalangalilwa B, Boatwright JS, Daru BH, Maurin O, van der Bank M (2013) Phylogenetic position and revised classification of *Acacia s.l.* (Fabaceae: Mimosoideae) in Africa, including new combinations in *Vachellia* and *Senegalia*. Bot J Linn Soc 172:500–523

Le Floc'h E, Grouzis M (2003) *Acacia raddiana*, un arbre des zones arides à multiples usages. In: Grouzis M, Le Floc'h E (eds) Un arbre au désert: Accacia raddiana. IRD, Paris, pp 21–58

Lee HL, Jansen RK, Chumley TW, Kim KJ (2007) Gene relocations within chloroplast genomes of *Jasminum* and *Menodora* (Oleaceae) are due to multiple, overlapping inversions. Mol Biol Evol 24:1161–1180

Lewis GP, Schrire B, MacKinder B, Lock M (2005) Legumes of the world. Royal Botanic Gardens, Kew

Lohse M, Drechsel O, Kahlau S, Bock R (2013) OrganellarGenomeDRAW—a suite of tools for generating physical maps of plastid and mitochondrial genomes and visualizing expression data sets. Nucleic Acids Res 41:W575–W581

Luckow M, Miller J, Murphy DJ, Livshultz T (2003) A phylogenetic analysis of Mimosoideae (Leguminosae) based on chloroplast DNA sequence. In: Klitgaard BB, Bruneau B (eds) Advances in legume systematics, chap. 10. The Royal Botanic Garden, Kew, pp 197–220

Magee AM, Aspinal S, Rice DW, Cusack BP, Sémon M, Perry AS, Stefanović S, Milbourne D, Barth S, Palmer JD, Gray JC, Kavanagh TA, Wolfe KH (2010) Localized hypermutation and associated gene losses in legume chloroplast genomes. Genome Res 20:1700–1710

Maslin BR (2008) Generic and subgeneric names in *Acacia* following retypification of the genus. Muelleria 26:7–9

Maslin BR, Miller J, Seigler DS (2003) Overview of the generic status of *Acacia* (Leguminosae: Mimosoideae). Aust Syst Bot 16:1–18

Migliore J, Baumel A, Juin M, Médail F (2012) From Mediterranean shores to central Saharan mountains: key phylogeographical insights from the genus *Myrtus*. J Biogeogr 39:942–956

Migliore J, Baumel A, Juin M, Fady B, Roig A, Duong N, Médail F (2013) Surviving in mountain climate refugia: new insights from the genetic diversity and structure of the relict shrub *Myrtus nivellei* (Myrtaceae) in the Sahara Desert. PLoS One 8:e73795

Murphy DJ (2008) A review of the classification of *Acacia* (Leguminosae, Mimosoideae). Muelleria 26:10–26

Navascués M, Emerson BC (2005) Chloroplast microsatellites: measures of genetic diversity and the effect of homoplasy. Mol Ecol 14:1333–1341

Odee DW, Telford A, Wilson J, Gaye A, Cavers S (2012) Plio-Pleistocene history and phylogeography of *Acacia senegal* in dry woodlands and savannahs of sub-Saharan tropical Africa: evidence of early colonisation and recent range expansion. Heredity 109:372–382

Odee DW, Wilson J, Omondi S, Perry A, Cavers S (2015) Rangewide ploidy variation and evolution in *Acacia senegal*: a north–south divide? AoB Plants 7:plv011

Osmondi SF, Machua J, Gicheru J, Hanaoka S (2015) Isolation and characterization of microsatellite markers for *Acacia tortilis* (Forsk.) Hayne. Conserv Genet Resour 7:529–531

Payne WA, Williams JH, Moussa KAM, Stern RD (1998) Crop diversification in the Sahel through use of environmental changes near *Faidherbia albida* (Del) A Chev. Crop Sci 38:1585–1591

Pérez-Jiménez M, Besnard G, Dorado G, Hernández P (2013) Varietal tracing of virgin olive oils based on plastid DNA variation profiling. PLoS One 8:e70507

Petit RJ, Kremer A, Wagner DB (1993) Finite island model for organelle and nuclear genes in plants. Heredity 71:930–641

Pichot C, Maâtaoui M, Raddi S, Raddi P (2001) Surrogate mother for endangered *Cupressus*. Nature 412:39

Powell W, Morgante M, McDevitt R, Vendramin G, Rafalski J (1995) Polymorphic simple-sequence repeat regions in chloroplast genomes: applications to the population genetics of pines. Proc Natl Acad Sci U S A 92:7759–7763

Quézel P (1965) La végétation du Sahara: du Tchad à la Mauritanie. Gustav Fischer Verlag, Stuttgart 333 p

Quézel P (1997) High mountains of the Central Sahara: dispersal, speciation, origin and conservation of the flora. In: Barakat HN, Hegazy AK (eds) Desert conservation and development. Metropole, Cairo, pp 159–175

Rambaut A (2014) FigTree 1.4.2. Available from http://tree.bio.ed.ac.uk

Reboud X, Zeyl C (1994) Organelle inheritance in plants. Heredity 72: 132–140

Roberts DG, Forrest CN, Denham AJ, Ayre DJ (2016) Varying levels of clonality and ploidy create barriers to gene flow and challenges for conservation of an Australian arid-zone ecosystem engineer, *Acacia loderi*. Biol J Linn Soc 118:330–343

Rockenbach K, Havird JC, Monroe JG, Triant DA, Taylor DR, Sloan DB (2017) Positive selection in rapidly evolving plastid-nuclear enzyme complexes. Genetics. doi:10.1534/genetics.116.188268

Ross JH (1979) A conspectus of the African Acacia species. Mem Bot Surv South Afr 44:1–155

Ross JH (1981) An analysis of the African Acacia species: their distribution, possible origins and relationships. Bothalia 13:389–413

Sahki R, Boucheneb N, Sahki A (2004) Guide des principaux arbres et arbustes du Sahara Central (Ahaggar et Tassili). Ed. INRF, Alger, 142 p

Schuelke M (2000) An economic method for the fluorescent labelling of PCR fragments. Nat Biotechnol 18:233–234

Schwarz EN, Ruhlman TA, Bailey CD, Jansen RK (2015) Plastid genome sequences of legumes reveal parallel inversions and multiple losses of *rps16* in papilionoids. J Syst Evol 53:458–468

Seigler DS, Ebinger JE (2009) New combinations in the genus *Senegalia* (Fabaceae: Mimosoideae). Phytologia 91:26–30

Sloan DB, Triant DA, Forrester NJ, Bergner LM, Wu M, Taylor DR (2014) A recurring syndrome of accelerated plastid genome evolution in the angiosperm tribe Sileneae (Caryophyllaceae). Mol Phylogenet Evol 72:82–89

Stamatakis A (2014) RAxML version 8: a tool for phylogenetic analysis and post-analysis of large phylogenies. Bioinformatics 30:1312–1313

Straub SCK, Parks M, Weitemier K, Fishbein M, Cronn RC, Liston A (2012) Navigating the tip of the genomic iceberg: next-generation sequencing for plant systematics. Am J Bot 99:349–364

Taylor CL, Barker NP (2012) Species limits in *Vachellia* (*Acacia*) *karroo* (Mimosoideae: Leguminoseae): evidence from automated ISSR DNA "fingerprinting". South Afr J Bot 83:36–43

Thompson GD, Bellstedt DU, Richardson DM, Wilson JRU, Le Roux JJ (2015) A tree well travelled: global genetic structure of the invasive tree *Acacia saligna*. J Biogeogr 42:305–314

Wang L, Wuyun TN, Du HY, Wang DP, Cao DM (2016) Complete chloroplast genome sequences of *Eucommia ulmoides*: genome structure and evolution. Tree Genet Genomes 12:12

Weising K, Gardner RC (1999) A set of conserved PCR primers for the analysis of simple sequence repeat polymorphisms in chloroplast genomes of dicotyledonous angiosperms. Genome 42:9–19

Williams AV, Boykin LM, Howell KA, Nevill PG, Small I (2015) The complete sequence of the *Acacia ligulata* chloroplast genome reveals a highly divergent *clpP1* gene. PLoS One 10:e0125768

Williams AV, Miller JT, Small I, Nevill PG, Boykin LM (2016) Integration of complete chloroplast genome sequences with small amplicon datasets improves phylogenetic resolution in *Acacia*. Mol Phylogenet Evol 96:1–8

Winters G, Shklar G, Korol L (2013) Characterizations of microsatellite DNA markers for *Acacia tortilis*. Conserv Genet Resour 5: 807–809

Xu JH, Liu Q, Hu W, Wang T, Xue Q, Messing J (2015) Dynamics of chloroplast genomes in green plants. Genomics 106:221–231

Yang Z (2007) PAML 4: phylogenetic hypothesis by maximum likelihood. Mol Biol Evol 24:1586–1591

Yang Z, Nielsen R, Goldman N, Krabbe Pedersen AM (2000) Codon-substitution models for heterogeneous selection pressure at amino acid sites. Genetics 155:431–449

Zhang J, Nielsen R, Yang Z (2005) Evaluation of an improved branch-site likelihood method for detecting positive selection at the molecular level. Mol Biol Evol 22:2472–2479

Author : Céline VAN DE PAER

Title : Structural diversity and contrasted evolution of cytoplasmic genomes in flowering plants: a phylogenomic approach in Oleaceae

Supervisor : Guillaume BESNARD

Abstract :

In plants, the structural dynamics and concerted evolution of nuclear and cytoplasmic genomes are poorly understood. The objective of this thesis was to study the structural diversity and evolution of mitogenomes and plastomes in the family Oleaceae with a phylogenomic approach. First, we assembled mitogenomes from low-coverage sequencing data obtained from live and herbarium material. Considerable structural variation of mitogenomes was observed in the olive, and a chimeric gene potentially associated to a type of male sterility was detected. Finally, we studied the evolution of plastomes and mitochondrial genes in the Oleaceae. Accelerated evolution of plastomes was observed in two independent lineages. This change of evolutionary rate could be the consequence of an occasional transmission of plastids with pollen, modifying selective pressures on some genes.

Key-words : Oleaceae**,** *Olea europaea***,** mitogenome, paternal leakage, phylogenomic

Auteur : Céline VAN DE PAER

Titre : Diversité structurelle et évolution contrastée des génomes cytoplasmiques des plantes à fleurs : une approche phylogénomique chez les Oleaceae

Directeur de thèse : Guillaume BESNARD

Lieu et date de soutenance : Université Paul Sabatier, Toulouse, le 19 décembre 2017

Résumé :

Chez les plantes, la dynamique structurelle et l'évolution concertée des génomes nucléaire et cytoplasmiques restent peu documentées. L'objectif de cette thèse était d'étudier la diversité structurelle et l'évolution des mitogénomes et des plastomes chez les Oleaceae à l'aide d'une approche de phylogénomique. Nous avons d'abord assemblé des mitogénomes à partir de données de séquençage de faible couverture, obtenues à partir de matériel frais et d'herbier. Une grande variation de structure du mitogénome a été observée chez l'olivier, et un gène chimérique potentiellement associé à un type de stérilité mâle a été identifié. Enfin, nous avons étudié l'évolution des plastomes et des gènes mitochondriaux chez les Oleaceae. Une accélération de l'évolution du plastome a été observée dans deux lignées indépendantes. Ce changement de trajectoire évolutive pourrait être la conséquence d'une transmission occasionnelle de plastes par le pollen, modifiant les pressions sélectives sur certains gènes.

Mots-clés : Oleaceae**,** *Olea europaea***,** mitogenome, paternal leakage, phylogénomique

Discipline administrative : Ecologie - Evolution

Intitulé et adresse du laboratoire : Laboratoire Ecologie & Diversité Biologique (EDB)
UMR 5174 (CNRS/UPS/ENFA), Université Paul Sabatier, Bâtiment 4R1
118 route de Narbone, 31062 Toulouse cedex 9, FRANCE