



UNIVERSITE DES ANTILLES ET DE LA GUYANE
Faculté des Sciences exactes et naturelles

École doctorale pluridisciplinaire :
Santé, Environnement et Sociétés dans les Amériques

THESE

pour obtenir le grade de
Docteur de l'Université des Antilles et de la Guyane
en Génétique Écologique

par

Stéphanie BARTHE

***Signatures de transitions démographiques
des arbres de la Forêt Tropicale Humide
du plateau des Guyanes***

sous la direction d'Ivan SCOTTI
et Bruno HERAULT

soutenue le 14 décembre 2012
à Kourou

N°: [2012AGUY0568]

Jury

Jérôme ORIVEL	DR2, CNRS, EcoFog	Président du jury
Arnaud ESTOUP	DR2, INRA, Supagro Montpellier	Rapporteur
Bruno FADY	DR2, INRA, URFM	Rapporteur
Alain ROUSTEAU	MCF, Université des Antilles-Guyane	Examineur
Benoît de THOISY	CR1, Institut Pasteur de Guyane	Examineur
Ivan SCOTTI	DR2, INRA, EcoFog	Directeur de Thèse

À Maxou-bidou ...

Remerciements

Mes premiers remerciements vont tout naturellement à mes directeurs de thèse. Tout d'abord à Ivan Scotti qui après m'avoir accueilli en tant que stagiaire m'a offert l'opportunité de mener une thèse sur un sujet qui m'a passionné. Merci pour la confiance que tu m'as accordé, pour ton encadrement pendant ces quatre années et pour toutes les connaissances que tu as su me transmettre. Merci de m'avoir poussé à dépasser mes limites, tout en ayant toujours été présent pour me rattraper avant la chute. Je te suis également reconnaissante d'avoir su respecter mon tempérament introverti. Je remercie chaleureusement mon co-directeur de thèse, Bruno Hérault, dont l'esprit critique m'a permis de développer une certaine rigueur. Merci pour ton aide précieuse en statistiques et en cette fin de thèse, sans oublier ta bonne humeur sur le terrain. Mes remerciements vont aussi à Caroline Scotti-Saintagne qui m'a fait prendre goût à la génétique évolutive et sans qui cette belle aventure n'aurait jamais pu commencer. Merci pour ton aide, tes conseils et ton immense gentillesse.

Je tiens à remercier Eric Marcon et Marc Gibernau pour m'avoir accueilli au sein de l'UMR Ecofog et pour avoir mis à ma disposition tous les moyens nécessaires au bon déroulement de ma thèse. Merci à Marc pour son soutien moral de fin de thèse.

Mes remerciements vont également à Hélène Fréville dont les conseils m'ont permis de faire le point et de finir cette thèse dans les temps. Je remercie mes rapporteurs Arnaud Estoup et Bruno Fady, mes examinateurs Alain Rousteau, Benoît de Thoisy et Jérôme Orivel d'avoir accepté d'évaluer mon travail et de participer à ma soutenance.

Un grand merci à tous les membres du laboratoire de génétique écologique et notamment à Valérie et Eliane qui font vivre ce lieu avec tant de bonne humeur et d'énergie. Votre aide m'a juste été indispensable et votre soutien professionnel et personnel m'a fait chaud au cœur. Je remercie également Saint-Omer pour avoir veillé sur moi en forêt. Les courbatures, les pieds meurtris après des journées de marche, les guêpes, les orpailleurs et le boa constrictor ne resteront que des bons souvenirs. Merci à Mélanie qui a participé dans le cadre de son stage à une partie de cette thèse, ça a été un plaisir de t'encadrer. Merci à Delphine, Pierre, William, Maxime, Louise, Alex, Vincent, Caro D. pour votre aide et nos échanges au quotidien. Je remercie également Stéphane Traissac, Céline Leroy, Angélique Quilichini, Christopher Baraloto, Heidy Schiman, Benoît Burban, de façon plus générale les membres des autres équipes de recherche ainsi que Josie, Carole, Louise B. et toute l'équipe des ressources humaines pour les nombreux services rendus. Merci à tous les anciens et nouveaux thésards, notamment Jérémie, Greg, Caroline P-J, Tatiane, Mélaïne et plus spécialement Quentin avec qui j'ai effectué le sprint final des dernières semaines.

Je tiens à remercier toute l'équipe d'Archéologie de Guyane. Merci à Sylvie Jérémie et Gérard Migeon de m'avoir consacré de leur temps pour me faire découvrir leur discipline. Un grand merci à Martijn van den Bel d'avoir été mon binôme de terrain et ami, de m'avoir préparé de délicieux sandwichs crackers-pâté, de m'avoir suivi dans

mes chutes au fond des criques et plus sérieusement de m'avoir apporté ton aide sur toute la partie occupation anthropique de ma thèse. Je tiens également à remercier Clara Samuelian et Jeanne Brancier pour leurs participations aux prospections et les bons moments que nous avons partagés.

Merci à mes amis d'« Asutanon Upiniamon », Olivia, Carlos, les jeunes mariés Nathalie et Cleoson, Louisia, Isabelle et tous les autres pour les aventures inoubliables que nous avons vécu ensemble ; en espérant qu'il y en ait encore de nombreuses. Merci à la troupe de métropole, dont ma jolie blonde Estelle pour avoir toujours gardé le contact, ma chéwi douce qui m'accueille dans sa tribu à chacun de mes retours, Mimile et Djo pour lesquels mon amitié ne sera jamais effacée par la distance.

Je remercie ma famille aux mille couleurs, qui en font sa force. Un immense merci à ma ch'tite maman et mon papa de cœur pour tout le soutien, le réconfort et l'amour que vous m'apportez. Je remercie également mon père pour ses attentions et ses nombreuses prières. Merci à ma Mémère, ma Ploum Ploum et mon petit Pô Tam pour les instants de bonheur que vous me faites vivre et à mon frère pour la fierté que je peux lire sur ton visage. Je remercie également mes beaux-parents qui m'ont accueilli dans leur famille, Petit Poulet, Brandon, Finéasse et Shanice dont les inventions insolites me font toujours autant rire.

Et enfin, merci pour tout Doudou. Ta présence et ton soutien sans faille ont pour moi une valeur inestimable. Mo kontan to.

Résumé

La distribution actuelle des espèces d'arbres des forêts tropicales humides amazoniennes est le résultat de processus dynamiques sur lesquels les perturbations passées, d'origine climatique ou anthropique, peuvent avoir jouées un rôle majeur. Des outils génétiques actuels associés à des méthodes d'analyses puissantes permettent d'inférer et de dater l'empreinte génomique des événements démographiques subits par les populations naturelles. Le but de cette thèse est de contribuer à une vision d'ensemble de l'évolution démographique de la forêt tropicale humide du plateau des Guyanes pendant le Quaternaire. J'ai d'abord étudié les propriétés des microsatellites (répétitions simple de séquences, SSR) et les limites de leurs applications en tant que marqueurs moléculaires en phylogéographie. Plusieurs allèles SSR contenu dans diverses populations de trois genres d'arbres différents ont été séquencés et le polymorphisme, porté par chaque partie des fragments ADN contenant des SSR, a été étudié. J'ai observé que la quantité de variation génétique était aussi importante dans les régions encadrant le microsatellite que dans le SSR lui-même. Ces régions ont contribué significativement à l'information phylogénétique et ont parfois été la principale source de différenciation entre les individus et les populations. Cette expérience m'a permis de choisir la meilleure stratégie à adopter pour traiter les données de variation génétique des SSR dans des modèles démographiques. J'ai ensuite utilisé ces marqueurs SSR et des méthodes ABC pour étudier l'impact local des occupations humaines précolombiennes sur les forêts en Guyane française. Les changements démographiques subis par des populations de quatre espèces d'arbres ont été comparés entre cinq sites occupés et quatre sites non perturbés. Les perturbations naturelles et induites par l'homme ont contribué à la mise en place de la diversité observée actuellement, comme le suggère des signatures d'effet fondateur détectées sur les sites perturbés ou non. Néanmoins, la taille des populations fondatrices semble être proportionnelle à l'impact humain estimé, indiquant que les perturbations anthropique ont induit un remplacement des populations plus important que les processus écologiques naturels. Pour terminer, j'ai mené une étude de démographie comparative d'espèces d'arbres tropicaux à l'échelle régionale de la Guyane française. Cinq scénarios démographiques ont été testés pour neuf populations de huit espèces en utilisant des microsatellites nucléaires et des séquences chloroplastiques. Les changements démographiques les plus récents détectés convergent vers une signature générale d'expansion de la forêt humide depuis probablement la dernière glaciation. Cela peut suggérer, soit que la composition des forêts du plateau des Guyanes était différente, ou que le couvert forestier était moins étendu. Ce résultat constitue le premier indice de changement passé, probablement d'origine climatique, subi par les communautés forestières du plateau des Guyanes ; qui n'avait jusque-là pas pu être testé avec des méthodes standards en raison d'un manque de fossiles. En conclusion, j'ai recueilli des preuves génétiques que les perturbations, à différents échelles de temps, ont marqué la structure et la composition génétique des forêts du plateau des Guyanes. L'application de ces méthodes à d'autres régions Néotropicales pourrait contribuer, à l'avenir, à reconstruire l'histoire écologique de la forêt amazonienne de façon détaillée.

Abstract

The current tree species distribution in Amazonian tropical rainforests is the outcome of dynamic processes in which past perturbations, climatic or anthropogenic, may have played a major role. Currently available genetic tools associated with powerful analytical methods allow inferring and dating the genomic imprint of demographic events undergone by natural populations. The aim of my thesis is to contribute to an overview of the demographic evolution of the tropical rainforest in the Guiana shield during the Quaternary. I have first studied the properties of microsatellites or Simple sequence repeat (SSR) and the limits in their applications as molecular markers in phylogeography. Several SSR alleles in multiple populations of three divergent tree genera have been sequenced and the sources of polymorphisms, carried by each portion of SSR-containing DNA amplicons, have been broken apart. I have observed that the amount of variation was as large in flanking regions as in the SSR itself. The former contributed significantly to the phylogenetic information and sometimes was the main source of differentiation among individuals and populations. This experiment allowed me to choose the best strategy to treat SSR variation in demographic modeling. I have subsequently used SSR markers and ABC methods to study the impact of pre-Columbian human settlements at the local level on rainforests in French Guiana. Demographic changes undergone by populations of four tree species were compared between five settled and four undisturbed sites. Both natural turnover and human-induced disturbances appeared as shaping currently observed genetic diversity, with signatures of founder effect both at disturbed and undisturbed sites. Nevertheless, the size of founding populations appeared to be proportional to the estimated human impact, suggesting that human-made disturbance caused deeper population turnover than background ecological processes. Finally, I have conducted a comparative demography study of tropical tree species at the regional level across French Guiana. Five demographic scenarios have been tested for nine populations from eight species using both nuclear microsatellites and chloroplastic sequences. The most recent demographic signals detected converge on a global signature of expansion of the rainforest since probably the last glaciation. This finding can suggest either that the composition of the Guiana shield rainforests was different, or that its global extent was smaller, during earlier epochs of the Pleistocene than today. This result is the first firm indication of changes undergone by the forest communities of the Guiana shield in the distant climatic past – something which could not be tested with standard methods due to almost complete lack of fossil evidence. In conclusion, I have gathered genetic evidence that disturbances, both in historical and geological times, have marked the structure and composition of forests in the Guiana shield. The extensive application of these methods to other regions of the Neotropics may contribute, in the future, to the detailed reconstruction of the ecological history of Amazonian forests.

Table des matières

Chapitre I	Introduction	1
I.1	Contexte de perturbations passées à l'échelle mondiale.....	3
I.2	Perturbations en Amazonie pendant le Tertiaire et le Quaternaire	7
I.2.1	Formation de l'Amazonie.....	7
I.2.2	Le plateau des Guyanes	9
I.2.3	Changements climatiques du Quaternaire	13
I.2.4	Conséquences des perturbations climatiques sur la forêt amazonienne.....	17
I.3	Perturbations en Guyane française pendant l'Holocène	21
I.3.1	Types d'occupations précolombiennes.....	23
I.3.2	Artefacts de l'occupation précolombienne humaine.....	25
I.3.3	Conséquences des perturbations anthropiques sur la forêt.....	27
I.4	Inférence de l'histoire écologique des communautés forestières du plateau des Guyanes.....	31
I.4.1	De l'histoire démographique des espèces à l'histoire écologique des communautés	31
I.4.2	Généétique et histoire démographique.....	33
I.4.3	Inférences statistiques des événements démographiques	41
I.5	Objectifs.....	47
Chapitre II	Étude de l'information phylogénétique portée par un allèle microsatellite	49
Chapitre III	Impact de l'occupation humaine précolombienne sur la forêt amazonienne	99
Chapitre IV	Histoire écologique des forêts tropicales humides au Quaternaire	135

Chapitre V	Discussion générale	169
V.1	Études démographiques et histoire écologique des forêts du plateau des Guyanes.....	170
V.1.1	Une empreinte durable des perturbations.....	171
V.1.2	Une empreinte controversée des perturbations.....	173
V.2	Apport de cette étude démographique dans la reconstruction de l'histoire de l'Amazonie	175
V.3	Généralisation des études démographiques des communautés.....	176
Bibliographie	179
Annexes	191

Tables des illustrations

Table des Figures

Figure 1.	Échelle des temps géologiques	2
Figure 2.	Répartition mondiale des glaces et de la végétation au dernier maximum glaciaire 20 000 ans BP	2
Figure 3.	Reconstruction de l'histoire des forêts d'Europe (A) et d'Amérique du Nord (B) au Quaternaire	4
Figure 4.	Carte du Nord de l'Amérique du Sud comprenant le bassin amazonien	6
Figure 5.	Cartes paléogéographiques de la modification de la croûte terrestre sud-américaine au cours des temps géologiques	6
Figure 6.	Carte des reliefs et de l'hydrographie du plateau des Guyanes	8
Figure 7.	Distribution annuelle des précipitations (A) et des températures (B) actuelles au nord de l'Amérique du Sud.....	10
Figure 8.	Cartes des refuges présumés au nord de l'Amérique du Sud pendant le dernier maximum glaciaire	16
Figure 9.	Modélisation de la paléo-végétation au Quaternaire	18
Figure 10.	Carte des reliefs et des précipitations en Guyane française.....	20
Figure 11.	Carte des Refuges forestiers présumés en Guyane française pendant le dernier maximum glaciaire	20
Figure 12.	Répartition des sites archéologiques précolombiens découverts en Guyane	22
Figure 13.	Traces d'occupation humaine précolombienne en Guyane française	24
Figure 14.	Hypothèse de l'évolution des végétations guyanaises au cours de l'Holocène ...	26
Figure 15.	Approche de démographie comparée interspécifique.....	30
Figure 16.	Conséquences de la dérive génétique sur la fréquence des allèles d'une population en fonction de sa taille N	34
Figure 17.	Distribution des fréquences alléliques dans une population à l'équilibre, après expansion ou contraction démographique	36
Figure 18.	Principe de base de la coalescence	38
Figure 19.	Conséquence d'un changement démographique (A) ponctuel et drastique ou (B) exponentiel sur la topologie des arbres de coalescence.....	40
Figure 20.	Principe de l'inférence des paramètres d'une population par le calcul Bayésien approché (ABC)	44

Figure 21. Séquence microsatellite entière.....	50
Figure 22. Histoire de la forêt amazonienne en relation avec les perturbations du Tertiaire à nos jours.....	171

Table des encadrés

Encadré 1. Les principaux modèles de mutation	32
---	----

Chapitre I

Introduction

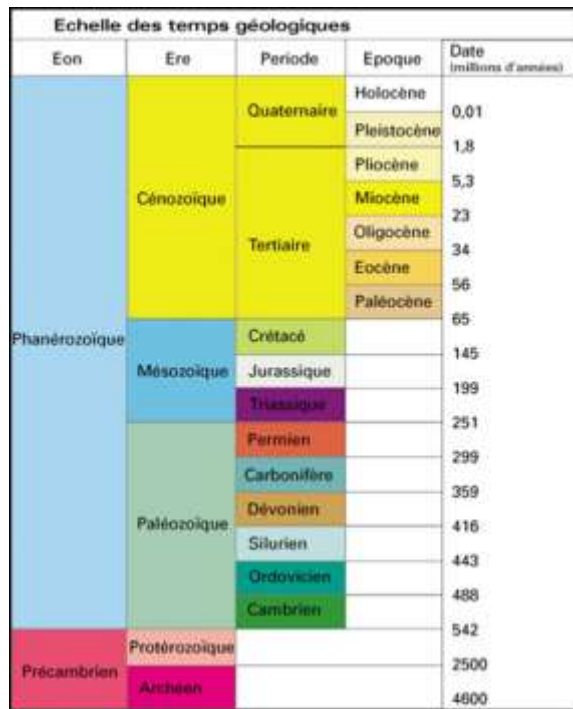


Figure 1. Échelle des temps géologiques. Source : www.sciencesnaturelles.be

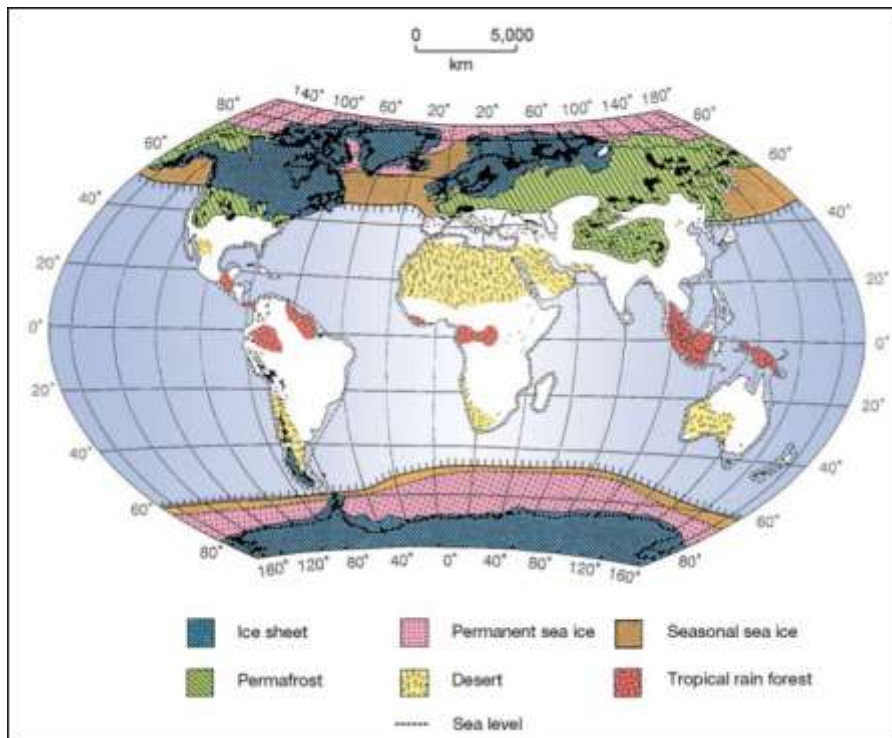


Figure 2. Répartition mondiale des glaces et de la végétation au dernier maximum glaciaire 20 000 ans BP, d'après Hewitt (2000). Les glaces, le permafrost, le niveau des mers, les déserts et les blocs de forêt tropicale sont indiqués sur la figure.

I.1 Contexte de perturbations passées à l'échelle mondiale

De la sortie des eaux des premières plantes à leur distribution actuelle, de nombreux changements ont contribué à façonner la végétation au cours du temps. Si la dislocation de la Pangée donnant le Gondwana et la Laurasia a modifié la dissémination des plantes continentales pendant l'ère Secondaire (Figure 1), des événements climatiques ont également participé à l'évolution des forêts. Le climat globalement plus chaud sur l'ensemble de la planète jusqu'au Pliocène a permis, par exemple en Europe, le développement d'une végétation tropicale à sub-tropicale jusqu'à la fin de l'Éocène, puis les refroidissements successifs ont modifié la composition de la flore et le type de végétation. Progressivement, les forêts tropicales ont laissé la place à des formations tempérées. La période Quaternaire est également riche en événements climatiques. Huit cycles glaciaires/interglaciaires, d'environ 100 000 ans chacun mais d'intensités différentes, se sont succédés au Pléistocène (EPICA, 2004). Le dernier maximum glaciaire (DMG) est survenu il y a 18 000-20 000 ans laissant place il y a 12 000 ans à la période interglaciaire actuelle. Le refroidissement global de la Terre pendant le DMG a entraîné des changements importants de végétation (Figure 2, Hewitt (2000)).

Les glaces ont alors recouvert une grande partie des continents européen et nord-américain alors que les zones équatoriales et la plupart de l'hémisphère sud ont été épargnées. En Europe, les forêts tempérées se sont principalement contractées dans 3 zones refuges situées autour du bassin méditerranéen (Balkans, Sud de l'Italie, Péninsule ibérique) et à la fonte des glaciers la recolonisation s'est effectuée du Sud vers le Nord (Taberlet *et al.*, 1998; Hewitt, 2000; Petit *et al.*, 2002; 2003). Des refuges dits « cryptiques », situés plus au Nord, semblent également avoir subsisté pendant cette période glaciaire (Petit *et al.*, 2008; Provan et Bennett, 2008) (Figure 3.A). Bien que l'Amérique du Nord soit un continent de même taille et placé à la même latitude que l'Europe, son histoire depuis le DMG est différente (Hewitt, 2000). La présence de barrières naturelles délimitant le nord du sud de l'Amérique du Nord, à la différence du continent européen où elles sont est-ouest, expliquent en partie que les refuges nord, en bord de glaciers, aient eu une histoire comparable à celle de l'Europe alors que les zones étudiées plus au sud attestent de conditions plus stables (Grivet *et al.*, 2006). À l'Ouest du continent, le climat du Quaternaire ainsi que des événements d'orogénèse ont contribué à la mise en place de multiples refuges

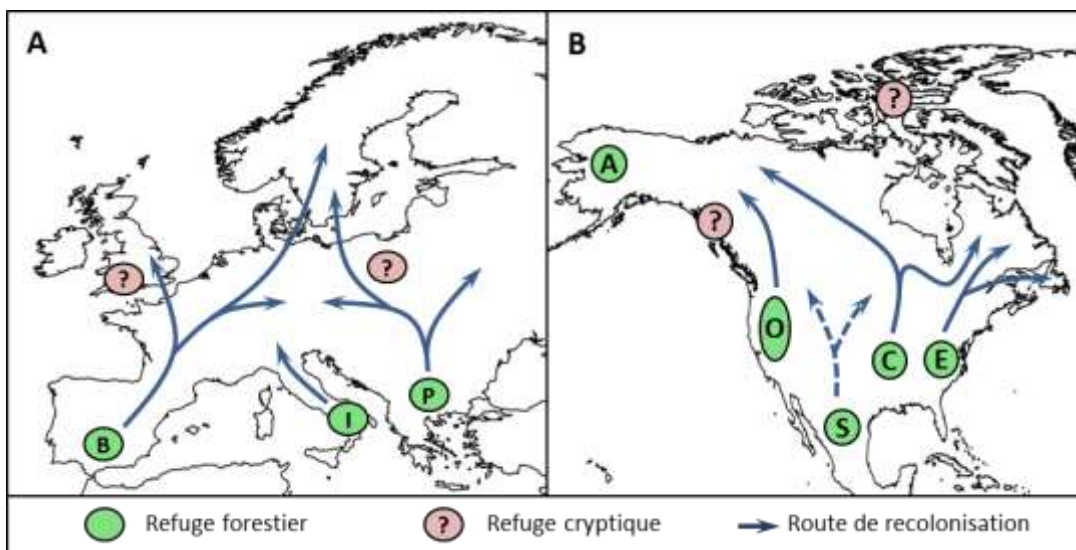


Figure 3. Reconstruction de l'histoire des forêts d'Europe (A) et d'Amérique du Nord (B) au Quaternaire. Les lettres indiquent les refuges. En Europe : Balkans (B), Italie (I) et Péninsule ibérique (P) ; refuges d'Amérique du Nord : Est (E), Centre (C), Ouest (O), Sud (S) et en Alaska (A). Les refuges et les voies de recolonisation ont été reconstruits d'après Hewitt (2000), Petit *et al.* (2003), Provan et Bennett (2008) pour l'Europe ; Fedorov et Stenseth (2002), Jaramillo-Correa *et al.* (2004), Waltari *et al.* (2007) et Shafer *et al.* (2010) pour l'Amérique du Nord.

(Shafer *et al.*, 2010) et à la distribution actuelle des espèces (Brunsfield *et al.*, 2001). Des refuges en Alaska (Anderson *et al.*, 2006) ou « cryptiques » (Shafer *et al.*, 2010) ont également été identifiés (Figure 3.B). Même si pendant le DMG, les glaciers ne sont pas descendus jusqu'aux zones tropicales, ces régions ont cependant connu des modifications climatiques. En Afrique centrale, le climat plus sec et plus froid, confirmé par la présence à faible altitude d'une végétation montagnarde entre 30 000 et 10 000 ans BP (Maley, 1991; Servant *et al.*, 1993), aurait entraîné des contractions et une fragmentation de la forêt humide dans des zones refuges situées en Guinée équatoriale, au Cameroun, au Gabon, au Congo et en République démocratique du Congo (Maley, 1991; 1996). Les routes de recolonisation des populations forestières restent encore floues même si des patrons génétiques communs entre espèces confirment l'existence de certains refuges (Dauby *et al.*, 2010; Lowe *et al.*, 2010).

Tous les peuplements forestiers n'ont pas connu les mêmes perturbations au cours des temps géologiques et n'ont donc pas tous la même histoire. Dans ce contexte de changements environnementaux passés mondiaux, je me suis intéressée plus particulièrement aux perturbations ayant affecté et construit l'histoire des forêts tropicales humides amazoniennes du plateau des Guyanes.



Figure 4. Carte du Nord de l'Amérique du Sud comprenant le bassin amazonien.

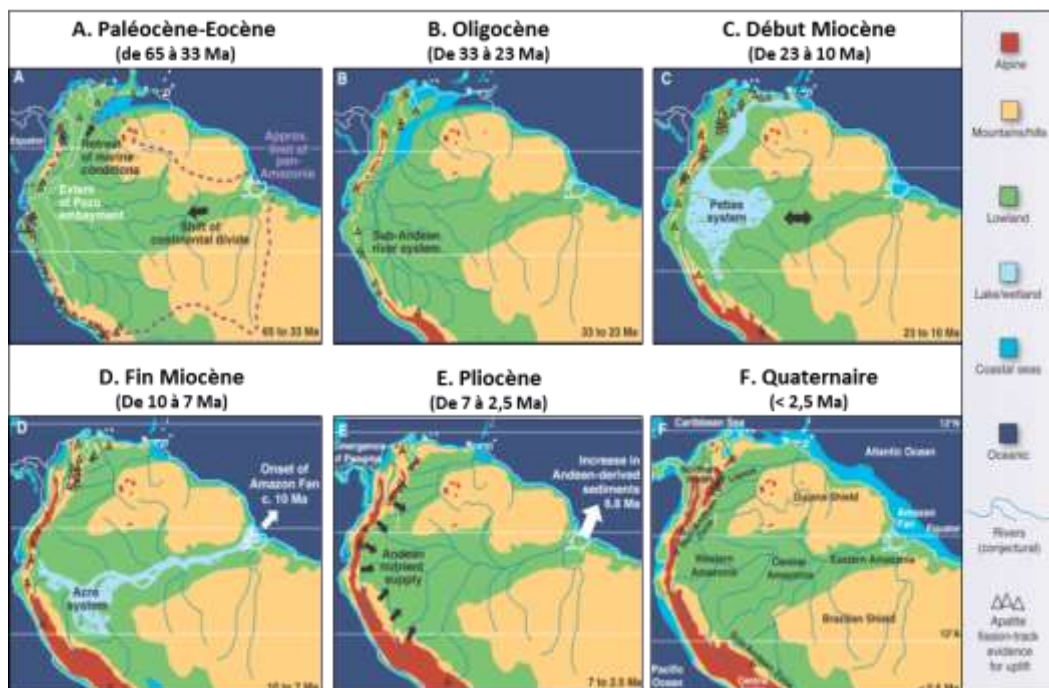


Figure 5. Cartes paléogéographiques de la modification de la croûte terrestre sud-américaine au cours des temps géologiques, d'après Hoorn *et al.* (2010). (A) Extension de l'Amazonie. Rupture des plaques du Pacifique et début du soulèvement des Andes. (B) Suite du soulèvement des Andes avec un drainage est-ouest du bassin amazonien. (C) Formation des montagnes dans le centre et le nord des Andes (~12 Ma) et d'une zone humide. (D) Soulèvement du nord des Andes restreint à la partie pan-Amazonnienne (E) Disparition de la zone humide et expansion des terres de forêts humides ; clotûre de l'isthme du Panama. (F) Quaternaire.

I.2 Perturbations en Amazonie pendant le Tertiaire et le Quaternaire

I.2.1 Formation de l'Amazonie

Le bassin amazonien désigne le bassin hydrographique desservi par le fleuve Amazone. Il s'étend sur le Brésil, la Bolivie, le Pérou, l'Équateur, la Colombie, le Venezuela et les trois Guyanes. Cette région est essentiellement composée de forêts de plaine entourées pour moitié par des reliefs : la chaîne de montagne des Andes à l'Ouest et le plateau des Guyanes au Nord (Figure 4). L'Amazonie représente la plus grande superficie de forêt tropicale humide au monde et abrite une extraordinaire diversité faunistique et floristique. On dénombre 1 104 espèces ligneuses sur une parcelle de 25 hectares de forêt humide en Équateur (Valencia *et al.*, 2004), environ 6 500 espèces au Guyana, 5 100 au Suriname et 5 400 en Guyane française, soit au total 9 500 espèces pour les trois Guyanes (Boggan *et al.*, 1997).

La formation de l'Amazonie est liée à la mise en place du continent sud-américain et aux événements géo-climatiques qui y sont liés. Suite à la dislocation de la Pangée (135 à 100 Ma), l'ouverture de l'Atlantique et la zone de subduction le long de la plaque Pacifique ont non seulement entraîné la séparation de l'Amérique du sud et de l'Afrique mais également la déformation de la croûte terrestre amazonienne. Le soulèvement des Andes s'est effectué d'Ouest en Est en plusieurs étapes. Il a débuté à la fin de l'Éocène (34 Ma) et a duré jusqu'au Pliocène (2,5 Ma). Ces événements d'orogénie ont modulé l'hydrographie du bassin Amazonien jusqu'à la mise en place du drainage ouest-est qu'elle connaît actuellement. Couplée à des changements climatiques, la formation de la cordillère des Andes a été à l'origine d'incursions marines à la fin de l'Éocène (33 Ma) et de la mise en place du lac Pebas pendant le Miocène (entre 23 et 10 Ma). La réorganisation de la plaque a également entraîné la fermeture de l'Isthme du Panama au Pliocène, il y a environ 3,5 Ma (Figure 5).

Tous ces événements ont influencé la répartition et la diversification des espèces végétales et animales (Hoorn *et al.*, 2010). Les végétaux de la famille des *Rubiaceae*, par exemple, auraient colonisé le continent nord-américain à partir de l'Europe en empruntant la route boréo-tropicale au Paléocène puis seraient arrivés en Amérique du sud en passant par des chaînes d'îles occasionnelles au début de l'Éocène. Leur distribution spatiale et leur

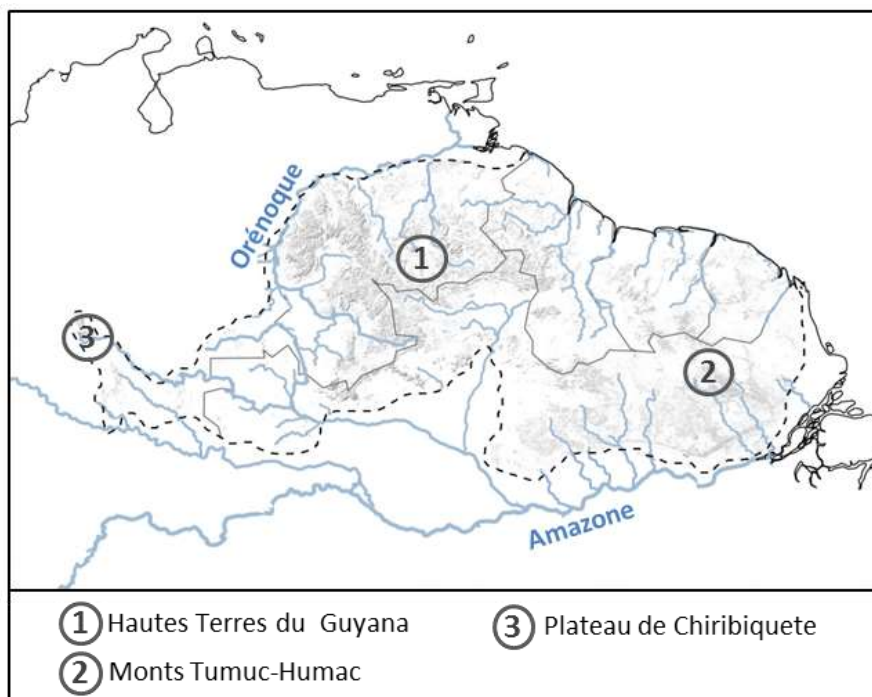


Figure 6. Carte des reliefs et de l'hydrographie du plateau des Guyanes.

structuration génétique actuelle illustrent le rôle de barrières géologiques que le soulèvement des Andes ou les incursions marines ont eu par le passé (Antonelli *et al.*, 2009). Plus généralement, si les Andes semblent avoir bloqué le flux de gènes de certaines espèces (Dick et Heuertz, 2008), cela ne semble pas être le cas pour d'autres espèces (Scotti-Saintagne *et al.*, 2012a; 2012b).

I.2.2 Le plateau des Guyanes

I.2.2.1 Formation et sols

Si l'Ouest de l'Amazonie a connu une activité géologique et tectonique intense pendant le Tertiaire, le plateau des Guyanes ou bouclier guyanais est quant à lui resté plutôt stable. Ce socle granitique a été formé au Précambrien et est daté entre 2,5 et 1,9 milliard d'années. Il couvre le nord du Brésil, les trois Guyanes et la partie orientale de la Colombie et du Venezuela. Située au niveau de l'Équateur, légèrement au nord de la plaque tectonique sud-américaine et au sud de la tri-jonction avec les plaques Caribéenne et nord-américaine (Figure 4), cette région est dépourvue d'activité sismique et volcanique depuis le Tertiaire. La topographie à grande échelle du plateau des Guyanes peut être délimitée en 3 zones montagneuses séparées par des forêts de plaine : les Hautes Terres du Guyana à l'est du Venezuela jusqu'au centre du Guyana, les monts Tumuc-Humac formant un arc de massifs au sud du Suriname et de la Guyane, et le plateau de Chiribiquete à l'Ouest est situé en Colombie (Figure 6). Les plaines côtières atlantiques forment une ceinture pouvant s'étendre jusqu'à 80 km à l'intérieur des Terres. Elles sont constituées d'argile consolidée et de limons déposés au cours des 10 000 dernières années par les rivières et fleuves qui drainent l'intérieur du bouclier et répandent leurs sédiments dans l'Atlantique. Deux systèmes hydrographiques majeurs séparent le plateau des Guyanes. Le bassin nord amazonien draine le fleuve Amazone et tous ses affluents des régions du Sud et de l'Ouest du plateau alors que le bassin des Guyanes entraîne toutes les eaux au nord du plateau dont le fleuve Orénoque jusqu'à l'Atlantique. De manière générale, les sols amazoniens sont acides (pH entre 3,5 et 6), pauvres en éléments nutritifs disponibles tels que le phosphore et le calcium. Ils contiennent souvent des grandes quantités d'aluminium et ont une faible capacité d'échange cationique (Hammond, 2005).

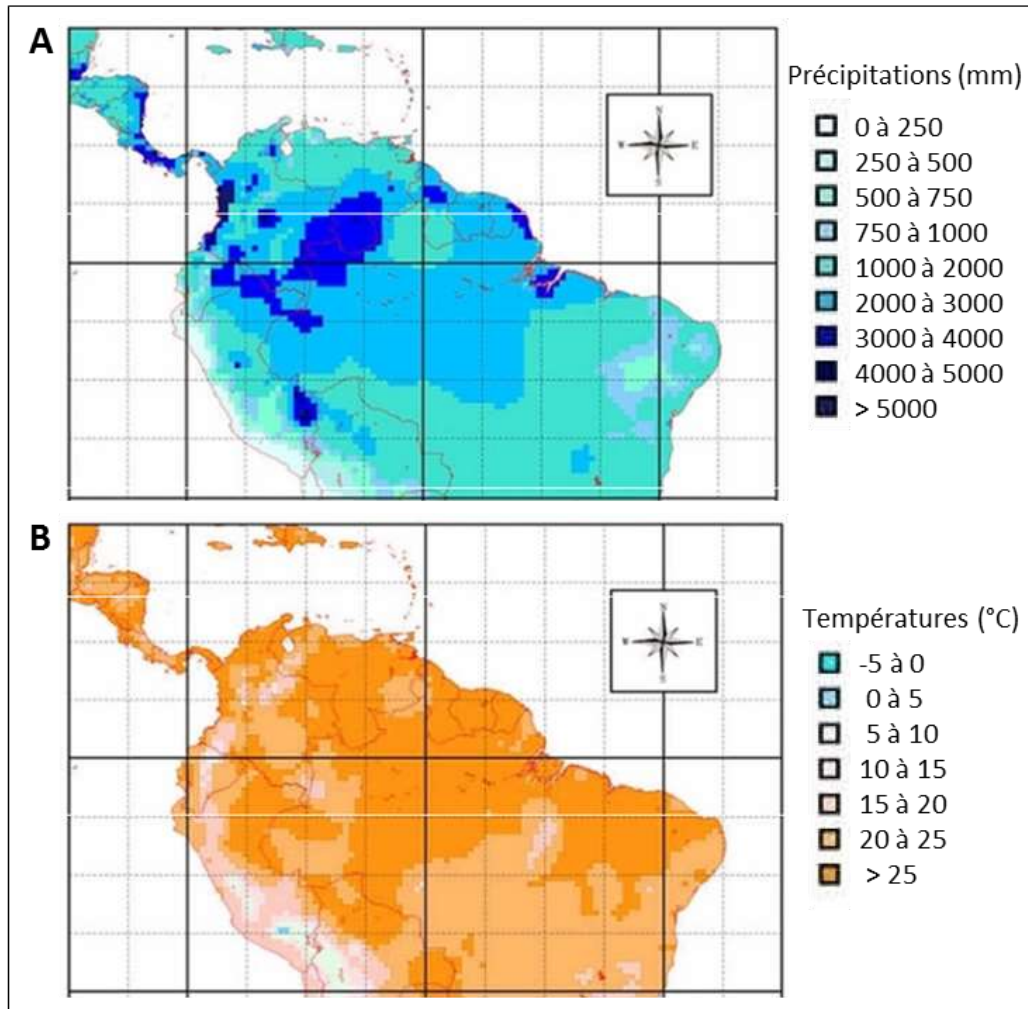


Figure 7. Distribution annuelle des précipitations (A) et des températures (B) actuelles au nord de l'Amérique du Sud, adapté par Amui-Vedel *et al.* (2009) d'après Legates et Wilmott (1990).

I.2.2.2 Climat actuel

Les conditions climatiques de l'ensemble du bouclier guyanais sont également très contrastées. Si l'Ouest du plateau, l'embouchure de l'Amazone ou encore le Nord-Est de la Guyane française présentent de fortes précipitations (>3 000mm par an), d'autres régions, au contraire, constituent un corridor de savanes caractérisé par un climat relativement sec (Figure 7.A). La saisonnalité de ces précipitations varie également d'une localisation à une autre selon trois grands modèles. Les saisons des pluies au nord et au sud du plateau des Guyanes sont uni-modales et se déroulent respectivement de Mai à Juin et de Novembre à Janvier, alors que les régions situées au nord-est de la bande côtière Atlantiques connaissent un premier pic de précipitations en même temps que les régions du Nord puis un second pic moins intense de Novembre à Janvier, correspondant à la saison des pluies des régions du sud (Molinier *et al.*, 1996; Hammond, 2005). Si les précipitations du bouclier guyanais suivent un schéma saisonnier prévisible et consistant avec les trois modèles majeurs, les quantités de pluies varient considérablement suivant les localités et les années. Le plateau des Guyanes regroupe sur son territoire toute la gamme de températures existante sur l'ensemble de l'Amazonie (Figure 7.B). Les sites présentant les températures les plus basses, avec une moyenne minimum de 6 à 20°C, sont localisés principalement sur les trois régions montagneuses du bouclier guyanais et sont entourées par une large bande de forêt de plaine où les températures sont plus élevées avec une moyenne minimum de 21 à 30°C. Les températures journalières varient généralement entre 7 et 15°C et leur amplitude entre les saisons, de 2 à 5°C tout au plus, est très peu marquée (Hammond, 2005).

I.2.2.3 Mécanismes de régulation du climat

Les différences de précipitations et de températures sur le plateau des Guyanes reflètent l'existence de divers facteurs forçant le climat. Le principal mécanisme atmosphérique contribuant aux pluies tropicales dans cette région est la circulation thermique de Hadley. La rencontre des vents des anticyclones des Açores et de Sainte-Hélène donne naissance à une ceinture dépressionnaire appelée la Zone Intertropicale de Convergence (ZIC). Le déplacement de cette formation de cumulonimbus de part et d'autre de l'Équateur entraîne des modifications de précipitation. Les différentes inclinaisons de l'axe terrestre par rapport au soleil le long de son orbite modifient la zone d'insolation maximum au cours de l'année donc les zones d'eaux chaudes, faisant ainsi varier la position de la ZIC et créant la saisonnalité des précipitations. La distribution géographique des températures de surface (SST) des océans Atlantique et Pacifique affecte donc la variation

des pluies (Good *et al.*, 2008) à travers leur impact sur les échanges océan-atmosphère. Outre le rayonnement solaire, des changements de vitesse des vents, de la géométrie des côtes et de la résonance des vagues ont été identifiés comme des phénomènes affectant ces SST. Associés à la circulation de Hadley, les événements ENSO (El Niño-Southern Oscillation) modulent la variabilité du climat en Amazonie (Molinier *et al.*, 1999). Le phénomène El Niño est une anomalie de la circulation de Walker (ou oscillation australe) qui entraîne un déplacement des zones de précipitations à l'est de l'océan Pacifique empêchant ainsi la remontée des eaux froides le long de la côte sud-américaine. L'ENSO amplifie la variation saisonnière des pluies mais n'altère pas la saisonnalité. Pendant les événements El Niño, les précipitations sont moins importantes pendant la saison humide mais plus abondantes pendant la saison sèche alors que lors des événements La Niña les saisons sèches et humides sont amplifiées (de Souza *et al.*, 2000).

Des facteurs terrestres tels que la variation de la topographie, l'albédo de surface ou le couvert forestier interagissent également avec les conditions hydrologiques locales et la circulation atmosphérique globale et entraînent des variations de précipitations à plus petites échelles spatiales. Les modifications topographiques importantes augmentent les forces de friction et entraînent une remontée et une formation rapide des nuages de pluies. L'effet des montagnes du Guyana sur les précipitations est bien sûr minime comparé à l'impact des pentes Andéennes. Les grandes dépressions topographiques telles que l'embouchure de l'Amazone peuvent également créer de vastes sources d'évaporation. L'évapotranspiration du couvert de végétation ou l'évaporation directe de la canopée et des surfaces d'eaux permettent le maintien des fortes précipitations dans certaines zones. La brise marine le long des côtes atlantiques affecte également le régime de pluies en intervenant sur les échanges terres-océans jouant ainsi sur les différences de températures entre ces deux compartiments (Hammond, 2005).

Les variations des pluies à travers le plateau des Guyanes s'expliquent par des événements océaniques et continentaux, constituant un gradient méridional, et par la variation du pic d'insolation, constituant un gradient zonal. À l'est du bouclier guyanais, les influences océaniques ont un rôle important dans une région très sensible aux changements de processus atmosphère-océan, tel que l'ENSO. Au centre, les montagnes du Guyana atténuent les fortes pluies locales. À l'ouest, l'influence de la ZIC est plus limitée et le cycle de précipitations est soumis à l'influence continentale des Andes caractérisée par une forte rugosité, une forte évapotranspiration et un faible albédo. Les variations de températures à travers le plateau des Guyanes sont, elles, largement influencées par l'insolation journalière

et saisonnière, la topographique, l'altitude, la proximité avec l'Atlantique, l'augmentation des gaz à effet de serre, l'interaction de tous ces facteurs avec le mouvement de la ZIC provoquant des changements de couverture nuageuse et les incursions de fronts froids polaires (Hammond, 2005; Ledru *et al.*, 2005).

I.2.3 Changements climatiques du Quaternaire

Si de multiples facteurs atmosphériques, océaniques et continentaux influencent les régimes journaliers et annuels des précipitations et des températures, les changements climatiques s'opèrent eux à une plus grande échelle spatiale. Ces processus reflètent les caractéristiques de notre planète, du Soleil et la variation cyclique de la relation astronomique entre ces deux astres, appelée cycles de Milankovitch et dont les trois paramètres sont l'obliquité, la précession et l'excentricité. L'obliquité est un changement cyclique de l'angle de d'inclinaison de l'axe de rotation. L'inclinaison terrestre est actuellement de $23,5^\circ$ par rapport au plan de son ellipse et varie de 22° à $24,5^\circ$ tous les 40 000 ans. Cet angle d'inclinaison détermine l'intensité de la saisonnalité et modifie également les limites des climats mais il n'a pas d'impact sur la zone équatoriale pour laquelle l'angle n'a pas d'incidence. La précession est un changement graduel d'orientation de l'axe de rotation et représente le mouvement cinétique de la Terre lorsqu'elle tourne sur elle-même. Le changement d'axe influence la durée des saisons et la modification de la distance au Soleil influence l'intensité des saisons. Un cycle complet varie entre 18 000 et 23 000 ans. L'excentricité est une description de l'orbite terrestre qui varie d'une forme quasiment circulaire (excentricité faible), comme c'est le cas actuellement, à une forme plus elliptique (excentricité élevée) suivant deux cycles de 100 000 et 412 000 ans. La forme de l'orbite terrestre influence donc la distance entre la Terre et le Soleil faisant ainsi varier la quantité d'insolation reçue (Hammond, 2005).

Considéré indépendamment, chaque cycle entraîne relativement peu de changement mais ensemble ils produisent des scénarios contrastés importants. Ces trois paramètres astronomiques sont à l'origine des épisodes d'expansion et de retrait des glaces que la Terre a connus au Quaternaire. Les variations observées dans les données sédimentaires (Olsen et Kent, 1996; Scarparo Cunha et Koutsoukos, 2001), isotopiques (Rial, 1999; Stirling *et al.*, 2001) et palynologiques (Mayle *et al.*, 2000; Bush *et al.*, 2002) en Amazonie attestent du rôle des cycles de Milankovitch sur le climat. Les changements de configuration de ces cycles ont eu un impact plus important dans l'Hémisphère Nord, notamment sur la régulation de la

température ambiante et la formation de glace polaire. Même si les changements d'insolation n'ont eu qu'un faible effet sur les tropiques par le passé, le gradient d'insolation latitudinal a tout de même été modifié changeant l'équilibre des processus océan-atmosphère et affectant la localisation saisonnière et l'intensité de la ZIC ainsi que les anomalies de SST. La reconstruction paléo-climatique des Néotropiques effectuée par Martin *et al.* (1997) indique une remontée latitudinale de +10° de la limite sud de la ZIC depuis la fin du Pléistocène attribuable à l'action des forces astronomiques. Les données des paléo-limons de deux lacs amazonien situés à l'ouest du plateau de Guyanes montrent également des variations de concentration en Potassium. Ces résultats attestent de fluctuations des niveaux marins et lacustres à l'échelle des événements de précession (cycle de 19-22 ka) et d'un refroidissement au DMG du nord-ouest de l'Amazonie (Bush *et al.*, 2002) qui ont accompagné le déplacement de la ZIC. Une position de la ZIC plus au Nord conduit également à des incursions polaires plus durables dans le sud de l'Amazonie et une poussée d'air froid plus fréquente le long de l'Équateur entraînant une réduction des températures annuelles dans ces régions (Hammond, 2005). Les données palynologiques attestent d'une baisse des températures de -4 à -6°C en Amazonie (Markgraf, 1989; Colinvaux *et al.*, 1996) et la reconstruction climatique de Clark *et al.* (1999) indique, sur le plateau des Guyanes, une différence de température de -4 (-2) à 0 (+4)°C entre l'été boréal du DMG et l'été actuel alors que les températures de l'ouest amazonien et d'Amérique centrale auraient été plus froides. Le taux d'humidité (précipitation-évaporation) à cette période témoigne quant à lui d'un climat plus sec dans le bouclier guyanais comparé aux régions voisines. Des sédiments, situés le long d'un corridor de savanes au Brésil, ont permis d'évaluer cette diminution des précipitations à 500 mm par an pendant le DMG (Absy *et al.*, 1991). De plus, les changements de niveau de la mer liés à la formation et à la fonte des glaciers ont sans doute influencé les climats amazoniens passés en modifiant notamment l'albédo des surfaces ou les zones d'évaporation.

Les Néotropiques n'ont donc pas répondu uniformément aux changements climatiques du passé, caractérisés par un ralentissement des circulations de Hadley et Walker et/ou une forte altération de la migration de la ZIC. Au DMG, le nord-ouest de l'Amazonie, avec une influence plutôt continentale, semble avoir subi une réduction des précipitations plus modérée mais une baisse de température plus importante que le plateau des Guyanes où la sécheresse semble avoir été sévère (Hammond, 2005).

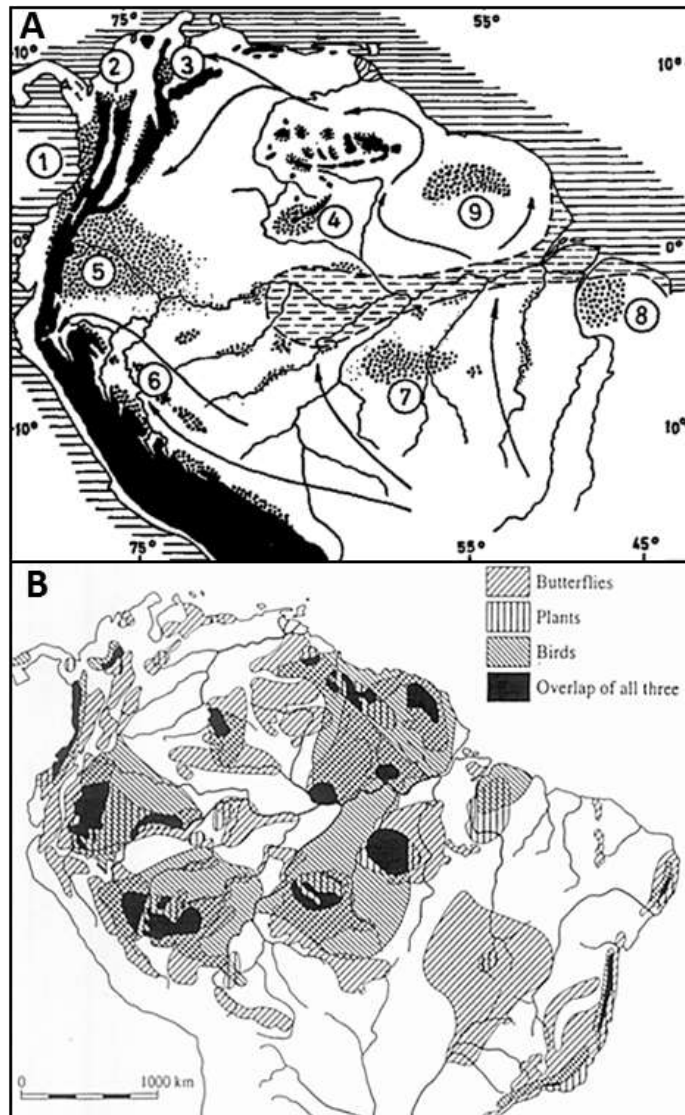


Figure 8. Cartes des refuges présumés au nord de l'Amérique du Sud pendant le dernier maximum glaciaire. Les zones refuges ont été déterminées par (A) Haffer (1969) dans sa théorie originelle à partir de la distribution d'oiseaux et adaptées par (B) Haffer et Prance (2001) d'après la superposition des patrons d'endémismes de plantes, de papillons et d'oiseaux néotropicaux de Brown (1987).

I.2.4 Conséquences des perturbations climatiques sur la forêt amazonienne

Le climat au DMG, globalement plus sec et plus froid qu'aujourd'hui, a modulé la forêt tropicale humide. Plusieurs hypothèses ont été proposées pour expliquer la répartition des espèces à cette période et les conséquences de ces modifications climatiques sur leur distribution actuelle.

I.2.4.1 La théorie des refuges

Dans sa théorie des refuges, Haffer (1969) fait l'hypothèse d'une division de la forêt amazonienne en plusieurs petites forêts, isolées les unes des autres par des milieux ouverts pendant les périodes climatiques sèches du Pléistocène. Les forêts ayant subsistées auraient servi de « zones refuges » pour un grand nombre de populations d'animaux forestiers, qui ont ensuite évoluées indépendamment les unes des autres pendant les périodes d'isolation géographique. Ces forêts isolées se seraient à nouveau rejoint pendant les périodes climatiques humides, lorsque les milieux ouverts seraient redevenus forestiers, permettant aux populations des zones refuges d'étendre leur aire de distribution. Ces ruptures et réunifications des diverses forêts en Amazonie auraient probablement eu lieu plusieurs fois pendant le Quaternaire et auraient permis une différenciation rapide de la faune et de la flore forestière amazonienne à des temps géologiques très récents. En croisant les données de précipitations annuelles et les patrons de distribution génétique de plusieurs oiseaux, Haffer (1969) a délimité neuf zones supposées refuges au DMG (Figure 8) dont deux majeures situées sur le plateau des Guyanes. Si cette hypothèse des refuges en Amazonie a longtemps été supportée (Prance, 1973; Brown *et al.*, 1974; Haffer, 1997) et l'est encore actuellement par quelques études génétiques (Vargas-Ramírez *et al.*, 2010), elle est aujourd'hui remise en discussion (Bush et De Oliveira, 2006).

I.2.4.2 L'hypothèse de transition floristique

La principale critique vis-à-vis de l'hypothèse des refuges est le manque de données sédimentaires supportant une contraction de la forêt amazonienne due à un climat globalement plus sec au DMG (Bush et De Oliveira, 2006). Des études palynologiques attestent par contre de la présence au DMG d'espèces montagnardes dans des zones actuellement occupées par la forêt tropicale humide (Ledru *et al.*, 2001) impliquant une diminution des températures de 4 à 7°C dans diverses régions du Panama

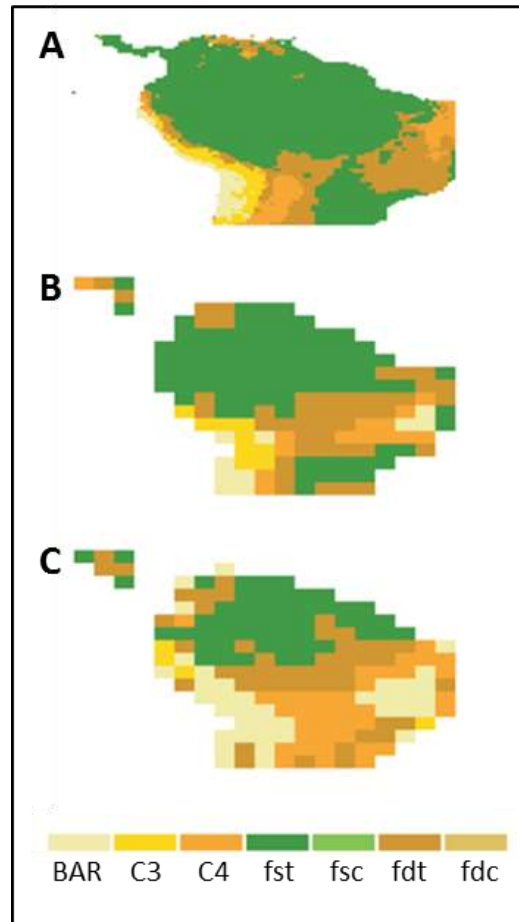


Figure 9. Modélisation de la paléo-végétation au Quaternaire d'après Mayle *et al.* (2004). Distribution de la végétation amazonienne à (A) l'époque pré-industrielle, (B) l'Holocène moyen et (C) au dernier maximum glaciaire. C3 : plantes en C3, C4 : plantes en C4, fst : forêt sempervirente tropicale, fsc : forêt sempervirente de conifères, fdt : forêt décidue tropicale, fdc : forêt décidue de conifères.

(Bush et Colinvaux, 1990), de l'Équateur (Bush *et al.*, 1990) et du Pérou (Colinvaux *et al.*, 1996). Au DMG la forêt amazonienne aurait donc été composée d'espèces décidues supportant un climat plus froid, remplacées ensuite par des espèces tropicales lors du réchauffement global au début de l'Holocène (Colinvaux *et al.*, 2000; Bush *et al.*, 2004). Ainsi durant les périodes les plus sèches, même si la composition forestière a pu varier en réponse aux changements climatiques, le couvert forestier ne se serait pas fragmenté (Bush *et al.*, 2002). Le modèle de dynamique de végétation développé par Sternberg (2001), décrivant l'équilibre savane-forêt en fonction des précipitations et de la durée de la saison sèche, corrobore les analyses de pollen du bassin amazonien et atteste de la stabilité passée du couvert forestier tropical malgré des changements climatiques.

Étant donné la variabilité régionale des conditions environnementales, climatiques et des forces les régulant, les climats passés du bassin amazonien n'ont pas été non plus homogènes (Colinvaux, 1987) et n'ont donc peut-être pas impacté la forêt de façon identique. La modélisation de la paléo-végétation en fonction du climat et de la concentration en CO₂, réalisée par Mayle *et al.* (2004), atteste du maintien d'une forêt humide dense sempervirente pendant le DMG au nord-ouest du bassin amazonien alors que le sud a été recouvert de forêts décidues ou semi-décidues sèches et que les savanes se sont étendues en bordure du bassin (Figure 9). Ce modèle est en accord avec plusieurs études réalisées à travers l'Amazonie. Les travaux de Colinvaux *et al.* (1996) et Bush *et al.* (2002) à partir de données de pollen montrent que le nord-ouest a été continuellement recouvert de forêt pendant les 170 000 dernières années alors que d'autres sites situés plus en marge du bassin amazonien révèlent des changements de distribution forêts-savanes (Mayle *et al.*, 2000) et la présence de forêts décidues (Burbridge *et al.*, 2004) entre le DMG et aujourd'hui. Sur le plateau des Guyanes, les quelques études sédimentaires réalisées sur les Hautes Terres du Guyana attestent d'un climat plus sec et plus froid il y a 11 700 ans BP (Montoya *et al.*, 2011). La reconstruction paléo-écologique effectuée par Nogué *et al.* (2009) indique que la végétation de haute altitude du mont Tepuis est restée inchangée depuis 4 300 ans BP et les mesures de variations en ¹³C du sol des travaux de Freycon *et al.* (2010) suggèrent que les forêts humides du nord-est de la Guyane française ont bien résisté aux conditions climatiques sèches au cours de l'Holocène.

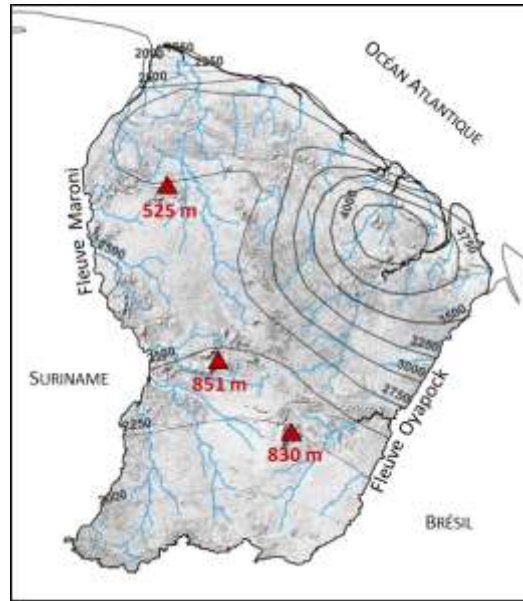


Figure 10. Carte des reliefs et des précipitations en Guyane française.

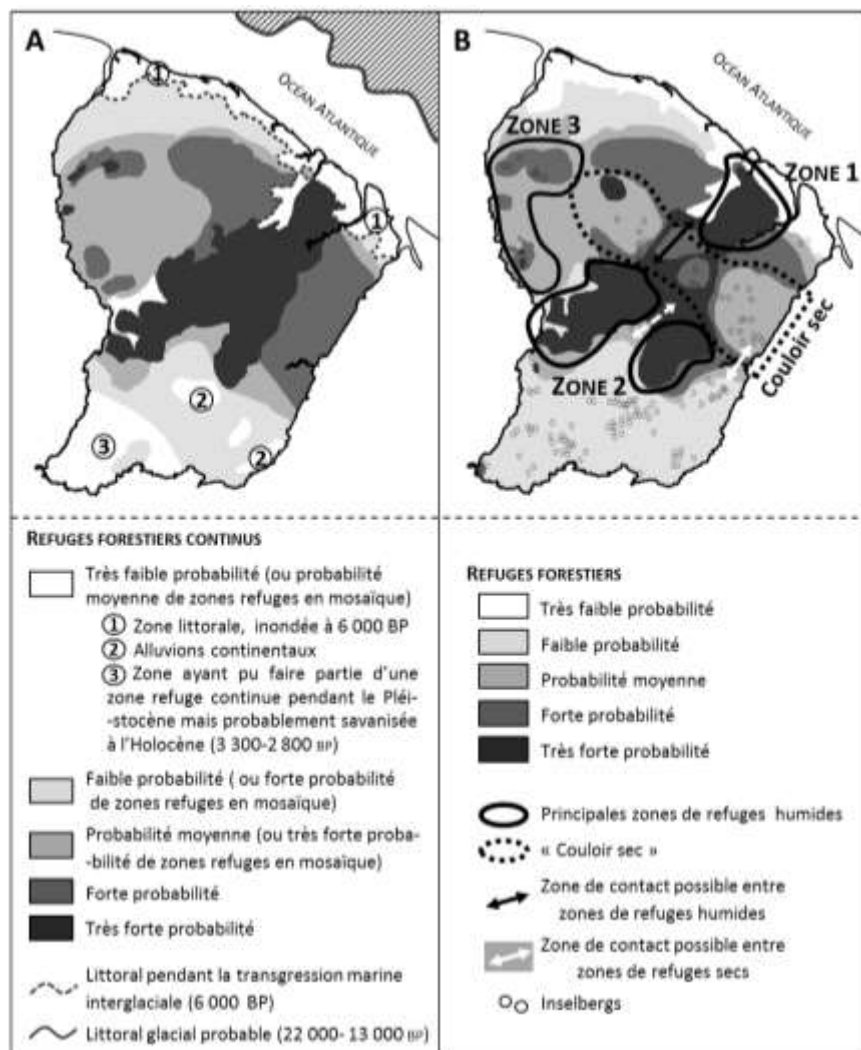


Figure 11. Carte des Refuges forestiers présumés en Guyane française pendant le dernier maximum glaciaire, d'après (A) de Granville (1982) et (B) Tardy (1998).

I.3 Perturbations en Guyane française pendant l'Holocène

La Guyane française fait partie du plateau des Guyanes et s'étend sur environ 90 000 km². Ses limites "naturelles" les plus évidentes sont les grands fleuves Oyapock et Maroni. Au sud, la ligne de partage des eaux entre le système hydrographique guyanais et les affluents rive gauche de l'Amazonie n'est pas toujours évidente et est jalonnée seulement par quelques hauteurs telles que les monts Tumuc-Humac et de vieux inselbergs qui n'ont jamais constitué un obstacle aux échanges humains (Mazières, 2006). Les précipitations moyennes sont comprises entre 2000 et 4000 mm et sont réparties selon un gradient ouest-est croissant (Figure 10). La température moyenne annuelle de 26,1°C ne présente que très peu de variation au cours de l'année mais une amplitude journalière plus marquée en saison sèche (Wagner, 2011).

Si peu de données permettent de reconstruire le climat Guyanais au DMG, de Granville (1982) a associé la distribution des précipitations actuelles aux données géographiques, topographiques, pédologiques, géomorphologiques et biogéographiques en Guyane française afin de déterminer des zones refuges plus précises dans cette région entre 22 000 et 13 000 ans BP (Figure 11). D'après ces résultats, les populations forestières ne se sont pas regroupées sur les côtes atlantiques et dans le sud de la Guyane au DMG mais la bande allant de Kaw à Saül montre une forte probabilité d'avoir été une zone refuge. Grâce à une étude anthracologique, Tardy (1998) a affiné la localisation de ces refuges en éliminant la zone des Nouragues où de nombreux feux attestent d'un couloir sec au DMG. Les régions nord-est et centrale ont, elles, été confirmées comme zones refuges.

Les études anthracologiques (Servant et Servant-Vildary, 1996; Tardy, 1998) et palynologiques (Ledru *et al.*, 2001) ponctuelles permettent également d'attester de phases climatiques plus sèches pendant l'Holocène. Les paléo-feux datés entre le début et le milieu de l'Holocène (entre 12 800 et 6 000 BP) ont une origine naturelle certaine (Charles-Dominique *et al.*, 1998; Mayle *et al.*, 2000; Mayle et Power, 2008; Power *et al.*, 2010) et ont pu être corrélés aux événements des glaciations (Tardy, 1998), alors que les phases d'incendies plus récentes sont attribuées à l'action conjuguée de l'activité climatique et anthropique (Servant et Servant-Vildary, 1996; Tardy, 1998; Bush *et al.*, 2008).

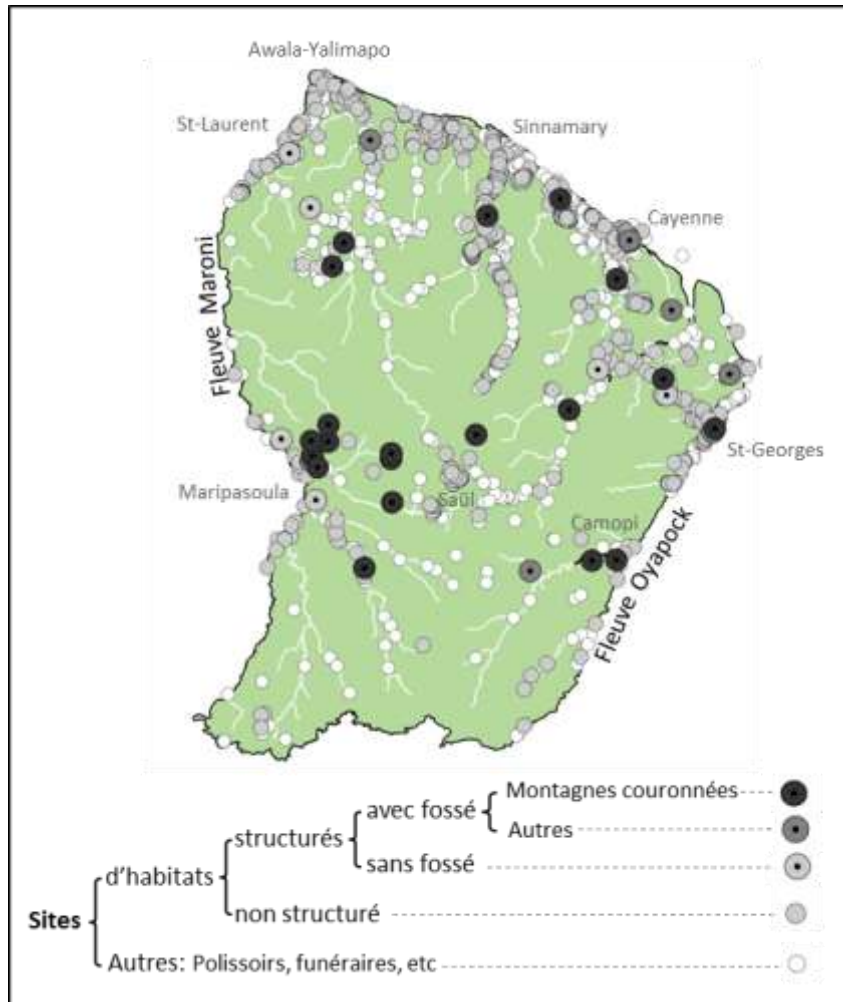


Figure 12. Répartition des sites archéologiques précolombiens découverts en Guyane jusqu'en 2009. Données: SRA DRAC Guyane.

I.3.1 Types d'occupations précolombiennes

Le peuplement de l'Amérique par l'Homme est à l'heure actuelle un sujet très sensible. Une théorie longtemps admise supposait une première arrivée sur le continent depuis l'Asie par le détroit, libre de glace, de Béring vers 14 000-12 000 BP, période marquée par une phase glaciaire ayant abaissée le niveau de la mer. Les découvertes de sites antérieurs à ces dates ont depuis remis en cause cette théorie et ont permis d'avancer l'hypothèse de plusieurs vagues de colonisations par voies maritimes et/ou par voies terrestres actuellement sous le niveau de la mer (Lavallée, 1995). En Guyane française, le Paléoindien est peu représenté (Boomert, 1980) et est suivi par l'époque archaïque également mal connue. La première occupation pré-céramique mise au jour en Guyane française a été découverte sur le Plateau des mines à l'Ouest du département. Des objets lithiques ainsi que des amas de galets de quartz constitués de pierres chauffées et de charbons de bois ont été retrouvés et datés d'environ 7 000 ans BP, plaçant ce site parmi les plus anciens connus à ce jour sur le plateau des Guyanes (Mestre et Delpech, 2008). Les deux sites archéologiques fournissant les plus vieilles céramiques, datées d'au moins 4 000 BP, sont situés le long du Bas Maroni (Chemin Saint-Louis, van den Bel *et al.* (2011)) et au sommet de petites collines formées au Pléistocène-Holocène et délimitées par des savanes. À partir de 2000 BP, les sites archéologiques sont retrouvés sur tout le territoire guyanais (Vacher *et al.*, 1998) et des sites datés dès 1 000 BP sont trouvés sur les cordons sableux de l'Holocène récent ou sur des digues plus élevées (Rostain, 1994). L'apparition de la céramique et des outils abrasifs est plus ou moins liée à la domestication de plantes telles que le maïs dont la présence a clairement été identifiée sur le littoral (McKey *et al.*, 2010; Pagán Jiménez, 2011; van den Bel *et al.*, 2011). Le développement de cette agriculture reflète également une évolution dans la complexité des sociétés précolombiennes. Les populations humaines composées probablement de groupes nomades de chasseurs-cueilleurs se sont au cours du temps sédentarisées (Rostain, 2008).

Les prospections archéologiques en Guyane, loin d'établir un état des lieux complet de l'occupation amérindienne ancienne sur tout le territoire, ont permis de mettre à jour certains sites précolombiens (Migeon, 2009) (Figure 12). La détection d'un site dépend bien souvent de l'effort de fouille réalisé ainsi que de son accessibilité, comme le prouve la forte densité de sites découverts le long de la côte ou des fleuves. Le plus faible nombre d'indices

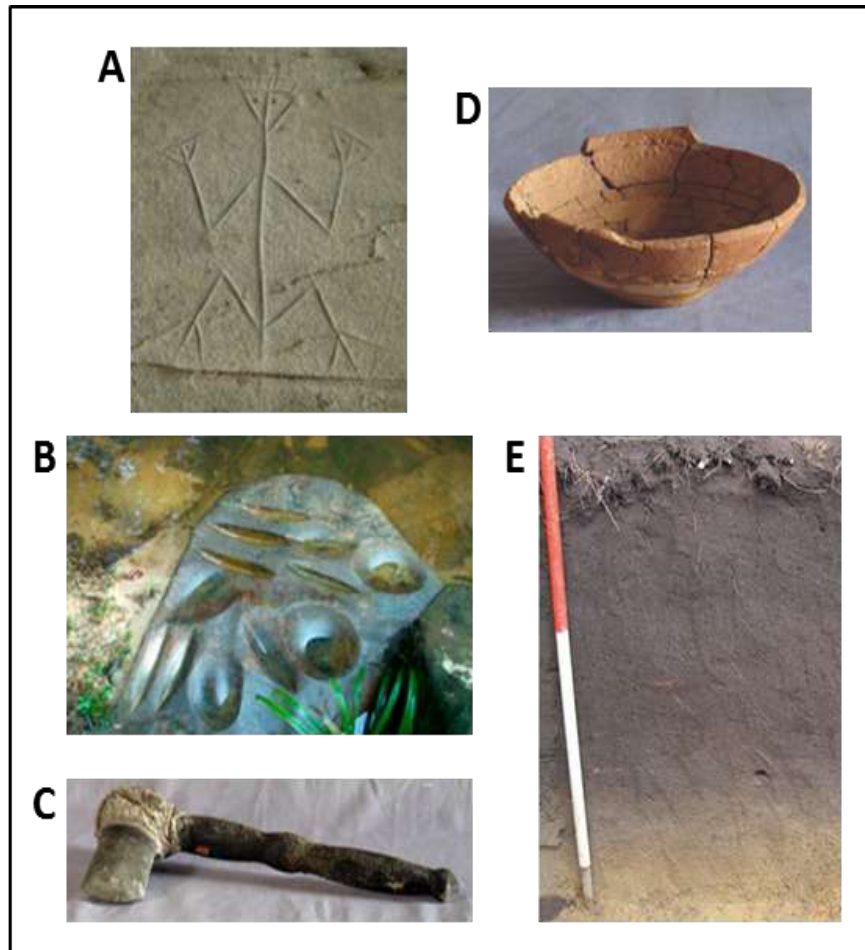


Figure 13. Traces d'occupation humaine précolombienne en Guyane française. (A) Roches gravées de La Carapa à Kourou. Photo personnelle. (B) Polissoir en cupule et fuseau, crique Arca à la Trinité. SRA Guyane. (C) Hache emmenchée, bois, résine et pierre verte découverte sur le fleuve Approuague, culture et date indéterminée. SRA Guyane. (D) Ecuelle incomplète restaurée trouvée sur le barrage de Petit Saut à Sinnamary. Culture et date indéterminée. SRA Guyane. Les photos B, C et D sont tirées de Migeon (2009). (E) Terra preta situé sur le site de fouille archéologique du chemin Saint-Louis à Saint-Laurent du Maroni. Photo tirée du rapport d'opération de van den Bel *et al.* (2011).

décelés au sud de la Guyane n'est donc pas représentatif de l'intensité d'occupation passé de cette région mais reflète plutôt le caractère inaccessible du territoire. L'urbanisme moderne croissant sur la bande côtière et le développement de l'archéologie préventive depuis 2002 a, au contraire, permis de révéler de nombreux sites et de reconstruire plus précisément l'occupation du littoral guyanais par quatre cultures précolombiennes différentes (Rostain, 1994; Vacher *et al.*, 1998). Aux sites d'habitats côtiers s'ajoutent d'autres sites dont de nombreux polissoirs, principalement situés sur le littoral et au bord des fleuves, et les roches gravées. Parmi les sites d'habitats forestiers, on distingue les sites de plein-air, les sites en bord de fleuve ou près des criques, les sites en grottes ou sous abris, principalement funéraires, et les sites fossoyés. Ces derniers sont des espaces caractérisés par des reliefs ou microreliefs d'origine anthropique, notamment des fossés dont on peut distinguer deux types d'aménagements. Les sites de type «éperon barré» sont souvent situés sur un sommet dont l'accès n'est possible que par un seul versant en pente douce et barré par la construction d'un fossé. Les montagnes couronnées sont, elles, des sites localisés sur un mont mais dont le fossé, probablement à usage défensif, entoure partiellement ou totalement la montagne ou délimite une enceinte construite sur une partie du plateau (Petitjean Roget, 1991; Mazière et Mazière, 1994). Outre les montagnes couronnées, les autres habitations précolombiennes étaient construites avec des poteaux en bois aujourd'hui disparus mais dont les trous laissés sont des indices de la taille et de l'intensité d'occupation d'un site.

I.3.2 Artefacts de l'occupation précolombienne humaine

Lors d'une prospection pédestre en forêts, l'habitat précolombien se manifeste le plus souvent par les trouvailles de quelques tessons et pièces lithiques en surface ou dans les mottes de racines. Ils représentent des indices tangibles d'une occupation humaine avant l'arrivée des européens, appelés artefacts, et peuvent être de plusieurs natures. Par exemple, les pièces lithiques incluent l'utilisation de la pierre sous toutes ses formes : des roches gravées à la construction d'outils divers en passant par les peintures rupestres, l'assemblage de pierres ou les polissoirs. Les fragments de céramique et plus généralement, les restes de métaux travaillés ou de matériaux organiques tels que les restes végétaux, animaux ou moins fréquemment humains mais aussi des charbons sont également des preuves d'une activité anthropique (Figure 13 A-D). Les témoins des différentes étapes de la mise en place des populations humaines sont bien souvent difficiles d'accès car si certains

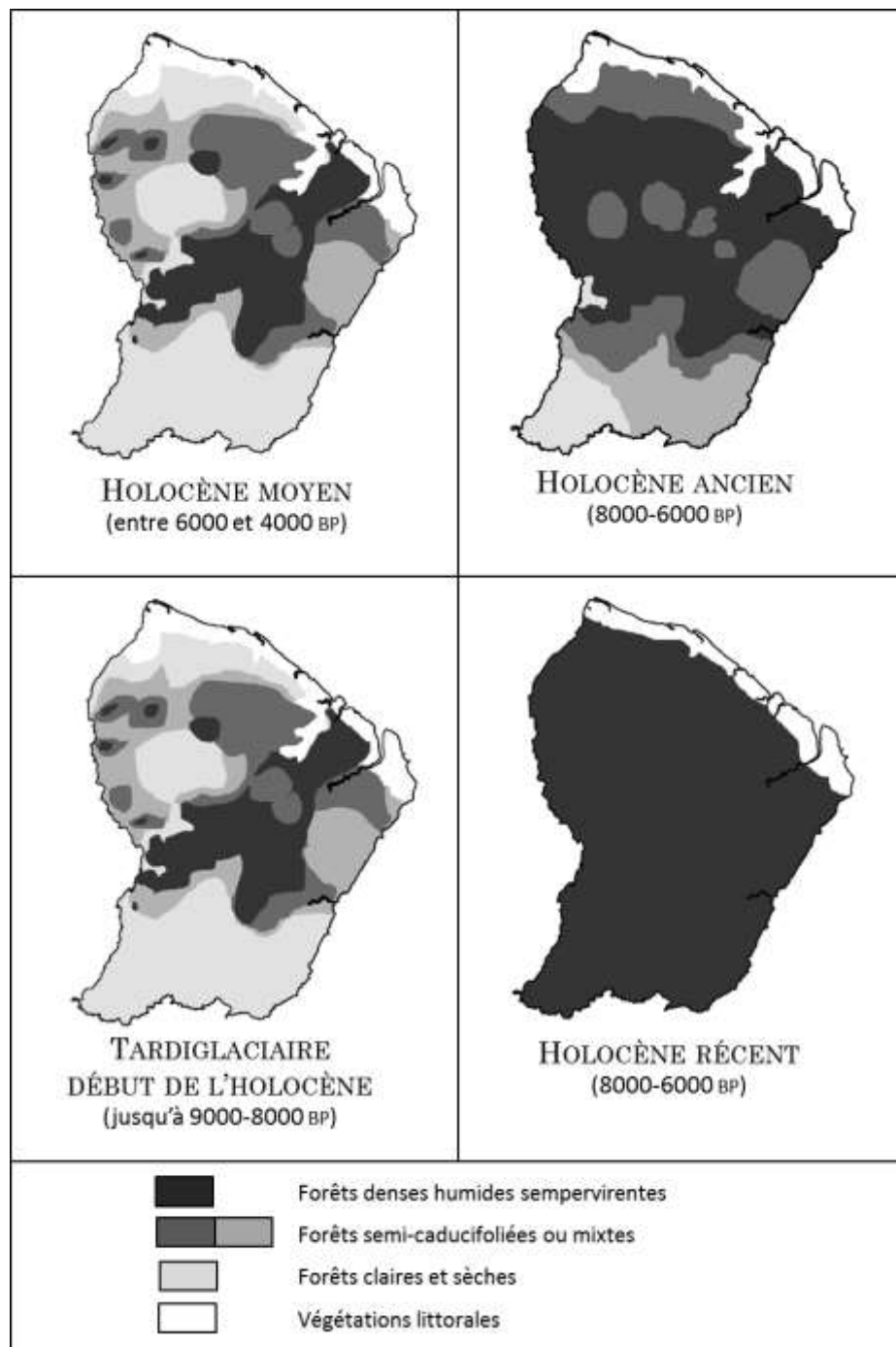


Figure 14. Hypothèse de l'évolution des végétations guyanaises au cours de l'Holocène, d'après Tardy (1998).

sont facilement détectables et datables d'autres le sont beaucoup moins (Hammond, 2005). Ainsi, la majorité du matériel lithique et des céramiques nécessitent la présence d'autres indicateurs tels que des charbons pour pouvoir permettre une datation AMS (Spectromètre de masse couplé à un accélérateur de particules) et déterminer la chronologie des événements. Les peintures rupestres ne connaissent pas ce problème grâce à la nature végétale de leurs pigments mais il n'en existe aucune preuve en Guyane notamment à cause de la forte porosité des roches dans cette région.

L'utilisation des terres par ces populations humaines ont également laissé des traces indélébiles sur les écosystèmes présents. La Terre noire ou *Terra preta* (Figure 13.E) est un type de sol sombre transformé chimiquement par les activités humaines et les intempéries amenant à un remaniement et un enrichissement des sédiments en matière organique sur les lieux d'habitation (Sombroek, 1966 d'après Eden *et al.* (1984)). L'augmentation en nutriments tels que l'azote, le phosphore, le potassium et le calcium en font un sol d'une incroyable fertilité. D'autres modifications liées à l'agriculture, telles que les champs drainés surélevés dans des régions inondées saisonnièrement, ont été étudiées en Guyane (McKey *et al.*, 2010) et sont probablement le reflet d'une augmentation des populations, d'une sédentarisation (Rostain, 2008) dans des zones peu favorables (Lombardo *et al.*, 2011).

I.3.3 Conséquences des perturbations anthropiques sur la forêt

Le climat en Guyane française pendant l'Holocène se rapprochait sûrement davantage des conditions actuelles tout en présentant certaines variations de température et plus encore de précipitations. Les contraintes climatiques locales ont ainsi pu conduire à moduler le couvert et la composition forestière guyanaise. En se basant sur des données paléo-environnementales, climatiques et des connaissances phytogéographiques locales, Tardy (1998) a émis des hypothèses concernant la répartition des différents types de végétations depuis les forêts sempervirentes denses jusqu'aux forêts sèches en passant par les formations semi-décidues, depuis le début de l'Holocène jusqu'à nos jours (Figure 14). Selon ce schéma, le maintien des conditions du DMG et la crise climatique sèche du début de l'Holocène (9 000-8 000 BP) auraient entraîné de nombreux incendies et la mise en place de forêts majoritairement semi-décidues et sèches alors que la forêt dense humide sempervirente aurait été restreinte à des zones refuges. Les conditions plus humides à l'Holocène moyen (8 000-6 000 BP) auraient été favorables au développement de la forêt

dense sempervirente même si une petite crise climatique semble avoir eu lieu autour de 7 500 BP. Le climat sec à l'Holocène moyen (6 000-4 000 BP) aurait provoqué des incendies et un recul général de la forêt humide. Enfin, le retour des conditions humides à l'Holocène récente (après 2000 BP) aurait permis la mise en place de la forêt dense humide jusqu'aux limites actuelles avec cependant de petites fluctuations sèches de courte durée et des incendies localisés influencés également par l'activité précolombienne. L'impact de ces occupations humaines sur leur environnement est encore peu connu et la connaissance du rayon et de l'intensité d'action de ces populations sur la forêt qui les entourait alors est une information qui fait cruellement défaut. On sait, par exemple, que les populations précolombiennes consommaient et cultivaient depuis l'Holocène moyen certaines plantes dont les patates douces, les « topinambour-pays » de la famille des Marantaceae, plus surprenant encore le maïs et plus tard la courge, le piment et les haricots (McKey *et al.*, 2010; van den Bel *et al.*, 2011). On peut également supposer une utilisation de certains végétaux à des fins médicinales (Grenand *et al.*, 2004) ou autres. Le défrichage nécessaire à la mise en place des habitats, l'existence d'abattis ou l'agriculture sur brûlis sont autant de pratiques qui ont pu modifier des zones forestières entières.

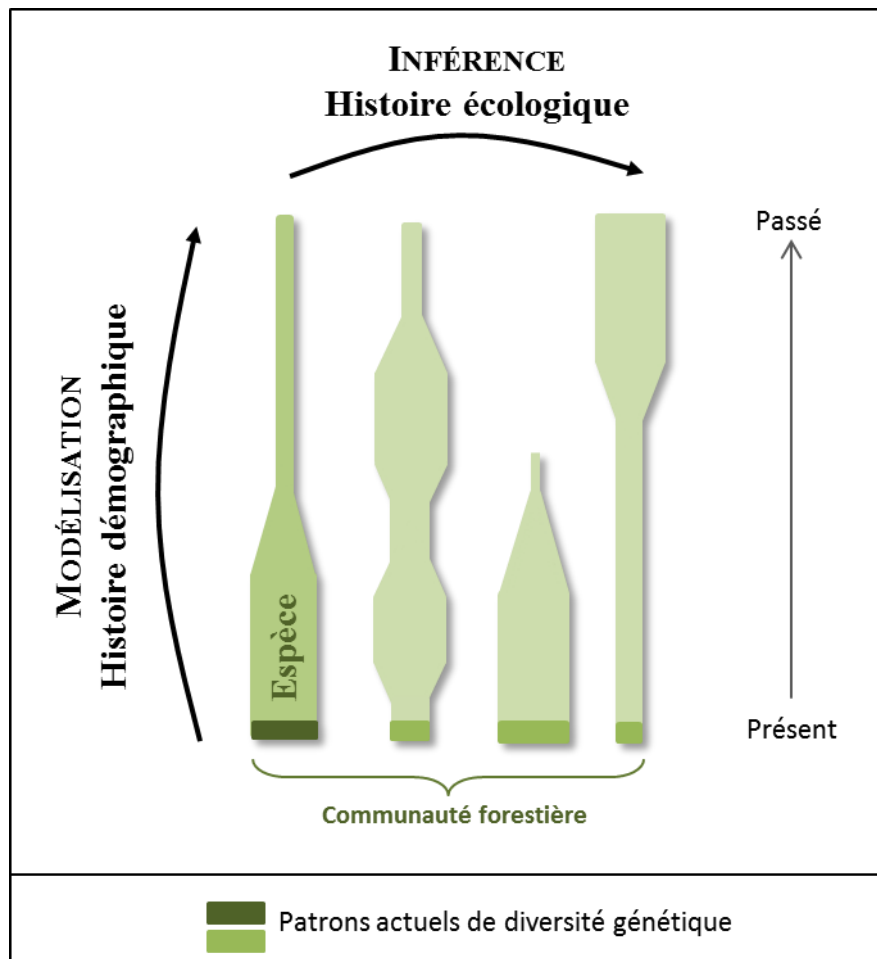


Figure 15. Approche de démographie comparée interspécifique.

I.4 Inférence de l'histoire écologique des communautés forestières du plateau des Guyanes

I.4.1 De l'histoire démographique des espèces à l'histoire écologique des communautés

La reconstruction de l'histoire des forêts tropicales humides en Amazonie repose sur des indices sédimentaires, palynologiques et à une plus faible échelle de temps archéologiques qui présentent toutes des faiblesses. Ledru *et al.* (1998) a ainsi montré que les dépôts correspondant au DMG manquaient sur la majorité des sédiments analysés en Amazonie. Cette lacune peut être interprétée comme une preuve de diminution de la biomasse due à un climat plus sec et peut également traduire une impossibilité de se prononcer quant à la végétation présente à cette période, car si aucun dépôt sédimentaire lacustre ou marin ne s'effectue, aucun pollen existant ne pourra s'y déposer. Même lorsque les données palynologiques sont disponibles, elles ne semblent pas arriver à un consensus pour décrire les climats et la végétation en Amazonie au Quaternaire et restent peu nombreuses pour le plateau des Guyanes. Un autre point faible est le fort taux de décomposition de la matière organique lié au climat saturé d'humidité et à l'acidité des sols amazoniens qui n'améliore pas la conservation des fossiles. Ceci explique la quasi-inexistence des restes végétaux et la disparition des traces d'occupation humaine dont les artefacts archéologiques sont eux aussi très mal conservés (Mazière et Mazière, 1994). Les habitats précolombiens, principalement composés de bois et de végétaux (Migeon, 2009), et les ossements sont rapidement éliminés. Les céramiques et les pièces lithiques restent les principaux vestiges à partir desquels une reconstitution de l'occupation humaine ancienne guyanaise peut être envisageable mais ne renseignent pas directement sur la composition et la présence du couvert forestier.

Une approche alternative aux méthodes précédentes est l'utilisation des patrons actuels de diversité génétique pour modéliser l'histoire démographique passée d'une population d'arbres (Figure 15). Les histoires démographiques de plusieurs espèces permettront à leur tour d'inférer l'histoire écologique de la communauté forestière à partir des paramètres des populations tels que la variation de leurs tailles efficaces au cours du temps. Cette inférence logique nécessite donc que la taille efficace d'une population soit le reflet de sa taille réelle. La variation démographique locale d'une espèce permet de rendre

Encadré 1. Les principaux modèles de mutation

Modèle à K allèles (KAM) (Kimura, 1968) :

Les mutations apparaissent en fréquence égales et peuvent créer K allèles différents de façon équiprobables.



Modèle de mutation des allèles infinis (IAM) (Kimura, 1964) :

Généralisation du modèle KAM ($K \in [1, +\infty]$). Chaque mutation crée un nouvel allèle qui n'existait pas auparavant dans la population.



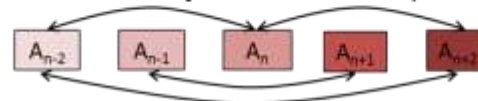
Modèle de mutation pas à pas strict (SMM) (Ohta et Kimura, 1973) :

Chaque mutation ajoute ou retranche un motif composé de plusieurs nucléotides.



Modèle de mutation généralisé (GSM) (Zhivotovsky *et al.*, 1997; Estoup *et al.*, 2002) :

Chaque événement mutationnel ajoute ou retranche plus d'un motif nucléotidique.



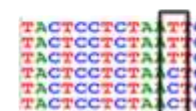
Modèle de mutation à deux phases (TPM) (Di Rienzo *et al.*, 1994) :

Généralisation du GSM. Chaque mutation ajoute ou enlève un motif nucléotidique (SMM) avec une probabilité p ou ajoute ou retranche X motifs (GSM) avec une probabilité $1-p$. X est tiré d'une loi géométrique.



Modèle des sites infinis (ISM) (Kimura, 1969) :

Chaque nouvelle mutation affecte un site différent de la séquence nucléotidique.



compte de son histoire à plus large échelle et rejoint dans ce sens le domaine de la phylogéographie qui permet d'élucider les processus historiques anciens dans leur contexte géographique grâce à des méthodes phylogénétiques (Avisé *et al.*, 1987; Avisé, 1989).

I.4.2 Génétique et histoire démographique

I.4.2.1 Évolution neutre du génome

La théorie neutre de l'évolution (Kimura, 1968; 1985; 1986) suggère que la majorité des changements évolutifs moléculaires résulte de la dérive aléatoire de mutations neutres ou presque neutres. Cette théorie ne nie pas le rôle de la sélection naturelle dans le cours de l'évolution adaptative, mais elle suppose que seule une fraction limitée des changements évolutifs de l'ADN est adaptative par nature tandis que la plupart des substitutions moléculaires sont phénotypiquement silencieuses et n'ont pas d'impact sur la fitness. Elles n'exercent donc aucune influence significative sur la survie et la reproduction et dérivent aléatoirement à travers les espèces.

Mutations neutres

Une mutation génétique est une modification de la séquence ADN pouvant affecter un seul nucléotide (SNP, Single Nucleotide Polymorphism) ou plusieurs bases de l'ADN simultanément. Parmi ces mutations, les substitutions impliquent le remplacement d'un nucléotide par un autre, les insertions-délétions (indels) représentent l'ajout ou la suppression d'une ou plusieurs bases et les inversions provoquent un changement de l'ordre des nucléotides. Toutes ces modifications entraînent la coexistence dans une population de plusieurs allèles d'une région génomique donnée, appelée locus, qui définit la notion de polymorphisme. Divers modèles de mutation ont été proposés pour décrire l'évolution de ce polymorphisme (Encadré 1). Aux mutations sélectivement neutres qui n'ont aucune influence significative sur la survie et la reproduction d'un organisme, on peut ajouter les mutations délétères qui se comportent comme des allèles quasiment neutres quand leur fréquence est faible. Par contre, à fréquence élevée ils sont effectivement contre-sélectionnés, ce qui les empêche de se fixer (Kimura, 1986).

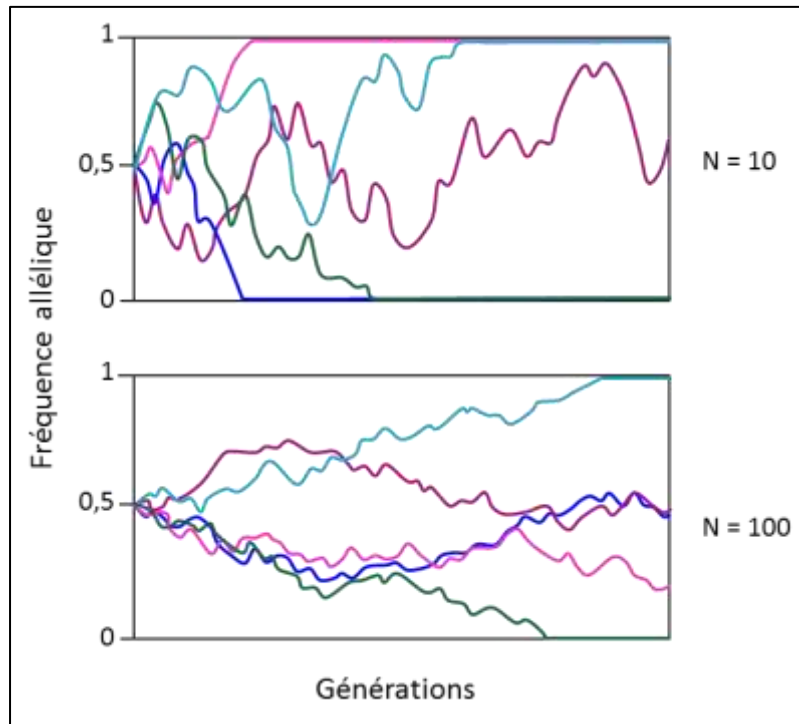


Figure 16. Conséquences de la dérive génétique sur la fréquence des allèles d'une population en fonction de sa taille N . Résultats de 5 simulations de l'évolution de la fréquence d'un allèle, fixée initialement à 0.5, au cours des générations pour deux populations de tailles différentes ($N=10$ et $N=100$). La dérive génétique fixe plus rapidement les allèles dans des populations de petites tailles.

Dérive génétique aléatoire

L'évolution des mutations neutres est déterminée par la dérive génétique aléatoire. Les mutants neutres étant sélectivement identiques aux formes préexistantes desquelles ils dérivent, leur devenir est donc le fruit du hasard et leur fréquence augmente ou décroît de façon aléatoire au fil du temps. Si de nombreux gamètes mâles et femelles sont engendrés à chaque génération, seul un nombre restreint est utilisé pour donner naissance aux individus de la génération suivante. Bien qu'une écrasante majorité des mutants neutres est éliminée par hasard, une minorité se fixera en définitive dans la population (Kimura, 1985) avec une probabilité $\frac{1}{2N}$ ($2N$ étant le nombre total de gamètes d'une population haploïde). Le changement aléatoire de fréquence des gènes entre générations est appelé dérive génétique. Cette dérive est non-directionnelle, n'est pas confinée au cas de neutralité sélective et dépend fortement de l'échantillonnage aléatoire des gamètes lors de la reproduction, lui-même lié à la taille de la population. Plus la population est petite, plus les effets de l'échantillonnage aléatoire sont importants et plus la fixation ou la perte d'allèles est rapide et aboutit à une réduction de la diversité génétique dans la population (Figure 16). À plus ou moins long terme, la dérive aléatoire seule tend à fixer des allèles dans la population mais cet effet est contré par l'apport constant de nouvelles mutations permettant le maintien du polymorphisme, et créant ainsi un équilibre mutation-dérive.

Taille efficace

Dans les populations naturelles, l'effectif de la population qui détermine le rythme de la dérive génétique n'est pas forcément égal à sa taille réelle. On utilise donc la taille efficace d'une population (Wright, 1931) qui correspond au nombre d'individus d'une population « idéale » de Wright-Fisher pour laquelle le degré de diversité génétique est équivalent à celui de la population réelle. La taille efficace est égale à la taille réelle lorsque les populations sont panmictiques, de taille constante, avec des générations non-chevauchantes, sans migration, ni sélection et un taux constant de mutations neutres (Wright, 1929; Fisher, 1930). Mais en pratique la taille efficace est souvent plus faible que la taille observée car elle est influencée par plusieurs facteurs tels qu'un sex ratio déséquilibré, une fertilité variable entre individus, des fluctuations de populations ou encore des croisements entre petits groupes (Ridley, 2004).

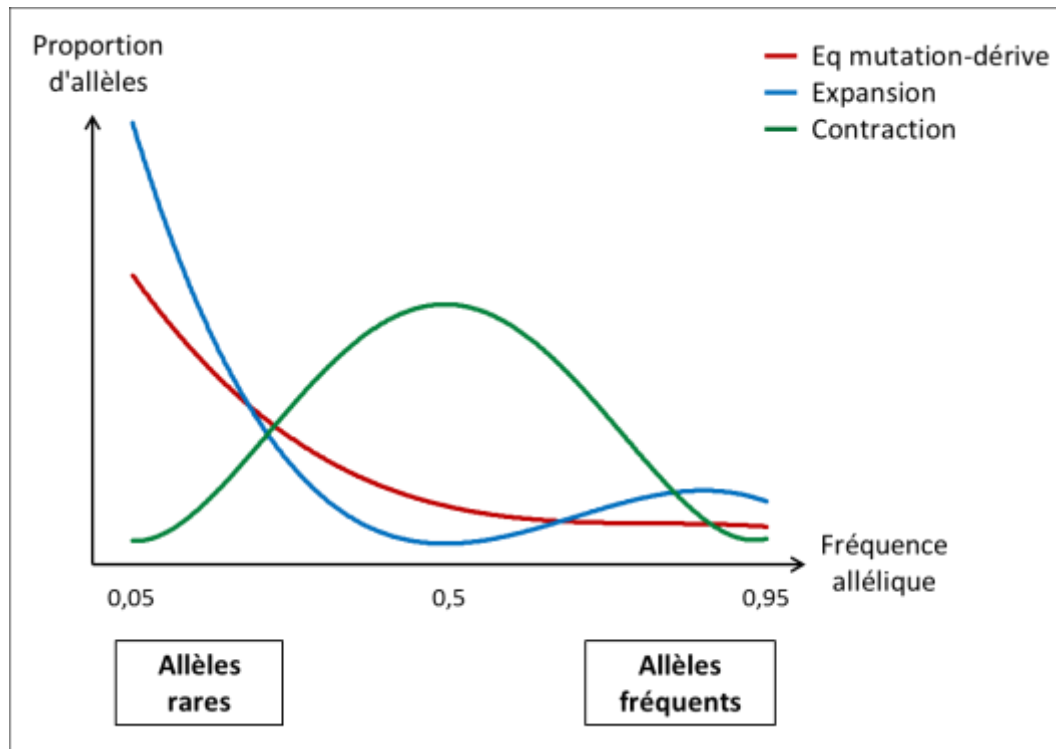


Figure 17. Distribution des fréquences alléliques dans une population à l'équilibre, après expansion ou contraction démographique.

Marqueurs moléculaires neutres

De nombreuses régions du génome ne sont pas soumises à l'effet de la sélection et sont dites neutres. La variabilité génétique que l'on peut observer entre ces séquences reflète donc uniquement la dynamique passée des populations étudiées (Milligan *et al.*, 1994). Cependant tous les génomes neutres n'évoluent pas de la même manière et chaque marqueur moléculaire, correspondant à une région ciblée dans l'ADN, a ses propres caractéristiques. L'ADN nucléaire comporte des régions hautement répétées plus connues sous le nom de microsatellites ou SSR (en anglais Simple Sequence Repeat). Ce sont des répétitions en tandem de motifs nucléotidiques, longues de 1 à 6 paires de bases (Kashi *et al.*, 1997; Toth *et al.*, 2000). Les microsatellites sont présents dans tout le génome mais sont plus abondants dans les régions non-codantes des génomes eucaryotes où ils sont très nombreux et très dispersés (Kashi *et al.*, 1997). Les principaux modèles de mutation utilisés pour décrire ces marqueurs sont le modèle pas à pas strict (SMM), le modèle généralisé (GSM) et le modèle à deux phases (TPM) (Voir Encadré 1). Le fort taux de mutation de ces régions, de l'ordre de 10^{-2} à 10^{-6} nucléotides par génération (Thuillet *et al.*, 2002) en fait un marqueur codominant adapté pour retracer l'histoire démographique récente des populations. Au contraire, l'absence de recombinaison du génome chloroplastique, due à une transmission clonale par la mère pour la plupart des angiospermes (Birky, 1995), ainsi que son faible taux de mutation de l'ordre de 10^{-9} nucléotides par génération (Wolfe *et al.*, 1987) font des régions neutres de cet ADN d'excellents outils pour étudier les processus démographiques anciens.

I.4.2.2 Effets des changements démographiques sur le génome

Comparaison au modèle neutre

Les marqueurs moléculaires neutres, affranchis de l'effet de la sélection, sont largement utilisés pour étudier les changements démographiques au cours du temps. Ces derniers vont tendre à éloigner les distributions des fréquences alléliques ou les indices de diversité génétique d'une population de ceux attendus pour une population de Wright-Fisher, démographiquement stable, à l'équilibre mutation-dérive (Figure 17). Une réduction rapide de la taille efficace par un goulot d'étranglement ou un effet fondateur aura pour conséquence une diminution des allèles rares et une augmentation des allèles en fréquence intermédiaire par rapport à l'attendu (Ewens, 1972; Watterson, 1978). Inversement, l'expansion d'une population entraîne une augmentation de la taille efficace qui diminue la

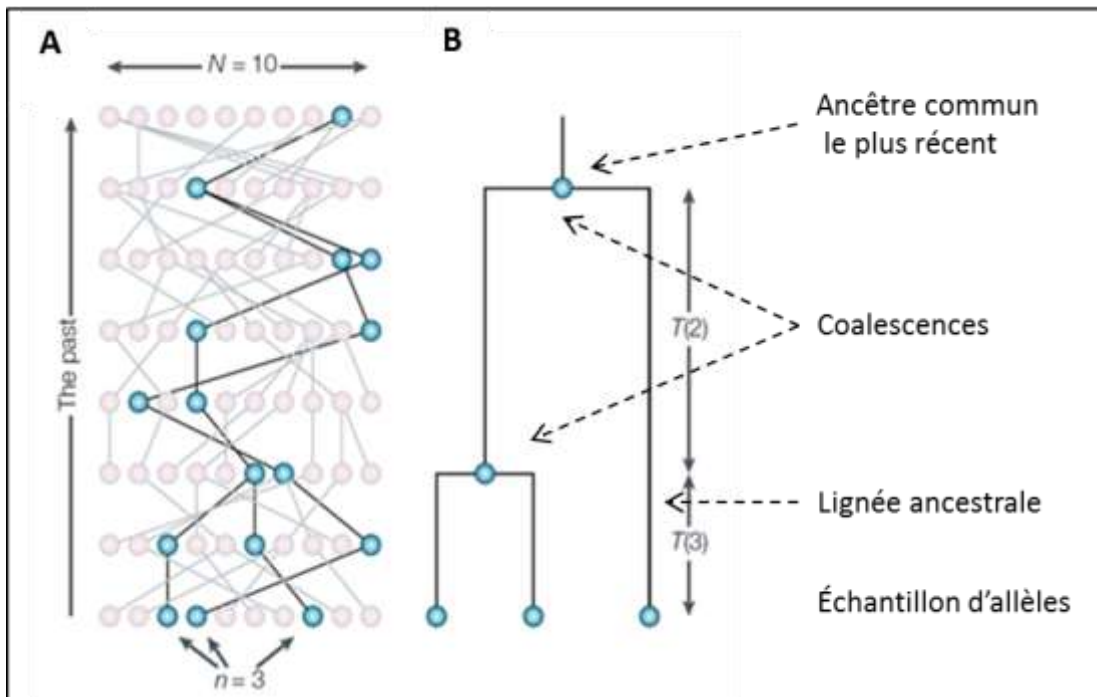


Figure 18. Principe de base de la coalescence. (A) Généalogie complète d'une population de 10 copies d'allèles d'après Rosenberg et Nordborg (2002). Les lignes noires retracent les ancêtres des 3 lignées échantillonnées jusqu'à leur unique ancêtre commun. (B) Généalogie simplifiée des 3 lignées échantillonnées. $T(2)$ et $T(3)$ sont les temps de coalescence.

dérive génétique (Kimura et Crow, 1964) et augmente la probabilité d'apparition de nouveaux allèles neutres. On observe donc une fréquence plus importante d'allèles rares (Ewens, 1972; Watterson, 1978) ainsi qu'un excès d'homozygotie (Watterson, 1986) par rapport aux populations stationnaires. L'indice D de (Tajima, 1989) est un des tests de neutralité qui compare la différence entre le nombre de sites polymorphes et la moyenne des différences entre paires de séquences ADN. Comme mentionné précédemment, l'expansion de population influence ces deux estimateurs en augmentant les sites polymorphes alors que l'excès d'allèles rares modifie peu la diversité.

Changements démographiques et Coalescence

La théorie de la coalescence (Kingman, 1982) permet de reconstruire rétrospectivement les arbres généalogiques des copies d'allèles d'une population supposée à l'équilibre de Wright-Fisher. Les lignées alléliques coalescent, c'est-à-dire fusionnent, au fur et à mesure que l'on remonte dans le temps jusqu'à leur ancêtre commun le plus récent (MRCA, de l'anglais "most recent common ancestor") (Figure 18 ; Rosenberg et Nordborg (2002)). Les mutations le long des branches de l'arbre créent du polymorphisme alors que d'autres allèles sont perdus par dérive génétique aboutissant à l'extinction de lignées. La probabilité de coalescence à chaque génération, d'une population diploïde à l'équilibre et de taille N, suit une loi géométrique de paramètre $\frac{1}{2N_i}$. Ainsi plus la taille d'une population est petite plus la probabilité de coalescence est forte et plus le temps de coalescence jusqu'au MRCA est réduit. La dernière coalescence de l'arbre occupe la plus grande partie du temps total de coalescence car la probabilité de fusion entre 2 lignées est plus faible qu'entre plusieurs lignées, diminuant ainsi la nécessité de grands échantillons pour préciser le temps depuis le MRCA.

La taille des populations au cours du temps module donc la topologie des arbres généalogiques d'allèles. En considérant des variations démographiques drastiques, on peut représenter cette transition par deux tailles de populations différentes entre elles au temps t_1 mais constantes avant cette date et entre t_1 et t_0 (Figure 19.A). Une expansion instantanée de population aura donc pour effet d'augmenter la taille de la population ($N_0 > N_1$), de diminuer la probabilité des événements de coalescence entre t_1 et t_0 et ainsi de rallonger les branches terminales de l'arbre généalogique. La perte d'individu, au contraire,

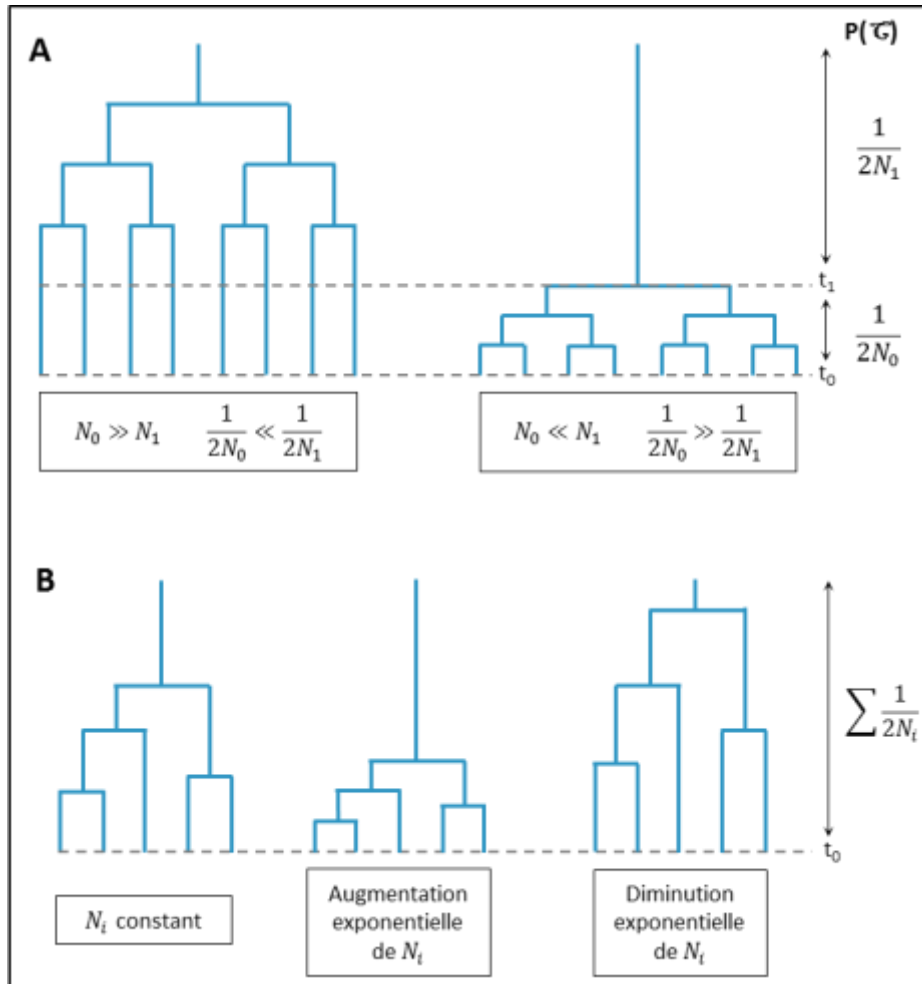


Figure 19. Conséquence d'un changement démographique (A) ponctuel et drastique ou (B) exponentiel sur la topologie des arbres de coalescence. (A) N_0 taille constante de la population entre t_0 et t_1 et N_1 taille constante de la population avant t_1 . N_0 et N_1 sont différents. (B) La taille de la population N varie au cours du temps. $P(\mathcal{T})$ représente la probabilité de coalescence.

réduira la diversité génétique existante à t_1 et la probabilité de coalescence sera plus importante en raison du faible effectif (Harpending *et al.*, 1998).

Dans les phases de contraction-expansion ou lors d'évènements démographiques continus, la violation du modèle de Wright-Fisher, concernant la constance des effectifs de population, entraîne une modification de la loi de probabilité des évènements de coalescence qui devient alors $\sum_{i=1}^t \left(\frac{1}{2N_i}\right)$. Lors d'une expansion de population, cette moyenne harmonique est influencée par les petites valeurs d'effectifs N_i au cours du temps. La probabilité de coalescence sera donc forte, elle se traduira par un raccourcissement des branches de l'arbre et un ancêtre commun plus récent que celui d'une population restée constante (Kuhner, 2008) (Figure 19.B). Des tests de neutralité basés sur le coalescent utilisent la diversité le long des généalogies pour détecter des écarts au modèle de Wright-Fisher. L'indice D' de Fu et Li (1993), par exemple, compare le polymorphisme récent des branches terminales avec les mutations plus anciennes présentes à la racine de l'arbre. Une expansion de population va alors produire un excès d'allèles rares dans les branches terminales des généalogies.

I.4.3 Inférences statistiques des évènements démographiques

I.4.3.1 Modélisation du coalescent

Les approches de coalescence sont des outils puissants pour modéliser la généalogie d'allèles et l'analyse génétique des processus démographiques. La topologie d'un arbre est d'abord simulée pour une population de taille N jusqu'à ce que tous les allèles coalescent jusqu'à l'ancêtre commun. Les mutations sont ensuite distribuées selon une loi de Poisson avec une probabilité μ et injectent du polymorphisme dans la population simulée. Enfin, des allèles sont tirés aléatoirement dans la population modélisée comme c'est le cas lorsque nous échantillons une population réelle.

I.4.3.2 Statistiques Bayésiennes

Les méthodes bayésiennes (Bayes et Price, 1763) utilisent la vraisemblance pour inférer de façon performante les paramètres d'un modèle en incorporant l'information a priori de ces paramètres $P(\theta)$. La probabilité postérieure des paramètres θ sachant les données D peut être calculée par le théorème de Bayes suivant :

$$P(\theta|\mathcal{D}) = \frac{P(\mathcal{D}|\theta) \times \pi(\theta)}{P(\mathcal{D})} \quad 1$$

avec $P(\mathcal{D}|\theta)$ la probabilité des données sachant le paramètre qui est la fonction de vraisemblance, $\pi(\theta)$ la loi à priori des paramètres et $P(\mathcal{D})$ la probabilité des données.

La probabilité des données devient alors :

$$P(\mathcal{D}) = \int P(\mathcal{D}|\theta) \pi(\theta) d\theta \quad 2$$

Dans le cas d'un modèle de coalescence, il est ainsi possible d'étudier la distribution de probabilité des topologies de l'arbre généalogique T et d'estimer les paramètres θ de ce modèle (Stephens, 2001) tels que la taille efficace des populations N_e , le taux de mutations μ mais aussi le taux de croissance, la date des événements démographiques ou le flux de gènes (Beaumont et Rannala, 2004). La vraisemblance des paramètres se traduit dans ce cas par l'estimation de la distribution de la probabilité conjointe des données D et des paramètres θ en prenant en compte la distribution de probabilité des topologies T et se traduit de la façon suivante :

$$L(\theta) = P(\mathcal{D}|\theta) = \int P(\mathcal{D}|T, \theta) P(T|\theta) dT \quad 3$$

avec $P(\mathcal{D}|T, \theta)$ la probabilité des données sachant la topologie et les paramètres, $P(T|\theta)$ la probabilité *a priori* de la topologie sachant l'information des paramètres qui représente un *hyperprior*. Les deux dernières probabilités sont calculables directement à partir de la théorie du coalescent (Stephens, 2001).

Avantages et inconvénients

Le Bayésien permet d'explorer efficacement l'espace des topologies d'arbre et celui des paramètres grâce notamment à l'utilisation de chaînes de Markov couplées à la méthode de Monte Carlo (MCMC). La chaîne de Markov parcourt l'espace multidimensionnel des paramètres. Ce processus sans mémoire, lorsqu'il atteint sa "phase stationnaire", est guidé par la méthode de Monte Carlo pour estimer la distribution de probabilité postérieure en échantillonnant les valeurs de ces paramètres de façon périodique. L'approximation de la distribution sera d'autant plus exacte que le nombre de pas effectué par la chaîne de Markov sera élevé. Il est tout de même nécessaire de faire varier T et θ de manière efficace afin de bien mélanger toutes les combinaisons de paramètres possibles ce qui est extrêmement lourd du point de vue du calcul. La fonction de vraisemblance devient également plus difficile à définir et à calculer lorsque le modèle se complexifie et que le nombre de paramètres à estimer et le nombre d'allèles du jeu de données augmentent. Il est également indispensable d'exécuter un grand nombre de chaînes de Markov afin d'éviter les

optimums locaux et s'assurer de la stabilité des résultats obtenus. Enfin le choix des priors des paramètres est important car si il n'est pas assez informatif certaines régions de l'espace des paramètres seront explorées inutilement entraînant une augmentation des temps de calcul déjà important.

I.4.3.3 Calcul Bayésien Approché (ABC)

Dans le contexte des méthodes ABC, de l'anglais Approximate Bayesian Computation (Beaumont *et al.*, 2002), l'inférence des paramètres d'un modèle ne repose plus sur la loi a posteriori $P(\theta|D)$ (Équation 1) mais sur la loi a posteriori partielle (Doksum et Lo, 1990) suivante :

$$P(\theta|S_{obs}) = \frac{P(S_{obs}|\theta) \times \pi(\theta)}{P(S_{obs})} \quad 4$$

avec S_{obs} un vecteur de dimension d des statistiques descriptives.

Algorithme de rejet

Les méthodes ABC prennent leur origine dans l'algorithme de rejet (Tavaré *et al.*, 1997; Pritchard *et al.*, 1999) dont la démarche générale est la suivante (Figure 20) :

1. La première étape consiste à formuler un modèle démographique ou évolutif et de le simuler un grand nombre de fois en utilisant le coalescent. Les paramètres de ce modèle, tels que la taille efficace des populations N_e , le taux de mutations μ mais aussi la date des évènements démographiques, de divergence entre populations ou le flux de gènes, sont tirés dans des distributions définies au préalable.
2. Un échantillonnage est effectué à l'intérieur des populations simulées. Il doit contenir le même nombre d'individu ou de copies d'allèles présents dans le jeu de données réel qui n'est en fait que l'équivalent d'un échantillonnage dans une population naturelle.
3. Des "statistiques descriptives" S sont calculées pour les échantillons simulés ainsi que pour le vrai jeu de données. Ces statistiques doivent contenir le maximum d'information de l'échantillon en restant sous une forme la plus simple possible. Elles peuvent par exemple résumer les propriétés de diversité d'un échantillon, sa structuration ou encore son écart au modèle de Wright-Fisher.
4. La comparaison entre les statistiques descriptives des simulations de coalescence et des vraies données permettent de construire un « cercle de proximité » δ , autour des valeurs observées, correspondant au seuil d'acceptation au-delà duquel les simulations sont

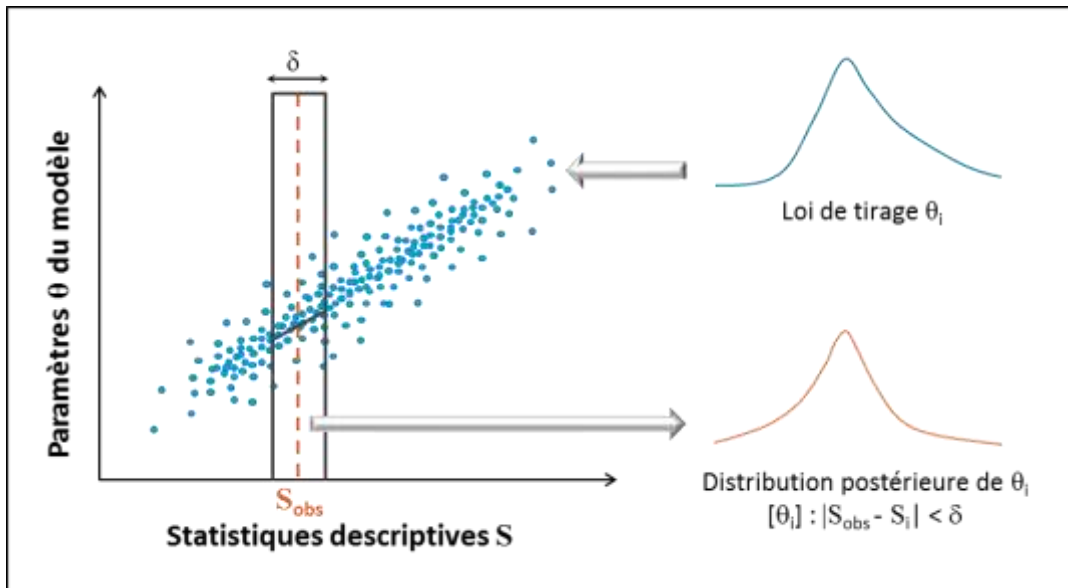


Figure 20. Principe de l'inférence des paramètres d'une population par le calcul Bayésien approché (ABC), adapté et simplifié d'après Beaumont (2010).

Le nuage de point représente les simulations de coalescence en fonction de la valeur des paramètres du modèle, tirée au préalable dans une distribution, et de la valeur des statistiques descriptives S qui en découle. Le « cercle de proximité » δ , autour des statistiques descriptives observées dans la population réelle S_{obs} , indique l'ensemble des simulations retenues pour calculer la distribution postérieure des paramètres θ .

rejetées. En général, une simple distance Euclidienne est calculée sur les statistiques descriptives normalisées et le seuil est défini tel qu'une petite fraction de simulations (0,1-3%), correspondant aux plus petites distances, soit retenue.

5. Les valeurs des paramètres θ ayant servi à modéliser les simulations retenues à l'étape 4 permettent de construire les distributions postérieures de ces paramètres ($[\theta] : |S_{obs} - S_i| < \delta$) et d'établir leur intervalle de crédibilité.

Méthodes avec ajustement

Une importante limite de ces algorithmes basés sur le rejet se pose lorsque la dimension du jeu de statistiques descriptives augmente. Beaumont *et al.* (2002) ont donc proposé une méthode d'estimation de la densité conditionnelle, basée sur un algorithme de régression linéaire locale permettant d'ajuster l'écart entre les statistiques descriptives simulées et observées. Dans cette approche, le choix du taux d'acceptation δ implique un compromis biais-variance : l'augmentation de δ réduit la variance grâce à un échantillonnage plus large permettant d'effectuer la régression, mais augmente aussi le biais lié aux écarts de linéarité notamment. Blum et François (2010) ont à leur tour apporté des améliorations supplémentaires en utilisant une régression non linéaire.

Développement de l'ABC

À ces méthodes ABC reposant directement sur un algorithme de rejet, sont venues s'ajouter des méthodes imitant les MCMC qui tiennent compte, dans la règle d'acceptation de l'algorithme de Metropolis-Hastings, de la distance entre les statistiques descriptives simulées et observées (Marjoram *et al.*, 2003; Bortot *et al.*, 2007). D'autres méthodes basées sur des algorithmes séquentiels par Monte Carlo combinent, elles, les méthodes de rejet et un échantillonnage séquentiel important (Sisson *et al.*, 2007; Beaumont *et al.*, 2009; Del Moral *et al.*, 2012).

L'ABC permet également, suivant la même démarche que celle décrite précédemment, de simuler et de comparer différents modèles afin de tester différentes hypothèses alternatives d'un processus. La sélection du modèle le plus représentatif du jeu de données réel nécessite d'en évaluer sa robustesse. D'autres étapes de validation et de vérification, notamment dans l'estimation des paramètres, sont nécessaires à la fiabilité des résultats et des conclusions tirées. (Beaumont, 2010; Bertorelle *et al.*, 2010; Csilléry *et al.*, 2010).

Avantages et inconvénients

L'ABC permet d'inférer des paramètres de populations naturelles sans avoir à calculer une fonction de vraisemblance souvent compliquée et offre la possibilité de tester plusieurs modèles démographiques et évolutifs (Beaumont, 2010). L'augmentation de l'utilisation de ces méthodes a conduit au développement de logiciels prenant en charge toutes les étapes, de la simulation du modèle à la distribution postérieure des paramètres en passant par la sélection entre modèles (DIY ABC, Cornuet *et al.* (2008) ; PopABC, Lopes *et al.* (2009) ; ABCtoolbox, Wegmann *et al.* (2010)). Les analyses sont ainsi facilitées d'autant plus qu'elles ne nécessitent que le calcul des statistiques descriptives, des distances Euclidiennes et des intervalles de confiance des paramètres. Ces calculs n'étant pas basés sur des chaînes de Markov, le problème de convergence ne se pose pas. Dans les cas où la fonction de vraisemblance peut être formulée et évaluée efficacement il n'y a aucun avantage à utiliser l'ABC comme alternative mais dans les cas contraires cette méthode permet de réduire le temps de calcul et de traiter des jeux de données très importants. Cependant, le choix des statistiques descriptives reste crucial et souvent problématique car si ces statistiques ne sont pas assez sensibles aux variations des paramètres du modèle elles n'apporteront aucune information et ne reflèteront pas ces paramètres. A une grande quantité de statistiques non informatives créant un bruit de fond qui peut entraîner un biais de l'estimation des paramètres, il est préférable d'utiliser moins de statistiques mais plus descriptives. De plus, l'absence de méthodes MCMC, dans la plupart des algorithmes, suppose une exploration aléatoire donc moins efficace de l'espace des paramètres nécessitant ainsi la simulation de nombreux coalescents pour pouvoir espérer explorer cet espace de façon convenable, notamment pour un modèle complexe (Beaumont, 2010; Bertorelle *et al.*, 2010; Csilléry *et al.*, 2010).

I.5 Objectifs

Le principal objectif de cette thèse est de caractériser, à travers l'étude de la démographie d'arbres forestiers tropicaux, l'impact de diverses perturbations naturelles et anthropiques afin de contribuer au développement d'une vision générale de l'évolution passée de cet écosystème sur le plateau des Guyanes. Une première hypothèse est que la démographie historique de la forêt à un endroit donné est le reflet de son évolution générale au cours du temps. Si tel est le cas, la reconstruction de son histoire nécessite que les traces de son passé soient gardées en mémoire. Une deuxième hypothèse est que les changements à faible échelle de temps et d'espace influencent localement la démographie des communautés alors que des perturbations plus anciennes et géographiquement plus importantes déterminent l'histoire régionale des forêts. L'échelle de temps considérée et les outils génétiques utilisés pour les étudier prennent alors tout leur sens. De même, lors de la mise en place de la stratégie d'échantillonnage, le contexte de changements globaux doit être gardé à l'esprit dans l'étude des différences locales et vice versa. Une dernière hypothèse est que toutes les espèces d'un écosystème ne réagissent pas de façon identique à une même perturbation. Cette supposition implique une modification des assemblages forestiers par le passé mais souligne surtout la nécessité de diversifier les espèces étudiées afin d'avoir une représentation fidèle de l'histoire des forêts tropicales guyanaises.

Dans le panel de marqueurs moléculaires disponibles pour les études d'évolution des populations, les microsatellites nucléaires montrent beaucoup d'avantages. La première partie de ma thèse vise donc à analyser le polymorphisme porté par chaque région de ce marqueur pour étudier ses propriétés générales. Pour cela, le séquençage de régions répétées de 3 genres d'arbres différents : *Jacaranda*, *Citrus* et *Quercus* a été réalisé. La connaissance des limites des microsatellites, notamment dans leur application en phylogéographie, est essentielle pour déterminer la meilleure stratégie à adopter pour la modélisation démographique utilisée par la suite.

La colonisation de la Guyane par l'homme à l'Holocène laisse supposer que cette occupation précolombienne a eu un impact sur les populations forestières locales. Pour tester cette hypothèse dans la deuxième partie de ma thèse, des marqueurs microsatellites et des méthodes ABC ont été utilisées pour retracer et comparer l'histoire récente des populations de 4 espèces d'arbres tropicaux entre plusieurs sites perturbés ou non par cette ancienne présence humaine. Si cette perturbation a laissé une trace au sein des génomes forestiers, il serait intéressant de pouvoir quantifier l'intensité des impacts anthropiques sur

ces forêts et les différencier des processus naturels. Des perturbations encore plus importantes, telles que les changements climatiques du Quaternaire, ont également dû moduler la distribution des forêts tropicales humide à l'échelle de la Guyane entière. Ainsi, dans la troisième partie de ma thèse j'ai étudié la démographie historique régionale de 8 espèces d'arbres tropicaux. L'analyse de données chloroplastiques, en plus des marqueurs microsatellites utilisés jusque-là, a permis d'étudier l'impact d'évènements récents et ancien. De plus, la comparaison de plusieurs scénarios avec les méthodes d'ABC a permis de tester plusieurs hypothèses et de donner une indication sur la signature démographique des forêts guyanaises.

Pour terminer, les principaux résultats de cette thèse seront rappelés et mis en relation afin d'esquisser un tableau général de l'impact des perturbations passées sur les forêts tropicales humides du plateau des Guyanes

Chapitre II

Étude de l'information phylogénétique portée par un allèle microsatellite

façon appropriée dans les chapitres suivants. Pour cela, des allèles de 1 à 3 marqueurs microsatellites ont été séquencés et génotypés pour des populations de 3 genres bien distincts d'arbres : *Citrus*, *Jacaranda* et *Quercus*.

Le séquençage a révélé des mutations d'un seul nucléotide (SNP) et/ou la présence d'insertion-délétion à l'intérieur mais aussi en dehors de la région répétée pouvant aboutir à plusieurs cas d'homoplasie de taille chez les 3 espèces étudiées. Des allèles ayant des histoires évolutives différentes seront donc traités comme identiques dans des analyses phylogénétiques, si la seule taille du fragment est prise en compte. Les patrons de diversité obtenus avec la variation des tailles des fragments, la variation du nombre de répétitions de la région répétée et le polymorphisme observé dans les régions encadrant le microsatellite ont ensuite analysés par des tests de corrélation des distances génétique des allèles des individus. Les différentes informations génétiques portées par chaque région d'un fragment microsatellite n'apparaissent que partiellement corrélées. Ainsi, la taille allélique reste une bonne approximation du nombre de répétitions malgré la présence de bruit de fond créé par d'autres mutations. Cependant, l'information génétique contenue dans cette région répétée n'est pas corrélée à l'information présente dans les séquences qui l'encadrent. La structuration génétique ainsi que des arbres phylogénétiques entre populations de *Jacaranda* ont été établis pour les variations génétiques de chaque région microsatellite avec plusieurs mesures de distance génétique (F_{ST} et N_{ST} pour les séquences, F_{ST} et R_{ST} pour les tailles alléliques et le nombre de répétition). Les résultats de ces analyses indiquent que l'information portée par la région répétée n'est pas représentative de l'information contenue dans le fragment microsatellite entier, ce qui implique l'existence d'autres sources de variabilité, décrites plus correctement par des modèles de mutation qui s'appliquent aux séquences ADN. En d'autres termes, la principale information phylogéographique contenue dans les données de tailles alléliques ne provient pas de la variation du nombre de répétitions mais du polymorphisme des régions qui encadrent cette région répétée.

Toutes les régions d'un marqueur microsatellite ne portent pas les mêmes informations génétiques et n'évoluent pas selon les mêmes modèles de mutation. Sachant que la principale source d'information phylogénétique portée par un allèle microsatellite provient des séquences encadrant la région répétée, l'application du modèle de mutation SMM aux données de tailles alléliques est fortement remise en question et l'analyse de ces données par des mesures de distance telles que le F_{ST} semble préférable. D'après ces résultats, le modèle SMM n'a, dans les chapitres suivants, jamais été retenu pour simuler les arbres de coalescence des copies d'allèles utilisés en ABC.

Always Look on Both Sides: Phylogenetic Information Conveyed by Simple Sequence Repeat Allele Sequences

Stéphanie Barthe¹, Felix Gugerli², Noelle A. Barkley³, Laurent Maggia⁴,
Céline Cardi⁵, Ivan Scotti⁶

- 1 Unité Mixte de Recherche EcoFoG (“Ecologie des forêts de Guyane”), UAG (University of French West Indies and French Guiana), Kourou, French Guiana
- 2 Biodiversity and Conservation Biology Research Unit, WSL (Swiss Federal Research Institute for Forest, Snow and Landscape Research), Birmensdorf, Switzerland
- 3 Plant Genetic Resources Conservation Unit, USDA-ARS (United States department of agriculture – Agricultural Research Service), Griffin, GA, USA
- 4 Unité Mixte de Recherche AGAP (“Amélioration génétique et adaptation des plantes méditerranéennes et tropicales”), Institut Agronomique néo-Calédonien, Nouméa, New Caledonia
- 5 Unité Mixte de Recherche AGAP (“Amélioration génétique et adaptation des plantes méditerranéennes et tropicales”), CIRAD (Centre de coopération internationale en recherche agronomique pour le développement), Montpellier, France
- 6 Unité Mixte de Recherche EcoFoG (« Ecologie des forêts de Guyane »), INRA (Institut National de la Recherche Agronomique), Kourou, French Guiana

Article publié dans le Journal PloSOne:

Barthe, S., F. Gugerli, N. A. Barkley, L. Maggia, C. Cardi and I. Scotti (2012). Always look on both sides: Phylogenetic information conveyed by Simple Sequence Repeat allele sequences. *PLoS One* 7: e40699.

Abstract

Simple sequence repeat (SSR) markers are widely used tools for inferences about genetic diversity, phylogeography and spatial genetic structure. Their applications assume that variation among alleles is essentially caused by an expansion or contraction of the number of repeats and that, accessorially, mutations in the target sequences follow the stepwise mutation model (SMM). Generally speaking, PCR amplicon sizes are used as direct indicators of the number of SSR repeats composing an allele with the data analysis either ignoring the extent of allele size differences or assuming that there is a direct correlation between differences in amplicon size and evolutionary distance. However, without precisely knowing the kind and distribution of polymorphism within an allele (SSR and the associated flanking region (FR) sequences), it is hard to say what kind of evolutionary message is conveyed by such a synthetic descriptor of polymorphism as DNA amplicon size. In this study, we sequenced several SSR alleles in multiple populations of three divergent tree genera and disentangled the types of polymorphisms contained in each portion of the DNA amplicon containing an SSR. The patterns of diversity provided by amplicon size variation, SSR variation itself, insertions/deletions (indels), and single nucleotide polymorphisms (SNPs) observed in the FRs were compared. Amplicon size variation largely reflected SSR repeat number. The amount of variation was as large in FRs as in the SSR itself. The former contributed significantly to the phylogenetic information and sometimes was the main source of differentiation among individuals and populations contained by FR and SSR regions of SSR markers. The presence of mutations occurring at different rates within a marker's sequence offers the opportunity to analyse evolutionary events occurring on various timescales, but at the same time calls for caution in the interpretation of SSR marker data when the distribution of within-locus polymorphism is not known.

Table 1. Glossary

Amplicon	Product of a DNA amplification reaction.
SSR	Simple sequences repeat (or “microsatellite”). Tandem repeat of simple di- to hexa-nucleotide sequence motifs.
FR	Flanking region. DNA sequences appearing in an amplicon on either side of the SSR sequence.
Indel	A sequence gap in a DNA sequence alignment caused by an insertion or deletion mutation.
SNP	Single nucleotide polymorphism. DNA polymorphism that involves a change in a single base of a DNA sequence.
SSR locus	A specific genomic region consisting of SSR (microsatellite) DNA and its flanking regions (FRs).

Introduction

Simple sequence repeats (SSRs; see glossary, Table 1) can conveniently be used as genetic markers owing to their polymorphism and relative ease of interpretation [1]. SSRs are widely used for, e.g., reconstructing phylogenetic relationships [2], for analysing spatial genetic structure among and within populations [1], and for detecting and explaining patterns linked to habitat fragmentation and gene flow [3]. In many studies, SSR markers demonstrated recent expansions or bottlenecks in various plant [4,5], animal [6-8] and human populations [9,10].

SSRs are often expected to mutate following the stepwise mutation model (SMM) [11], whereby mutations alter the length of the repeat either by adding or by deleting a single repeat unit at a fixed rate [12]. More elaborate mutational processes, such as the two-phase model (TPM; [13]) and the generalized stepwise model (GSM; [14,15]), allow for multi-step mutations. Departures from these models can however occur [16-20]. Moreover, SSR alleles are generally scored as the length (in base pairs) of PCR amplicons (see glossary, Table 1) containing the SSR, with differences in amplicon size taken to represent differences in repeat number in the SSR. Amplicon size includes of course SSR repeat number plus the length of the flanking region (FR; see glossary, Table 1). These chunks of sequence can be and actually often are polymorphic and may contain both single nucleotide polymorphisms (SNPs; see glossary, Table 1) and insertions/deletions (indels; see glossary, Table 1) [18,21-24]. Indels which have been demonstrated to occur in the FR sequences of some SSR alleles clearly contribute to total amplicon size, but they are interpreted under the SMM as (false) variation in SSR length. SNPs contribute to the total sequence variance of SSR loci (see glossary, Table 1) that goes largely unnoticed when only amplicon size is recorded (which is typical of most SSR studies). As a consequence, SSR amplicon size data are prone to a particular form of size homoplasy [15], i.e. equally sized alleles may have different sequences and therefore be evolutionarily different. This casts doubt on the evolutionary and population-genetic inferences that can be drawn from estimates of divergence obtained from SSR amplicon size data. It is therefore necessary to examine the kinds of evolutionary information conveyed in SSR alleles, to assess their consequences on fragment size-based evolutionary inferences, and to find ways to take advantage of this so far unexploited source of variability.

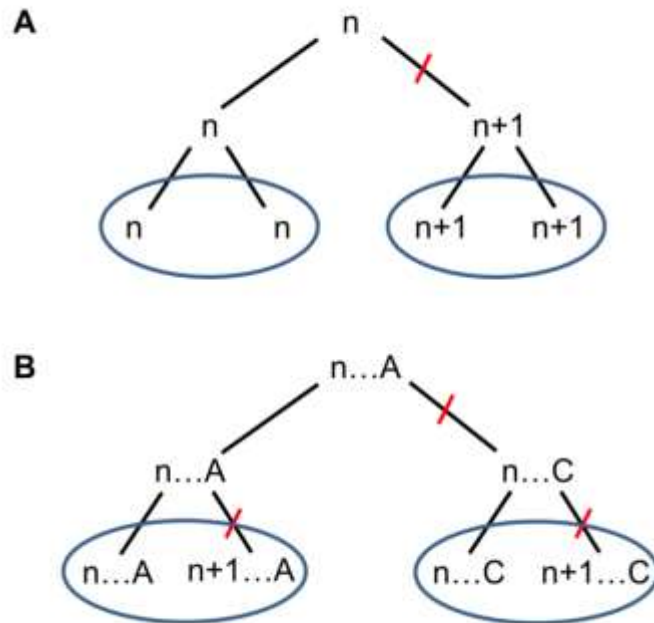


Figure 1. Two alternative genealogies for simple sequence repeat (SSR) alleles containing the same number of repeats. SSR alleles are indicated by their number of repeats (n , $n+1$). The A/C letters indicate a SNP in the flanking regions. Red bars correspond to mutational events in the flanking region sequence or in the number of SSR repeats. (A) No sequence information: deduced genealogy of observed data (third line) groups alleles together according to their number of repeats and involves a single SSR mutation. Genealogy is recent. (B) Consideration of sequence information: deduced genealogy involves a SNP mutation and two SSR mutations (alternative topology will involve two identical and independent substitutions at the same nucleotide site and a unique SSR mutation, which is less likely). Genealogy is ancient and SSR alleles do not group according to their numbers of repeats.

Figure 1 succinctly shows how information in FR sequence variation may change one's view of SSR marker evolutionary information content. Figure 1A shows a "plain" SSR locus, where a mutation (bar on the right branch) changes the number of repeats from n to $n+1$. The four observed alleles are then linked by a simple genealogy where alleles with the same number of repeats are evolutionarily closer. This is the classical view of SSR allele evolution under repeat-number mutational models. Figure 1B shows four alleles with the same numbers of repeats as in Figure 1A, but in this case information from a SNP in the FR (indicated by the "...A/...C" symbols) is added. Here, alleles with different numbers of repeats share the same SNP. Given that SNP mutation rates are generally lower than those of SSRs, the most-likely genealogy is one in which divergence occurs first at the SNP (mutation in the upper right branch) and then in SSRs within flanking-region haplotypes (mutations in the lower branches). In this case, three mutations have occurred, the genealogy is expected to be older than in 1A, and the alleles are grouped based on their FR sequences and not based on the number of repeats. A simulation study [25] revealed that SSR and SNP sources of variation provide independent evolutionary information, so it is reasonable to study the role of SNP variation in SSR flanking regions. We therefore asked the following questions: How much does amplicon size variation reflect variation in SSR repeat number? How important is the contribution of the various sources of sequence variation in total SSR diversity? Do these extra-SSR polymorphisms carry a different population and phylogenetic differentiation signal than the SSRs themselves? If so, how do the different signals combine? How does one best interpret SSR data, once these departures from the SMM have been acknowledged and assessed? To answer these questions, we sequenced SSR amplicons obtained from three divergent genera of long-lived angiosperm tree species to inspect the distribution of polymorphisms within and around SSRs and to evaluate the impact of each source of sequence variation on the detection of evolutionary and population genetic signals.

Material and Methods

Choice of data sets

Three data sets were utilized in this study. The structures of the three data sets analysed here varied in the following ways: the *Citrus* data represent sequences of individual trees each belonging to a different species, variety or cultivar; the *Jacaranda* data were collected in four different sampling sites considered as four distinct populations; and the oak data came from a single mixed stand of sessile/pedunculate oaks (*Quercus robur*, *Q. petraea*). The *Citrus*, *Jacaranda*, and *Quercus* data sets were submitted to partially different types of analyses according to the structure of each data set.

Given the taxonomy and prevailing theory on many *Citrus* species being derived by natural hybridization from ancestral species, alleles were chosen to minimize known hybrids from the data set of Barkley et al. [26,27], who examined the genetic diversity of 370 *Citrus* accessions. The individual identity of each accession (as provided by the University of California, Riverside citrus database, <http://www.citrusvariety.ucr.edu/>) is given in the Supplementary Table 1 of Barkley et al. [26]. The *Jacaranda* data were obtained from three sites in French Guiana (Counami, Paracou and Saint-Laurent) and one in Brazil (Tapajos) (Table S1). The *Quercus* alleles were taken from genetic analyses of *Q. petraea* and *Q. robur* in a mixed oak stand in Switzerland [28,29]. The SSR markers used for this study were: cAGG9, CCT01 and GT03 for *Citrus* [26,30]; Jc3H10, Jc3F4 and Jc3H10 for *Jacaranda*; QrZAG30 for *Quercus* [31]. Allele sequencing for the *Citrus* and *Quercus* data sets was as reported in Barkley et al. [26] and Gugerli et al. [29].

Molecular methods

Jacaranda markers were obtained following the method developed for SSR markers in tropical crops by Billotte et al. [32]. Ten nanograms of DNA was digested with *Rsa*I, followed by an enrichment step in (GT)_n and (GA)_n repeats by hybridization of cleaved fragments with biotin-labelled (CA)₈ and (CT)₈ oligoprobes and capture of the selected sequences with streptavidincoated magnetic beads. Those fragments were ligated into a pGEM-TEasy vector (Promega, Madison, WI), and plasmids were then used to transform competent XL-1 Blue strain *Escherichia coli* (Stratagene, La Jolla, CA). Positive colonies were tested using colony polymerase chain reaction (PCR) to check the presence of inserts. Amplified fragments were then transferred from an agarose gel onto a N+ Hybond membrane for SSR presence screening by hybridization with labeled γ -³²P (GT)₁₅ and (GA)₁₅ oligoprobes. One hundred and thirteen bacterial clones containing plasmids with inserts

that gave strong hybridization signals, with sizes ranging from 500 to 1000bp, were selected and cultured. Plasmids were extracted and sequenced using the universal T7 primer on an automated ABI 377 capillary sequencer (Applied Biosystems, Foster City, CA). After discarding duplicates, hybrid clones and clones with the SSR region too close to either end of the sequence, thirteen sequences were suitable for primer design and allowed a successful design using Oligo 3 software (Molecular Biology Insights Inc, USA). Upon screening for polymorphism and clarity of PCR patterns, three primer pairs were chosen for subsequent analyses. Primer sequences, annealing temperatures and GenBank/EMBL accession numbers are shown in Table S2. Dried cambium discs and leaf samples were flash-frozen in liquid nitrogen and later ground to a powder using a mortar and pestle. DNA extractions were performed by following Colpaert et al. [33]. PCRs for the detection of SSR polymorphisms were carried out in a 12 μ L volume containing 6 μ L 20-fold diluted DNA, 1x *Taq* buffer, 0.26 mM of dNTP, 0.03 U/ μ L *Taq* DNA polymerase (all products from Invitrogen, Carlsbad, CA) and 0.54 μ M of each primer (MWG Biotech, Ebersberg, Germany). For Jc3A10 primers, 0.3% BSA was added. An initial denaturation at 94°C for 5 min was followed by 35 cycles of 94°C for 30 s, annealing temperature for 30 s and 72°C for 30 s, and a final extension at 72°C for 10 min. Genotyping was performed using fluorescently labeled primers (PET, 6FAM and NED) in the previous PCR protocol and fragments were separated on an ABI 3130XL capillary sequencer (Applied Biosystems, Foster City, CA) using ABI POP4 and Applied Biosystems LIZ-500 as internal standard, following the manufacturer's instructions. DNA was bidirectionally sequenced directly from PCR amplification products. When necessary, gametic phase was determined in heterozygotes by cloning amplicons and sequencing at least one allele per individual following the TA cloning kit protocol from Invitrogen (Carlsbad, CA). PCR products were ligated with the plasmid pCR®2.1 and used to transform competent DH5 cells. After an overnight incubation at 37°C, white colonies were isolated. The presence of an SSR allele was checked by PCR with universal M13 primers. PCRs consisted of 1 μ L of cultivated colonies, 1x *Taq* buffer, 0.125 mM dNTPs, 0.025 U/ μ L *Taq* DNA polymerase (all products from Invitrogen, Carlsbad, CA) and 0.83 μ M forward and reverse primers. Cycling conditions consisted of 94°C for 5 min; 35 cycles of: 94°C for 1 min, 50°C for 1 min, 72°C for 2 min; and one cycle of 72°C for 5 min.

PCR products

Table 2. Characteristics of simple sequence repeat (SSR) data sets of the three taxa, *Citrus* (C), *Jacaranda* (J) and *Quercus* (Q).

Data set	Type of data	SSR locus	Size (bp) ^a	N ^b	Reference
Set C	Collection of provenances	cAGG9	123	34	[30]
		CCT01	167	34	[27]
		GT03	177	36	[27]
Set J	Four populations	Jc3A10	180	100	This study
		Jc3F4	144	101	This study
		Jc3H10	215	96	This study
Set Q	One mixed population	QrZAG30	255	47	[31]

^a Total length of the amplicon consensus sequence in base pairs.

^b Total number of analysed alleles

were checked on a 4% agarose gel in 0.5X TAE. Plasmids were sequenced with the ABI BigDye® Terminator V3.1 kit (Applied Biosystems, Foster City, CA) following the manufacturer's protocol. Cloning products were diluted 1:40, purified with ExoSap-IT (USB Corporation, Cleveland, OH), and separated on an ABI 3130XL capillary sequencer (Applied Biosystems, Foster City, CA). All sequences were aligned and edited using CodonCode Aligner V1.6.3 (Codoncode Corporation, Dedham, MA).

Data analyses

For each data set (*Citrus*, *Jacaranda*, and *Quercus*), each allele was characterised by the following information: (a) amplicon size variation, (b) SSR variation, (c) FR sequence variation and (d) amplicon sequence variation. Each of these portions of information conveyed by SSR marker data was analysed as a separate source of variation. Results were compared across data sets to assess their evolutionary information content. Indels were coded as SNP mutations to represent them as single mutational events. Levels of polymorphism (number of alleles (A) or haplotypes (h), Nei's genetic diversity (H_e), number of SNPs and indels) were recorded for each source of variation. Linkage disequilibrium (LD) was computed for FR sequence variation.

Data were analysed in two ways: (a) for all data sets, at the level of individual alleles; (b) for *Jacaranda* only, at the population level.

(a) At the individual-allele level, matrices of pairwise genetic distances were computed between individuals and between alleles based on each source of variation. For amplicon length, distance is represented by difference in length; for SSR length, distance is the difference in number of repeat units; for whole amplicon sequence, distance is the total number of differences, including SNPs, indels and number of repeat units; for flanking region sequences, distance is the total number of SNPs and indels. The correlation between matrices was tested by a Mantel test. LD between all pairs of polymorphic sites in FR sequences was tested with a Markov chain Monte Carlo (MCMC) procedure of 10,000 steps and a burn-in phase of 1,000 steps (default values). Pairwise genetic distances for amplicon and FR sequence variation were computed based on the number of mutational steps between sequences. Pairwise genetic distances for SSR variation and amplicon size variation data were computed as Euclidian distances (genetic distances based on the SSR variation were not computed for *Citrus* loci CCT01 and GT03, which displayed extremely complex repeat sequences; see Figure S1). Most of the above calculations were performed with

Table 3. Polymorphism at each simple sequence repeat (SSR) locus differentiated by segments of the DNA amplicon. SSR1, SSR2 and SSR3: respectively first, second and third SSR occurring in each amplicon (see Supplementary figure S1 for details on each amplicon's sequence); FR: flanking regions; H_e : Nei's genetic diversity; A: number of alleles; h: number of haplotypes. Data sets: C, *Citrus*; J, *Jacaranda*; Q, *Quercus*.

SSR locus (data set)	Amplicon size variation		Amplicon sequence variation		SSR1 variation			SSR2 variation			SSR3 variation			FR sequence variation					
	H_e	A	H_e	h	Repeat unit	H_e	A	Repeat unit	H_e	A	Repeat unit	H_e	A	Sequence length (bp)	Numb er of SNPs	SNPs per 100 bp	Number of indels	H_e	h
cAGG9 (C)	0.772	4	0.868	11	(GAG) _n	0.772	4	(GAA) _n	0.485	4	-	-	-	82	5	6.10	1	0.542	8
CCT01 (C)	0.758	4	0.966	20	(TCC) _n	-	-	-	-	-	-	-	-	140	17	12.14	0	0.861	14
GT03 (C)	0.783	5	0.900	12	(GT) _n	-	-	(AT) _n	-	-	-	-	-	138	12	8.70	1	0.808	9
Jc3A10 (J)	0.856	1 3	0.937	43	(CT) _n	0.769	12	(CA) _n	0.700	6	-	-	-	116	2	1.72	1	0.306	5
Jc3F4 (J)	0.912	2 0	0.920	32	(GA) _n	0.853	15	-	-	-	-	-	-	104	3	2.88	4	0.662	11
Jc3H10 (J)	0.856	1 3	0.922	40	(CT) _n	0.748	19	(CACG) _n	0.476	3	(CGCACA) _n	0.527	4	111	1	0.90	2	0.409	2
QrZAG30 (Q)	0.966	2 3	0.983	31	(GA) _n	0.915	16	-	-	-	-	-	-	185	14	7.57	15	0.922	18

ARLEQUIN versions 3.11 and 3.5.1.2 [34]. Genetic distance matrices for SSR variation and amplicon size variation data were computed with an ad-hoc routine in R [35], and correlations between genetic distance matrices were computed with the `mantel.test` function of the R package “ape”.

(b) At the population level, for *Jacaranda*, global and pairwise F_{ST} values [36] were computed for all sources of variation; R_{ST} values [37] were computed on sources of variability comprising amplicon size variation and SSR variation; N_{ST} values [38] were computed on amplicon and FR sequence variations. In N_{ST} calculations on amplicon sequence variation, mutations in repeat number were weighted less than substitutions (weight = 0.05) to account for their faster mutation rate. Weight choice was based on the observation that SSRs have, in our *Jacaranda* sample, approximately ten times more alleles than each SNP (see Results section). Considering that rare alleles might be missing from our sample, we estimated a ratio of 1:20 in allele richness between SNPs and SSRs and used this ratio to establish the relative weight. A hierarchical analysis of molecular variance (AMOVA) was performed to partition genetic variance into within- and among-population components. The significance of variance components was tested by 1,000 permutations [39]. These analyses were performed with ARLEQUIN 3.11 and 3.5.3.1. Locus-by-locus pairwise F_{ST} , R_{ST} , and N_{ST} matrices were used to construct population-level consensus UPGMA (Unweighted Pair Group Method with Arithmetic Mean) cladograms by averaging genetic distance information of three loci with SplitsTree4 v4.6 [40,41].



Figure 2. Alignment of a subset of DNA fragments at simple sequence repeat (SSR) loci Jc3H10 (*Jacaranda copaia*) (A) and QrZAG30 (*Quercus robur*) (B). Numbers in the top line indicate positions relative to the consensus sequence of (A) 190 bp and (B) 247 bp. Bold nucleotides in brackets indicate SSR motifs and their number of repeats. Dashes indicate gaps, highlighted nucleotides in green (light grey*) indicate indels of one to three bases, highlighted nucleotides in red (dark grey*) mark mutations from one base to another, and yellow (light grey*) boxes indicate groups of insertion/deletions longer than three bases and considered as a single mutational event.* In shades-of-gray printouts.

Results

Distribution of polymorphism

In order to evaluate homoplasy and the variation in phylogenetic signal from different regions of SSR alleles (repeat region and FR), sequence data produced from three divergent tree genera were evaluated. Forty-six, 101, and 47 samples constituted the *Citrus*, *Jacaranda*, and *Quercus* data sets, respectively. Three, three and one SSRs were genotyped and sequenced for the *Citrus*, *Jacaranda* and *Quercus* data sets, respectively. The lengths of the sequenced fragments varied between 123 and 255bp (Table 2). Whenever possible, multiple alleles of the same fragment size were genotyped and sequenced to estimate the degree of sequence divergence among equally sized alleles (“size homoplasy”). This includes both variation in SSR repeat number (that is taken into account by the SMM) and variation in FR sequences, which is the focus of our study.

Levels of polymorphism (number of alleles/haplotypes, Nei’s genetic diversity, number of SNPs and indels) for each amplicon and each source of variation are displayed in Table 3. Large amounts of variability were observed, as expected, in the amplicon sequences. SNPs were found to interrupt the repeat in all three taxa (imperfect SSRs: 33 sequences, data not shown). Sequences with SNPs within the repeat were excluded from subsequent analyses owing to the complexity of the mutation model of SSR repeats containing SNPs. Four loci had two or more repeats (compound SSRs). The highest number of alleles and Nei’s genetic diversity were detected for amplicon sequence variation, then for amplicon size variation and lowest was found for SSR variation (Table 3). FR sequence variation showed the lowest diversity. Indels were observed in six loci out of seven, while SNP density varied between 0.90 and 12.14 SNPs per 100bp.

An example of the complexity of the polymorphisms observed is shown for markers Jc3H10 and QrZAG30 (*Jacaranda* and *Quercus* data set, Figure 2; the complete alignments for all sequenced amplicons are given in Figure S1). Three indels of several nucleotides (between 12 and 15 bases) were detected in the FR sequences of the compound marker Jc3H10. One was upstream of the first SSR, the second was between the repeat motifs and the last was detected after the SSR repeats. In the oak marker QrZAG30, in addition to SNPs, a large 66-nucleotide indel was observed, with SNPs occurring among alleles carrying the DNA fragment involved in the indel. Several cases of size homoplasy were detected, such as: (i) indels in the FRs compensating differences in number of repeats,

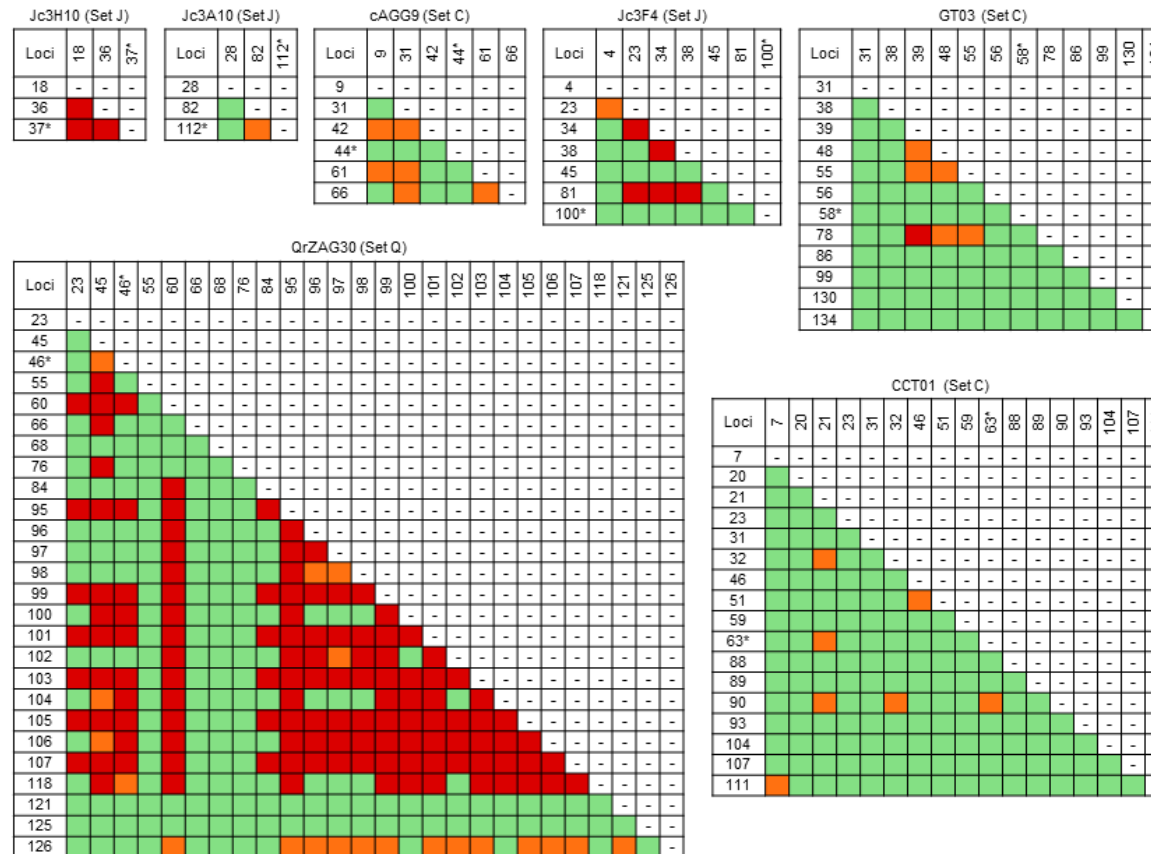


Figure 3. Linkage Disequilibrium (LD) for pairs of polymorphisms in the flanking regions. Dark cells contain LD values significantly different from zero. An asterisk next to a locus name indicates that the SSR repeat(s) is (are) located between that locus and the next. Colouring (shading) indicates the degree of significance of the test: green (light grey*), $P > 0.05$; orange (grey*), $0.05 < P < 0.001$; red (dark grey*), $P < 0.001$. * In shades-of-gray printouts.

(ii) compound SSRs with the same number of repeats, but composed of different numbers of repeats in different motifs, (iii) SSRs with the same amplicon size and number of repeats but with SNPs in the FR sequences or interrupting the repeat (Figure S1). For the three data sets, 50 alleles with the same amplicon sizes out of 68 (74%) showed size homoplasy (Table S3).

Linkage disequilibrium

LD was tested for all pairs of SNP and indel sites within the FR sequence variation of each marker (Figure 3). For the *Citrus* data set, only 19 pairs out of 217 (8.8%) showed significant LD. For the three SSR loci, LD was very irregular and there was no relationship between LD and sequence distance (Mantel tests, $P > 0.05$ for all loci). For the *Jacaranda* data set, ten pairs out of 27 (37%) showed significant LD. For marker Jc3F4, the SNP at position 81 was in disequilibrium with two indels and one SNP, all located upstream of the SSR. LD was strongest for Jc3H10, where all three pairs were significant. The relationship between LD and sequence distance was significant (Mantel test, $P = 0.023$). Tests of correlation between LD and distance in bp could not be performed for Jc3A10 and Jc3H10 because only three nucleotide sites were polymorphic (three LD values). For the *Quercus* data set (marker QrZAG30), 146 pairs of SNPs and indels out of 325 (44.9%) showed significant LD. In particular, 13 or 14 contiguous SNPs and indels located downstream of the SSR at positions 84–107 formed a disequilibrium block, itself in weaker disequilibrium with upstream polymorphic sites at positions 23, 45 and 46 and another SNP downstream the SSR (position 60).

Correlation between genetic distance matrices

We tested whether different sources of variation conveyed the same information on relatedness among alleles. To do so, genetic distances were computed for each pair of individuals with each source of variation (SSR variation, amplicon size, FR sequence, and amplicon sequence variation) within each marker, and the correlation between genetic distance matrices was tested by a Mantel test. Among the 36 pairs of matrices, 27 (75%) showed a significant correlation (Figure S2). Genetic distance matrices based on SSR variation and on FR sequence variation were the least correlated (one significant pair out of five). Amplicon sequence variation and amplicon size variation as well as SSR variation and

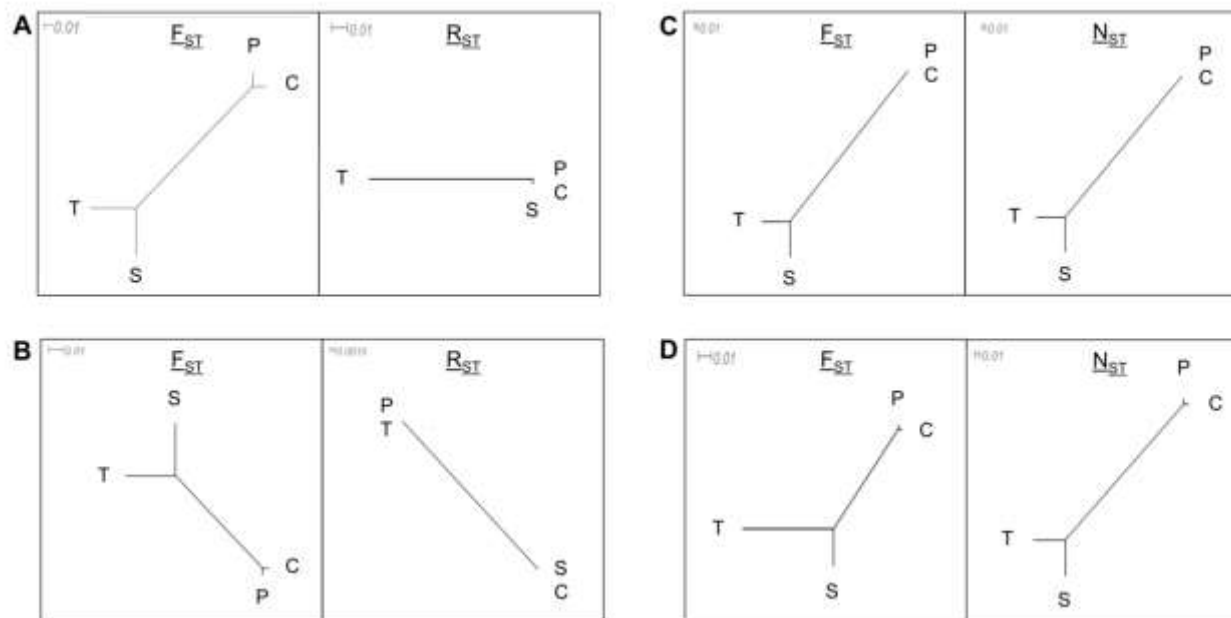


Figure 4. Phylogenetic trees of *Jacaranda* populations based on different components of simple sequence repeat (SSR) data. SSR variation (A), amplicon size variation (B), flanking region (FR) sequence variation (C) and amplicon sequence variation (D). Each type of data was analysed according to a pair of suitable genetic distance estimators: F_{ST} , suitable for loci following the infinite allele model (IAM); R_{ST} , for loci following the stepwise mutation model (SMM); and N_{ST} , for loci following the infinite site model (ISM) model. There are four geographic populations: Counami (C), Paracou (P) and Saint-Laurent (S) in French Guiana; Tapajos (T) in Brazil. Note that scales are not the same among trees.

amplicon size variation, were correlated for all markers. The matrices computed on amplicon sequence variation were correlated to the remaining matrices in 95% of the cases (18 out of 19). Conversely, FR sequence variation alone predicted other matrices in only 53% of the cases.

Population differentiation

For the *Jacaranda* data set, we tested population genetic differentiation for each source of variation by computing F_{ST} , R_{ST} and N_{ST} as appropriate (see Methods; Figure 4). All F_{ST} values were significant. FR sequence variation showed the largest value (0.42) and SSR variation the smallest (0.05). R_{ST} values were very small (<0.02) and non-significant when computed on both amplicon size variation and SSR variation. On the contrary, N_{ST} values were large and significant, with values of 0.50 for FR sequence variation and 0.36 for amplicon sequence variation. In the phylogenetic trees obtained from each combination of source of variation/genetic distance measures, F_{ST} and N_{ST} provided the same topology for all cases they were applied to, indicating a more or less close link between pairs of populations Tapajos/Saint-Laurent and Counami/Paracou (Figure 4). R_{ST} grouped Paracou with Tapajos and Counami with Saint-Laurent (for amplicon size variation) or separated the three Guiana shield populations from Tapajos (for SSR variation; Figure 4).

Discussion

Distribution of polymorphism

SSR polymorphisms are widely used to infer population history and biogeographic patterns. These inferences rely on assumptions about SSR mutational models, and departures from such models are likely to bias divergence and diversity estimates. SSR alleles have been demonstrated to contain other sources of polymorphism in addition to variation in repeat number. In this study we have characterised the contribution of molecular variation occurring outside SSR repeats to SSR marker variability. Whole allele sequencing from SSR markers was analyzed from three tree genera (*Citrus*, *Jacaranda* and *Quercus*). A total of seven (dinucleotide and trinucleotide) SSR markers were targeted, including compound and imperfect markers. More genetic diversity was found for amplicon sequence variation than for amplicon size variation, suggesting that much of data polymorphism is neglected when SSRs are described through amplicon size alone. On average, 7.7 SNPs and 3.4 indels were detected in the flanking sequences of each marker. Genetic diversity (H_e) ranged between 0.306 and 0.922 for the FRs. LD among SNPs and indels in the FR was generally significant. These results demonstrate that SSR alleles are very often riddled with abundant non-SSR indels, plus vast amounts of SNPs both within and near the repeat itself. Similar genetic variation was previously observed at intra- or inter-specific levels [18,42-44] and does not seem to be exceptional in plants. SSR allele mutations are clearly not restricted to the hypervariable tandem repeat region, and polymorphisms outside SSR motifs are also present in animals [22,45,46] and humans [21]. Hidden indels can directly confound the estimation of repeat number from amplicon size (“allele size homoplasy”). While undetected SNPs do not directly influencing repeat number estimates, they do cause haplotypes with different evolutionary histories to be treated as if they were identical. Both forms of homoplasy can have profound consequences on the way data should be interpreted.

Correlation of polymorphism at the individual-allele level

Our results show that pieces of genetic information carried by the different portions of an amplicon are only partially correlated. In particular, blocks of linkage disequilibrium never cover the whole amplicon, and matrices of allele-level pairwise genetic distances built on each source of variation are not always well correlated. In other terms, the different sources of variation are not all equally capable to represent the “true” allele genealogy. Not surprisingly, whole-amplicon sequences were the most parsimonious predictors of

relatedness described by the different parts of the amplicons in all our data sets; conversely, FR sequences alone did not adequately synthesize the information from the whole amplicon sequence. Correlation between genetic distances as obtained from FR sequence variation and from SSR variation was globally weak (Figure S2). This recalls the finding of Payseur and Cutter [25] which stated that coalescence times for SSRs and linked SNPs were uncorrelated. Thus, analysing all parts of SSR amplicons increases the amount of available, phylogenetically independent information, with the additional advantage of providing data from DNA regions with different mutation rates. Nevertheless, the correlation analysis shows that the best predictor of amplicon size is SSR length. This indicates that amplicon size is a reliable first-approximation proxy for SSR repeat length, even though extensive homoplasy blurs the correlation.

Correlation of polymorphism at the population level

The consequences – and usefulness – of hidden sequence variation are shown when we compare how sources of variation perform in detecting population divergence. This analysis was performed in the *Jacaranda* data set comprising four populations. In this example, two alternative patterns were expected for population relatedness: (a) a strictly geographical clustering (the three Guiana shield populations cluster and are separated from the Amazonian population), and (b) a pattern derived from independent results on chloroplast DNA divergence (Counami and Paracou populations form a group, Tapajos and Saint-Laurent form another group (Caroline Scotti-Saintagne, INRA, UMR « Ecologie des forêts de Guyane », article under second round of revision for *Journal of Biogeography*). Our results support the latter hypothesis, as proven by the convergent topology of trees obtained with F_{ST} and N_{ST} independently of the source of variation. The purely geographical hypothesis, on the other hand, is only supported by SSR variation in combination with R_{ST} (although with a non-significant global R_{ST} value). R_{ST} estimates have larger variance than F_{ST} and N_{ST} , particularly when based on small numbers of loci, and R_{ST} values obtained here may be unreliable. Nevertheless, the pattern of R_{ST} differentiation closely follows geographical distance. This may indicate that repeat number variation has arisen locally from small numbers of founding alleles and reflects recent population divergence, while variation in flanking regions may follow larger phylogeographic patterns and reflect deeper divergence. If R_{ST} estimated can be trusted, all sources of variation converge with independent, chloroplast-based results, except SSR variation itself. Therefore, information derived from the “true” nature of SSR data (SSR variation) is actually not representative of the information conveyed by alleles as a whole. This finding strongly hints that other

sources of variability, more correctly described by mutation models that apply to DNA sequences, actually provide the bulk of information carried by SSRs. In other words, phylogeographic information carried by amplicon length is not provided by the SSR (which amplicon length is intended to represent) but by variation in flanking regions. Note that this contrasts with analyses carried out at the individual level (see above), where information on genetic distance match when measured as amplicon size and as SSR size. This issue raises the question of the application of the SMM, or other step-based mutational models, to SSRs, as already discussed by Colson and Goldstein [47] and Ellegren [12]. At the same time, these results substantiate the hypothesis of Cornuet *et al.* [48], who postulated that not all SSRs would evolve under the same mutational model; additional mutational models are required to appropriately handle variation in SSR data produced from different genera. Moreover, variation in the FRs contributed significantly to the phylogenetic signal and sometimes represented the main source of differentiation among individuals and populations, as shown by *Jacaranda* SSR markers and by the *Citrus* SSR marker GT03 [27].

Conclusion

Given the complexity of amplicon sequences described here, one may wonder how to properly exploit SSR data. The weight of polymorphisms other than variation in repeat number cannot and should not be overlooked: identity in allele size does not necessarily indicate identity in sequence content or the number of repeats within same sized alleles. When considering individual alleles, amplicon size correlates well with SSR repeat number. Nevertheless, even if repeat number could be obtained directly, it may provide only weak and inconsistent signals of population genetic differentiation, as shown by the *Jacaranda* data set (but larger numbers of markers may reduce estimation variance and mitigate this problem). Taking sequence variation into account actually adds a significant piece of information to phylogenetic or phylogeographic reconstruction. Population structuring emerges more clearly on the basis of sequence data, and the combination of sequences and SSR variation provides higher resolution. Consequently, genetic distance measures assuming the SMM or related models should be restricted to perfect SSRs with invariable FRs. On the other hand, measures such as F_{ST} (for amplicon size data) and N_{ST} (for amplicon sequences) seem to be the tool of choice for the analysis of the average, not-so-ideal SSR markers that a population geneticist meets in his or her everyday work.

Acknowledgments

We wish to thank Maristerra Lemes for providing a part of the samples, Raphaël Leblois and Caroline Scotti-Saintagne for fruitful discussions and for help with preliminary data analyses, and Valérie Troispoux and Saint-Omer Cazal for technical support.

References

1. Tautz D (1989) Hypervariability of simple sequences as a general source for polymorphic DNA markers. *Nucleic Acids Res* 17: 6465-6471.
2. Takezaki N, Nei M (1996) Genetic distances and reconstruction of phylogenetic trees from microsatellite DNA. *Genetics* 144: 389-399.
3. Rodriguez H, Geistlinger J, Berlyn G, Kahl G, Weising K (2000) Characterization of novel microsatellite loci isolated from the tropical dioecious tree *Simarouba amara*. *Mol Ecol* 9: 489-504.
4. England PR, Usher AV, Whelan RJ, Ayre DJ (2002) Microsatellite diversity and genetic structure of fragmented populations of the rare, fire-dependent shrub *Grevillea macleayana*. *Mol Ecol* 11: 967-977.
5. Taylor DR, Keller SR (2007) Historical range expansion determines the phylogenetic diversity introduced during contemporary species invasion. *Evolution* 61: 335-345.
6. Estoup A, J Wilson IJ, Sullivan C, Cornuet J-M, Moritz C (2001) Inferring population history from microsatellite and enzyme data in serially introduced cane toads, *Bufo marinus*. *Genetics* 159: 1671-1687.
7. Randi E, Tabarroni C, Rimondi S, Lucchini V, Sfougaris A (2003) Phylogeography of the rock partridge (*Alectoris graeca*). *Mol Ecol* 12: 2201-2214.
8. Bonhomme M, Blancher A, Cuartero S, Chikhi L, Crouau-Roy B (2008) Origin and number of founders in an introduced insular primate: estimation from nuclear genetic data. *Mol Ecol* 17: 1009-1019.
9. Reich DE, Goldstein DB (1998) Genetic evidence for a Paleolithic human population expansion in Africa. *Proc Nat Acad Sci USA* 95: 8119-8123.
10. Liu H, Prugnolle F, Manica A, Balloux F (2006) A geographically explicit genetic model of worldwide human-settlement history. *Am J Hum Genet* 79: 230-237.
11. Kimura M, Ohta T (1978) Stepwise mutation model and distribution of allelic frequencies in a finite population. *Proc Nat Acad Sci USA* 75: 2868-2872.
12. Ellegren H (2004) Microsatellites: simple sequences with complex evolution. *Nat Rev Genet* 5: 435-445.
13. Di Rienzo A, Peterson AC, Garza JC, Valdes AM, Slatkin M, et al. (1994) Mutational processes of simple-sequence repeat loci in human populations. *Proc Nat Acad Sci USA* 91: 3166-3170.
14. Zhivotovsky LA, Feldman MW, Grishchkin SA (1997) Biased mutations and microsatellite variation. *Mol Biol Evol* 14: 926-933.
15. Estoup A, Jarne P, Cornuet J-M (2002) Homoplasy and mutation model at microsatellite loci and their consequences for population genetics analysis. *Mol Ecol* 11: 1591-1604.
16. Primmer CR, Ellegren H (1998) Patterns of molecular evolution in avian microsatellites. *Mol Biol Evol* 15: 997-1008.
17. Symonds VV, Lloyd AM (2003) An analysis of microsatellite loci in *Arabidopsis thaliana*: Mutational dynamics and application. *Genetics* 165: 1475-1488.
18. Curtu A-L, Finkeldey R, Gailing O (2004) Comparative sequencing of a microsatellite locus reveals size homoplasy within and between European oak species (*Quercus* spp.). *Plant Mol Biol Rep* 339-346: 22.

19. Stajner N, Jakse J, Kozjak P, Javornik B (2005) The isolation and characterisation of microsatellites in hop (*Humulus lupulus* L.). *Plant Sci* 168: 213-221.
20. Xie H, Sui Y, Chang F-Q, Xu Y, Ma R-C (2006). SSR allelic variation in almond (*Prunus dulcis* Mill.). *Theor Appl Genet* 112: 366-372.
21. Grimaldi M-C, Crouau-Roy B (1997) Microsatellite allelic homoplasy due to variable flanking sequences. *J Mol Evol* 44: 336-340.
22. Orti G, Pearse DE, Avise JC (1997) Phylogenetic assessment of length variation at a microsatellite locus. *Proc Nat Acad Sci USA* 94: 10745-10749.
23. Mogg R, Batley J, Hanley S, Edwards D, O'Sullivan H, et al. (2002) Characterization of the flanking regions of *Zea mays* microsatellites reveals a large number of useful sequence polymorphisms. *Theor Appl Genet* 105: 532-543.
24. Lia VV, Bracco M, Gottlieb AM, Poggio L, Confalonieri VA (2007) Complex mutational patterns and size homoplasy at maize microsatellite loci. *Theor Appl Genet* 115: 981-991.
25. Payseur BA, Cutter AD (2006) Intergrating patterns of polymorphism at SNPs and STRs. *Trends Genet* 22: 424-429.
26. Barkley NA, Roose ML, Krueger RR, Federici CT (2006) Assessing genetic diversity and population structure in a citrus germplasm collection utilizing simple sequence repeat markers (SSRs). *Theor Appl Genet* 112: 1519-1531.
27. Barkley NA, Krueger RR, Federici CT, Roose ML (2009) What phylogeny and gene genealogy analyses reveal about homoplasy in citrus microsatellite alleles. *Plant Syst Evol* 282: 71-86.
28. Gugerli F, Walser J-C, Douvani K, Holderegger R, Finkeldey R (2007) Coincidence of small-scale spatial discontinuities in leaf morphology and nuclear microsatellite variation of *Quercus petraea* and *Q. robur* in a mixed forest. *Ann Bot* 99: 713-722.
29. Gugerli F, Brodbeck S, Holderegger R (2008) Insertions–deletions in a microsatellite flanking region may be resolved by variation in stuttering patterns. *Plant Mol Biol Rep* 26: 255-262.
30. Kijas JMH, Thomas MR, Fowler JCS, Roose ML (1997) Integration of trinucleotide microsatellites into a linkage map of *Citrus*. *Theor Appl Genet* 94: 701-706.
31. Kampfer S, Lexer C, Glössl J, Steinkellner H (1998) Characterization of (GA)_n microsatellite loci from *Quercus robur*. *Hereditas* 129: 183-186.
32. Billotte N, Lagoda PJJ, Risterucci A-M, Baurens F-C (1999) Microsatellite-enriched libraries: applied methodology for the development of SSR markers in tropical crops. *Fruits* 54: 277-288.
33. Colpaert N, Cavers S, Bandou E, Caron H, Gheysen G, et al. (2005) Sampling tissue for DNA analysis of trees: Trunk cambium as an alternative to canopy leaves. *Silvae Genet* 54: 265-269.
34. Excoffier L, Lischer HEL (2010) Arlequin suite ver 3.5: a new series of programs to perform population genetics analyses under Linux and Windows. *Mol Ecol Res* 10: 564-567.
35. R Development Core Team (2011) R: A language and environment for statistical computing. v. 2.14.1 R Foundation for Statistical Computing, Vienna, Austria. Available: <http://www.R-project.org>. Accessed 2012 June 18.
36. Weir BS, Cockerham CC (1984) Estimating F-statistics for the analysis of population structure. *Evolution* 38: 1358-1370.

37. Michalakis Y, Excoffier L (1996) A generic estimation of population subdivision using distances between alleles with special reference for microsatellite loci. *Genetics* 142: 1061-1064.
38. Pons O, Petit RJ (1996) Measuring and testing genetic differentiation with ordered versus unordered alleles. *Genetics* 144: 1237-1245.
39. Excoffier L, Smouse PE, Quattro JM (1992) Analysis of molecular variance inferred from metric distances among DNA haplotypes: Application to human mitochondrial DNA restriction data. *Genetics* 131: 479-491.
40. Huson DH (1998) SplitsTree: analysing and visualizing evolutionary data. *Bioinformatics* 14: 68-73.
41. Huson DH, Bryant D (2006) Application of phylogenetic networks in evolutionary studies. *Mol Bio Evol* 23: 254-267.
42. Peakall R, Gilmore S, Keys W, Morgante M, Rafalski A (1998) Cross-species amplification of Soybean (*Glycine max*) simple sequence repeats (SSRs) within the genus and other legume genera: Implications for the transferability of SSRs in plants. *Mol Biol Evol* 15: 1275-1287.
43. Saha MC, Mian MAR, Eujayl I, Zwonitzer JC, Wang L, et al. (2004) Tall fescue EST-SSR markers with transferability across several grass species. *Theor Appl Genet* 109: 783-791.
44. McDonald MJ, Wang W-C, Huang H-D, Leu J-Y (2011) Clusters of nucleotide substitutions and insertion/deletion mutations are associated with repeat sequences. *PLoS Biol* 9: e1000622. doi: 10.1371/journal.pbio.1000622
45. Brohede J, Ellegren H (1999) Microsatellite evolution: polarity of substitutions within repeats and neutrality of flanking sequences. *Proc R Soc Lond B*, 266: 825–833.
46. Shepherd LD, Lambert DM (2005) Mutational bias in penguin microsatellite DNA. *J Hered* 96: 566-571.
47. Colson I, Goldstein DB (1999) Evidence for complex mutations at microsatellite loci in drosophila. *Genetics* 152: 617–627.
48. Cornuet J-M, Beaumont MA, Estoup A, Solignac M, Carey PD (2006) Inference on microsatellite mutation processes in the invasive mite, *Varroa destructor*, using reversible jump Markov chain Monte Carlo. *Theor Popul Biol* 69: 129-144.

ACCCATTTTATCTACTGATCTGATTTTCAGCCAACAATGATGAAGTTCGGAGCA
ACCCATTTTATCTACTGATCTGATTTTCAGCCAACAATGATGAAGTTCGGAGCA
ACCCATTTTATCTACTGATCTGATTTTCAGCCAACAATGATGAAGTTCGGAGCA
ACCCATTTTATCTACTGATCTGATTTTCAGCCAACAATGATGAAGTTCGGAGCA
ACCCATTTTATCTACTGATCTGATTTTCAGCCAACAATGATGAAGTTCGGAGCA
ACCCATTTTATCTACTGATCTGATTTTCAGCCAACAATGATGAAGTTCGGAGCA
ACCCATTTTATCTACTGATCTGATTTTCAGCCAACAATGATGAAGTTCGGAGCA

CCT01

1225_Cmax_pum_Hun TCAACACCTCGAACAGAAGGTTCTCGTCCAGATTTGTGAAGCCAGTTTCTTCGGATTGCTCTCCCTCTTCCTCCACATCCACCTCCTCCTCTGTGTTGACTAACTTCCAAGTGAAAA
2578_Cmax_pum_FlS TCAACACCTCGAACAGAAGGTTCTCGTCCAGATTTGTGAAGCCAGTTTCTTCGGATTGCTCTCCCTCTTCCTCCACCTCCACCTCCTCCTCTGTGTTGACTAACTTCCAAGTGAAAA
2341F_Cmax_pum_KLY TCAACACCTCGAACAGAAGGTTCTCGTCCAGATTTGTGAAGCCAGTTTCTTCGGATTGCTCTCCCTCTTCCTCCACCTCCACCTCCTCCTCTGTGTTGACTAACTTCCAAGTGAAAA
2355F_Cmax_pum_KaP TCAACACCTCGAACAGAAGGTTCTCGTCCAGATTTGTGAAGCCAGTTTCTTCGGATTGCTCTCCCTCTTCCTCCACCTCCACCTCCTCCTCTGTGTTGACTAACTTCCAAGTGAAAA
3066F_Cmax_pum_Sou TCAACACCTCGAACAGAAGGTTCTCGTCCAGATTTGTGAAGCCAGTTTCTTCGGATTGCTCTCCCTCTTCCTCCACCTCCACCTCCTCCTCTGTGTTGACTAACTTCCAAGTGAAAA
2341R_Cmax_pum_KLY TCAACACCTCGAACAGAAGGTTCTCGTCCAGATTTGTGAAGCCAGTTTCTTCGGATTGCTCTCCCTCTTCCTCCACCTCC---TCCTCCTCCTCTGTGTTGACTAACTTCCAAGTGAAAA
3793F_Csp_pap_Unn TCAACACCTCGAACAGAAGGTTCTCGTCCAGATTTGTGAAGCCAGTTTCTTCGGATTGCTCTCCCTCTTCCTCCACCTCC---TCCTCCTCCTCTGTGTTGACTAACTTCCAAGTGAAAA
3237F_Fjap_kum_Unn TCAACACCTCGAACAGAAGGTTCTCGTCCAGATTTGTGAAGCCAGTTTCTTCGGATTGCTCTCCCTCTTCCTCCACCTCC---TCCTCCTCCTCTGTGTTGACTAACTTCCAAGTGAAAA
448_Cmax_pum_Moa TCAACACCTCGAACAGAAGGTTCCCGTCCAGATTTGTGAAGCCAGTTTCTTCGGATTGCTCTCCCTCTTCCTCCACCTCC---TCCTCCTCCTCTGTGTTGACTAACTTCCAAGTGAAAA
3055F_Cmed_citr_Ben TCAACACCTCGAACAGAAGGTTCTCGTCCAGATTTGTGAAGCCAGTTTCTTCGGATTGCTCTCCCTCTTCCTCCACCTCC---TCCTCCTCCTCTGTGTTGACTAACTTCCAAGTGAAAA
3816F_Cret_man_KiU TCAACACCTCGAACAGAAGGTTCTCGTCCAGATTTGTGAAGCCAGTTTCTTCGGATTGCTCTCCCTCTTCCTCCACCTCC---TCCTCCTCCTCTGTGTTGACTAACTTCCAAGTGAAAA
644F_Cmax_pum_Phi TCAACACCTCGAACAGAAGGTTCTCGTCCAGATTTGTGAAGCCAGTTTCTTCGGATTGCTCTCCCTCTTCCTCCACCTCC---TCCTCCTCTCTGTGTTGACTAACTTCCAAGTGAAAA
3066R_Cmax_pum_Sou TCAACACATCGAACAGAAGGTTCTCGTCCAGATTTGTGAAGCCAGTTTCTTCGGATTGCTCTCCCTCTTCCTCCACCTCC---TCCTCC---TCTGTTGACTAACTTCCAAGTAAAA
3055R_Cmed_cit_Ben TCAACACCTCGAACAGAAGGTTCTCGTCCAGATTTGTGAAGCCAGTTTCTTCGGATTGCTCTCCCTCTTCCTCCACCTCC---TCCTCC---TCTGTTGACTAACTTCCAAGTAAAA
644R_Cmax_pum_Phi TCAACACCTCGAACAGAAGGTTCTCGTCCAGATTTGTGAAGCCAGTTTCTTCGGATTGCTCTCCCTCTTCCTCCACCTCC---ACCTCCTCC---TCTGTTGACTAACTTCCAAGTGAAAA
661_Cmed_cit_Ind TCAACACCTCGAACAGAAGGTTCTCGTCCAGATTTGTGAAGCCAGTTTCTTCGGATTGCTCTCCCTCTTCCTCCACCTCC---TCCTCC---TCTGTTGACTAACTTCCAAGTGAAAA
3237R_Fjap_kum_Unn TCAACACCTCGAACAGAAGGTTCTCGTCCAGATTTGTGAAGCCAGTTTCTTCGGATTGCTCTCCCTCTTCCTCCACCTCC---TCCTTC---TCTGTTGACTAACTTCCAAGTGAAAA
300_Cret_man_Par TCAACACCTCGAACAGAAGGTTCTCGTCCAGATTTGTGAAGCCAGTTTCTTCGGATTGCTCTCCCTCTTCCTCCACCTCC---TCCTCC---TCTGTTGACTAACTTCCAAGTGAAAA
3816R_Cret_man_KiU TCAACACCTCGAACAGAAGGTTCTCGTCCAGTTTGTGAAGCCAGTTTCTTCGGATTGCTCTCCCTCTTCCTCCACCTCC---TCCTCC---TCTGTTGACTAACTTCCAAGTGAAAA
2554_Ptri_tri_Bar TCAACACCTCGAACAGAAGGTTCTCGTCCAGACTTGTGAAGCCAGTTTCTTCGGATTGCTCTCTCTCTTCCTCCACCTCC---TCCTCC---TCTGTTGACTAACTTCCAAGTGAGAA
3056_Csp_pap_Unn TCAACACCTCGAACAGAAGGTTCTCGTCCAGATTTGTGAAGCCAGTTTCTTCGGATTGCTCTCCCTCTTCCTCCACCTCC---TCC-----TCTGTTGACTAACTTCCAAGTGAAAA
3780_Chal_cithyb_Unn TCAACACCTCGAACAGAAGGTTCTCGTCCAGATTTGTGAAGCCAGTTTCTTCGGATTGCTCTCCCTCTTCCTCCACCTCC---TCC-----TCTGTTGACTAACTTCCAAGTGAAAA
3793R_Csp_pap_Unn TCAACACCTCGAACAGAAGGTTCTCGTCCAGATTTGTGAAGCCAGTTTCTTCGGATTGCTCTCCCTCTTCCTCCACCTCC---TCC-----TCTGTTGACTAACTTCCAAGTGAAAA
3797_Csp_pap_Unn TCAACACCTCGAACAGAAGGTTCTCGTCCAGATTTGTGAAGCCAGTTTCTTCGGATTGCTCTCCCTCTTCCTCCACCTCC---TCC-----TCTGTTGACTAACTTCCAAGTGAAAA

TTTTCAAAAACCTCATTGCTGCAATTCCTTGTGCTAGCATGTGGG
TTTTCAAAAACCTCATTGCTGCAATTCCTTGTGCTAGCATGTGGG
TTTTCAAAAACCTCATTGCTGCAATTCCTTGTGCTAGCATGTGGG
TTTTCAAAAACCTCATTGCTGCAATTCCTTGTGCTAGCATGTGGG
TTTTCAAAAACCTCATTGCTGCAATTCCTTGTGCTAGCATGTGGG

-----CACACACACACACA-----TAG
-----CACACACACACACACA-----TAG
-----CACACACACACACACA-----TAG
-----CACACACACACACACACA-----TAG
-----CACACACACACACACA-----TAG
-----CACACACACACACACA-----TAG
-----CACACACACACACACA-----TAG
TCTCTCTCTCTCACACACACACACACA-----TAG
-----CACACACACACACACA-----TAG
-----CACACACACACACACA-----TAG
-----CACACACACACACACA-----TAG
-----CACACACACACACACA-----TAG
-----CACACACACACACACA-----TAG
-----CACACACACACACACA-----TAG
-----CACACACACACACACA-----TAG
-----CACACACACACACACA-----TAG
-----CACACACACACACACA-----TAG
-----CACACACACACACACA-----TAG
-----CACACACACACACACA-----TAG
-----CACACACACACACACA-----TAG
-----CACACACACACACACA-----TAG
-----CACACACACACACACA-----TAG
-----CACACACACACACACA-----TAG
-----CACACACACACACACA-----TAG
-----CACACACACACACACACATAG
-----CACACACTCACACACACACACA-----TAG
-----CACACACACACACACACA-----TAG
-----CACACACACACACA-----TAG
-----CACACACACACACACACA-----TAG
-----CACACACACACACACACA-----TAG
-----CACACACACACACACACA-----TAG
-----CACACACACACACACACA-----TAG
-----CACACACACACACACACA-----TAG
-----CACACACACACACACACA-----TAG
-----CACACACACACACACACA-----TAG
-----CACACACACACACACACA-----TAG
-----CACACACACACACACACATAG
-----CACACACACACACACA-----TAG

TATCCATAATGCTCATTGTGGG-----CAATAGCCACGTATAGGG
TATCCATAATGCTCATTGTGGG-----CAATAGCCACGTATAGGG
TATCCATAATGCTCATTGTGGG-----CAATAGCCACGTATAGG-
TATCCATAATGCTCATTGTGGG-----CAATAGCCACGTATAGGG
TATCCATAATGCTCATTGTGGGATGTT-----ACTTTATCCA-TAATGCTCATTGTGGGCAATAGCCACATATAGGG
TATCCATAATGCTCATTGTGGGATGTT-----ACTTTATCCA-TAATGCTCATTGTGGGCAATAGCCACATATAGGG
TATCCATAATGCTCATTGTGGGATGTT-----ACTTTATCCA-TAATGCTCATTGTGGGCAATAGCCACATATAGGG
TATCCATAATGCTCATTGTGGGATGTT-----ACTTTATCCA-TAATGCTCATTGTGGGCAATAGCCACATATAGGG
TATCCATAATGCTCATTGTGGGATGTT-----ACTTTATCCA-TAATGCTCATTGTGGGCAATAGCCACATATAGGG
TATCCATAATGCTCATTGTGGGATGTT-----ACTTTATCCA-TAATGCTCATTGTGGGCAATAGCCACATATAGGG
TATCCATAATGCTCATTGTGGGATGTT-----ACTTTATCCA-TAATGCTCATTGTGGGCAATAGCCACATATAGGG
TATCCATAATGCTCATTGTGGGATGTT-----ACTTTATCCA-TAATGCTCATTGTGGGCAATAGCCACATATAGGG
TATCCATAATGCTCATTGTGGGATGTT-----ACTTTATCCA-TAATGCTCATTGTGGGCAATAGCCACATATAGGG
TATCCATAATGCTCATTGTGGGATGTT-----ACTTTATCCA-TAATGCTCATTGTGGGCAATAGCCACATATAGGG
TATCCATAATGCTCATTGTGGGATGTT-----ACTTTATCCA-TAATGCTCATTGTGGGCAATAGCCACATATAGGG
TATCCATAATGCTCATTGTGGGATGTT-----ACTTTATCCA-TAATGCTCATTGTGGGCAATAGCCACATATAGGG
TATCCATAATGCTCATTGTGGGATGTT-----ACTTTATCCA-TAATGCTCATTGTGGGCAATAGCCACATATAGGG
TATCCATAATGCTCATTGTGGGATGTTTTTTTTTTTGGATAAAAAAACAACACAGAGAAGCTTT---CATTAAT-CTGATTTGTGGGCAATAGCCACATATAGGG
TATCCATAATGCTCATTGTGGGATGTT-----ACTTTATCCA-TAATGCTCATTGTGGGCAATAGCCACATATAGGG
TATCCATAATGCTCATTGTGGGATGTTTTTTTTTTTGGATAAAAAAACAACACAGAGAAGCTTT---CATTAAT-CTGATTTGTGGGCAATAGCCACATATAGGG

cAGG9 (Set C)					Jc3A10 (Set J)				
	SSR	Amplicon size	FR sequence	Amplicon sequence		SSR	Amplicon size	FR sequence	Amplicon sequence
SSR	-	-	-	-	SSR	-	-	-	-
Amplicon size	3381	-	-	-	Amplicon size	99287	-	-	-
FR sequence	438	1716	-	-	FR sequence	4723	9135	-	-
Amplicon sequence	3976	15234	2114	-	Amplicon sequence	140905	221320	15362	-

CCT01 (Set C)					Jc3F4 (Set J)				
	SSR	Amplicon size	FR sequence	Amplicon sequence		SSR	Amplicon size	FR sequence	Amplicon sequence
SSR	-	-	-	-	SSR	-	-	-	-
Amplicon size	-	-	-	-	Amplicon size	20396	-	-	-
FR sequence	-	1923	-	-	FR sequence	28857	63307	-	-
Amplicon sequence	-	8139	4190	-	Amplicon sequence	208873	437446	79496	-

GT03 (Set C)					Jc3H10 (Set J)				
	SSR	Amplicon size	FR sequence	Amplicon sequence		SSR	Amplicon size	FR sequence	Amplicon sequence
SSR	-	-	-	-	SSR	-	-	-	-
Amplicon size	-	-	-	-	Amplicon size	536254	-	-	-
FR sequence	-	6670	-	-	FR sequence	36187	78765	-	-
Amplicon sequence	-	32022	10612	-	Amplicon sequence	71197314	2418089	326451	-

QrZAG30 (Set Q)				
	SSR	Amplicon size	FR sequence	Amplicon sequence
SSR	-	-	-	-
Amplicon size	131723	-	-	-
FR sequence	106440	474646	-	-
Amplicon sequence	232872	812074	863056	-

Figure S2. Z-statistics and summary of Mantel tests for the correlations of pairs of genetic distance matrices within species. Colouring (shading) indicates the level of significance of the test: green (light grey*), $P > 0.05$; orange (grey*), $0.05 < P < 0.001$; red (dark grey*), $P < 0.001$. Identity of DNA fragment sections is reported: simple sequence repeat (SSR) variation; amplicon size variation, flanking region (FR) sequence variation, and amplicon sequence variation. Marker names are displayed above each panel and data sets refer to *Citrus* (C), *Jacaranda* (J) and *Quercus* (Q). Sequences with single nucleotide polymorphisms (SNPs) within the repeat were excluded from subsequent analyses because of the complexity to describe the mutation model of SSR repeats bracketing SNPs.

Table S1. List and details of the sampled sites for the *Jacaranda* (J) data set

Data Set	Country	Location	Latitude	Longitude
Data J	French Guiana	Counami	5.35°N	52.18°E
	French Guiana	Paracou	5.25°N	52.92°E
	French Guiana	Saint-Laurent	5.49°N	54.04°E
	Brazil	Tapajos	3.07°S	55.58°E

Table S2. Description, amplification conditions and polymorphism of the nuclear simple sequence repeats (SSRs) analysed in *Jacaranda copaia*.

Name	Ta (°C) ^a	Repeat Motif	Size (bp) ^b	Primer sequence	n ^c	GenBank/EMBL accession number
<i>Jc3A10</i>	57°C	(GA)4(CT)8(CA)11	242	F: AATTTACAGCGGCTCCTC R: CCCTTCTCTTTCTCGTGTCG	201	JN661809
<i>Jc3F4</i>	60°C	(GA)16	226	F:CCAGGCAAAGCATCGTAAAT R: GCACATTAATGGCGTCGTC	201	JN661810
<i>Jc3H10</i>	60°C	(TC)18 AT(AC)5	239	F:CGTTCATGGTGTAGAACTTCA R: CCTCAACCAAGACACAGCAA	201	JN661811

^a Annealing temperature^b Total length of the amplicon sequence in base pairs^c Total number of analysed samples

Table S3. List and details of alleles for the three data sets. Amplicon sizes, frequencies and number of associated haplotypes are indicated for each allele.

SSR locus (data set)	Allelic size (bp)	Allelic size frequency	Number of haplotype
cAGG9 (C)	105	7	4
	114	8	2
	117	9	3
CCT01 (C)	158	7	3
	161	12	8
	164	8	6
	167	7	3
GT03 (C)	153	5	3
	167	9	2
	171	10	4
	173	10	2
Jc3A10 (J)	145	4	2
	147	24	3
	149	16	2
	151	17	10
	153	16	7
	155	8	5
	157	2	2
	159	5	5
Jc3F4 (J)	161	2	2
	113	4	2
	114	18	3
	116	2	2
	120	10	3
	122	8	3
	123	8	2
	130	2	2
Jc3H10 (J)	128	18	2
	130	2	2
	134	6	2
	136	2	2
	140	5	3
	141	20	2
	144	2	2
	146	2	2
	148	4	3
	150	5	2
QrZAG30 (Q)	152	2	2
	158	3	2
	168	2	2
	133	4	2
	159	3	2
	163	5	2
	165	3	2
	171	3	2
	173	4	2
	175	2	2
187	3	2	
201	2	2	
219	3	2	

Chapitre III

Impact de l'occupation humaine précolombienne sur la forêt amazonienne

L'approche méthodologique précédente, sur les marqueurs moléculaires microsatellites, m'a permis dans les chapitres suivants d'étudier, de façon la moins biaisée possible, l'évolution démographique passée de la Forêt Tropicale Humide amazonienne. La reconstruction de l'histoire écologique des forêts du plateau des Guyanes nécessite de prendre en compte tous les facteurs naturels ou non pouvant avoir affecté ces communautés à différentes échelles de temps et d'espace par le passé. Dans ce chapitre, je me suis intéressée plus particulièrement à la modification des forêts guyanaises par l'occupation humaine précolombienne durant l'Holocène. Ce domaine d'étude encore largement inexploré soulève de nombreuses interrogations. Comment les populations amérindiennes anciennes occupaient-elles le milieu forestier ? Ont-elles eu un impact assez important sur leur environnement pour laisser une trace encore visible sur les communautés forestières actuelles ? Si les recherches archéologiques apportent les premiers indices indispensables pour répondre à ces questions, elles ne permettent cependant pas d'inférer les changements de structure et la dynamique des peuplements forestiers liés à cette présence anthropique. Ces bases archéologiques ont été utilisées pour essayer de déterminer si les perturbations liées à l'occupation précolombienne avaient influencé la distribution actuelle des espèces au niveau du paysage (Annexe 2) et pouvoir étudier les changements démographiques associés à cette présence humaine pour reconstruire le passé récent de 4 espèces de la forêt guyanaise en utilisant les outils génétiques décrits dans le deuxième chapitre de cette thèse.

Une étude préliminaire de l'impact des perturbations précolombiennes au niveau du paysage actuel (Annexe 2) a été réalisée dans une zone forestière de 9 km² en Guyane française, sur laquelle deux sites structurés de type « montagne couronnée » avaient au préalable été découverts. Des prospections pédestres ont permis de relever des données spatiales, pédologiques, anthracologiques et d'abondance de 5 espèces de palmiers (*Astrocaryum paramaca*, *Astrocaryum sciophilum*, *Euterpe oleraceae*, *Oenocarpus bacaba* et *Oenocarpus bataua*) et de 4 arbres forestiers (*Carapa surinamensis*, *Dicorynia guianensis*, *Virola michelii* et *Vouacapoua americana*) qui ont ensuite été analysées à l'aide de modèles linéaires généralisés. Cette étude a permis de distinguer plusieurs catégories : les espèces sensibles à l'occupation humaine, qui ont été favorisées ou non par cette perturbation, et des espèces insensibles. Il serait également intéressant d'exploiter toutes les données spatialisées récoltées sur cette large zone d'étude avec des approches par système d'information géographique (SIG) qui permettraient de reconstruire les aires de circulation autour des sites occupés afin de délimiter plus finement les zones forestières perturbées.

Dans ce chapitre, l'impact de ces perturbations anthropiques sur la démographie locale passée a été étudié sur 4 espèces communes de la forêt guyanaise (*C. surinamensis*, *D. guianensis*, *V. michelii* et *V. americana*). Pour cela l'intensité de l'occupation précolombienne a été évaluée sur et autour de 5 sites archéologiques, comprenant les 2 montagnes couronnées étudiées précédemment, et l'absence de perturbations humaine a été confirmée sur 4 sites considérés comme témoins. Deux modèles démographiques ont ensuite été testés sur ces sites pour les populations des 4 espèces d'arbres forestiers étudiées. Ces modèles supposent un effet de fondation des populations forestières après l'abandon d'un site par les amérindiens et ainsi une divergence locale par rapport à la population régionale d'origine, ce qui ne devrait pas être le cas des populations de milieux non perturbés. Selon les espèces, entre 4 et 9 marqueurs moléculaires microsatellites ont été utilisés pour leur capacité à garder en mémoire l'histoire récente des populations. Les résultats tirés de l'étude méthodologique, au chapitre précédent, ont permis d'orienter le choix des modèles de mutations de ces marqueurs ainsi que le choix des statistiques descriptives de l'approche ABC utilisée (Excoffier *et al.*, 2000; Beaumont *et al.*, 2002). Cette méthode a permis de modéliser la généalogie des allèles, de dater les événements démographiques de fondation et de tester pour tous les sites étudiés deux scénarios avec et sans divergence entre la population locale et sa population régionale d'origine non structurée. Les sites archéologiques étudiés montrent des intensités différentes d'occupation par l'homme car si peu d'artefacts ont été retrouvés sur certains sites, suggérant plutôt une activité agricole sous couvert forestier, d'autres sont plus fortement marqués par la présence humaine et font souvent partie d'un réseau plus ou moins grand de sites occupés par le passé. Pour tous les sites perturbés ou non et toutes les espèces, le scénario de divergence génétique entre les populations locales et régionales est l'hypothèse la plus supportée. Ce résultat suggère que l'action de perturbations locales, naturelles et aléatoires imposant des fluctuations continues des tailles efficaces des populations ne peut pas être différenciée de la dynamique forestière après une perturbation par l'homme. Cependant, comparé aux sites témoins, la taille efficace des populations fondatrices d'arbres semble être plus importante sur les sites impactés par les perturbations anthropiques alors que les sites faiblement perturbés montrent des populations de plus petite taille.

Ces résultats suggèrent que les occupations humaines les plus intenses ont impacté des zones forestières plus importantes que les trouées occasionnées par des perturbations naturelles. De plus, le nombre d'espèces ayant une plus grande population fondatrice sur les sites archéologiques semble augmenter avec l'intensité d'occupation de ces sites suggérant

qu'une activité anthropique plus intense modifie les processus démographiques d'un plus grand nombre d'espèces et a donc un impact plus sévère sur l'ensemble de la forêt. Même si de nombreuses améliorations sont encore à apporter, la détection de l'impact des occupations humaines précolombiennes ouvre la voie au développement futur de méthodes génétiques permettant de détecter des perturbations passées anthropiques.

Impact of pre-Columbian settlement on the population dynamics of Amazonian trees: a genetic approach

Stéphanie Barthe¹, Mélanie Chazalet¹, Martijn van den Bel², Caroline Scotti-Saintagne³, Laurent Maggia^{4,5}, Céline Cardi⁴, Bruno Hérault^{1,6}, Ivan Scotti²

1 Université des Antilles et de la Guyane, UMR EcoFoG (« Ecologie des forêts de Guyane »), Campus Agronomique, BP 709, 97387 Kourou, French Guiana (France)

2 INRAP, Base archéologique de Guyane, 842 chemin Saint-Antoine, 97300 Cayenne, French Guiana (France)

3 INRA, UMR EcoFoG (« Ecologie des forêts de Guyane »), Campus Agronomique, BP 709, 97387 Kourou, French Guiana (France)

4 CIRAD, UMR AGAP, F-98800 Nouméa, Nouvelle Calédonie (France)

5 IAC, BP 18239, Nouméa Sud, F-98857, New Caledonia (France)

6 CIRAD, UMR EcoFoG (« Ecologie des forêts de Guyane »), Campus Agronomique, BP 709, 97387 Kourou, French Guiana (France)

Article en préparation

Abstract

Past geological events and climatic disturbances impacted the composition of forest ecosystems and the demography of tree species in the Amazonian rainforest. During the Holocene, pre-Columbian human settlements, dating to the Ceramic Age, altered surrounding forests by foraging and cultivation, and are therefore expected to have left their marks on the recent demography of local forest tree populations. By matching archaeological data with estimates of the extent of past demographic disturbances in forest tree populations, we have tested whether it is possible to detect the footprint of human activity on current genetic diversity patterns at the landscape level by indirect modelling methods. To do this, we have applied Approximate Bayesian Computations to genetic marker (Single Sequence Repeat, SSR) data to identify the most likely scenario for local and global population history in four common forest tree species. Recent founder events were detected both at control (undisturbed) and archaeological (disturbed) sites, suggesting that rainforest undergo local, natural population turnover. Disturbance dates were close for control and archaeological sites. Nevertheless, comparison of past population effective sizes suggests that forest patches growing in abandoned settlement sites were founded by larger populations than under natural turn-over, indicating that larger portions of the forest were removed by human activities than by natural processes. In conclusion, natural and human-induced forest turnover processes were similar in nature, but the impact on forest community was stronger for anthropogenic than for natural events. This novel but exploratory multi-disciplinary study could help in the future to develop an accessible method to detect past human disturbance without heavy and expensive analyses. To do this, work is still needed to validate disturbance indexes, such as the effective population size ratio of impacted and not impacted forests. Precautions must be taken because undetectable human impact in natural sites cannot be excluded.

Keywords

Amazonian rainforest, Pre-Columbian settlement, Human disturbances, Holocene, Nuclear microsatellites, Coalescent simulations, Approximate Bayesian Computation, Divergence model

Introduction

Tropical ecosystems have been affected by climatic changes during the glacial phases of the Pleistocene (Colinvaux *et al.*, 2000; Bush *et al.*, 2004; Mayle *et al.*, 2004; Urrego *et al.*, 2009). Even if the nature of these disturbances is controversial (Bush and De Oliveira, 2006; Rull, 2011), they shaped the current Amazonian rainforest and imposed demographic transitions of plant (Caron *et al.*, 2000; Dutech *et al.*, 2003) and animal (Flanagan *et al.*, 2004; Quijada-Mascareñas *et al.*, 2007; Solomon *et al.*, 2008; de Thoisy *et al.*, 2010) populations. In the Holocene, charcoal analyses suggested disturbances, such as paleo-fires, related to drier climate periods (Charles-Dominique *et al.*, 1998; Mayle *et al.*, 2000; Mayle and Power, 2008; Power *et al.*, 2010) combined with pre-Columbian human activities such as foraging, agriculture and the establishment of permanent or transient settlements (Servant and Servant-Vildary, 1996; Tardy, 1998; Bush *et al.*, 2008). Occasional pre-ceramic occupation in French Guiana dates back to about 7000 BP (Mestre and Delpech, 2008) whereas the oldest ceramics and the presence of domesticated plants (*i.e.*, maize and sweet potatoes) date at least back to about 4000 BP (van den Bel *et al.*, 2011; Pagán Jiménez, 2011). Ring-ditched sites as well large black earth (“terra preta”) sites are known from the beginning of the first millennium in both the interior and the coastal fringe. The presence of extensive drained fields in French Guiana for period from AD 900 onwards, which were at first recorded for Suriname (Boomert, 1977, 1978, 1980; Versteeg, 1985), is generally associated with a population increase and the rise of chiefdoms (Rostain, 2008), although recent studies contradict the link between drained-field agriculture and human population expansion (Lombardo *et al.*, 2011). Whatever its amplitude, this surge in human population density suddenly ceased around the twelfth century AD, but its effects on natural forest communities may be long-lasting, as observed in temperate ecosystems (Dupouey *et al.*, 2002). Phylogeographic and demographic studies have shown that past disturbances leave footprints on genetic diversity patterns in tree populations (Ledig *et al.*, 1999; Heuertz *et al.*, 2006; Jaramillo-Correa *et al.*, 2006; Pyhäjärvi *et al.*, 2007; Petit *et al.*, 2008; Wachowiak *et al.*, 2011; Scotti-Saintagne *et al.*, 2012). In particular, the study of genetic signatures of human disturbances in forest tree populations has a long history, dating back at least to Ledig (1992). Neutral and highly polymorphic genetic markers such as microsatellites, with a mutation rate of about 10^{-2} to 10^{-6} nucleotides per generation (Thuillet *et al.*, 2002), offer the opportunity to analyse evolutionary events occurring over evolutionarily recent time

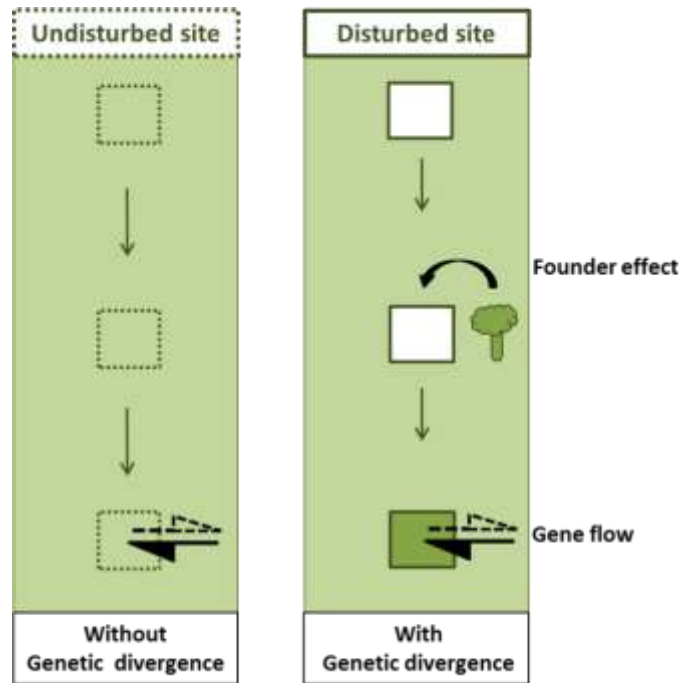


Figure1. Demographic models applied to genetic data from undisturbed and disturbed sites. Divergence models between local patch are represented by a square and its surrounding forest in light green (light grey*). Undisturbed site is a light green (light grey*) square dotted and disturbed site is a white square in bold. Due to large differences in population size, realised gene flow will be probably be much stronger from the forest to the local patch (solid arrow) than the other way round, making patch-to-forest gene flow effectively negligible (dotted arrow).

*In shades-of-gray printouts

spans, but at the same time call for caution in the interpretation of results when the origin of within-locus polymorphism is not known (Barthe *et al.*, 2012). The coalescent process (Kingman, 1982) is a powerful tool to model neutral molecular polymorphism (Nordborg, 2001), to detect demographic changes and to date population divergence. Approximate Bayesian Computation (ABC) (Beaumont *et al.*, 2002) can be used to infer coalescent parameters and choose between complex scenarios (Beaumont, 2010).

In this study, we modelled the impact of pre-Columbian settlements on the forest dynamics as a local extinction-recolonisation process, and asked whether it influenced demography of tree populations. We postulate that forest patches of unknown size were removed by human activities (*e.g.* settlement, agriculture) and those trees recolonised the cleared sites after they were abandoned. Provided that disturbances were large and ancient enough to leave a signature in genetic diversity patterns, then local forest populations may bear the footprint of founding events, which should not appear in the surrounding, undisturbed forests (Figure 1). We therefore tested the hypothesis that tree populations on archaeological sites have undergone different demographic events when compared to historically undisturbed forests. The results of this interdisciplinary study may pave the way to detect the impact of past disturbances in currently undisturbed ecosystems.

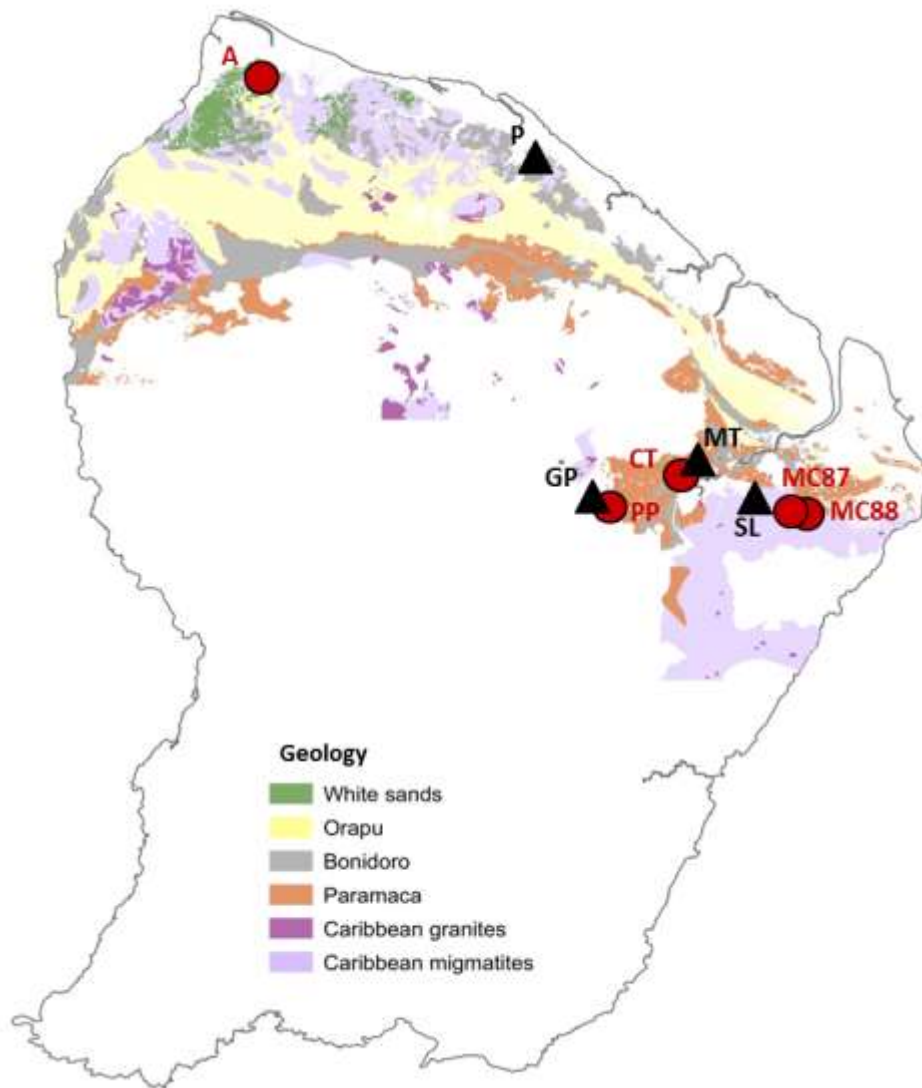


Figure 2. Localisation of studied sites in French Guiana. Archaeological sites disturbed by human settlement are represented by red circles and undisturbed control sites by black triangles. Sample ID are indicated for each sites and details are given Table 1.

Material and Methods

Demographic model

The demographic model applied here is depicted in Figure 1. After the abandonment of a human settlement (*i.e.* an identifiable archaeological site) by human populations, a patch of land becomes available for forest regeneration. The new tree populations growing in the patch will have higher chances to be founded by the nearest trees than by distant trees, and will be the outcome of local founder events. The new forest patch will not be homogeneous with the larger gene pool represented by the ‘surrounding forest’. After recolonisation, gene flow occurring between the new patch and the surrounding forest will tend to erase divergence over time. On the contrary, at undisturbed sites we can expect to observe no genetic divergence between local population and the surrounding forest; the forest growing at any undisturbed sampled site is just part of the larger continuous forest and should not have undergone any founder effect.

Choice of sites and species; sampling

Five archaeological sites disturbed by human settlement and four undisturbed sites, each one closer than 30 Km to at least one archaeological site, except for the site Paracou, and in a similar geological and topographical context, taken as control sites (Figure 2 and details in Table 1). Archaeological sites were ranked according to the intensity of past human disturbance determined by the pedological analysis, the presence and quantity of artefacts, their spatial extent and the presence or absence of human development as ditches on each site. Manual surveys of 1.5 by 1.5m were performed to the number of 10 for the site MC87, 2 for MC88 and 5 for A. On the site CT, 6 trenches of 3.5m wide and 1.7m deep were opened using a mechanical shovel and no survey was conducted for the site PP.

Four common tropical trees were studied: *Carapa surinamensis* (Miq.) (Meliaceae), *Dicorynia guianensis* (Amshoff) (Caesalpiniaceae), *Virola michelii* (Heckel) (Miristicaceae) and *Vouacapoua americana* (Aubl.) (Caesalpiniaceae). These species have a moderate to strong preference for colonising open habitats (Molino and Sabatier, 2001; Vincent *et al.*, 2011), but at the same time they are persistent components of mature forests. They were therefore chosen as ‘markers’ because they can react to the availability of forest gaps, but –unlike pure pioneer species – they do not disappear after forest cover has been reconstituted.

Table 1. List and details of the sampled sites for the four studied species. *Carapa surinamensis* (*Cs*), *Dicorynia guianensis* (*Dg*), *Virola michelii* (*Vm*) and *Vouacapoua americana* (*Va*).

Site ID ^a	Site name	Site state ^b	Latitude	Longitude	N ^c			
					<i>Cs</i>	<i>Dg</i>	<i>Vm</i>	<i>Va</i>
A	Acarouany	S	5°32'N	53°48'W	25	48	52	36
P	Paracou	C (A)	5°16'N	52°55'W	30	157	28	24
CT	Crique Tortue	S	4°11'N	52°26'W	37	57	29	78
MT	Montagne Tortue	C (CT)	4°12'N	52°26'W	14	29	30	12
MC87	Montagne couronnée 87	S	4°03'N	52°05'W	53	30	55	29
MC88	Montagne couronnée 88	S	4°03'N	52°04'W	1	52	18	43
SL	Saut Lavilette	C (MC87-88)	4°06'N	52°12'W	33	30	24	25
PP	Nouragues Petit Plateau	S	4°05'N	52°41'W	28	33	12	27
GP	Nouragues Grand Plateau	C (PP)	4°05'N	52°40'W	65	281	21	273

^a Sample ID used for genetic analyses

^b S (Settled site: archaeological site disturbed by human settlement), C (Control site: undisturbed site).
Settled sites corresponding control sites are shown in brackets.

^c Number of genotyped samples per species

Samples of cambium of 10 to 50 individuals per species were collected at each archaeological site. Individuals of each tree species were collected, first on the settled site and then around it in the area considered to have been affected by the human disturbance. For control sites, an equivalent number of samples were collected on areas roughly equivalent to the size of disturbed sites (for *D. guianensis* at the Paracou site, the data from Latouche-Hallé *et al.* (2003) were used). Between 90 to 230 further samples per species were collected throughout French Guiana for *C. surinamensis*, *D. guianensis* and *V. michelii* to obtain a regional genetic pool that was used to identify the ‘surrounding forest’ (Figure 1) in the demographic model (for *V. americana*, the regional pool was represented by the data from Dutech *et al.* (2003)).

Molecular methods

For all samples, total genomic DNA extraction were performed by following Colpaert *et al.* (2005) or with the Qiagen DNeasy® 2006 protocol with a DNeasy 96 plant kit (Qiagen, Hilden, Germany) according to the manufacturer’s protocol. DNA quality was checked by spectrophotometry or by agarose gel electrophoresis.

Nuclear microsatellite markers (nSSR) were selected as follows (Table S1): five loci for *C. surinamensis* (Dayanandan *et al.*, 1999; Vinson *et al.*, 2005) and *D. guianensis* (Latouche-Hallé *et al.*, 2003); seven markers for *V. michelii* (Draheim *et al.*, 2009); seven markers for *V. americana*, of which five drawn from Dutech *et al.* (2000), and two (Va2G5 and Va3G10) obtained following the same methodology as in Barthe *et al.* (2012). Markers, primer sequences, annealing temperatures and GenBank/EMBL accession numbers are shown in Table S2. The four *C. surinamensis* markers: Cg1L, Cg5, Cg7 and Cg17, were amplified according the protocol of the Type-it® multiplex-PCR kit (Qiagen, Hilden, Germany). PCRs for *C. surinamensis* marker Cg01 and for *D. guianensis*, *V. michelii* and *V. americana* markers were performed respectively in 10, 12.5, 15 and 12.5 µL volumes containing 4, 6.5, 3 and 6.5 µL of 50-fold diluted DNA, 1x *Taq* buffer, 0.1 mM of dNTP, 0.04 U/µL *Taq* DNA polymerase (all products from Invitrogen, Carlsbad, CA) and 0.2µM of each primer (MWG Biotech, Ebersberg, Germany). For *V. michelii* SSRs, 0.7mM MgCl₂ was added. PCR programs were as follows. For *V. michelii* markers: initial denaturation at 94°C for 5 min, 35 cycles of [30 s at 94°C, 40 s at 52°C], 1 min at 72°C and a final extension at 72°C for 10 min. Thermal cycle for *C. surinamensis* markers Cg1L, Cg5, Cg7 and Cg17 followed the Type-it® protocol (see above). For marker Cg01 (*C. surinamensis*) and for *D. guianensis* and *V. americana* SSRs: a denaturation step at

Table 2. Statistic values and significance of the Kendall's test between genetic and geographical matrices every ten kilometers for each species.

Distance (km)	<i>C. surinamensis.</i>	<i>D. guianensis</i>	<i>V. michelii</i>	<i>V. americana</i>
90	-0.57	373	77	274
100	-0.39	493	81	439
110	-0.18	596	141	515
120	-0.05	2.10*	182	698
130	0.75	2.48*	206	2.46*
140	1.35	2.84**	317	2.07*
150	1.79	3.00**	384	2.21*
160	1.93	3.00**	493	2.92**
170	2.11*	2.96**	555	2.83**
180	1.91	2.79**	-0.53	2.93**
190	1.45	2.80**	-0.08	3.27***
200	1.54	3.03**	0.46	3.17***

Significance levels: * p<0.05; ** p<0.01; *** p<0.001

94°C for 5 min, 35 cycles (*C. surinamensis*) or 30 cycles (*D. guianensis* and *V. americana*) of [45 s at 94°C, 45 s at T_a (Table 2), 45 s at 72°C], final extension at 72°C for 10 min. Forward PCR primers were fluorescently labelled (with PET, 6FAM, NED and VIC). PCR fragments were separated on an ABI 3130XL capillary sequencer (Applied Biosystems, Foster City, CA) using ABI POP4 and Applied Biosystems LIZ-500 as internal standard, following the manufacturer's instructions. Genotypes were scored using the GENEMAPPER 4.0 software package (Applied Biosystems, Foster City, CA).

Data analyses

nSSR Bayesian clustering and isolation-by-distance analysis of background populations

Because *C. surinamensis* belongs to a species complex in which the species are difficult to discriminate morphologically, we applied the methods of Duminil *et al.* (2006) and Scotti-Saintagne *et al.* (2012), which rely on Bayesian clustering method implemented in STRUCTURE 2.3.1 (Pritchard *et al.*, 2000), to exclude from our data set all individuals not corresponding to the 'pure' *C. surinamensis* genetic cluster.

The coalescent model requires that populations are panmictic. Therefore, we analysed population structure at the regional scale to identify the most appropriate panmictic genetic cluster, to be used as 'surrounding forest' population in comparisons with disturbed and control sites (Figure 1). To do this, we proceeded as follows. We tested Isolation-by-distance (IBD) to determine the geographical scale at which significant population IBD could be detected: matrices of pairwise genetic and geographic Euclidean distances were computed using GENEPOP 4 (Rousset, 2008), and IBD was checked on increasingly larger geographical ranges, by steps of 10 Km (i.e. for ranges 0-50 Km, 0-60 Km, ... up to the maximum pairwise population distance). Correlation between genetic and geographical matrices was tested with a Kendall test (Hollander & Wolfe, 1973). The largest range not showing any significant IBD was recorded for each species. Half this range was interpreted as the distance defining the panmictic neighbourhood of a given population (that is, all populations within this radius from a given population belong to the same panmictic group, because their maximum pairwise distance falls within the threshold of detection of significant IBD). The 'surrounding forest' population for each site was then defined by all samples collected within its panmictic neighbourhood radius.

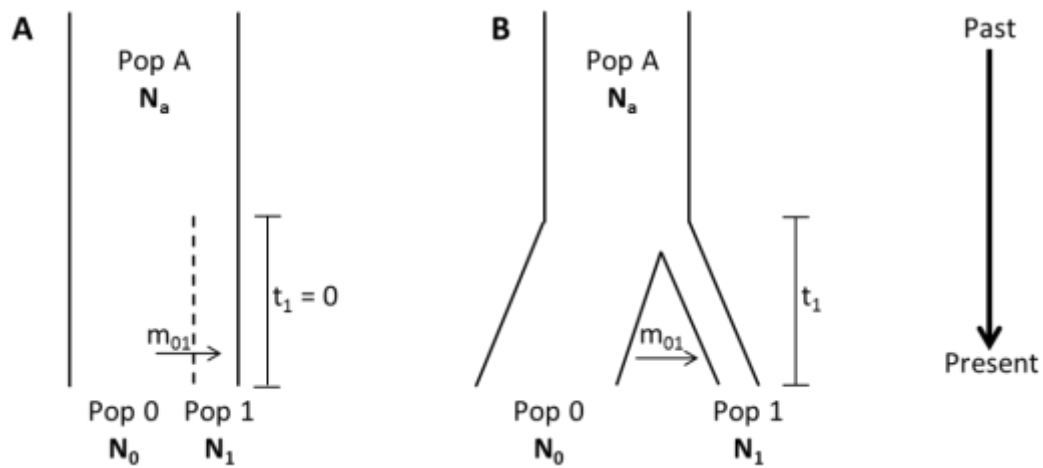


Figure 3. Mathematical modelling and population parameters of the two tested hypotheses. Divergence models between background (Pop 0) and local (Pop 1) populations from an ancestral population (Pop A). (A) Background hypothesis (H_0) without genetic divergence. (B) Disturbance hypothesis (H_i) with founder effect and genetic differentiation between the two current populations. Population parameters are effective population sizes (N_0 , N_1 and N_a), gene-flow from background to local population (m) and times of population splitting (t_i).

Coalescent modelling

Approximate Bayesian Computation (for recent reviews see Bertorelle *et al.*, 2010; Csilléry *et al.*, 2010) is commonly used to track demographic history of populations and estimate population parameters (Hudson, 1991; Nordborg, 2001). We performed coalescent simulations with SIMCOAL2 software (Excoffier *et al.*, 2000; Laval & Excoffier, 2004). Data for seven unlinked nuclear microsatellites were generated under a generalized stepwise mutation model (GSM) with different mutation rates between loci. Two models with gene flow were tested with the ABC approach, the background hypothesis (H_0) without genetic divergence between local and background populations and the disturbance hypothesis (H_1) with divergence and founder effect of the local population by few individuals from the background population (Figure 3), corresponding respectively to the ‘Undisturbed site’ and ‘Disturbed site’ scenarios described in Figure 1. Populations of 30 alleles (or 12 alleles when fewer than 15 genotypes were available), were modelled. Priors were chosen as follows: effective size of background population (N_0) was sampled from a Gamma (mean= 1.2×10^5 ; shape= 2×10^5), effective size of local population (N_1) from a Uniform[2, N_0], mutation rates (mu) per generation from a Uniform[10^{-8} , $\theta/4N_0$] with θ the population diversity parameter and $\theta/4N_0$ between 10^{-8} and 10^{-2} , migration rate from background to local population (m_{01}) from exponential distribution with rate = 10^5 ; time of population splitting (t_i) in generations from a Gamma distribution with a first quartile between 0 and 30 000 (for the disturbance hypothesis H_1) or fixed at 0 (for the background hypothesis H_0). We performed 1,000,000 simulations with values randomly drawn from the priors. To choose the most informative summary statistics, we tested the correlation between several summary statistics and population parameters. The following statistics were chosen: (i) population-level heterozygosity H_e per locus, (ii) population-level allele richness A per locus and (iii) population level demographic test g (Reich & Goldstein, 1998) per locus. These summary statistics were calculated using R routines (R Development Core Team, 2011) for simulated data and for observed data sampling of 15 or 6 individuals.

Calculations of posterior distributions of population parameters were performed on the 1% or 10,000 best scenarios using the ABCEST2 software (Excoffier *et al.*, 2005). To identify the most plausible ABC scenario (H_0 vs H_1) for each site, we applied the *direct* approach of Cornuet *et al.* (2008) based on the proportion of datasets obtained for each model in the closest n_8 simulations. To test differences in the distributions of the distances generated by the two scenarios, we used the non-parametric rank test of

Table 3. Tests of Mann-Whitney between the two hypotheses on each site and F_{ST} values between each site and its background population for the four studied species.

Site ID	Site state ^a	<i>C. surinamensis</i>			<i>D. guianensis</i>			<i>V. michelii</i>			<i>V. americana</i>		
		W ^b	Means ^c	F_{ST}	W ^b	Means ^c	F_{ST}	W ^b	Means ^c	F_{ST}	W ^b	Means ^c	F_{ST}
A	S	1E04***	$H_0 > H_1$	0.049	1E04***	$H_0 > H_1$	0.000	1E04***	$H_0 > H_1$	0.029	1E04***	$H_0 > H_1$	-
P	C	1E04***	$H_0 > H_1$	-0.010	1E04***	$H_0 > H_1$	-	1E04***	$H_0 > H_1$	0.013	1E04***	$H_0 > H_1$	0.018
CT	S	1E04***	$H_0 > H_1$	0.003	1E04***	$H_0 > H_1$	0.017	1E04***	$H_0 > H_1$	0.021	1E04***	$H_0 > H_1$	-
MT	C	6.6E03***	$H_0 > H_1$	0.003	1E04***	$H_0 > H_1$	0.028	1E04***	$H_0 > H_1$	0.021	8.1E03***	$H_0 > H_1$	-
MC87	S	1E04***	$H_0 > H_1$	0.051	1E04***	$H_0 > H_1$	0.018	1E04***	$H_0 > H_1$	0.020	1E04***	$H_0 > H_1$	-
MC88	S	-	-	-	1E04***	$H_0 > H_1$	0.040	1E04***	$H_0 > H_1$	0.019	1E04***	$H_0 > H_1$	-
SL	C	1E04***	$H_0 > H_1$	-0.015	1E04***	$H_0 > H_1$	0.036	8.5E03***	$H_0 > H_1$	0.068	1E04***	$H_0 > H_1$	-
PP	S	1E04***	$H_0 > H_1$	0.045	1E04***	$H_0 > H_1$	0.081	7.1E03***	$H_0 > H_1$	0.046	1E04***	$H_0 > H_1$	-
GP	C	1E04***	$H_0 > H_1$	-0.013	1E04***	$H_0 > H_1$	0.065	8.1E03***	$H_0 > H_1$	0.057	1E04***	$H_0 > H_1$	-

^a S (Settled site: archeological site disturbed by human settlement), C (Control site: Undisturbed site)

^b Wilcoxon's statistic. Significance levels: * $p < 0.05$; ** $p < 0.01$; *** $p < 0.001$

^c Results between the first 100 Euclidean distance means of background (H_0) and disturbance (H_1) hypotheses

Mann-Whitney (Hollander & Wolfe, 1973) to compare Euclidean distance means of the 100 closest simulations to the observed data. Median posterior values were provided by ABCEST2 by weighted linear regression and were transformed using a $\log(\tan)$ transformation to keep posteriors within the boundaries of the priors. Parameter mode values were calculated using a quantile classification creating classes of roughly the same number of individuals. We chose a quantile of 0.01 corresponding to 100 simulations. Difference between effective population sizes on disturbed and control sites were tested by a Mann-Whitney's test. To quantify the impact of past human disturbance, we looked at the relationship between counts of population pairs having larger, identical and smaller founder population effective size, on one hand, and disturbance intensity levels, defined as above, on the other hand. The site Acarouany (A) was removed from this analysis because it was compared with a relatively remote control site, thus possibly biasing the comparison with other sites.

Table 4. Mann-Whitney tests between local effective sizes of disturbed sites by human settlement and undisturbed control sites.

Pair of S-C sites ^a	<i>C. surinamensis</i>		<i>D. guianensis</i>		<i>V. michelii</i>		<i>V. americana</i>	
	W ^b	N difference ^c	W ^b	N difference ^c	W ^b	N difference ^c	W ^b	N difference ^c
A-P	3869.5 ***	N _s <N _c	21258 ***	N _s <N _c	2693369 ***	N _s >N _c	29397 ***	N _s >N _c
PP-GP	9577 ***	N _s >N _c	8864.5 ***	N _s <N _c	21750 ns	N _s =N _c	18085 ***	N _s <N _c
CT-MT	47.5 ***	N _s <N _c	41961 ***	N _s >N _c	1331 ***	N _s <N _c	13 ***	N _s <N _c
MC88-SL	-	-	22047 ns	N _s =N _c	5388 ns	N _s =N _c	20107 ns	N _s =N _c
MC87-SL	5581 ns	N _s =N _c	21912 ns	N _s =N _c	8234 ***	N _s >N _c	23780 **	N _s >N _c

^a Pair of Settled and corresponding Control sites

^b Wilcoxon's statistic. Significance levels: * p<0.05; ** p<0.01; *** p<0.001

^c Effective size difference of local populations on settled (N_s) and control (N_c) sites

Results

Archaeological surveys

The Acarouany site (A), revealed a small concentration of artefacts upon a large plateau of six hectares. Combined archaeological and pedological data suggest that this plateau was cultivated and had a very small Late Ceramic Age habitat located at the plateau edge (due to settlement differences compared to the other sites and distance from its control site, this site was not included in the semi-quantitative analysis of human disturbances). Crique Tortue (CT) has been identified as a Late Ceramic habitation site with an important black layer (*terra preta*) and large quantities of ceramic and lithic material, without evidence of a major settlement. The Petit Plateau (PP) site resembles the CT site and is also identified as a hilltop habitation site. Studies of Tardy *et al.* (1998) and Migeon *et al.* (2005) attested to the presence of other sites around the PP. The very large number of fragments or whole potteries, as well as soil enrichment in organic and mineral elements, found on MC87 and MC88, shows an intense settlement at these two sites. The intensity of the impact at these sites is further confirmed by the presence of two to three metre-deep circular ditches. In addition, our pedestrian surveys revealed a larger network of sites probably interrelated by “roads”, which likely increased forest disturbance around the settlements. Based on these results, sites can be ranked from less to more intensely disturbed in the following order: A, PP and CT, MC88 and MC87.

Background genetic structure

Spatial background structure was studied for each species by inspecting Isolation-by-Distance (IBD) patterns. Results of Kendall tests, presented Table 2, allowed the identification of distance thresholds for the absence of IBD at 160 km for *C. surinamensis*, 110 km for *D. guianensis* and 120 km for *V. americana*. No IBD was detected for *V. michelii*. Background genetic pools were defined for each species and each population based on these distances.

Model choice, parameter estimation and comparison

For all sites (both control and disturbed) and all species, the H_1 (disturbance) scenario was preferred over H_0 according to Mann-Whitney tests (Table 3). Divergence times (t_i) between local and background populations were close to zero generations in all cases with very large confidence intervals. Migration rates (m_{oi}) ranged between $8 \cdot 10^{-07}$ and $1 \cdot 10^{-06}$ migrants per generation and per allele (Table S3).

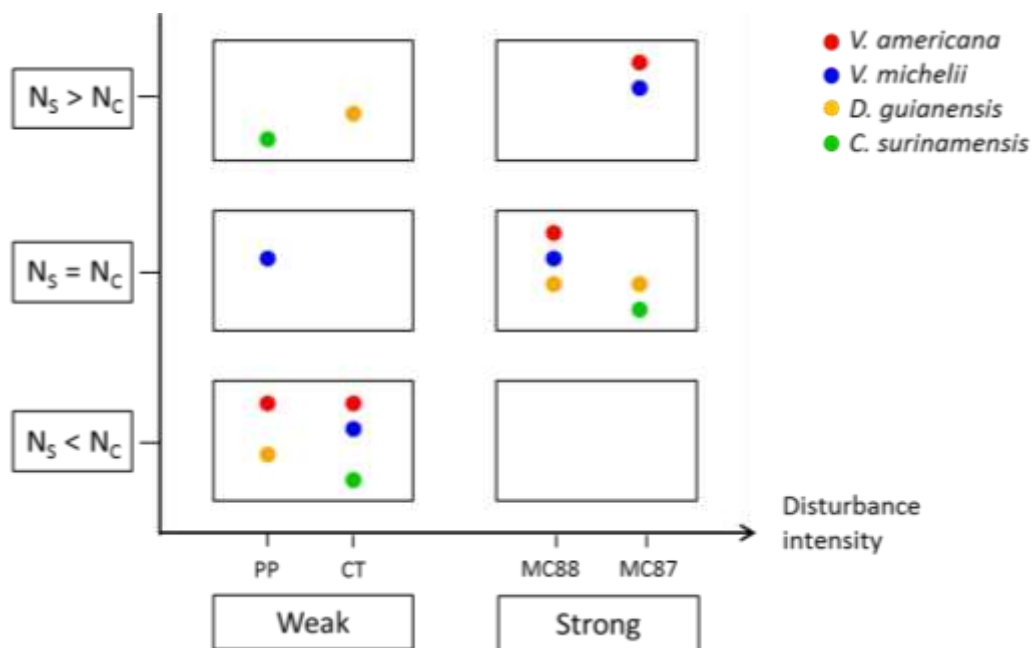


Figure 4. Effective size difference of local populations on settled (N_s) and control (N_c) sites as a function of intensity of pre-Columbian human impact.

However, migration rate posterior distributions fitted closely prior distributions, indicating poor estimation power.

Local and background effective population size estimates are shown in Table S3 and ranged between 10 and 50 individuals for *C. surinamensis*, between 50 and 100 for *D. guianensis*, between 90 and 700 for *V. michelii* and between 30 and 100 individuals for *V. Americana*, all confound sites. Effective founding population sizes at archaeological sites and corresponding control sites were compared based on the shape of posterior distributions. Six population pairs out of 19 (31.5%) had larger effective population sizes at disturbed sites than at corresponding control sites, six pairs (31.5%) showed no difference, and seven pairs (37%) had smaller effective population sizes at disturbed sites than at corresponding control sites (Table 4). At heavy-impact sites, such as MC87 and MC88, effective founder population size was higher than, or equal to, control site population size for all species; conversely, effective founder population size at the lightest-impact sites (Petit Plateau (PP) and Crique Tortue (CT)), was mainly smaller than at the corresponding control site (Figure 4). *V.michelii* showed larger or equal founder effective size for disturbed than for control populations for four site-pairs out of five, and smaller founder effective size for disturbed than for control populations in one case. The three remaining species all have two pairs of sites with smaller population sizes for disturbed sites. *C. surinamensis* and *D. guianensis* showed only one pair of sites with larger population sizes for disturbed sites and *V. americana* two pairs of sites.

Discussion

The demography of wild populations living in natural ecosystems is expected to be affected by the disturbances caused by human activity. Whereas the consequences of contemporary disturbances can be followed directly, the direct and indirect impacts of human settlements having occurred in past times are hard to evaluate. Here, we have applied analytical methods permitting to model past demographical processes occurring in natural populations to infer the changes undergone by forest ecosystems in association with the presence of pre-Columbian human settlements. The combination of coalescent models and ABC allows efficient modelling of past demographic population history, as shown by studies on plant (François *et al.*, 2008; Ingvarsson, 2008; Burgarella *et al.*, 2012), animal (Legrand *et al.*, 2009; Wegmann and Excoffier, 2010; Lye *et al.*, 2011; Thalmann *et al.*, 2011; Purcell *et al.*, 2012) and human populations (Laval *et al.*, 2010; Ray *et al.*, 2010; Batini *et al.*, 2011; Veeramah *et al.*, 2012). The ideas of inferring past habitat or ecosystem changes from the demography of relevant species (Quéméré *et al.*, 2012) or to link past demographic changes to human activity by inference methods (Salmona *et al.*, 2012) have been applied before; in this study, we introduce the possibility to follow demographic change at a geographically restricted scale (*i.e.* landscape level).

Natural rainforest population dynamics

All undisturbed control sites of all studied species show genetic divergence from the surrounding background population. This could be explained by the effect of natural local disturbances, inducing random, recurrent fluctuations in population effective sizes and possibly recurrent population turnover with meta-population dynamics (Hanski and Gaggiotti, 2004). Such recurrent founder effects are consistent with patch dynamics observed for *D. guianensis* (Kokou, 1994) and *V. americana* (Traissac, 2003). Although *C. surinamensis* and *V. michelii* are not strictly aggregative species, they may exhibit aggregated distribution under special conditions, such as intense waterlogging for *C. surinamensis* (Collinet, 1997; Forget, 1997; Traissac, 2003; Dessard *et al.*, 2004) or environmental disturbances for *V. michelii* (Dessard *et al.*, 2004). Our results clearly indicate a local and natural turn-over of the rainforest. The recent, but not very precisely estimated, divergence times can therefore be explained by recent colonization events due to the continuous dynamics in time and space for the studied species. Estimates of divergence times are very dependent on migration rate estimates, but gene flow was poorly estimated here. This may explain the great inaccuracy of the divergence time estimate.

Anthropogenic impact on rainforest population dynamics

The disturbance hypothesis also emerged as the best model to explain the observed diversity on archaeological sites for all species. Moreover, estimates of time since population divergence, which are close to zero for control and disturbed sites alike, reveals that pre-Columbian human disturbances cannot be told apart from natural turn-over or that these disturbances are too recent to be detected at the tree population dynamic level. In other words, even after pre-Columbian human settlement, the nature of local rainforest dynamics is the same as in undisturbed forest. Nonetheless, founder population effective size at settled sites tends to be higher than, or equal to, undisturbed sites, suggesting that forest patches growing in abandoned settlements are founded by larger population than in natural turn-over. Moreover, the number of species showing larger or equal founding populations at archaeological sites increases with the strength of signs of human occupation (Figure 4) while population sizes are smaller on lightest-impact site. This suggests that low-intensity human disturbance have a less important or more localized impact on populations compared to a natural turn-over that change the entire tree community. Conversely, stronger anthropogenic disturbances have a larger impact on demographic processes in the forest as a whole.

Species sensitivity to disturbance

V. michelii is the species showing the highest number of disturbed sites with larger founding populations than at control sites. This may indicate that species showing an aggregated distribution only under disturbed conditions, such as *V. michelii* (Dessard *et al.*, 2004), are good indicator species of past human settlement, although *C. surinamensis*, which also has an occasionally aggregated nature dependent upon disturbance (Dessard *et al.*, 2004), does not show demographic effects as clear as *V. michelii*. Strictly aggregative species (*D. guianensis* and *V. americana*), do not consistently detect disturbances, probably because their natural patch expansion dynamics, resting on recurrent dramatic founder events, intrinsically tends to decrease effective population sizes.

Limitations of the study

Although the scenarios that we have modelled are as simple as possible, thus making the smallest possible set of assumptions on actual forest dynamics, lack of comprehensive knowledge of medium- to long-term forest dynamics at the landscape and regional scale

prevented us from precisely modelling our system. This is evident in the fact that estimation of absolute values for several parameters is rather poor, and in some cases dependent on other parameters, although they are theoretically independent. This is particularly the case of background (regional) population effective size, which tends to depend upon the estimation of effective size of the local population it is modelled with (Table S4). Nevertheless, this should not affect arguments based on the comparison of estimates of disturbed and control local founder population sizes, which are the main evidence used in this study. These limitations prevent us from comparing archeological dating of human settlement and genetic dating of impacted tree population divergence. In this perspective, it will be interesting to refine site recolonisation dating, for example by testing more refined ABC models with different fixed divergence times or by using full Bayesian methods.

Finally, the effect of individual species properties on their behaviour as indicators of historical disturbance may be further studied, and the methods presented here may be applied to a larger panel of species, to obtain a more complete picture of the impact on demography of forest disturbances.

Conclusion

This fully interdisciplinary work shows the power of the application of a combined archaeological and genetic strategy to the characterisation of human-induced impact on natural ecosystems. Thanks to the analytical power conveyed by molecular markers and statistical modelling methods, it was possible to infer – despite the limits of this pioneering study – trends of demographic disturbance in natural tree populations in association with the presence of human populations. These results open the way to the definition of quantitative or semi-quantitative indicators of historical ecosystem disturbance at the landscape level. Remarkably, such indicators may be independent of historical data, and rely only on present-day patterns of genetic diversity, analogously to indicators based on present-day plant community composition (Dupouey *et al.*, 2002).

We found that after pre-Columbian human disturbances, rainforest underwent the same kind of local dynamic than under natural turn-over. In this respect, Ceramic age human populations of the Guiana shield may be said not to have had too large an impact on their ecosystem. Nevertheless, larger or longer-lasting settlements caused more intense disturbances, the footprints of which are still visible today in patterns of genetic variation in wild tree populations, indicating that even relatively modest human activity can leave permanent environmental marks.

Acknowledgments

We wish to thank Cyril Dutech, Henri Caron, William Montaigne, Maxime Casalis, Luc Allard for providing a part of the data, Valérie Troispoux, Saint-Omer Cazal and the entire team of the Couac project for technical support.

Funding

Financial support to the COUAC project was provided by the CNRS programme Amazonie. SB's Doctoral program has been co-funded by the European Social Fund and by the INRA "research package" awarded to Christopher Baraloto. Sampling was partially funded by the 2011 Nouragues Research Grants program. The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

Author contribution

SB, BH and IS conceived the experiments. MV, BH prospected on archeological sites. LM and CC isolated SSR markers for *V. Americana*. SB, MC obtained SSR genotypes. SB, CS, MC and IS analysed the data. SB, MV, BH and IS wrote the manuscript.

References

- Barthe, S., F. Gugerli, N. A. Barkley, L. Maggia, C. Cardi and I. Scotti (2012). Always look on both sides: Phylogenetic information conveyed by Simple Sequence Repeat allele sequences. *PLoS One* **7**: e40699.
- Batini, C., J. Lopes, D. M. Behar, F. Calafell, L. B. Jorde, L. Van der Veen, L. Quintana-Murci, G. Spedini, G. Destro-Bisol and D. Comas (2011). Insights into the demographic history of African pygmies from complete mitochondrial genomes. *Molecular Biology and Evolution* **28**: 1099-1110.
- Beaumont, M. A. (2010). Approximate Bayesian Computation in Evolution and Ecology. *Annual Review of Ecology, Evolution and Systematics* **41**: 379-405.
- Beaumont, M. A., W. Zhang and D. J. Balding (2002). Approximate Bayesian Computation in Population Genetics. *Genetics* **162**: 2025-2035.
- Bertorelle, G., A. Benazzo and S. Mona (2010). ABC as a flexible framework to estimate demography over space and time: some cons, many pros. *Molecular Ecology* **19**: 2609-2625.
- Boomert, A. (1977). Prehistorie. *Encyclopedie van Suriname*. C. F. A. Bruijning and J. Voorhoev. Amsterdam, Elsevier: 506-517.
- Boomert, A. (1978). Prehistoric habitation mounds in the Canje River area? *Journal of Archaeology and Anthropology* **1**: 44-51.
- Boomert, A. (1980). Hertenrits: an Arauquinoid complex in North West Suriname. *Journal of Archaeology and Anthropology* **3**: 68-104.
- Burgarella, C., M. Navascués, M. Zabal-Aguirre, E. Berganzo, M. Riba, M. Mayol, G. G. Vendramin and S. C. Gonzales-Martinez (2012). Recent population decline and selection shape diversity of taxol-related genes. *Molecular Ecology* **21**: 3006-3021.
- Bush, M. B. and P. E. De Oliveira (2006). The rise and fall of the Refugial Hypothesis of Amazonian Speciation: a paleo-ecological perspective. *Biota Neotropica* **6**.
- Bush, M. B., M. R. Silman, C. Mc Michael and S. S. (2008). Fire, climate change and biodiversity in Amazonia: a Late-Holocene perspective. *Philosophical Transactions of the Royal Society B* **363**: 1795-1802.
- Bush, M. B., M. R. Silman and D. H. Urrego (2004). 48,000 years of climate and forest change in a biodiversity hot spot. *Science* **303**: 827-829.
- Caron, H., S. Dumas, G. Marque, C. Messier, E. Bandou, R. J. Petit and A. Kremer (2000). Spatial and temporal distribution of chloroplast DNA polymorphism in a tropical tree species. *Molecular Ecology* **9**: 1089-1098.
- Charles-Dominique, P., P. Blanc, D. Larpin, M.-P. Ledru, B. RiCra, C. Sarthou, M. Servant and C. Tardy (1998). Forest perturbations and biodiversity during the last ten thousand years in French Guiana *Acta Oecologica* **19**: 295-302.
- Colinvaux, P. A., P. E. De Oliveira and M. B. Bush (2000). Amazonian and neotropical plant communities on glacial time-scales: The failure of the aridity and refuge hypotheses. *Quaternary Science Reviews* **19**: 141-169.
- Collinet, F. (1997). *Essai de regroupements des principales espèces structurantes d'une forêt dense humide d'après l'analyse de leur répartition spatiale*. Thèse de Doctorat, Claude Bernard, Lyon I, France, 203.

- Colpaert, N., S. Cavers, E. Bandou, H. Caron, G. Gheysen and A. J. Lowe (2005). Sampling tissue for DNA analysis of trees: Trunk cambium as an alternative to canopy leaves. *Silvae Genetica* **54**: 265-269.
- Cornuet, J.-M., F. Santos, M. A. Beaumont, C. P. Robert, J.-M. Marin, D. J. Balding, T. Guillemaud and A. Estoup (2008). Inferring population history with DIY ABC: a user-friendly approach to approximate Bayesian computation. *Bioinformatics* **24**: 2713-2719.
- Csilléry, K., M. G. B. Blum, O. E. Gaggiotti and O. François (2010). Approximate Bayesian Computation (ABC) in practice. *Trends in Ecology & Evolution* **25**: 410-418.
- Dayanandan, S., J. Dole, K. Bawa and R. Kesseli (1999). Population structure delineated with microsatellite markers in fragmented populations of a tropical tree, *Carapa guianensis* (Meliaceae). *Molecular Ecology* **8**: 1585-1592.
- de Thoisy, B., A. G. Da Silva, M. Ruiz-Garcia, A. Tapia, O. Ramirez, M. Arana, V. Quse, C. Paz-y-Mino, M. Tobler, C. Pedraza and A. Lavergne (2010). Population history, phylogeography, and conservation genetics of the last Neotropical mega-herbivore, the lowland tapir (*Tapirus terrestris*). *BMC Evolutionary Biology* **10**: 278.
- Dessard, H., N. Picard, R. Pélissier and F. Collinet-Vautier (2004). Spatial patterns of the most abundant tree species. *Ecology and Management of a Neotropical Rainforest 1*: 177-190.
- Draheim, H., M. Cui and C. W. Dick (2009). Characterization of 14 microsatellite DNA markers for the tropical forest tree *Virola surinamensis* (Rol.) Warb. (Myristicaceae). *Molecular Ecology Resources*.
- Duminil, J., H. Caron, I. Scotti, S.-O. Cazal and R. J. Petit (2006). Blind population genetics survey of tropical rainforest trees. *Molecular Ecology* **15**: 3505-3513.
- Dupouey, J. L., E. Dambrine, J. D. Laffite and C. Moares (2002). Irreversible impact of past land use on forest soils and biodiversity. *Ecology* **83**: 2978-2984.
- Dutech, C., L. Amsellem, N. Billotte and P. Jarne (2000). Characterization of (GA)n microsatellite loci using an enrichment protocol in the neotropical tree species *Vouacapoua americana*. *Molecular Ecology* **9**: 1433-1449.
- Dutech, C., L. Maggia, C. Tardy, H. I. Joly and P. Jarne (2003). Tracking a genetic signal expansion of extinction-recolonization events in a neotropical tree species: *Vouacapoua americana* Aublet in French Guiana. *Evolution* **57**: 2753-2764.
- Excoffier, L., A. Estoup and J.-M. Cornuet (2005). Bayesian analysis of an admixture model with mutations and arbitrarily linked markers. *Genetics* **169**: 1727-1738.
- Excoffier, L., J. Novembre and S. Schneider (2000). SIMCOAL: a general coalescent program for the simulation of molecular data in interconnected populations with arbitrary demography. *The Journal of Heredity* **91**: 506-509.
- Flanagan, N. S., A. Tobler, A. Davison, O. G. Pybus, D. D. Kapan, S. Planas, M. Linares, D. Heckel and W. O. McMillan (2004). Historical demography of Müllerian mimicry in the neotropical *Heliconius* butterflies. *Proceedings of the National Academy of Sciences* **101**: 9704-9709.
- Forget, P.-M. (1997). Effect of microhabitat on seed fate and seedling performance in two rodent-dispersed tree species in rain forest in French Guiana. *Journal of Ecology* **85**: 693-703.
- François, O., M. G. B. Blum, M. Jakobsson and N. A. Rosenberg (2008). Demographic history of European populations of *Arabidopsis thaliana*. *PLoS Genetics* **4**.

- Hanski, I. and O. E. Gaggiotti (2004). *Ecology, genetics, and evolution of metapopulations*. Elsevier Academic Press, London.
- Heuertz, M., E. De Paoli, T. Kallman, H. Larsson, I. Jurman, M. Morgante, M. Lascoux and N. Gyllenstrand (2006). Multilocus patterns of nucleotide diversity, linkage disequilibrium and demographic history of Norway Spruce [*Picea abies* (L.) Karst]. *Genetics* **174**: 2095-2105.
- Hollander, M. and D. A. Wolfe (1973). *Nonparametric Statistical Methods*. John Wiley & Sons, New York.
- Hudson, R. R. (1991). Gene genealogies and the coalescent process. *Oxford surveys in Evolutionary Biology* **7**: 1-44.
- Ingvarsson, K. (2008). Multilocus patterns of nucleotide polymorphism and the demographic history of *Populus tremula*. *Genetics society of America* **180**: 329-340.
- Jaramillo-Correa, J. P., J. Beaulieu, F. T. Ledig and J. Bousquet (2006). Decoupled mitochondrial and chloroplast DNA population structure reveals Holocene collapse and population isolation in a threatened Mexican-endemic conifer. *Molecular Ecology* **15**: 2787-2800.
- Kingman, J. F. C. (1982). On the genealogy of large populations. *Journal of Applied Probability* **19**: 27-43.
- Kokou, K. (1994). Spatial evolution of *Angelica* aggregates (*Dicorynia guianensis*, Caesalpiniaceae) in the study device "natural forest" of Paracou in French Guiana. *Acta botanica gallica* **141**: 351-359.
- Latouche-Halle, C., A. Ramboer, E. Bandou, H. Caron and A. Kremer (2003). Nuclear and chloroplast genetic structure indicate fine-scale spatial dynamics in a neotropical tree population. *Heredity* **91**: 181-190.
- Laval, G. and L. Excoffier (2004). SIMCOAL 2.0: a program to simulate genomic diversity over large recombining regions in a subdivided population with a complex history. *Bioinformatics* **20**: 2485-2487.
- Laval, G., E. Patin, L. B. Barreiro and L. Quintana-Murci (2010). Formulating a historical and demographic model of recent human evolution based on resequencing data from noncoding regions. *PLoS One* **5**: e10284.
- Ledig, F. T. (1992). Human impacts on genetic diversity in forest ecosystems. *Oikos* **63**: 87-108.
- Ledig, F. T., M. Thompson Conkle, B. Bermejo-Velazquez, T. Eguiluz-Piedra, P. D. Hodgskiss, D. R. Johnson and W. S. Dvorak (1999). Evidence for an extreme bottleneck in a rare Mexican Pinyon: genetic diversity, disequilibrium, and the mating system in *Pinus maximartinezii*. *Evolution* **53**: 91-99.
- Legrand, D., M. I. Tenaillon, P. Matyot, J. Gerlach, D. Lachaise and M.-L. Cariou (2009). Species-wide genetic variation and demographic history of *Drosophila sechellia*, a species lacking population structure. *Genetics Society of America* **182**: 1197-1206.
- Lombardo, U., E. Canal-Beeby, S. Fehr and H. Veit (2011). Raised fields in the Bolivian Amazonia: a prehistoric green revolution or a flood risk mitigation strategy? *Journal of Archaeological Science* **38**: 502-512.
- Lye, G. C., O. Lepais and D. Goulson (2011). Reconstructing demographic events from population genetic data: the introduction of bumblebees to New Zealand. *Molecular Ecology* **20**: 2888-2900.

- Mayle, F. E., D. J. Beerling, W. D. Gosling and M. B. Bush (2004). Responses of Amazonian ecosystems to climatic and atmospheric carbon dioxide changes since the last glacial maximum. *Philosophical transactions of the royal society Lond. B* **359**: 499-514.
- Mayle, F. E., R. Burbridge and T. J. Killeen (2000). Millennial-scale dynamics of southern Amazonian rain forests. *Science* **290**: 2291-2294.
- Mayle, F. E. and M. J. Power (2008). Impact of a drier early–mid-Holocene climate upon Amazonian forests. *Philosophical transactions of the royal society B* **363**: 1829-1838.
- Mestre, M. and S. Delpech (2008). Saint-Laurent du Maroni. Plateau des Mines. *Rapport de fouille INRAP*.
- Migeon, G., M. Chevolut, G. Lemaire, M. Canolle, J. Lescaudey, M.-P. Lambert and M. Labat (2005). Rapport de prospection et de sondages archéologiques. Nouragues, Régina, Guyane. Cayenne, Guyane, DRAC, Service Régional de l'Archéologie.
- Molino, J.-F. and D. Sabatier (2001). Tree diversity in tropical rain forests: a validation of the intermediate disturbance hypothesis. *Science* **294**: 1702-1704.
- Nordborg, M. (2001). *Coalescent theory. Handbook of statistical genetics*. John Wiley & Sons, New York.
- Pagán Jiménez, J. R. (2011). Early phytocultural processes in the pre-Colonial Antilles. *Essays in archaeology, ethnohistory & ethnography of the Amerindian Circum-Caribbean*. C. Hofman and A. van Duijvenbode. Sidestone Press, Leiden
- Petit, R. J., F. S. Hu and C. W. Dick (2008). Forests of the past: a window to future changes. *Science* **320**: 1450-1452.
- Power, M. J., M. B. Bush, H. Behling, S. P. Horn, F. E. Mayle and D. H. Urrego (2010). Paleofire activity in tropical America during the last 21 ka: A regional synthesis based on sedimentary charcoal *Pages news* **18**: 73-75.
- Pritchard, J. K., M. Stephens and P. Donnelly (2000). Inference of population structure using multilocus genotype data. *Genetics* **155**: 945-959.
- Purcell, K. M., N. Ling and C. A. Stockwell (2012). Evaluation of the introduction history and genetic diversity of a serially introduced fish population in New Zealand. *Biological invasions*.
- Pyhäjärvi, T., M. R. Garcia-Gil, T. Knurr, M. Mikkonen, W. Wachowiak and O. Savolainen (2007). Demographic History Has Influenced Nucleotide Diversity in European *Pinus sylvestris* Populations. *Genetics* **177**: 1713-1724.
- Quéméré, E., X. Amelot, J. Pierson, B. Crouau-Roy and L. Chikhi (2012). Genetic data suggest a natural prehuman origin of open habitats in northern Madagascar and question the deforestation narrative in this region. *Proceedings of the National Academy of Sciences* **109**: 13028-13033.
- Quijada-Mascarenas, J. A., J. E. Ferguson, C. E. Pook, M. Da Graça Salomao, R. S. Thorpe and W. Wüster (2007). Phylogeographic patterns of trans-Amazonian vicariants and Amazonian biogeography: the Neotropical rattlesnake (*Crotalus durissus* complex) as an example. *Journal of Biogeography* **34**: 1296-1312.
- R Development Core Team (2011) R: A language and environment for statistical computing. v. 2.14.1 R Foundation for Statistical Computing, Vienna, Austria. Available: <http://www.R-project.org>.
- Ray, N., D. Wegmann, N. J. R. Fagundes, S. Wang, A. Ruiz-Linares and L. Excoffier (2010). A statistical evaluation of models for the initial settlement of the American

- continent emphasizes the importance of gene flow with Asia. *Molecular Biology and Evolution* **27**: 337-345.
- Reich, D. E. and D. B. Goldstein (1998). Genetic evidence for a Paleolithic human population expansion in Africa. *Proceedings of the National Academy of Sciences of the USA* **95**: 8119-8123.
- Rostain, S. (2008). The archaeology of the Guianas: an overview. *Handbook of South American Archaeology*. H. Silverman and W. H. Isbell. Springer, New York
- Rousset, F. (2008). Genepop'007: a complete re-implementation of the Genepop software for Windows and Linux. *Molecular Ecology Resources* **8**: 103-106.
- Rull, V. (2011). Origins of biodiversity. *Science* **331**: 398-399.
- Salmona, J., M. Salamolard, D. Fouillot, T. Ghestemme, J. Larose, J.-F. O. Centon, V. Sousa, D. A. Dawson, C. Thebaud and L. Chikhi (2012). Signature of a pre-Human population decline in the critically endangered Reunion island endemic forest bird *Coracina newtoni*. *PLoS One* **7**: e43524.
- Scotti-Saintagne, C., C. W. Dick, H. Caron, G. G. Vendramin, E. Guichoux, A. Buonanici, C. Duret, P. Sire, R. Valencia, M. R. Lemes, R. Gribel and I. Scotti (2012). Phylogeography of a species complex of lowland Neotropical rain forest trees (*Carapa*, Meliaceae). *Journal of Biogeography*: 10.1111/j.1365-2699.2011.02678.x.
- Servant, M. and S. Servant-Vildary (1996). *Dynamique à long terme des écosystèmes forestiers intertropicaux*, Paris.
- Solomon, S. E., M. J. Bacci, J. J. Martins, G. G. Vinha and U. G. Mueller (2008). Paleodistributions and comparative molecular phylogeography of leafcutter ants (*Atta* spp.) provide new insight into the origins of Amazonian diversity. *PLoS One* **3**: e2738.
- Tardy, C. (1998). *Paléoincendies naturels, feux anthropiques et environnements forestiers de Guyane française du tardiglaciaire à l'holocène récent. Approches chronologique et anthracologique*. Thèse de Doctorat, Sciences et Techniques, Montpellier II, France, 343.
- Thalmann, O., D. Wegmann, M. Spitzner, M. Arandjelovic, K. Guschanski, C. Leuenberger, R. A. Bergl and L. Vigilant (2011). Historical sampling reveals dramatic demographic changes in western gorilla populations. *BMC Evolutionary Biology* **11**: 85.
- Thuillet, A.-C., D. Bru, J. David, P. Roumet, S. Santoni, P. Sourdille and T. Bataillon (2002). Direct Estimation of Mutation Rate for 10 Microsatellite Loci in Durum Wheat, *Triticum turgidum* (L.) Thell. ssp *durum* desf. *Molecular Biology and Evolution* **19**: 122-125.
- Traissac, S. (2003). *Dynamique spatiale de Vouacapoua americana (Aublet), arbre de forêt tropicale humide à répartition agrégée* Thèse de Doctorat, Claude Bernard, Lyon I, France, 221.
- Urrego, D. H., M. B. Bush, M. R. Silman, A. Correa-Metrio, M.-P. Ledru, F. E. Mayle, G. Paduano and B. G. Valencia (2009). Millennial-scale ecological changes in tropical South America since the Last Glacial Maximum. *Past climate variability in South America and surrounding regions*. F. Vimeux,
- van den Bel, M., N. Serrand, L. Bruxelles, P. Verdin, J. R. Pagán Jiménez, J. Brancier, C. Cammas, J. Birk, B. Glazer and S. Knippenberg (2011). Chemin Saint-Louis. *Rapport de fouille Inrap*.

- Veeramah, K. R., D. Wegmann, A. Woerner, F. L. Mendez, J. C. Watkins, G. Destro-Bisol, H. Soodyall, L. Louie and M. F. Hammer (2012). An early divergence of KhoeSan ancestors from those of other modern Humans is supported by an ABC-based analysis of autosomal resequencing data. *Molecular Biology and Evolution* **29**: 617-630.
- Versteeg, A. H. (1985). *The Prehistory of the Young Coastal Plain of West Suriname*. PhD Dissertation, University of Leiden, 35: 653-750.
- Vincent, G., J.-F. Molino, L. Marescot, K. Barkaoui, D. Sabatier, V. Freycon and J. Roelens (2011). The relative importance of dispersal limitation and habitat preference in shaping spatial distribution of saplings in a tropical moist forest: a case study along a combination of hydromorphic and canopy disturbance gradients. *Annals of Forest Science* **68**: 357-370.
- Vinson, C. C., V. C. R. Azevedo, I. Sampaio and A. Y. Ciampi (2005). Development of microsatellite markers for *Carapa guianensis* (Aublet), a tree species from the Amazon forest. *Molecular Ecology Notes* **5**: 33-34.
- Wachowiak, W., M. J. Salmela, R. A. Ennos, G. Iason and S. Cavers (2011). High genetic diversity at the extreme range edge: nucleotide variation at nuclear loci in Scots pine (*Pinus sylvestris* L.) in Scotland. *Heredity* **106**: 775-787.
- Wegmann, D. and L. Excoffier (2010). Bayesian inference of the demographic history of chimpanzees. *Molecular Biology and Evolution* **27**: 1425-1435.

Supporting Information

Table S1. Description, amplification conditions of two nuclear simple sequence repeats (SSRs) analyzed in *Vouacapoua americana*.

Species	SSR locus	Size (bp) ^a	Reference
Cs	Cg1L	156-160	Vinson (2005)
	Cg01	190-198	
	Cg17	106-122	Dayanandan <i>et al.</i> (1999)
	Cg5	202	
	Cg7	214-241	
Dg	Dg1	269	Latouche-Hallé <i>et al.</i> (2003)
	Dg3	252	
	Dg4	192	
	Dg5	122	
	Dg6	268	
Vm	Vmul2-65	158-192	Draheim <i>et al.</i> (2009)
	Vmul2-35	285-303	
	Vsur34	283-305	
	Vsur58	245-271	
	Vmul68	271-289	
	Vseb3	224-230	
	Vsur56	221-237	
Va	Wac1	240-300	Dutech <i>et al.</i> (2000)
	Wac3	240-260	
	Wac5	160-260	
	Wac13	140-150	This study
	Wac14	220-280	
	Va2G5	190-220	
	Va3G10	218-230	

^a Total length of the amplicon sequence in base pairs

Table S2. Characteristics of simple sequence repeat (SSR) data sets of the four species. *Carapa sp.* (Cs), *Dicorynia guianensis* (Dg), *Virola michelii* (Vm), *Vouacapoua americana* (Va).

Marker Name	Ta (°C) ^a	Pattern	Size (bp) ^b	Primer sequence	GenBank/EMBL accession number
Va2G5	59	(TC) ₁₈	190-220	F: ACGCAAAACGCAAGATTTTC R: ACCCTCACCTTCACCACATC	JX421698
Va3G10	59	(GT) ₁₄	218-230	F: GGGTCACCTCGGATACGATA R: GTAACCTCGCCCAGCCCTAA	JX413522

^a Annealing temperature

^b Total length of the amplicon sequence in base pairs

Table S3. Most probable values of population parameters for each site of the four studied species.

Site ID	Site state ^a	<i>C. surinamensis</i>			<i>D. guianensis</i>			<i>V. michelii</i>			<i>V. americana</i>		
		N_0 ^b	N_1 ^c	m_{01} ^d	N_0 ^b	N_1 ^c	m_{01} ^d	N_0 ^b	N_1 ^c	m_{01} ^d	N_0 ^b	N_1 ^c	m_{01} ^d
A	S	24	9	1.50E-07	99	63	1.85E-06	417	213	1.30E-06	58	68	1.15E-06
P	C	23	12	1.50E-07	99	83	1.00E-06	347	133	3.00E-07	86	53	4.00E-07
CT	S	12	6	6.50E-07	99	94	1.00E-06	346	135	3.00E-07	86	26	2.50E-07
MT	C	49	39	2.50E-07	95	66	1.00E-06	663	282	3.00E-07	113	81	8.00E-07
MC87	S	19	10	1.50E-07	95	70	3.50E-07	634	343	3.00E-07	62	70	3.00E-07
MC88	S	-	-	-	95	70	3.50E-07	356	325	1.30E-06	100	59	1.40E-06
SL	C	18	9	1.35E-06	96	70	2.00E-07	104	87	3.50E-07	97	58	3.00E-07
PP	S	19	11	1.50E-07	84	51	7.50E-07	114	90	3.50E-07	92	59	3.00E-07
GP	C	17	8	1.50E-07	93	68	3.50E-07	104	86	7.50E-07	92	81	3.00E-07

^a S (Settled site: archeological site disturbed by human settlement). C (Control site: Undisturbed site)

^b Effective size of background population

^c Effective size of local population

^d Migration rate from background to local population

Table S4. Mann-Whitney tests between background population effective sizes estimated with data sets including settled or undisturbed control sites.

Pair of S-C sites ^a	<i>C. surinamensis</i>		<i>D. guianensis</i>		<i>V. michelii</i>		<i>V. americana</i>	
	W ^b	N ₀ difference ^c	W ^b	N ₀ difference ^c	W ^b	N ₀ difference ^c	W ^b	N ₀ difference ^c
A-P	6749 **	N _{0s} >N _{0c}	19045 ns	N _{0s} =N _{0c}	2474881 ***	N _{0s} >N _{0c}	6949 ***	N _{0s} <N _{0c}
PP-GP	6996 **	N _{0s} >N _{0c}	13240 ***	N _{0s} <N _{0c}	22561 ***	N _{0s} >N _{0c}	18053 ns	N _{0s} =N _{0c}
CT-MT	347 ***	N _{0s} <N _{0c}	22274 ns	N _{0s} =N _{0c}	982094 ***	N _{0s} <N _{0c}	10476 ***	N _{0s} <N _{0c}
MC88-SL	-	-	20661 ns	N _{0s} =N _{0c}	242422 ***	N _{0s} >N _{0c}	20505 ns	N _{0s} =N _{0c}
MC87-SL	5486 ns	N _{0s} =N _{0c}	21035 ns	N _{0s} =N _{0c}	384002 ***	N _{0s} >N _{0c}	5061 ***	N _{0s} <N _{0c}

^a Pair of Settled and corresponding Control sites

^b Wilcoxon's statistic. Significance levels: * p<0.05; ** p<0.01; *** p<0.001

^c Effective size difference of background populations estimated with data sets of settled (N_{0s}) and control (N_{0c}) sites

Background population sizes, as estimated from the same background population data plus either a disturbed population or corresponding control sites, should theoretically be the same, but showed some inconsistencies. Among the 19 site-pairs, 12 (63%) showed a significant difference between the two estimates of background population effective sizes (six showing a higher estimate when analysing disturbed site data, and six when analysing control data). However, background effective size estimates were found in the same order of magnitude for each species whatever the local site used with which they were analysed (Table 3).

Chapitre IV

Histoire écologique des forêts tropicales humides au Quaternaire

Si des perturbations récentes, engendrées notamment par l'occupation amérindienne passée étudiée dans le chapitre précédent, ont pu influencer la dynamique locale des populations forestières, d'autres changements plus anciens et probablement plus importants ont pu moduler la distribution des forêts du plateau des Guyanes à plus large échelle spatiale. Les dernières grandes perturbations passées connues à ce jour sont les cycles glaciaires dont le dernier maximum glaciaire remonte à 2.5 Ma, caractérisé en Amazonie par une diminution de quelques degrés des températures et une diminution de la pluviométrie. L'objectif de ce chapitre a donc été de savoir si ces changements climatiques ont affecté la démographie des populations actuellement présentes, dans l'objectif de contribuer à reconstruire l'histoire écologique des forêts en Guyane française pendant le Quaternaire. Une étude de démographie comparée a donc été menée au niveau régional à partir des patrons actuels de diversité génétique de plusieurs régions du génome. Les marqueurs microsatellites, qui renseignent sur les événements démographiques récents des populations, ont été associés à deux séquences ADN chloroplastiques permettant, elles, de remonter plus loin dans le temps afin de couvrir une large partie de l'histoire passée des forêts jusqu'à nos jours. Les histoires démographiques de 8 espèces d'arbres tropicaux appartenant à diverses familles botaniques (*Carapa surinamensis*, *Dicorynia guianensis*, *Jacaranda copaia*, *Simarouba amara*, *Symphonia globulifera* et *Symphonia sp1.*, *Virola michelii* et *Vouacapoua americana*) ont été étudiées en échantillonnant 15 individus répartis sur l'ensemble de la Guyane afin de s'affranchir des histoires locales des populations. Cinq scénarios alternatifs ont été modélisés et testés par des méthodes ABC basées sur le coalescent (Cornuet *et al.*, 2008) pour lesquels le choix des modèles de mutation utilisés pour les marqueurs microsatellites et les statistiques descriptives retenues ont été influencés par l'étude menée au chapitre II.

Les modèles les mieux supportés sont : une expansion démographique d'un facteur 100 en moyenne pour 4 des espèces étudiées, une population constante pour 3 espèces, une contraction démographique faiblement supportée pour *D. guianensis* uniquement et un scénario plus compliqué d'une contraction suivi d'une expansion plus récente pour la population Ouest de *V. americana*. Même si les événements démographiques n'ont pas pu être datés précisément pour pouvoir être associés directement aux perturbations du Quaternaire, le nombre de générations auquel ils ont été estimés avoir eu lieu est approximativement du même ordre de grandeur entre les espèces. Le signal démographique général qui ressort de cette étude est une augmentation des populations forestières au cours du temps allant dans le sens de la théorie des refuges de Haffer (1969), avec cependant des

espèces dont les populations sont restées constantes jusqu'à aujourd'hui. Cependant, les histoires démographiques de toutes les espèces étudiées ne convergent pas vers une contraction générale des populations au dernier maximum glaciaire et amènent à envisager d'autres hypothèses, comme celle avancée par Colinvaux *et al.* (1996), supposant un changement des communautés forestières à cette période. Le manque de corrélation détecté entre les signaux de transitions démographiques et les traits fonctionnels indique que les stratégies écologiques des espèces ne permettent pas d'expliquer leur dynamique passée. L'application de ces études démographiques à d'autres régions des Néotropiques pourrait permettre, dans le futur, à améliorer nos connaissances sur l'histoire écologique des forêts amazoniennes.

**Ecological history of tropical rainforests:
Quaternary demographic changes in Guiana Shield
tree populations as inferred from current
genetic diversity patterns**

Stéphanie Barthe¹, Bruno Hérault^{1,2}, Caroline Scotti-Saintagne³,
Ivan Scotti³

1 UAG, UMR EcoFoG (« Ecologie des forêts de Guyane »), Campus Agronomique, BP 709,
97387 Kourou, French Guiana (France)

2 CIRAD, UMR EcoFoG (« Ecologie des forêts de Guyane »), Campus Agronomique, BP 709,
97387 Kourou, French Guiana (France)

3 INRA, UMR EcoFoG (« Ecologie des forêts de Guyane »), Campus Agronomique, BP 709,
97387 Kourou, French Guiana (France)

Article en préparation

Abstract

The way past climatic and geological disturbances influenced the composition of Neotropical forests is hotly debated. During the Quaternary, rainfall and/or temperature changes having occurred during and immediately after the Last Glacial Maximum (LGM) are thought to have strongly affected the geographical distribution and local abundance of tree species. The paucity of the fossil record in these areas, though, prevents a direct description of such processes. To describe community-level historical trends in forest composition, we turned therefore to inferential methods based on the reconstruction of past demographic changes. In particular, we have studied the past history of rainforests in French Guiana over a time scale of several thousand years, through the application of Approximate Bayesian Computation (ABC) to diversity data at nuclear (microsatellite) and chloroplast (sequence) loci. Five demographic scenarios have been tested for nine populations from eight species. The most recent demographic signals detected converge on an overall signature of expansion of rainforest populations, presumably having occurred after the LGM. This finding suggests either that the composition of the rainforests was different, or that its global extent was smaller, during earlier epochs of the Pleistocene than today. Therefore, the currently observed composition of forests of the Guiana shield may be recent; further studies of the same kind will help to reconstruct the ecological history of Amazonian forests.

Keywords

Guiana shield rainforest history, Approximate Bayesian Computation, chloroplast DNA, climatic changes, demographic inferences, nuclear microsatellites

Introduction

South America underwent several geologic and climatic changes that may have shaped the current composition of Amazonian rainforests (Antonelli *et al.*, 2009; Hoorn *et al.*, 2010). In Europe and North America – and to some extent in Africa – the biogeographic history of forest species since the Last Glacial Maximum (LGM, 18,000 – 20,000 years BP) is well understood (Taberlet *et al.*, 1998; Hewitt, 2000, 2004; Petit *et al.*, 2002; Hu *et al.*, 2009; Dainou *et al.*, 2010). In South America, the past distribution and the demographical history of rainforest communities is still relatively unknown (Hewitt, 2004). Unfortunately, the paucity of fossil data due to high decay rates of dead organisms and to poor sedimentary soil in this area makes it difficult to reconstruct past ecosystems. From the last available evidence, the distribution of closed forests in the region has continually expanded and contracted over the last several million years in response to climate variations, including climate impacts on sea level, rainfall and temperature (Hammond, 2005). Available biogeographical, geomorphological and palaeological data have suggested that rainforests were fragmented into several isolated geographical areas (Haffer, 1969; Prance, 1982), assuming a drier climate in Amazonia during the LGM (Ledru *et al.*, 1998). Drier climates at LGM are supported by several charcoal analyses suggesting Holocene drought-related palaeo-fires (Charles-Dominique *et al.*, 1998; Tardy, 1998; Mayle *et al.*, 2000; Bush *et al.*, 2008; Mayle and Power, 2008; Power *et al.*, 2010). Climate changes during Pleistocene glacials are thus likely to have imposed demographic transitions on plant and animal populations, provoked by large-scale fragmentation / recolonisation processes (Aide and Rivera, 1998; Caron *et al.*, 2000; Dutech *et al.*, 2004; Flanagan *et al.*, 2004; Quijada-Mascareñas *et al.*, 2007; Motamayor *et al.*, 2008; Solomon *et al.*, 2008; de Thoisy *et al.*, 2010; Scotti-Saintagne *et al.*, 2012a). The dry-climate-with-refugia hypothesis has nevertheless been challenged (Bush and De Oliveira, 2006). One alternative hypothesis, suggested by Amazonian pollen records (Colinvaux *et al.*, 1996), is that a 6°C cooling during the LGM may have caused a floristic transition, rather than rainforest fragmentation and replacement by open and drier tropical forests (Colinvaux and De Oliveira, 2000; Bush *et al.*, 2004).

All the studies cited above share the property of implicitly deducing past demographic processes, such as population contractions and expansions, from current genetic diversity distribution patterns, via the fundamental assumption that range and demographic size changes are correlated (Avice, 2000). Direct approaches to demographic process modelling

are therefore desirable to accurately describe the past ecological history of Neotropical forests.

Past changes in effective population size, can be detected using statistical approaches that rely on the comparison of observed patterns of genetic diversity at neutral molecular loci with the expectations provided by the neutral theory of molecular evolution (Kimura, 1985). Highly polymorphic genetic markers such as microsatellites (or Simple Sequence Repeat, SSR), with a mutation rate of about 10^{-2} to 10^{-6} nucleotides per generation (Thuillet *et al.*, 2002), offer the opportunity to analyse demographic events having occurred in the recent evolutionary past. On the other hand, chloroplast sequence markers, with a lower mutation rate of about 10^{-9} nucleotides per generation (Wolfe *et al.*, 1987) and no or rare recombination event (Birky, 1995), return more slowly at the equilibrium state after demographic change, allowing the detection of older demographic transitions. Genetic information carried by both diploid nuclear and haploid chloroplast markers allows describing and studying a wide range of past demographical events that may have affected current genetic diversity. A variety of statistics can be used to describe the properties of population-level diversity, including generalist descriptors of diversity such as allele richness and Nei's gene diversity (Nei, 1973) or more specialised statistics, designed specifically for sensitivity to departures from the neutral (mutation-drift) equilibrium. Of these statistics, some were specifically developed for the analysis of microsatellite data (e.g. Reich *et al.*, 1998) and others for sequence data (e.g. Rogers and Harpending, 1992; Tajima 1989a). Such statistics can be used, for greatly enhanced analytical power, in combination with simulations of the coalescent process (Kingman, 1982), a powerful tool to model neutral molecular polymorphism (Nordborg, 2001). Hypothesis testing, parameter estimation and demographic model selection can then be performed through Approximate Bayesian Computation (ABC; Beaumont *et al.*, 2002; Beaumont, 2010).

The goal of the present study is to test whether past demographic transitions can be detected for eight tree species belonging to present-day forests of the Guiana shield. Common trends will be sought as evidence of rainforest fragmentation and/or past reduction of global forest cover caused by dryer and/or cooler climate, while differences between demographic trends in ecologically different species could be considered as evidence of changes in forest composition.

Material and Methods

Choice of species

Eight common tropical trees, typical of forests from the Guiana shield, were studied: *Carapa surinamensis* Miq. (Meliaceae), *Dicorynia guianensis* Amshoff (Caesalpiniaceae), *Jacaranda copaia* (Aubl.) D. Don (Bignoniaceae), *Simarouba amara* Aubl. (Simaroubaceae), *Symphonia globulifera* L. f. (Clusiaceae) and the related subspecies *S. sp1* (Sabatier et al. 1997), *Virola michelii* Heckel (Myristicaceae) and *Vouacapoua americana* Aubl. (Caesalpiniaceae). These species are common in the Guiana Shield and show relatively ample variation in ecological strategies (Table S1).

Sampling and DNA extraction

Samples of leaf or cambium were collected from 24 to 356 individuals of each species across French Guiana (Table 1). Total genomic DNA extraction were performed by following Colpaert *et al.* (2005) or according to the Qiagen DNeasy® 2006 protocol with a DNeasy 96 plant kit (Qiagen, Hilden, Germany). DNA quality was checked by spectrophotometry or by agarose gel electrophoresis.

DNA amplification, genotyping and sequencing

Nuclear microsatellites

Nuclear genome genetic diversity was analyzed at four to nine microsatellite (nSSR) markers depending on the species (Table S2). Polymerase chain reactions (PCRs) for *S. amara* markers were performed in 15µL volume containing 6µL of 20-fold diluted DNA, 1x *Taq* buffer, 0.2mM of dNTP, 0.04 U/µL *Taq* DNA polymerase (all products from Invitrogen, Carlsbad, CA) and 0.75µM of each primer (MWG Biotech, Edersberg, Germany). An initial denaturation at 94°C for 5 min was followed by 30 cycles of 94°C for 30s, annealing temperature of 53°C for 30 s and 72°C for 30 s, and a final extension at 72°C for 5 min.

For the four microsatellites markers of *Symphonia globulifera* and *sp1*, PCR followed the M13-tailed primer PCR method (Schuelke, 2000). Amplifications were carried out in a total volume of 15 µL containing 8.2µL of 20-fold diluted DNA, 1x *Taq* buffer, 0.125mM of dNTP, 0.05 U/µL *Taq* DNA polymerase, 0.2µM of Forward M13-tailed and Reverse primers, 0.05µM of Forward primer (MWG Biotech, Edersberg, Germany). PCR conditions included a denaturation step at 94°C for 4 min, followed by 30 cycles of 45 s at 94°C, 45 s at annealing temperature primers, and 45 s at 72°C, followed by 8 cycles of 45 s at 94°C, 45 s at 53°C, and 45 s at 72°C, and a final extension step at 72°C for 10 min. In all the protocols

Table 1. Sample size per type of molecular markers and per type of analysis for the eight studied species.

Species	nSSR ^a	cpDNA ^b		nSSR + cpDNA
		trnH	trnC	
<i>C. surinamensis</i>	229	107	189	30
<i>D. guianensis</i>	87	67	24	15
<i>J. copaia</i>	108	145	145	30
<i>S. amara</i>	155	78	78	30
<i>S. globulifera</i>	37	124	122	15
<i>Symphonia sp1</i>	42	274	271	15
<i>V. michelii</i>	67	122	122	30
<i>V. americana</i>	356	75	34	15

^a Nuclear simple sequence repeat

^b chloroplast DNA

above, forward primers were fluorescently labelled (with PET, 6-FAM, NED or VIC) in the and fragments were separated on an ABI 3130XL capillary sequencer (Applied Biosystems, Foster City, CA) using ABI POP4 and Applied Biosystems LIZ-500 as internal standard, following the manufacturer's instructions. Genotypes were scored using the GENEMAPPER 4.0 software package (Applied Biosystems, Foster City, CA).

nSSRs of *C. surinamensis*, *D. guianensis* and *V. michelii* were amplified and genotyped as described by Barthe *et al.* (in prep¹). We used genotyping data of Scotti-Saintagne *et al.* (2012b) for *J. copaia* and data from Dutech *et al.* (2004) for *V. americana*.

Chloroplast sequences

Two universal chloroplast intergenic space regions, *trnH-psbA* and *trnC-yef6* were amplified, as described in Shaw *et al.* (2005), for all eight studied species. DNA sequencing was performed with the ABI BigDye Terminator® v3.1 cycle sequencing kit (Applied Biosystems, Foster city, CA) according to the manufacturer's protocol. PCR products were purified with ExoSAP-IT (USB, Cleveland, OH) and separated on an ABI 3130XL capillary sequencer (Applied Biosystems, Foster city, CA). DNA sequences were aligned with CODONCODE ALIGNER 1.6.3 (CodonCode Corporation, Dedham, MA) and edited with BIOEDIT 7.0.9.0 (Hall, 1999).

Data analyses

Bayesian clustering of nuclear data

To control for possible genetic sub-structuring within each species, we used Bayesian clustering analyses implemented in the program STRUCTURE 2.3.1 (Pritchard *et al.*, 2000) to probabilistically assign individuals to K clusters based on nuclear microsatellite genotypes. 37 to 356 individuals were analyzed depending of species (See Table 1). Parameter settings and choice of the most likely number of clusters K from ΔK developed by Evanno *et al.* (2005) were as described in Barthe *et al.* (in prep). Individuals were assigned with high probability (membership coefficients > 0.80) to one of the K clusters. These analyses were repeated several times for datasets with a complex genetic structure and each repeat is designated by a level of nesting. At a following nesting level, STRUCTURE analysis is repeated by eliminating individuals previously assigned to different clusters until the assignment probability is below the threshold of 0.8 for all remaining individuals.

¹ See Chapter 3 of this thesis. Barthe *et al.* Impact of pre-Columbian settlement on the population dynamics of Amazonian trees: a genetic approach.

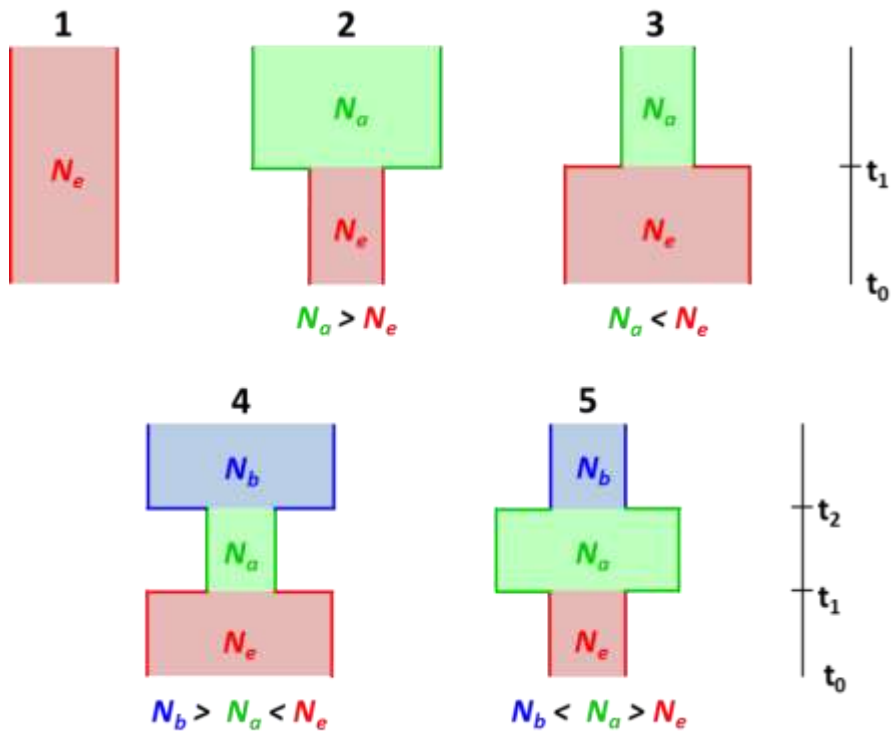


Figure 1. Scenarios explored with DIY ABC to explain observe genetic diversity of nine populations. t_0 is the present time and N_e is the present effective size of all scenarios. Scenario 1 is a model with constant population size N_e . Scenario 2 is a model with a population contraction and scenario 3 is a model with a growing population size (ancestral size N_a , time since the onset of the demographic event t_1). Scenario 4 is similar to scenario 3, but it includes an ancient bottleneck before expansion and scenario 5 is similar to scenario 2 with an ancient expansion before contraction (ancestral size before the more recent demographic event N_a , ancestral size before the older demographic events N_b , time since the onset of the more recent demographic event t_1 and time since the onset of the oldest demographic event t_2).

Choice of the chloroplastic mutation model

After removing nucleotide insertions-deletions in chloroplast sequence data, the best-fit nucleotide substitution model was inferred using the Akaike information criterion (AIC) implemented in the program MODELTEST 3.8 (Posada & Crandall, 1998; Posada & Buckley, 2004), with the starting tree obtained using the neighbour-joining algorithm in PAUP 4.10b (Swofford, 1991). 24 to 274 individuals depending of markers and species were used for this analysis (See Table 1). Hasegawa-Kishino-Yano (HKY; (Hasegawa *et al.*, 1985)) model described 11 out of 12 chloroplast data sets (six not polymorphic data sets were ignored) and was therefore chosen for both *trnH-psbA* and *trnC-ycf6* markers for all species.

Coalescent modelling

Demographic history was modelled through the coalescent-based Approximate Bayesian Computation ((ABC); (for recent reviews see Bertorelle *et al.*, 2010; Csilléry *et al.*, 2010), as implemented in the DIY ABC algorithm (Cornuet *et al.*, 2008). Five alternative historic demographic scenarios were tested in all species (Figure 1). Demographic parameters were drawn with the following distributions: Log-Normal (300; 250) for demographic event times of all species and Log-Normal (1000; 1000) for current or ancestral effective population sizes. After eliminating the genetic structuring within each species, based on clusters defined by STRUCTURE, individuals with both nSSR and chloroplast sequence data were sampled uniformly with respect to their geographical distribution. Thirty individuals were drawn from the complete data set for *C. surinamensis*, *J. copaia*., *S. amara*, *V. michelii*; fifteen individuals were drawn for *D. guianensis*, *S. globulifera*, *Symphonia sp1*, and two subgroups (East and West) for *V. americana*.

For nuclear microsatellites markers, locus numbers and allelic size range are reported in Table S2. We simulated nSSRs using a generalized stepwise mutation (GSM; Zhivotovsky *et al.*, 1997; Estoup *et al.*, 2002) model with two parameters: the mean mutation rate (μ_{SSR}) and the mean of the geometric distribution of the length in number of repeats of mutation events (P) were drawn respectively from Uniform[10^{-5} ; 10^{-3}] and Uniform[0.1; 0.3] distributions. Each locus was characterized by individual μ_i and P_i values respectively drawn from Gamma (mean = mean μ_{SSR} ; shape = 2) and Gamma (mean = P ; shape = 2) distributions. The two chloroplast sequence loci were assumed to follow the HKY evolutionary model in all species, as determined previously, with a fraction of constant site fixed to 0% and the shape parameter of the

Table 2. Clusters K of each studied species determined by the software STRUCTURE with the highest posterior probability ΔK . Nesting level indicates the analysis iteration.

Species	Nesting level	K	ΔK
<i>C. surinamensis</i>	1 st	5	6.5
	2 ^d	2	11
<i>D. guianensis</i>	1 st	3	82.5
	2 ^d	3	2.9
<i>J. copaia</i>	1 st	2	402.7
<i>S. amara</i>	1 st	7	3.4
	2 ^d	2	4.8
	3 ^d	4	3.5
<i>S. globulifera</i>	1 st	2	6.5
Symphonia sp1	1 st	4	3.6
<i>V. michelii</i>	1 st	3	5.6
	2 ^d	2	13
<i>V. americana</i>	1 st	2	843.8

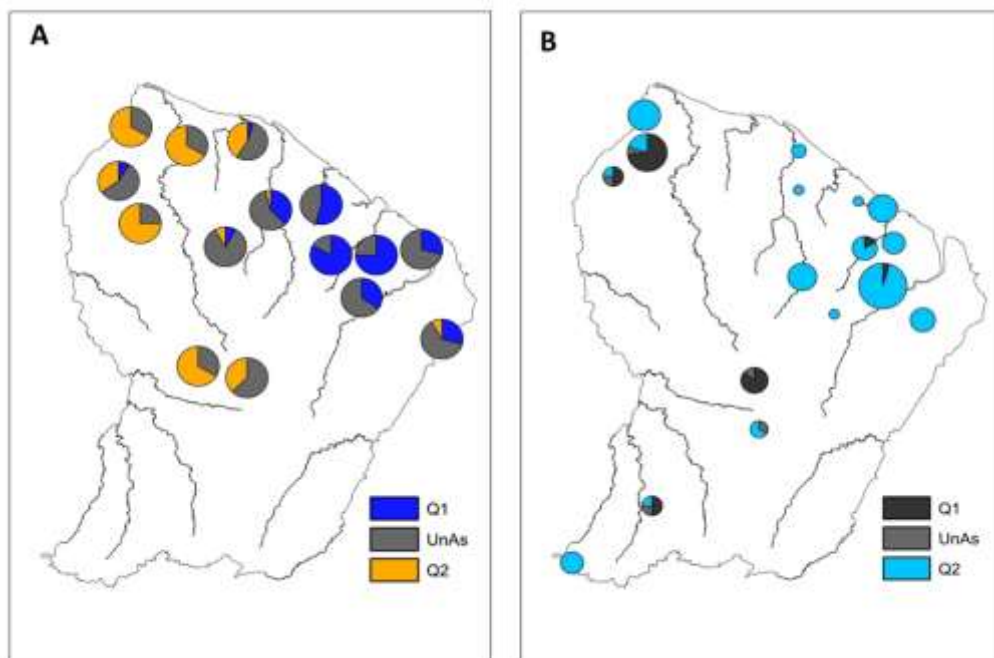


Figure 2. Geographic distribution of clusters inferred with Bayesian analysis for the two species *Vouacapoua americana* (K=2) (A) and *Jacaranda copaia* (K=2) (B) in French Guiana. Individuals were assigned by STRUCTURE with a membership coefficient higher than 0.80 to clusters Q1 or Q2 and “UnAs” refers to unassigned individuals. Circle sizes are proportional to the number of individuals.

Gamma distribution of mutations among sites equal to 2. The mean mutation rate per nucleotide and generation μ_{cpSeq} was drawn in a Uniform $[10^{-8}; 10^{-7}]$ and mutation rate of each locus was chosen from a Gamma (mean = μ_{cpSeq} ; shape = 2).

The summary statistics for nSSRs loci were the mean number of alleles, gene diversity (Nei, 1987) and allele size variance, except for the two data sets of *V. americana*, for which alleles are were not in coded by their length, but only by their identity (Dutech *et al.*, 2004). For chloroplast DNA sequence loci we chose three summary statistics: the number of distinct haplotypes, Tajima's D (Tajima, 1989b) and the mean of numbers of the rarest nucleotide at segregating sites. We simulated 2,000,000 data sets for each scenario. For model comparison, we estimated the posterior probabilities of the competing scenarios using a logistic regression method (Ghirotto *et al.*, 2010) on the 1% of simulated data sets closest to the observed data set (Cornuet *et al.*, 2008). Posterior parameter distributions were estimated under the best scenario using a linear regression on the 1% closest simulations and applying a *logit* transformation to parameter values (Cornuet *et al.*, 2008). Posterior distributions were obtained for three composite parameters: population diversity parameters $\theta_0 = N_0\mu_0$, $\theta_1 = N_1\mu_1$ and ratio $r_\theta = \theta_0/\theta_1$.

Since we analysed the same chloroplast sequences for all species, we could directly compare chloroplast θ and r_θ values across species; assuming that mutation rates are roughly the same for all species and constant over time, θ and r_θ values are estimators of effective population sizes and ratios of effective population sizes, respectively, that can be directly compared across species. On the other hand, SSR markers may have widely different mutation rates between species, so direct comparisons of θ values cannot be used to indirectly compare N_e values. The only meaningful across-species comparison that can be applied to nSSR data involves r_θ , which can again be taken as a proxy for ratio of effective population sizes.

Ecological characteristics

Finally, we tested correlations between values of θ ratios (r_θ) and (i) functional traits describing the life-history, leaf economic and stem economic strategy (Baraloto *et al.*, 2010) as well as (ii) demographic parameters (turn-over rates, growth potential and plasticity) derived from Hérault *et al.* (2011) and Aubry-Kientz *et al.* (in review).

Table 3. Most strongly supported demographic scenario for each species, associated posterior probability and parameter values. All values are provided with their confidence interval at 95%.

Species	Scenario	P^a	t^b	$\theta_{n(\text{present})}^c$	$\theta_{n(\text{past})}^c$	$\theta_{cp(\text{present})}^c$ ($\times 10^{-4}$)	$\theta_{cp(\text{past})}^c$ ($\times 10^{-4}$)	$r_{\theta_{cp}}^d$	$r_{\theta_n}^d$
<i>C. surinamensis</i>	Expansion	0.41 [0.39-0.42]	1140 [1-41300]	1.24 [0.25-23.7]	0.02 [0-0.24]	5.85 [0.51-62.8]	0.07 [0-1.73]	93.63 [5.41- $+\infty$]	93 [5.39- $+\infty$]
<i>J. copaia</i>	Expansion	0.45 [0.41-0.48]	2920 [1-43700]	23.5 [7.2-77.6]	1.15 [0-8.67]	21.2 [4.56-73.5]	0.89 [0-11.7]	24.27 [1.91- $+\infty$]	24.26 [1.91- $+\infty$]
<i>S. amara</i>	Expansion	0.42 [0.40-0.44]	950 [1-36900]	12 [2.4-74]	0.12 [0.001-1.17]	7.78 [0.93-57.9]	0.06 [0-1.11]	143.52 [6.95- $+\infty$]	145.41 [6.96- 33100]
<i>V. americana</i> (East)	Expansion	0.31 [0.31-0.32]	514 [1-30500]	1.9 [0.4-51.4]	0.17 [0.001-0.8]	1.98 [0.16-48.2]	0.12 [0-1.66]	27 [1.32- $+\infty$]	27.04 [1.31- 8300]
<i>S. globulifera</i>	Constant	0.45 [0.43-0.47]	-	22.7 [7.47-63.6]	-	40.9 [15.6-82.7]	-	1	1
<i>Symphonia sp1</i>	Constant	0.47 [0.46-0.49]	-	26.4 [8.33-67.2]	-	47.8 [19.3-86.1]	-	1	1
<i>V. michelii</i>	Constant	0.39 [0.35-0.43]	-	15 [6.91-37.8]	-	25.8 [8.17-65.6]	-	1	1
<i>D. guianensis</i>	Contraction	0.27 [0.26-0.27]	108 [1-26500]	0.43 [0.01-2.97]	3.62 [0.79-44]	0.31 [0-3.81]	3.32 [0.35-42.4]	0.08 [0-1.14]	0.08 [0.0007-1.14]
<i>V. americana</i> (West)	Contraction + Expansion	0.44 [0.41-0.47]	131 [1-10000] 11 [1-538]	0.61 [0.04-21.5]	5.2 [0-0.35]	3.6 [0.11-50.6]	23.9 [0.28-88.8]	0.14 [0.004-15.68]	0.14 [0.004-24.70]

^a P : Posterior probability of chosen scenario based on the logistic estimate. Logistic was performed on the 20 000 simulations closest to the real value.

^b t : Median of the time to demographic event

^c θ : Median of the diversity index ($\theta = Ne \times \mu$, with Ne the past or current effective size of population and μ the mutation rate).

^d r_{θ} : Median of the ratio θ_0 / θ_1 .

^{c,d} n : nuclear genome. cp : chloroplastic genome

Results

Population genetic structure

Population structure was analysed to identify genetically homogeneous groups complying with the coalescent assumption of panmictic Wright-Fisher populations (Table 2). *S. globulifera* and *S. sp1* showed a genetic structure with two and four clusters respectively with weak ΔK values (3.6 and 6.5 respectively), but no individual was assigned to a cluster with a membership coefficient greater than 0.80. Therefore, nuclear data of these two species were considered as not structured. A strong structure of two clusters was detected for *V. americana* ($\Delta K=843.8$) and *J. copaia* ($\Delta K=402.7$) (See Figure 2). The two *V. americana* clusters Q1 and Q2 were composed of 87 and 97 individuals respectively, with an East-West distribution. *J. copaia* structure comprised a large cluster Q1 with 80 individuals and a smaller cluster Q2 with only 23 individuals. The latter was removed from the data set used for further analyses. The analysis of *C. surinamensis*, *D. guianensis*, *S. amara* and *V. michelii* data were made in two or three levels of nesting because the first showed a more complex genetic structure ($K=3$ to $K=7$) with generally weak ΔK values (between 3.4 and 82.5) and few individuals assigned to a single cluster with a probability greater than 0.80.. The lack of individuals assigned with a high probability to the last nesting level indicated no structured populations for *C. surinamensis*, *D. guianensis*, *S. amara* and *V. michelii*. Therefore, we had a total of nine data sets representing eight species; only *V. americana* is composed of two populations, eastern and western.

Demographic transitions

The scenario with the highest posterior probability had constant population for *V. michelii* (P=0.39), *S. globulifera* (P=0.45) and *S. sp1* (P= 0.47) (Table 3). For *C. surinamensis*, *J. copaia*, *S. amara* and eastern *V. americana* populations, the expansion scenario was the most likely, with probabilities of 0.41, 0.45, 0.42 and 0.31, respectively. In these scenarios, demographic expansions of all species were dated between 514 and 2920 generations with very large confidence intervals up to 30,500 and 41,300 generations. The best scenario for *V. americana* Eastern population (P=0.44) including a contraction, with a mode for its age at 131 generations, followed by a recent expansion (age mode: 11 generations ago). The best scenario for *D. guianensis* shows a contraction dating back 11 generations, but was only weakly supported (P=0.27) and barely better supported than the constant population scenario (P=0.22) (for details see Table S4). Species affected by expansions had current chloroplast and nuclear population sizes between 20 to 140 times

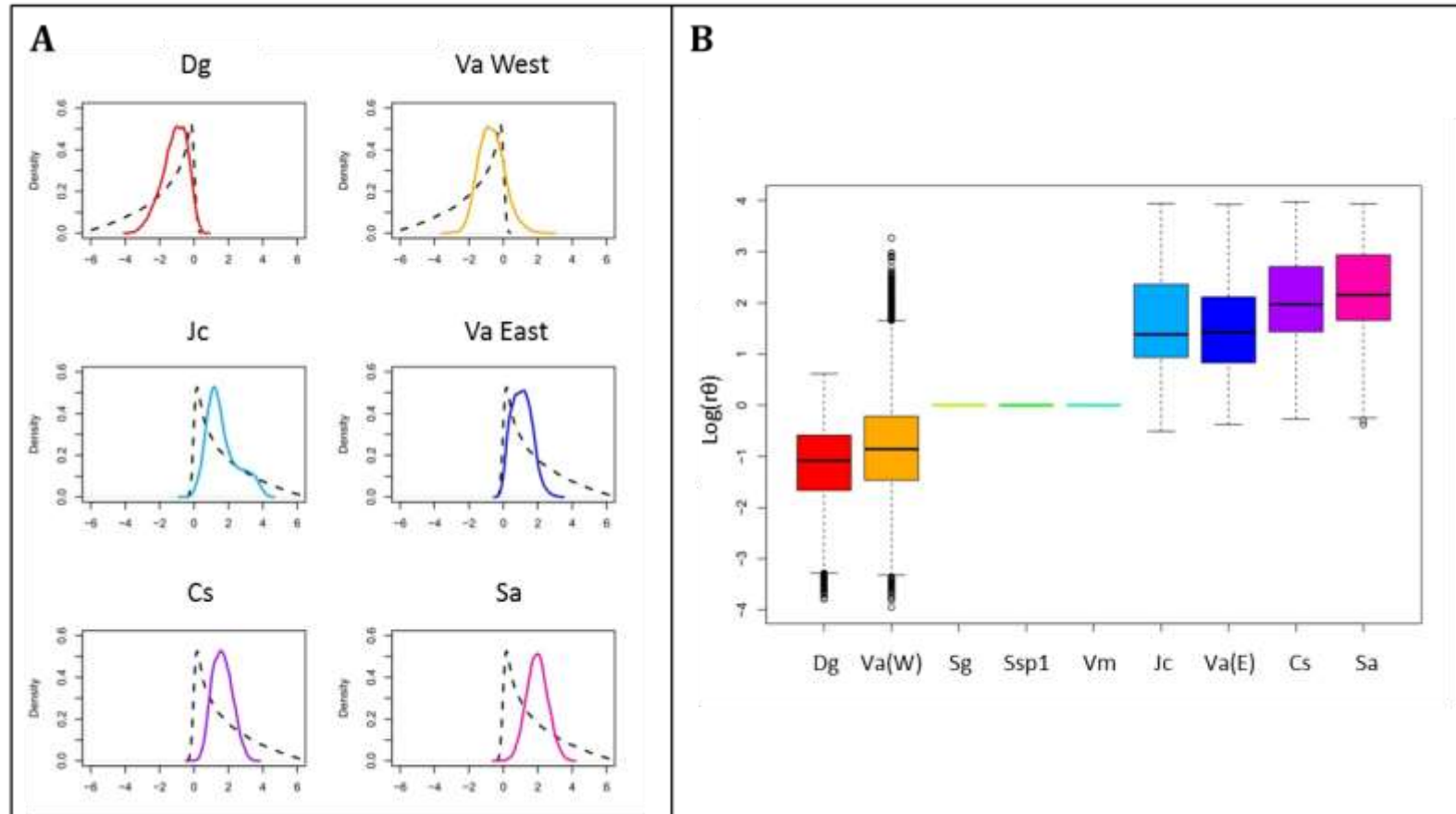


Figure 3. Posterior distributions (A) and value comparison (B) of the ratio of present and past diversity index θ , between studied species. *Carapa surinamensis* (Cs), *Dicorynia guianensis* (Dg), *Jacaranda copaia* (Jc), *Simarouba amara* (Sa), *Symphonia globulifera* (Sg), *Symphonia sp1* (Ssp1), *Virola michelii* (Vm), eastern *Vouacapoua americana* (Va(E)) and western *Vouacapoua Americana* (Va(W)).

larger than in the past (Table 3). *S. amara* showed the strongest expansion event and *J. copaia* the smallest, but r_{θ} values of the four species were not significantly different (Figure 3B). Contraction signals detected for *D. guianensis* and western *V. americana* showed population decreasing by a factor 10 or less (for the latter population, r_{θ} was computed with θ estimations of the current and the oldest populations).

To test the hypothetical link between the demographic transition signal and the ecological strategy, correlation between θ ratios and functional traits as well as demographic parameters was tested but none of the considered traits was significant (Table S4). This lack of significant correlation is probably due to a low statistical power of the data set based only on 9 r_{θ} values.

Discussion

Several past demographic scenarios of expansion, contraction and constant population were tested by coalescent modelling coupled with the ABC approach (François *et al.*, 2008; Ingvarsson, 2008; Legrand *et al.*, 2009; Lander *et al.*, 2011; Burgarella *et al.*, 2012). In our study, at least half of the species (five out of eight, if we include *V. americana*) showed a signature of demographic expansion with an increase of population effective size up to more than two orders of magnitude for both chloroplast and nuclear markers. Three species showed constant population size, and only one (*D. guianensis*) bears a signature of contraction, but with a very weak support. Globally, therefore, the trend that emerges from our results is one of constant or increased population sizes. Dates in generations (and associated confidence intervals) of detected expansions were of the same magnitude, maybe indicating concomitant expansion events for several species. Nevertheless, credible intervals are extremely large, so that making inferences about the timing (in generations) of forest expansion would be highly speculative. Imprecision of timing estimates may be due to limits in our simplistic model (*e.g.* population expansion may have been gradual rather than sudden) or to violations of model assumptions (*e.g.* non-overlapping generations, which are unlikely to be realistic for most trees). Furthermore, generation time is often complicated to estimate in trees (again partly because of overlapping generations), making it difficult to accurately place demographic events on an absolute time scale.

For at least two species, population expansions detected in this study are associated with phylogeographic patterns suggesting recolonisation events, thus supporting the idea that phylogeographic patterns are suggestive of population expansion. In particular, In *V. americana*, while the study of Dutech *et al.* (2004) did not find nuclear genetic divergence between putative biogeographical groups, the East-West structure detected here (Figure 2A) attests to the role of the Sinnamary River basin as biogeographical barrier (de Granville, 1982) to gene flow, consistently with chloroplast marker structuring described by Dutech *et al.* (2003). The past demographic expansion inferred for the eastern population of *Vouacapoua* supports the hypothesis of Dutech *et al.* (2003, 2004) of a reduction of range area and subsequent recolonisation from several different refugia (de Granville, 1982) during Holocene climate change. However, the more complex demographic history of the western population of *V. americana*, showing contraction followed by a more recent expansion, rather seems to describe a local patchy colonisation processes specific to this species (Forget *et al.*, 1999; Traissac, 2003). The population expansion detected for *J. copaia* is also consistent with results describing phylogeographic patterns (Scotti-Saintagne *et al.*,

2012b) for which the low chloroplast diversity and the nuclear genetic structuring suggest a population expansion across the entire Guiana shield and indicate that *Jacaranda* populations of French Guiana belong to a lineage derived from the Central Amazon refugia identified by Gentry (1982). The largest cluster studied here (Q1), distributed along the coast and in the eastern part of French Guiana (Figure 2B), belongs to the same nuclear genetic pool as Manaus (Brazil) (Scotti-Saintagne *et al.*, 2012b). Two waves of Pleistocene colonisation, from Tapajos and Manaus' regions, could explain the current population structure of this species in French Guiana. Preliminary microsatellite sequencing of *S. amara* showed a clinal South-West to North-East variation in allele size in French Guiana (Scotti-Saintagne *et al.*, in prep.). This putative directional evolution may reflect an increase in allele size from the “ancestral” to biogeographically derived populations (Vigouroux *et al.*, 2003) supporting the demographic expansion found for *S. amara*.

Three species remained constant over time. Two of them, *Symphonia globulifera* and *S. sp1* cannot be genetically differentiated and are likely to be two ecotypes of the same species having diverged very recently (Casalis, in prep), perhaps sharing most of their Pleistocene demographic history. The signal for *D. guianensis* favours contraction, but with a very weak support compared to other scenarios.

Overall, the most recent demographic signals detected converge on a global signature of expansion or stability of rainforest populations during the late Pleistocene and the Holocene. Species-specific biological traits did not explain differences in most-likely scenarios among species, suggesting that past climate changes affected the rainforest community as a whole, as opposed to favouring some species and penalising others based on their ecological characteristics. The strong demographic expansion detected for several species supports the hypothesis of a rainforest fragmentation, in agreement with Mayle and Beerling (2004), and/or past reduction of global forest cover, in agreement with Clapperton (1993), Anhuf *et al.* (2006) and Mayle and Power (2008), caused by dryer and/or cooler climate. This shrinkage phase would have been followed by recovery until the mid-Holocene, starting from putative refugia identified in South America (Haffer, 1969) and more particularly in French Guiana (de Granville, 1982). Alternatively, the presence of a group of species displaying constant population size may be explained by the hypothesis of change in the forest composition, rather than reduction, proposed by Colinvaux *et al.*, (1996) and supported by Mayle and Beerling (2004). This model suggests that glacial interval forests were comparable to modern forests but also included species now restricted to higher elevations. Species for which we show stable demography may have persisted in the forests,

in spite of community composition turnover, due to adaptation to colder and drier environments. However, we were unable to link demographic stability to any ecological trait, making it difficult to figure out what these adaptations may have been. Nevertheless, these interpretations should be taken with caution, because paleoecological inferences reported in the literature are mostly centred on the Amazonian basin and provide only marginal information on the Guiana shield, which may have undergone partially different changes.

Conclusion

A population-genetic approach, combining the analysis of diversity at nuclear SSR and chloroplast loci, has allowed us to detect a global signature of demographic expansion of some species from the Guiana shield rainforests. This suggests either that forest community composition was different, or that its global extent was smaller, during earlier epochs of the Pleistocene than today. This result is a direct, although inferential, evidence of changes undergone by the forest communities of the Guiana Shield in the distant climatic past – something which could not be precisely tested with standard historical methods due to the rareness of fossil evidence. The extensive application of this method to other regions of the Neotropics may contribute, in the future, to the detailed reconstruction of the ecological history of Amazonian forests.

Acknowledgments

We wish to thank Cyril Dutech, Henri Caron, William Montaigne, Maxime Casalis, Christine Le Mounier and Luc Allard for providing a part of the data, Valérie Troispoux and Saint-Omer Cazal for technical support.

Funding

Financial support was provided by the EU-funded INCO ‘SEEDSOURCE’ project and by the EU-funded PO-FEDER ‘ENERGIRAVT’ program. SB’s Doctoral program has been co-funded by the European Social Fund and by INRA. The funding agencies had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

Author contribution

SB, BH, CS and IS conceived the experiments. SB, CS obtained SSR genotypes and DNA sequences. SB, BH and IS analyzed the data. SB, BH and IS wrote the manuscript.

References

- Aide, T. M. and E. Rivera (1998). Geographic patterns of genetic diversity in *Poulsenia armata* (Moraceae): implications for the theory of Pleistocene refugia and the importance of riparian forest. *Journal of Biogeography* 25: 695-705.
- Anhuf, D., M. P. Ledru, H. Behling, F. W. Da Cruz Jr., R. C. Cordeiro, T. Van der Hammen, I. Karmann, J. A. Marengo, P. E. De Oliveira, L. Pessenda, A. Siffedine, A. L. Albuquerque and P. L. Da Silva Dias (2006). Paleo-environmental change in Amazonian and African rainforest during the LGM. *Palaeogeography, Palaeoclimatology, Palaeoecology* 239: 510-527.
- Antonelli, A., J. A. A. Johan A. A. Nylander, C. Persson and I. Sanmartin (2009). Tracing the impact of the Andean uplift on Neotropical plant evolution. *Proceedings of the National Academy of Sciences of the USA* 106: 9749-9754.
- Avise, J. C. (2000). *Phylogeography. The history and formation of species*. Harvard University Press, Cambridge, MA.
- Baraloto, C., C. E. T. Paine, L. Poorter, J. Beauchene, D. Bonal, A.-M. Domenach, B. Hérault, S. Patino, J.-C. Roggy and J. Chave (2010). Decoupled leaf and stem economics in rain forest trees. *Ecology Letters* 13: 1338-1347.
- Barthe, S., F. Gugerli, N. A. Barkley, L. Maggia, C. Cardini and I. Scotti (2012). Always look on both sides: Phylogenetic information conveyed by Simple Sequence Repeat allele sequences. *PLoS One* 7: e40699.
- Beaumont, M. A. (2010). Approximate Bayesian Computation in Evolution and Ecology. *Annual Review of Ecology, Evolution and Systematics* 41: 379-405.
- Beaumont, M. A., W. Zhang and D. J. Balding (2002). Approximate Bayesian Computation in Population Genetics. *Genetics* 162: 2025-2035.
- Bertorelle, G., A. Benazzo and S. Mona (2010). ABC as a flexible framework to estimate demography over space and time: some cons, many pros. *Molecular Ecology* 19: 2609-2625.
- Birky, C. W., Jr. (1995). Uniparental inheritance of mitochondrial and chloroplast genes: mechanisms and evolution. *Proceedings of the National Academy of Sciences of the USA* 92: 11331-11338.
- Burgarella, C., M. Navascués, M. Zabal-Aguirre, E. Berganzo, M. Riba, M. Mayol, G. G. Vendramin and S. C. Gonzales-Martinez (2012). Recent population decline and selection shape diversity of taxol-related genes. *Molecular Ecology* 21: 3006-3021.
- Bush, M. B. and P. E. De Oliveira (2006). The rise and fall of the Refugial Hypothesis of Amazonian Speciation: a paleo-ecological perspective. *Biota Neotropica* 6.
- Bush, M. B., P. E. De Oliveira, P. A. Colinvaux, M. C. Miller and J. E. Moreno (2004). Amazonian paleoecological histories: one hill, three watersheds. *Palaeogeography, Palaeoclimatology, Palaeoecology* 214: 359- 393.
- Bush, M. B., M. R. Silman, C. Mc Michael and S. S. (2008). Fire, climate change and biodiversity in Amazonia: a Late-Holocene perspective. *Philosophical Transactions of the Royal Society B* 363: 1795-1802.
- Caron, H., S. Dumas, G. Marque, C. Messier, E. Bandou, R. J. Petit and A. Kremer (2000). Spatial and temporal distribution of chloroplast DNA polymorphism in a tropical tree species. *Molecular Ecology* 9: 1089-1098.

- Charles-Dominique, P., P. Blanc, D. Larpin, M.-P. Ledru, B. RiCra, C. Sarthou, M. Servant and C. Tardy (1998). Forest perturbations and biodiversity during the last ten thousand years in French Guiana *Acta Oecologica* 19: 295-302.
- Clapperton, C. M. (1993). Nature of environmental changes in South America at the Last Glacial Maximum. *Palaeogeography, Palaeoclimatology, Palaeoecology* 101: 189-208.
- Colinvaux, P. A., P. E. De Oliveira and M. B. Bush (2000). Amazonian and neotropical plant communities on glacial time-scales: The failure of the aridity and refuge hypotheses. *Quaternary Science Reviews* 19: 141-169.
- Colinvaux, P. A., P. E. De Oliveira, J. E. Moreno, M. C. Miller and M. B. Bush (1996). A long pollen record from lowland Amazonia: Forest and cooling in glacial times. *Science* 274: 85-88.
- Colpaert, N., S. Cavers, E. Bandou, H. Caron, G. Gheysen and A. J. Lowe (2005). Sampling tissue for DNA analysis of trees: Trunk cambium as an alternative to canopy leaves. *Silvae Genetica* 54: 265-269.
- Cornuet, J.-M., F. Santos, M. A. Beaumont, C. P. Robert, J.-M. Marin, D. J. Balding, T. Guillemaud and A. Estoup (2008). Inferring population history with DIY ABC: a user-friendly approach to approximate Bayesian computation. *Bioinformatics* 24: 2713-2719.
- Croat, T. B. (1978). *Flora of Barro Colorado Island*.
- Csilléry, K., M. G. B. Blum, O. E. Gaggiotti and O. François (2010). Approximate Bayesian Computation (ABC) in practice. *Trends in Ecology & Evolution* 25: 410-418.
- Dainou, K., J.-P. Bizoux, J.-L. Doucet, G. Mahy, O. J. Hardy and M. Heuertz (2010). Forest refugia revisited: nSSRs and cpDNA sequences support historical isolation in a wide-spread African tree with high colonization capacity, *Milicia excelsa* (Moraceae). *Molecular Ecology* 19: 4462-4477.
- Dayanandan, S., J. Dole, K. Bawa and R. Kesseli (1999). Population structure delineated with microsatellite markers in fragmented populations of a tropical tree, *Carapa guianensis* (Meliaceae). *Molecular Ecology* 8: 1585-1592.
- de Granville, J. J. (1982). Rain forest and xeric flora refuges in French Guiana. *Biological diversification in the Tropics*. G. T. Prance. Columbia University Press, New York
- de Thoisy, B., A. G. Da Silva, M. Ruiz-Garcia, A. Tapia, O. Ramirez, M. Arana, V. Quse, C. Paz-y-Mino, M. Tobler, C. Pedraza and A. Lavergne (2010). Population history, phylogeography, and conservation genetics of the last Neotropical mega-herbivore, the lowland tapir (*Tapirus terrestris*). *BMC Evolutionary Biology* 10: 278.
- Degen, B., E. Bandou and H. Caron (2004). Limited pollen dispersal and biparental inbreeding in *Symphonia globulifera* in French Guiana. *Heredity* 93: 585-591.
- Dick, C. W., K. Abdul-Salim and E. Bermingham (2003). Molecular systematic analysis reveals cryptic tertiary diversification of a widespread tropical rain forest tree. *The American Naturalist* 162: 691-703.
- Draheim, H., M. Cui and C. W. Dick (2009). Characterization of 14 microsatellite DNA markers for the tropical forest tree *Virola surinamensis* (Rol.) Warb. (Myristicaceae). *Molecular Ecology Resources*.
- Dutech, C., L. Amsellem, N. Billotte and P. Jarne (2000). Characterization of (GA)_n microsatellite loci using an enrichment protocol in the neotropical tree species *Vouacapoua americana*. *Molecular Ecology* 9: 1433-1449.

- Dutech, C., H. I. Joly and P. Jarne (2004). Gene flow, historical population dynamics and genetic diversity within French Guianan populations of a rainforest tree species, *Vouacapoua americana*. *Heredity* 92: 69-77.
- Dutech, C., L. Maggia, C. Tardy, H. I. Joly and P. Jarne (2003). Tracking a genetic signal expansion of extinction-recolonization events in a neotropical tree species: *Vouacapoua americana* Aublet in French Guiana. *Evolution* 57: 2753-2764.
- Estoup, A., P. Jarne and J.-M. Cornuet (2002). Homoplasy and mutation model at microsatellite loci and their consequences for population genetics analysis. *Molecular Ecology* 11: 1591-1604.
- Evanno, G., S. Regnaut and J. Goudet (2005). Detecting the number of clusters of individuals using the software STRUCTURE: a simulation study. *Molecular Ecology* 14: 2611-2620.
- Flanagan, N. S., A. Tobler, A. Davison, O. G. Pybus, D. D. Kapan, S. Planas, M. Linares, D. Heckel and W. O. McMillan (2004). Historical demography of Müllerian mimicry in the neotropical *Heliconius* butterflies. *Proceedings of the National Academy of Sciences* 101: 9704-9709.
- Forget, P.-M., F. Mercier and F. Collinet (1999). Spatial patterns of two rodent-dispersed rain forest trees *Carapa procera* (Meliaceae) and *Vouacapoua americana* (Caesalpiniaceae) at Paracou, French Guiana. *Journal of Tropical Ecology* 15: 301-313.
- François, O., M. G. B. Blum, M. Jakobsson and N. A. Rosenberg (2008). Demographic history of European populations of *Arabidopsis thaliana*. *PLoS Genetics* 4.
- Gentry, A. H. (1982). Patterns of neotropical plant species diversity. *Evolutionary Biology* 15: 1-84.
- Ghirotto, S., S. Mona, A. Benazzo, F. Papparazzo, D. Caramelli and G. Barbujani (2010). Inferring genealogical processes from patterns of bronze-age and modern DNA variation in Sardinia. *Molecular Biology and Evolution* 27: 875-886.
- Haffer, J. (1969). Speciation in Amazonian forest birds. *Science* 165: 131-137.
- Hall, T. A. (1999). BioEdit: a user-friendly biological sequence alignment editor and analysis program for Windows 95/98/NT. *Nucleic Acids Symposium Series* 41: 95-98.
- Hammond, D. S. (2005). *Tropical forests of the Guiana shield. Ancient forests in a modern world*. CABI publishing, Wallingford, UK and Cambridge, MA, USA.
- Hardy, O. J., L. Maggia, E. Bandou, P. Breyne, H. Caron, M.-H. Chevallier, A. Doligez, C. Dutech, A. Kremer, C. Latouche-Hallé, V. Troispoux, V. Veron and D. B. (2006). Fine-scale genetic structure and gene dispersal inferences in 10 Neotropical tree species. *Molecular Ecology* 15: 559-571.
- Hasegawa, M., H. Kishino and T.-A. Yano (1985). Dating of the Human-ape splitting by a molecular clock of mitochondrial DNA. *Journal of Molecular Evolution* 22: 160-174.
- Hérault, B., B. Bachelot, L. Poorter, V. Rossi, F. Bongers, J. Chave, C. E. T. Paine, F. Wagner and C. Baraloto (2011). Functional traits shape ontogenetic growth trajectories of rain forest tree species. *Journal of Ecology* 99: 1431-1440.
- Hewitt, G. M. (2000). The genetic legacy of the Quaternary ice ages. *Nature* 405: 907-913.
- Hewitt, G. M. (2004). The structure of biodiversity – insights from molecular phylogeography. *Frontiers in Zoology* 1.
- Hoorn, C., F. P. Wesselingh, H. ter Steege, M. A. Bermudez, A. Mora, J. Sevink, I. Sanmartín, A. Sanchez-Meseguer, C. L. Anderson, J. P. Figueiredo, C. Jaramillo, D.

- Riff, F. R., Negri, H., Hooghiemstra, J., Lundberg, T., Stadler, T., Särkinen and A. Antonelli (2010). Amazonia through time: Andean uplift, climate change, landscape evolution, and biodiversity. *Science* 330: 927-931.
- Hu, F. S., A. Hampe and R. J. Petit (2009). Paleoeecology meets genetics: deciphering past vegetational dynamics. *Front Ecol Environ* 7.
- Ingvarsson, K. (2008). Multilocus patterns of nucleotide polymorphism and the demographic history of *Populus tremula*. *Genetics society of America* 180: 329-340.
- Kenfack, D. (2011). A synoptic revision of *Carapa* (Meliaceae). *Harvard Papers in Botany* 16: 171-231.
- Kimura, M. (1985). *The neutral theory of molecular evolution*. Cambridge University Press, Cambridge, UK.
- Kingman, J. F. C. (1982). On the genealogy of large populations. *Journal of Applied Probability* 19: 27-43.
- Lander, T. A., S. Oddou-Muratorio, H. Prouillet-Leplat and E. K. Klein (2011). Reconstruction of a beech population bottleneck using archival demographic information and Bayesian analysis of genetic data. *Molecular Ecology* 20: 5182–5196.
- Latouche-Halle, C., A. Ramboer, E. Bandou, H. Caron and S. Decroocq (2002). Isolation and characterization of microsatellite markers in the tropical tree species *Dicorynia guianensis* (Caesalpinaceae). *Molecular Ecology Notes* 2: 228-230.
- Ledru, M.-P., J. Bertaux and A. Sifeddine (1998). Absence of last glacial maximum records in lowland tropical forests. *Quaternary Research* 49: 233–237.
- Legrand, D., M. I. Tenaillon, P. Matyot, J. Gerlach, D. Lachaise and M.-L. Cariou (2009). Species-wide genetic variation and demographic history of *Drosophila sechellia*, a species lacking population structure. *Genetics Society of America* 182: 1197-1206.
- Mayle, F. E. and D. J. Beerling (2004). Late Quaternary changes in Amazonian ecosystems and their implications for global carbon cycling. *Palaeogeography, Palaeoclimatology, Palaeoecology* 214: 11-25.
- Mayle, F. E., R. Burbridge and T. J. Killeen (2000). Millennial-scale dynamics of southern Amazonian rain forests. *Science* 290: 2291-2294.
- Mayle, F. E. and M. J. Power (2008). Impact of a drier early–mid-Holocene climate upon Amazonian forests. *Philosophical transactions of the royal society B* 363: 1829-1838.
- Motamayor, J. C., P. Lachenaud, J. W. da Silva e Mota, R. Looor, D. N. Kuhn, J. S. Brown and R. J. Schnell (2008). Geographic and Genetic Population Differentiation of the Amazonian Chocolate Tree (*Theobroma cacao* L). *PLoS One* 3: e3311.
- Nei, M. (1973). Analysis of gene diversity in subdivided populations. *Proceedings of the National Academy of Sciences of the USA* 70: 3321-3323.
- Nei, M. (1987). *Molecular evolutionary genetics*.
- Nordborg, M. (2001). *Coalescent theory. Handbook of statistical genetics*. John Wiley & Sons, New York.
- Petit, R. J., S. Brewer, S. Bordacs, K. Burg, R. Cheddadi, E. Coart, J. Cottrell, U. M. Csaikl, B. van Dam, J. D. Deans, S. Espinel, S. Fineschi, R. Finkeldey, I. Glaz, P. G. Goicoechea, J. S. Jensen, A. O. Konig, A. J. Lowe, S. F. Madsen, G. Matyas, R. C. Munro, F. Popescu, D. Slade, H. Tabbener, S. G. M. de Vries, B. Ziegenhagen, J.-L. de Beaulieu and A. Kremer (2002). Identification of refugia and post-glacial

- colonisation routes of European white oaks based on chloroplast DNA and fossil pollen evidence. *Forest Ecology and Management* 156: 49-74.
- Porter, D. M. (1973). Flora of Panama family 90 : Simaroubaceae. *Annals of Missouri botanical garden* 60: 23-39.
- Posada, D. and T. R. Buckley (2004). Model selection and model averaging in phylogenetics: advantages of Akaike Information Criterion and Bayesian approaches over Likelihood Ratio Tests. *Systematic Biology* 53: 793-808.
- Posada, D. and K. A. Crandall (1998). MODELTEST: testing the model of DNA substitution. *Bioinformatics* 14: 817-818.
- Power, M. J., M. B. Bush, H. Behling, S. P. Horn, F. E. Mayle and D. H. Urrego (2010). Paleofire activity in tropical America during the last 21 ka: A regional synthesis based on sedimentary charcoal *Pages news* 18: 73-75.
- Prance, G. T. (1982). A review of the phytogeographic evidences for Pleistocene climate changes in the neotropics. *Annals of the Missouri Botanical Garden* 69: 594-624.
- Pritchard, J. K., M. Stephens and P. Donnelly (2000). Inference of population structure using multilocus genotype data. *Genetics* 155: 945-959.
- Quijada-Mascarenas, J. A., J. E. Ferguson, C. E. Pook, M. Da Graça Salomao, R. S. Thorpe and W. Wüster (2007). Phylogeographic patterns of trans-Amazonian vicariants and Amazonian biogeography: the Neotropical rattlesnake (*Crotalus durissus* complex) as an example. *Journal of Biogeography* 34: 1296-1312.
- Reich, D. E. and D. B. Goldstein (1998). Genetic evidence for a Paleolithic human population expansion in Africa. *Proceedings of the National Academy of Sciences of the USA* 95: 8119-8123.
- Rodriguez, H., J. Geistlinger, G. Berlyn, G. Kahl and K. Weising (2000). Characterization of novel microsatellite loci isolated from the tropical dioecious tree *Simarouba amara*. *Molecular Ecology* 9: 489-504.
- Rogers, A. and H. Harpending (1992). Population growth makes waves in the distribution of pairwise genetic differences. *Molecular Biology and Evolution* 9: 552-569.
- Sabatier, D. and M.-F. Prévost (1990). Quelques données sur la composition floristique et la diversité des peuplements forestiers de Guyane Française. *Revue Bois et Forêts des Tropiques* 219: 31-55.
- Schuelke, M. (2000). An economic method for the fluorescent labeling of PCR fragments. *Nature Biotechnology* 18: 233-234.
- Scotti-Saintagne, C., C. W. Dick, H. Caron, G. G. Vendramin, E. Guichoux, A. Buonanici, C. Duret, P. Sire, R. Valencia, M. R. Lemes, R. Gribel and I. Scotti (2012a). Phylogeography of a species complex of lowland Neotropical rain forest trees (*Carapa*, Meliaceae). *Journal of Biogeography* 10.1111/j.1365-2699.2011.02678.x
- Scotti-Saintagne, C., C. W. Dick, H. Caron, V. G. G., V. Troispoux, P. Sire, M. Casalis, A. Buonamici, R. Valencia, M. R. Lemes, R. Gribel and I. Scotti (2012b). Amazon diversification and cross-Andean dispersal of the widespread Neotropical tree species *Jacaranda copaia* (Bignoniaceae). *Journal of Biogeography* 10.1111/j.1365-2699.2012.02797.x
- Shaw, J., E. B. Lickey, J. T. Beck, S. B. Farmer, W. Liu, J. Miller, K. C. Siripun, C. T. Winder, E. E. Schilling and R. L. Small (2005). The tortoise and the hare II: relative utility of 21 noncoding chloroplast DNA sequences for phylogenetic analysis. *American Journal of Botany* 92: 142-166.

- Solomon, S. E., M. J. Bacci, J. J. Martins, G. G. Vinha and U. G. Mueller (2008). Paleodistributions and comparative molecular phylogeography of leafcutter ants (*Atta* spp.) provide new insight into the origins of Amazonian diversity. *PLoS One* 3: e2738.
- Swofford, D. L. (1991). PAUP: Phylogenetic Analysis Using Parsimony, Version 3.1 Computer program distributed by the Illinois Natural History Survey, Champaign, Illinois.
- Taberlet, P., L. Fumagalli, A.-G. Wust-Saucy and J.-F. Cosson (1998). Comparative phylogeography and postglacial colonization routes in Europe. *Molecular Ecology* 7: 453-464.
- Tajima, F. (1989). The effect of change in population size on DNA polymorphism. *Genetics* 123: 597-601.
- Tajima, F. (1989). Statistical method for testing the Neutral mutation hypothesis by DNA polymorphism *Genetics* 123: 585-595.
- Tardy, C. (1998). *Paléoincendies naturelles, feux anthropiques et environnements forestiers de Guyane française du tardiglaciaire à l'holocène récent. Approches chronologique et anthracologique*. Thèse de Doctorat, Sciences et Techniques, Montpellier II, France, 343.
- Thuillet, A.-C., D. Bru, J. David, P. Roumet, S. Santoni, P. Sourdille and T. Bataillon (2002). Direct Estimation of Mutation Rate for 10 Microsatellite Loci in Durum Wheat, *Triticum turgidum* (L.) Thell. ssp *durum* desf. *Molecular Biology and Evolution* 19: 122-125.
- Traissac, S. (2003). *Dynamique spatiale de Vouacapoua americana (Aublet), arbre de forêt tropicale humide à répartition agrégée* Thèse de Doctorat, Claude Bernard, Lyon I, France, 221.
- van Roosmalen, M. G. M. (1985a). *Fruits of the Guianan Flora*. Institute of Systematic Botany, Utrecht University press, Wageningen, The Netherlands.
- van Roosmalen, M. G. M. (1985b). Habitat preferences, diet, feeding strategy and social organization of the black spider monkey (*Ateles paniscus* Linnaeus 1758) in Surinam. *Acta Amazonica (supplement)* 15: 1-238.
- Vigouroux, Y., Y. Matsuoka and J. Doebley (2003). Directional evolution for microsatellite size in maize. *Molecular Biology and Evolution* 20: 1480-1483.
- Vinson, C. C., V. C. R. Azevedo, I. Sampaio and A. Y. Ciampi (2005a). Development of microsatellite markers for *Carapa guianensis* (Aublet), a tree species from the Amazon forest. *Molecular Ecology Notes* 5: 33-34.
- Vinson, C. C., A. C. Amaral, I. Sampaio and A. Y. Ciampi (2005b). Characterization and isolation of DNA microsatellite primers for the tropical tree, *Symphonia globulifera* Linn. f. *Molecular Ecology Notes* 5: 202-204.
- Wilson, T. K. (2004). Myristicaceae. *Flowering plants of the Neotropics*. N. Smith, S. A. Mori, D. W. Henderson and H. S. V. New York Botanical garden and Princeton University Press, Princeton, NJ
- Wolfe, K., W.-H. Li and P. M. Sharp (1987). Rates of nucleotide substitution vary greatly among plant mitochondrial, chloroplast, and nuclear DNAs. *Proceedings of the National Academy of Sciences of the USA* 84: 9054-9058.
- Zhivotovsky, L. A., M. W. Feldman and S. A. Grishchkin (1997). Biased mutations and microsatellite variation. *Molecular Biology and Evolution* 14: 926-933.

Supporting information

Table S1. Some of life history traits related to reproductive strategies and ecological characteristics of studied species.

Species	Family	Geographical Distribution ^a	Growth strategy ^{b*}	Mode of seed dispersal ^{c*}	Distance gene dispersal (m) [*]
<i>Carapa surinamensis</i>	Meliaceae	C, AB, (A)	H	R	182
<i>Dicorynia guianensis</i>	Caesalpiniaceae	G	HT	R, W	203
<i>Jacaranda copaia</i>	Bignoniaceae	CA to AB	P	W	1182
<i>Simarouba amara</i>	Simaroubaceae	C, AB	P **	Bi **	-
<i>Symphonia globulifera</i>	Clusiaceae	AB, A	H	AM	141
<i>Virola michelii</i>	Miristicaceae	AB	H	AM, Ba, R	505
<i>Vouacapoua americana</i>	Caesalpiniaceae	AB	T	R	176

^a Geographical distribution . A: Africa; AB: Amazonian basin; C: Caribbean; CA: Central America; G: Guiana Shield.

C. surinamensis (Sabatier and Prévost, 1990 ; Kenfack, 2011), *D. guianensis* (van Roosmalen, 1985a, Sabatier and Prévost, 1990), *Jacaranda copaia* (Croat, 1978), *S. amara* (Porter, 1973), *S. globulifera* (Dick *et al.*, 2003), *V. michelii* (Sabatier and Prévost, 1990; Wilson, 2004), *V. Americana* (van Roosmalen, 1985b)

^b Growth strategy .H: heliophilic, HT: hemi-tolerant, P:pioneer, T: shade tolerant

^c Seed dispersers. AM, nocturnal arboreal mammal; Ba: bats; Bi: Birds; M: monkeys; R: rodents; W: wind.

* (Hardy *et al.*, 2006)

** Mariwenn data base (<http://mariwenn.ecofog.gf>)

Table S2. Characteristics of simple sequence repeat (SSR) of each species. *Carapa surinamensis* (Cs), *Dicorynia guianensis* (Dg), *Jacaranda copaia* (Jc), *Simarouba amara* (Sa), *Symphonia globulifera* (Sg), *Symphonia sp1* (Ssp1), *Virola michelii* (Vm) and *Vouacapoua americana* (Va).

Species	Locus	Repeat	Allele size range (bp)	Reference
Cs	Cg5	(AG) _n (AAT) _n	178-179	Dayanandan <i>et al.</i> (1999)
	Cg7	(AG) _n	216-240	
	Cg17	(AG) _n	104-106	Vinson <i>et al.</i> (2005a)
	Cg1L	(AG) _n	162-168	
	Cg01	(GAA) _n	179-199	
Dg	Dg1	(CA) _n	253-268	Latouche-Hallé <i>et al.</i> (2002)
	Dg3	(AC) _n	257-274	
	Dg4	(GA) _n	191-207	
	Dg5	(CT) _n	118-123	
	Dg6	(CA) _n	264-290	
Jc	Jc3A10	(GA) _n [(CT) _n (CA) _n	220-272	Barthe <i>et al.</i> (2012)
	Jc3F4	(TA) _n [(GA) _n	200-238	
	Jc3H10	(CT) _n [(AC) _n	216-286	Scotti-Saintagne <i>et al.</i> (2012b)
	JACC11	(GT) _n	100-138	
	JACC2	(CT) _n (CA) _n	200-264	
	JACC21	(GT) _n	213-266	
	JACC22	(GT) _n	158-180	
	Jc3C5	(TC) _n [(CT) _n	130-162	
Jc3F11	(CT) _n	153-219		
Sa	Sa6	(AAT) _n [(AAT) _n	140	Rodriguez <i>et al.</i> (2000)
	Sa7	(GAA) _n	199-205	
	Sa27	(TTA) _n	197-247	
	Sa29	(TTA) _n	245-257	
Sg	Sg03	(CT) _n	335-353	Degen <i>et al.</i> (2004)
	Sg18	(CT) _n	306-332	Vinson <i>et al.</i> (2005b)
	Sg06	(CT) _n	206-254	
	Sg10	(CT) _n (CTAT) _n	163-221	
Ssp1	Sg03	(CT) _n	320-366	Degen <i>et al.</i> (2004)
	Sg18	(CT) _n	316-340	Vinson <i>et al.</i> (2005b)
	Sg06	(CT) _n	210-268	
	Sg10	(CT) _n (CTAT) _n	147-227	
Vm	Vmul2-65	(CT) _n (CA) _n	133-196	Draheim <i>et al.</i> (2009)
	Vsur2-35	(GA) _n	261-317	
	Vsur34	(CT) _n	350-417	
	Vsur58	(CT) _n (GTCT) _n (CTGT) _n (CTCTGT) _n (CT) _n	213-273	
	Vmul68-2	(AG) _n	340-413	
	Vseb3	(AG) _n	220-267	
	Vsur56	(CT) _n	200-234	
Va	Wac1	(GA) _n	-	Dutech <i>et al.</i> (2000)
	Wac3	(AG) _n	-	
	Wac5	(AG) _n	-	
	Wac13	(AG) _n	-	
	Wac14	(GA) _n [(GA) _n [(GA) _n	-	
	Wac7	(GA) _n	-	
	Wac10	(AG) _n	-	
	Wac12	(GA) _n [(GA) _n	-	
Wac16	(GA) _n [(AG) _n [(GA) _n	-		

Table S3. Posterior probability of each scenario per population determined by DIY ABC. Posterior probabilities are based on the logistic estimate performed on the 20 000 simulations closest to the real value. *Carapa surinamensis* (Cs), *Dicorynia guianensis* (Dg), *Jacaranda copaia* (Jc), *Simarouba amara* (Sa), *Symphonia globulifera* (Sg), *Symphonia sp1* (Ssp1), *Virola michelii* (Vm) eastern *Vouacapoua americana* (Va(E)) and western *Vouacapoua Americana* (Va(W)). Scenario 1: constant population, scenario 2: contraction, scenario3: expansion, scenario 4: contraction followed by an expansion, scenario 5: expansion followed by a contraction.

Scenario	Cs	Dg	Jc	Sa	Sg	Ssp1	Vm	Va(E)	Va(W)
1	0.063 [0.057-0.069]	0.222 [0.216-0.228]	0.154 [0.132-0.175]	0.042 [0.037-0.048]	0.448 [0.430-0.466]	0.475 [0.464-0.485]	0.392 [0.351-0.433]	0.174 [0.168-0.179]	0.051 [0.041-0.060]
2	0.0206 [0.018-0.023]	0.266 [0.260-0.273]	0.051 [0.040-0.062]	0.015 [0.012-0.018]	0.179 [0.165-0.192]	0.230 [0.221-0.239]	0.144 [0.117-0.170]	0.094 [0.089-0.098]	0.293 [0.266-0.321]
3	0.409 [0.393-0.425]	0.151 [0.146-0.157]	0.448 [0.412-0.484]	0.420 [0.403-0.438]	0.192 [0.177-0.207]	0.133 [0.126-0.141]	0.279 [0.239-0.319]	0.314 [0.308-0.321]	0.042 [0.033-0.050]
4	0.227 [0.213-0.240]	0.178 [0.173-0.184]	0.143 [0.117-0.170]	0.274 [0.258-0.290]	0.052 [0.044-0.060]	0.053 [0.048-0.058]	0.050 [0.033-0.067]	0.203 [0.198-0.209]	0.436 [0.405-0.467]
5	0.282 [0.267-0.296]	0.182 [0.176-0.188]	0.204 [0.176-0.232]	0.249 [0.234-0.264]	0.130 [0.118-0.142]	0.109 [0.102-0.115]	0.135 [0.107-0.162]	0.215 [0.209-0.221]	0.178 [0.156-0.200]

Table S4. Biological criteria related to the ability of species to grow and colonize. *Carapa surinamensis* (Cs), *Dicorynia guianensis* (Dg), *Jacaranda copaia* (Jc), *Simarouba amara* (Sa), *Symphonia globulifera* (Sg), *Symphonia sp1* (Ssp1), *Virola michelii* (Vm) and *Vouacapoua americana* (Va).

	Unit	Cs	Dg	Jc	Sa	Sg	Ssp1	Vm	Va	p-value *
Demographic parameters										
Turn-over	-	0.0102	0.0027	0.0061	0.0014	0.0177	0.0051	0.0019	0.0012	0.749
Kurtosis	-	0.73	0.87	0.73	-	0.81	0.83	0.54	0.78	0.791
Gmax	-	2.83	3.63	3.74	-	6.55	4.26	3.49	2.56	0.306
Life-history traits										
DBH95	mm	39.96	80	39.65	54.05	92	44.62	51	73	0.34
Hmax	m	35	50	28	37	26	26	38	42	0.915
Stem Economics										
Bark	mm	4.35	5.53	4.79	12.00	4.33	6.71	6.76	4.69	0.089
WdDens	g.cm ⁻³	0.55	0.65	0.38	0.37	0.56	0.56	0.47	0.69	0.206
WdMst	%	0.71	0.71	0.80	0.64	0.90	0.94	0.87	0.61	0.135
Leaf Economics										
CN	g.g ⁻¹	27.02	21.26	21.15	26.17	24.01	26.82	22.17	19.97	0.172
SLA	cm ² .g ⁻¹	10.40	11.89	13.25	8.60	10.23	10.29	11.12	14.24	0.186
Tough	N	2.56	1.74	0.90	1.12	1.29	1.13	0.91	1.49	0.51
Chlo	µg.mm ⁻²	88.70	76.55	84.58	104.54	90.83	96.37	85.48	55.53	0.251
C13	‰	-31.04	-29.34	-31.30	-28.64	-31.14	-30.96	-32.85	-31.84	0.153
TwigDens	g.cm ⁻³	0.52	0.55	0.39	0.45	0.61	0.65	0.47	0.64	0.285
TwigBark	mm	1.62	2.67	2.03	3.27	1.98	1.59	2.22	1.94	0.236
P (10 ⁻⁴)	µg.g ⁻¹	6.36	-	6.00	-	6.33	5.06	6.3	16.02	0.958
K (10 ⁻³)	µg.g ⁻¹	4.31	-	2.24	-	6.48	3.64	3.51	7.34	0.838
N	cg.g ⁻¹	0.018	0.024	0.025	0.019	0.020	0.018	0.024	0.026	0.297
LTD (10 ⁻⁵)	g.cm ⁻³	4.308	3.691	3.602	3.656	3.128	3.383	4.751	4.687	0.93

Turnover: turn-over rates, **Kurtosis:** ontogenetic plasticity in growth trajectory, **Gmax:** Maximum growth rates at mature developmental stages **DBH95:** 95th percentiles of the Diameter distribution, **Hmax:** Maximum height, **WdDens:** Trunk xylem density, **CN:** Foliar C:N, **SLA:** Specific Leaf area, **WdMst:** Trunk wood moisture content, **Tough:** Lamina toughness, **Chlo:** Lamina total chlorophyll, **Bark:** Trunk bark thickness, **C13:** Foliar ¹³C composition, **TwigDens:** Branch xylem density, **TwigBark:** Branch bark thickness, **P:** Foliar P_m, **K:** Foliar K_m, **N:** Foliar N_m, **LTD:** Leaf tissue density. *p-value of the test for correlation between θ ratios and trait values

Chapitre V

Discussion générale

Dans le cadre de cette thèse je me suis intéressée à l'histoire écologique passée des forêts humides du plateau des Guyanes à travers l'étude de l'histoire démographique des populations d'arbres. La discussion suivante permettra dans une première partie d'aborder l'aspect méthodologique des études démographiques menées dans cette thèse à des échelles temporelles et spatiales différentes pour ensuite caractériser l'impact des diverses perturbations naturelles et anthropiques sur la forêt guyanaise et formuler une hypothèse sur l'évolution passée de cet écosystème forestier. Deuxièmement, j'évoquerai en quoi les résultats de cette thèse sont innovants et apportent de nouveaux éléments à la compréhension de l'histoire écologique des forêts amazoniennes. Pour terminer, le développement futur de ces approches démographiques et les applications plus générales qui peuvent en être faites seront discutés.

V.1 Études démographiques et histoire écologique des forêts du plateau des Guyanes.

L'étude de l'histoire démographique des populations forestières par une approche génétique requiert une bonne connaissance des régions génomiques considérées pour pouvoir par la suite les utiliser de façon appropriée avec des méthodes puissantes d'inférence statistique. Je me suis plus particulièrement intéressée aux marqueurs microsatellites qui du fait de leur facilité d'utilisation et d'interprétation nous ont permis d'étudier la démographie passée des populations de plusieurs espèces. L'utilisation de ces marqueurs microsatellites, étudiés communément par leur taille allélique, nécessite quelques précautions. En effet, l'étude réalisée dans le deuxième chapitre de cette thèse a permis de confirmer la présence de polymorphisme à l'intérieur et en dehors de la région répétée de ce fragment ADN, qui atteste de mutations autres que les variations du nombre de répétitions. Le fait que deux allèles de même taille n'aient pas forcément deux séquences nucléotidiques identiques peut fausser les reconstructions généalogiques basées sur le coalescent que nous avons employées. Plus grave encore, l'utilisation des tailles alléliques avec des modèles de mutation propres aux régions répétées peut entraîner des erreurs lorsque d'autres mutations, évoluant différemment, sont comprises dans les fragments étudiés. Cela est d'autant plus vrai que la majorité du signal phylogéographique des marqueurs microsatellites étudiés est portée par l'information génétique présente dans les régions non répétées. Ainsi, les mesures de distance génétique assumant des modèles de

mutations de type « pas à pas » (SMM) doivent être restreintes aux microsatellites dits « parfaits » dont les séquences encadrant la région répétée ne sont pas polymorphes. Ceci étant rarement le cas, dans les études démographiques menées au troisième et quatrième chapitre de ma thèse, j'ai donc choisi de modéliser mes données microsatellites avec, dans la mesure du possible, des modèles de mutation autres que le SMM strict. Les méthodes de calcul Bayésien approché, utilisées dans ces études démographiques, ont également nécessité le choix de statistiques-descriptives représentant au mieux ces données moléculaires complexes. Pour cela, j'ai choisi, entre autres, des indices de diversité génétique ou des mesures de différenciation, telles que le F_{ST} , basées sur un modèle de mutation propre aux séquences.

V.1.1 Une empreinte durable des perturbations

Ces méthodes d'analyses m'ont permis d'étudier plus particulièrement les perturbations qui ont profondément modulé la distribution et la diversité génétique des communautés forestières du plateau des Guyanes au cours du temps (Figure 22).

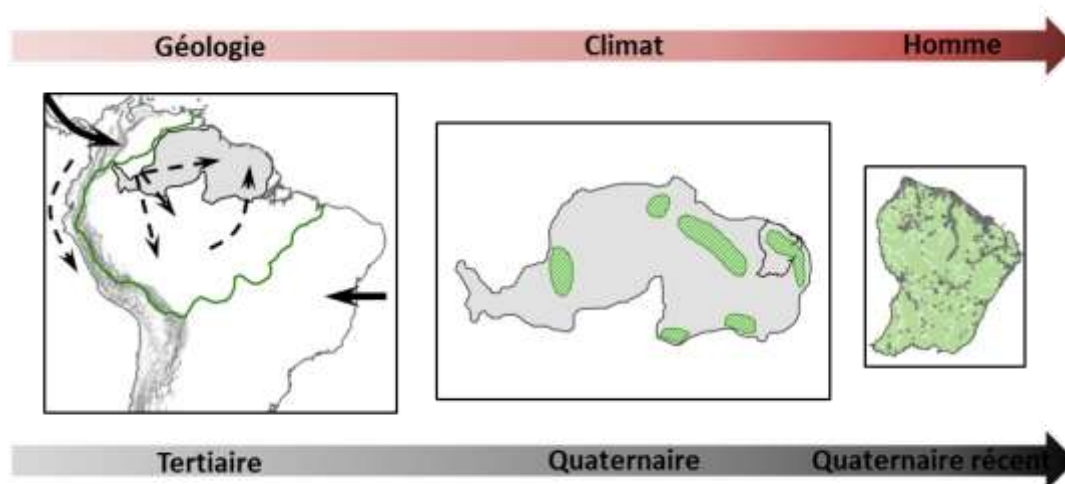


Figure 22. Histoire de la forêt amazonienne en relation avec les perturbations du Tertiaire à nos jours. (A) Arrivée des espèces tropicales (flèches pleines) et colonisation (flèches en pointillées) du continent sud-américain pendant le Tertiaire, adapté d'Antonelli *et al.* (2009). (B) Zones refuges du Quaternaire (en vert) sur le plateau des Guyanes, adaptées de Haffer et Prance (2001). (C) Sites archéologiques amérindiens découverts en Guyane française (DRAC Guyane).

Si les événements de tectonique des plaques et d'orogénèse du Tertiaire ont largement influencé la mise en place des espèces tropicales dans cette région, dans le cadre de cette thèse, j'ai réussi à détecter l'empreinte génétique laissée par des perturbations plus récentes

sur le génome des arbres forestiers en Guyane française. En effet, cette thèse a permis de mettre en évidence des signaux démographiques convergeant, pour la majorité des espèces étudiées, vers une expansion globale et plus ou moins simultanée des populations forestières au niveau régional (Chapitre III). Ces changements de taille des populations, encore visible aujourd'hui, attestent d'une empreinte durable laissée par des perturbations dont nous supposons une origine climatique. Le dernier maximum glaciaire du Pléistocène est le dernier événement climatique important connu pouvant expliquer les transitions démographiques régionales que nous avons détectées mais qui restent difficiles à dater avec précision. Ainsi, au Quaternaire moyen, l'étendue des communautés forestières a probablement été modifiée à large échelle spatiale, supportant l'hypothèse d'une réduction passée du couvert forestier amazonien due à un climat plus sec et/ou plus froid, suivie d'une recolonisation à partir de refuges supposés, identifiés en Amérique du Sud (Haffer, 1969) et plus précisément en Guyane française (de Granville, 1982; Tardy, 1998). Dans le troisième chapitre de cette thèse, des changements démographiques plus locaux et plus récents ont également été détectés. Un résultat majeur et inattendu de l'étude menée est l'existence d'une divergence génétique locale des populations forestières pouvant s'expliquer par un remplacement naturel et constant des populations à l'échelle du paysage qui a entraîné des changements démographiques. Cette étude suggère également que si la présence humaine précolombienne n'a pas modifié cette dynamique forestière, il semble cependant qu'elle ait eue un impact sur la taille efficace des populations participant au renouvellement naturel des forêts. Les travaux de thèse, effectués par Durrieu de Madron (1993) avaient déjà mis en évidence que la taille des trouées forestières naturelles ou d'exploitations conditionnait le type de réponse des arbres : du remplacement local des individus à la mise en place d'une succession végétale. La tendance générale qui se dégage de cette étude est que les perturbations anthropiques peu intenses ont eu un impact plus restreint sur les populations d'arbres que le renouvellement naturel, alors que sur les sites fortement marqués par la présence humaine, faisant notamment partie d'un réseau de sites, l'impact a probablement eu un rayon d'action plus important et plus large que celui de la dynamique naturelle des populations. La confirmation de cette tendance nécessite à l'heure actuelle de pouvoir évaluer l'ampleur de la perturbation exercée par ces occupations anthropiques passées et dans ce cadre, le développement des méthodes de représentation des aires de circulation passées semble une bonne alternative aux fouilles archéologiques approfondies souvent destructrices.

V.1.2 Une empreinte controversée des perturbations

Les études réalisées dans cette thèse suggèrent que les perturbations climatiques anciennes mais également anthropiques ou naturelles très récentes ont écrit l'histoire écologique des forêts en modifiant la dynamique des populations à plusieurs échelles spatiales et ont laissé une trace encore décelable sur les forêts du plateau des Guyanes. Cette empreinte semble cependant controversée. Premièrement, l'approche génétique utilisée dans le troisième chapitre indique que l'action de différents facteurs locaux sur les populations peut entraîner des patrons démographiques similaires, empêchant ainsi de détecter avec certitude des perturbations anthropiques. L'histoire démographique d'une espèce peut être difficile à déterminer car non seulement la dernière perturbation peut modifier ou effacer les empreintes des événements précédents mais l'intervention simultanée de plusieurs perturbations peut également complexifier les patrons génétiques laissés dans le génome des arbres.

Deuxièmement, les résultats de cette thèse suggèrent que les perturbations n'ont pas affecté la démographie passée de toutes les espèces de façon identique. Si des espèces comme le *Virola michelii* apparaissent comme étant de bonnes indicatrices de perturbations locales, le *Carapa surinamensis* semble lui y être insensible (Chapitre III). En plus des divergences démographiques locales entre espèces, nous avons mis en évidence des différences interspécifiques d'histoires démographiques au niveau de la Guyane (Chapitre IV). L'existence de quelques espèces dont la taille de population semble être restée constante au cours du temps vient nuancer l'histoire écologique des forêts guyanaises et laisse supposer un changement de composition des communautés forestières au Quaternaire (Colinvaux *et al.*, 1996), plutôt qu'une réduction drastique et globale du couvert forestier supportée par les expansions détectées pour certaines espèces. Même si aucune preuve n'a pu être apportée au cours de cette thèse, les divers patrons démographiques pourraient s'expliquer de façon générale par les différences de comportement écologique des espèces, donc de réponses à des changements, alors qu'au niveau local d'autres facteurs tels que des différences d'utilisation des espèces forestières par les populations précolombiennes entrent en jeu.

L'empreinte controversée, observée sur les populations actuelles d'arbres tropicaux, est le reflet d'une histoire écologique complexe de ces forêts, composée d'histoires démographiques différentes entre espèces et influencée par plusieurs perturbations locales et/ou globales, simultanées et/ou échelonnées dans le temps. Afin d'affiner les signaux démographiques détectés et la précision des paramètres estimés, dont la datation des

événements démographiques, il serait donc nécessaire de complexifier les modèles testés pour mieux représenter les dynamiques forestières réelles. Par exemple, l'information démographique régionale, qui était encore inconnue lors de la réalisation du troisième chapitre de cette thèse, pourrait être aujourd'hui ajoutée aux modèles de divergence de population des espèces étudiées. De plus, les scénarios testés dans le quatrième chapitre pourraient être améliorés en modélisant des événements démographiques exponentiels continus plus réalistes que des expansions ou des contractions soudaines. L'utilisation des méthodes bayésiennes complètes a également été envisagée et testée pour toutes les études démographiques réalisées. Cependant, le temps important de convergence des chaînes de Markov associé au nombre d'espèces et/ou de sites étudiés ont été ici un frein pour mener à bien les analyses dans un temps raisonnable.

Pour finir, l'étude de l'histoire démographique d'espèces supplémentaires telles que des espèces de savanes ainsi que la multiplication des zones Néotropicales étudiées permettraient également d'améliorer la compréhension de l'histoire écologique des forêts tropicales amazoniennes. Si nous nous basons sur les expansions de population détectées pour plusieurs espèces au niveau régional dans notre étude, on peut facilement supposer que les savanes se sont, elles, réduites proportionnellement pour laisser la place aux forêts. Ceci confirmerait alors la théorie des refuges de Haffer (1969) et la présence de savanes séparant des zones de forêts relictuelles au Quaternaire. Mais si au contraire, nous considérons les résultats des populations régionales restées constantes au cours du temps, on peut alors s'attendre à ce que l'étendue des savanes n'ait pas été modifiée, supportant alors l'hypothèse de Colinvaux *et al.* (1996) et attestant de la présence continue d'un couvert forestier et des flux de gènes associés malgré les perturbations liées aux glaciations.

V.2 Apport de cette étude démographique dans la reconstruction de l'histoire de l'Amazonie

Même si l'histoire des forêts guyanaises semble être controversée entre espèces et que de nombreuses améliorations peuvent encore être apportées, les études démographiques menées dans cette thèse ont permis d'enrichir les maigres connaissances existantes sur le plateau des Guyanes (Tardy, 1998; Caron *et al.*, 2000; Dutech *et al.*, 2004). Le développement des approches démographiques de cette thèse répond à un besoin de compréhension de la dynamique passée des forêts tropicales qui reste encore un mystère non résolu à des grandes échelles de temps par les méthodes existantes et non étudié au niveau local. Plusieurs facteurs sont responsables de cette difficulté à reconstruire l'histoire des forêts tropicales. Premièrement, les changements climatiques du Quaternaire, qui ont entraîné une réduction importante du couvert forestier des régions tempérées et ont laissé une empreinte génétique forte et homogène sur ces communautés, ont été moins drastiques à de plus basses latitudes où, comme nous l'avons vu, l'histoire des forêts est plus contrastée. Ainsi, si les études démographiques des populations forestières de l'hémisphère Nord (Heuertz *et al.*, 2006; Pyhäjärvi *et al.*, 2007; Ingvarsson, 2008; Vendramin *et al.*, 2008; Holliday *et al.*, 2010; Wachowiak *et al.*, 2011) apportent uniquement des précisions sur l'histoire déjà bien connue de ces régions, elles prennent toute leur importance pour reconstruire la dynamique passée dans le cas des forêts tropicales. L'étude des histoires démographiques de plusieurs espèces a été un point fort de cette thèse pour inférer l'histoire écologique générale de cette communauté forestière et a permis de rendre compte de son évolution complexe au cours du temps, ce qui n'aurait pas été possible avec une étude mono-spécifique.

Deuxièmement, les études démographiques menées ont également permis de palier un manque de données palynologiques et fossiles généralement utilisées pour déterminer les changements climatiques et forestiers. Si ces données sont largement disponibles dans la plupart des régions du monde, le plateau des Guyanes semble faire exception, notamment à cause d'une lacune sédimentaire (Ledru *et al.*, 1998) ou encore du faible nombre de prélèvements effectués dans cette région (Ledru *et al.*, 2001; Nogué *et al.*, 2009). Les approches démographiques ont également eu l'avantage de nous permettre de travailler à des échelles spatiales et temporelles beaucoup plus fines et de considérer directement les processus démographiques étudiés, au contraire des approches phylogéographiques qui utilisent de larges patrons de distribution génétique spatialisés pour inférer ces processus.

V.3 Généralisation des études démographiques des communautés

Le développement croissant des méthodes de séquençage à haut débit et l'augmentation de la disponibilité de l'information génétique qui en découle ne peut que renforcer la puissance et la fiabilité des approches démographiques futures. L'étude de Schönberg *et al.* (2011) est un exemple d'utilisation de ces techniques de nouvelle génération pour séquencer le génome mitochondrial entier de l'homme et permettre des inférences démographiques par le calcul Bayésien approché. Baird *et al.* (2008) ont, eux, développé une approche de séquençage de seconde génération permettant la détection et le typage simultané de milliers de SNP dans tout le génome. Cette technique de sites de restriction associés à des régions ADN (marqueurs RAD) se révèle de faible coût et applicable à des organismes non-modèles.

L'acquisition de l'information génétique et son utilisation pour des approches démographiques multi-espèces pourrait dans le futur se développer pour :

- (i) Être utilisée conjointement avec des études phylogéographiques grâce notamment au développement des méthodes bayésiennes ou ABC permettant de prendre en compte des paramètres démographiques et de divergence entre populations ou entre espèces (Laval *et al.*, 2010; Batini *et al.*, 2011; Thalmann *et al.*, 2011; Veeramah *et al.*, 2012).
- (ii) Identifier des espèces indicatrices d'occupations anthropiques, comme cela a déjà été fait avec la composition des communautés végétales en Europe (Dupouey *et al.*, 2002) et plus généralement développer des indicateurs de perturbations locales tels que les tailles efficaces des populations utilisées dans cette thèse.
- (iii) Reconstruire de façon détaillée et à moindre coût l'histoire écologique de n'importe quelle communauté.

Connaitre le passé pour mieux prédire l'avenir

Les approches démographiques ont de nombreuses applications directes mais peuvent également être intégrées dans des domaines plus larges tels que la modélisation. L'intégration des histoires démographiques couplées à l'information sur les perturbations climatiques et/ou anthropiques passées permettrait de modéliser le destin démographique

des espèces, et donc la distribution future des forêts, en réponse à des changements globaux. Cette approche permettrait de valider les modèles écologiques existants (Thuiller *et al.*, 2008) qui n'utilisent généralement que la présence/absence ou l'abondance des espèces actuelles pour représenter les processus démographiques (Araujo et Guisan, 2006; Austin, 2007). Le développement de tels modèles implique la réalisation d'études démographiques qui nécessite relativement peu de moyens et se résume à un échantillonnage et des analyses moléculaires et statistiques. Si les modèles prédictifs des climats sont également déjà disponibles (IPCC, 2007), la plus grande difficulté sera l'obtention de données paléoclimatiques dans des régions du monde peu étudiées.

En Guyane française, si la diminution prévue des précipitations (Cox *et al.*, 2004; Petit et Prudent, 2008) a les mêmes conséquences que la sécheresse passée nous pouvons nous attendre à une réduction du couvert forestier identique à celle détectée dans cette thèse. Cependant, l'augmentation future des températures annuelles de 3,3°C en moyenne sur ce territoire (Cox *et al.*, 2004; IPCC, 2007), qui diffère des conditions plus froides du Pléistocène, et l'impact anthropique en rien comparable à celui des populations précolombiennes passées (Nepstad *et al.*, 2008) laissent envisager des conséquences différentes sur les forêts gyanaises, que seuls des modèles intégrant les informations tirées de l'histoire démographique, de la dynamique des peuplements et du potentiel d'adaptation au milieu permettraient de prévoir : sachant comment les espèces ont répondu aux perturbations par le passé nous pouvons légitimement nous demander quel sera le potentiel d'adaptation de ces forêts face à des changements futurs aussi rapides. Même si l'évolution génomique des arbres est lente, en raison de leur exceptionnelle longévité, il faut espérer que leur grande diversité génétique et la taille de leur population leur permettront de répondre rapidement et de s'adapter en quelques générations (Petit *et al.*, 2008) à de nouvelles conditions.

Bibliographie

- Absy, M. L., A. Cleef, M. Fournier, L. Martin, M. Servant, A. Sifeddine, M. Ferreira da Silva, F. Soubiès, K. Suguio, B. Turcq and T. Van Der Hammen (1991). Mise en évidence de quatre phases d'ouverture de la forêt dense dans le sud-est de l'Amazonie au cours des 60 000 dernières années. Première comparaison avec d'autres régions tropicales. Paris, *Compte rendu de l'Académie des Sciences de Paris* **312**: 673-678.
- Amui-Vedel, A.-M., R. Travela, L. Blanc and P. Palany (2009). Impact of climate change on the ecosystems of the Guiana shield: summary of knowledge and conservation challenges. French Guiana, Cirad-WWF.
- Anderson, L. L., F. S. Hu, D. M. Nelson, R. J. Petit and K. N. Paige (2006). Ice-age endurance: DNA evidence of a white spruce refugium in Alaska. *Proceedings of the National Academy of Sciences* **103**: 12447-12450.
- Antonelli, A., J. A. A. Johan A. A. Nylander, C. Persson and I. Sanmartin (2009). Tracing the impact of the Andean uplift on Neotropical plant evolution. *Proceedings of the National Academy of Sciences of the USA* **106**: 9749-9754.
- Araujo, M. B. and A. Guisan (2006). Five (or so) challenges for species distribution modelling. *Journal of Biogeography* **33**: 1677-1688.
- Austin, M. (2007). Species distribution models and ecological theory: A critical assessment and some possible new approaches. *Ecological modelling* **200**: 1-19.
- Avise, J. C. (1989). Gene trees and organismal histories: a phylogenetic approach to population biology. *Evolution* **43**: 1192-1208.
- Avise, J. C., J. Arnold, R. M. Ball, E. Bermingham, T. Lamb, J. E. Neigel, C. A. Reeb and N. C. Saunders (1987). Intraspecific phylogeography: the mitochondrial DNA bridge between population genetics and systematics. *Annual Review of Ecology and Systematics* **18**: 489-522.
- Baird, N. A., P. D. Etter, T. S. Atwood, M. C. Currey, A. L. Shiver, Z. A. Lewis, E. U. Selker, W. A. Cresko and E. A. Johnson (2008). Rapid SNP discovery and genetic mapping using sequenced RAD markers. *PLoS ONE* **3**: e3376.
- Batini, C., J. Lopes, D. M. Behar, F. Calafell, L. B. Jorde, L. Van der Veen, L. Quintana-Murci, G. Spedini, G. Destro-Bisol and D. Comas (2011). Insights into the demographic history of African pygmies from complete mitochondrial genomes. *Molecular Biology and Evolution* **28**: 1099-1110.
- Bayes, M. and M. Price (1763). An essay towards solving problem in the doctrine of chances. *Philosophical Transactions of the Royal Society B* **53** 370-418.
- Beaumont, M. A. (2010). Approximate Bayesian Computation in Evolution and Ecology. *Annual Review of Ecology, Evolution and Systematics* **41**: 379-405.
- Beaumont, M. A., J.-M. Cornuet, J. M. Marin and C. P. Robert (2009). Adaptive approximate Bayesian computation. *Biometrika* **96**: 983-90.
- Beaumont, M. A. and B. Rannala (2004). The bayesian revolution in genetics. *Genetics* **5**: 251-261.
- Beaumont, M. A., W. Zhang and D. J. Balding (2002). Approximate Bayesian Computation in Population Genetics. *Genetics* **162**: 2025-2035.
- Bertorelle, G., A. Benazzo and S. Mona (2010). ABC as a flexible framework to estimate demography over space and time: some cons, many pros. *Molecular Ecology* **19**: 2609-2625.

- Birky, C. W., Jr. (1995). Uniparental inheritance of mitochondrial and chloroplast genes: mechanisms and evolution. *Proceedings of the National Academy of Sciences of the USA* **92**: 11331-11338.
- Blum, M. G. B. and O. François (2010). Non-linear regression models for Approximate Bayesian Computation. *Statistics and computing* **20**: 63-73.
- Boggan, J., V. Funk, C. Kelloff, M. Hoff, G. Cremers and C. Feuillet (1997). Checklist of the plants of the Guianas (Guyana, Surinam, French Guiana). Washington, DC, US, National museum of Natural History, Smithsonian Institution.
- Boomert, A. (1980). Hertenrits: an Arauquinoid complex in North West Suriname. *Journal of Archaeology and Anthropology* **3**: 68-104.
- Bortot, P., S. G. Coles and S. A. Sisson (2007). Inference for stereological extremes. *Journal of the American Statistical Association* **102**: 84-92.
- Brown, K. S. J. (1987). Conclusions, synthesis, and alternative hypotheses. *Biogeography and Quaternary history in tropical America*. T. C. Whittmore and D. T. Plance. Oxford, Clarendon Press: 175-196.
- Brown, K. S. J., P. M. Sheppard and J. R. G. Turner (1974). Quaternary refugia in tropical America: evidence from race formation in *Heliconius* butterflies. *Proceedings of the Royal Society of London B* **187**: 369-378.
- Brunsfeld, S. J., J. Sullivan, D. E. Soltis and P. S. Soltis (2001). Comparative phylogeography of north-western North America: a synthesis. *Integrating ecology and evolution in a spatial context*. J. S. J. Antonovics, Oxford: Blackwells: 319-339.
- Burbridge, R. E., F. E. Mayle and T. Killeen (2004). Fifty-thousand-year vegetation and climate history of Noel Kempff Mercado National Park, Bolivian Amazon. *Quaternary Research* **61**: 215-230.
- Bush, M. B. and P. A. Colinvaux (1990). A pollen record of a complete glacial cycle from lowland Panama. *Journal of Vegetation Science* **1**: 105-118.
- Bush, M. B., P. A. Colinvaux, M. C. Wiemann, D. R. Piperno and K.-B. Liu (1990). Late Pleistocene temperature depression and vegetation change in Ecuadorian Amazonia. *Quaternary Research* **34**: 330-345.
- Bush, M. B. and P. E. De Oliveira (2006). The rise and fall of the Refugial Hypothesis of Amazonian Speciation: a paleo-ecological perspective. *Biota Neotropica* **6**.
- Bush, M. B., M. C. Miller, P. E. De Oliveira and P. A. Colinvaux (2002). Orbital forcing signal in sediments of two Amazonian lakes. *Journal of Paleolimnology* **27**: 341-352.
- Bush, M. B., M. R. Silman, C. Mc Michael and S. S. (2008). Fire, climate change and biodiversity in Amazonia: a Late-Holocene perspective. *Philosophical Transactions of the Royal Society B* **363**: 1795-1802.
- Bush, M. B., M. R. Silman and D. H. Urrego (2004). 48,000 years of climate and forest change in a biodiversity hot spot. *Science* **303**: 827-829.
- Caron, H., S. Dumas, G. Marque, C. Messier, E. Bandou, R. J. Petit and A. Kremer (2000). Spatial and temporal distribution of chloroplast DNA polymorphism in a tropical tree species. *Molecular Ecology* **9**: 1089-1098.
- Charles-Dominique, P., P. Blanc, D. Larpin, M.-P. Ledru, B. RiCra, C. Sarthou, M. Servant and C. Tardy (1998). Forest perturbations and biodiversity during the last ten thousand years in French Guiana *Acta Oecologica* **19**: 295-302.
- Clark, P. U., R. B. Alley and D. Pollard (1999). Northern hemisphere ice-sheet influences on global climate change. *Science* **286**: 1104-1111.

- Colinvaux, P. A. (1987). Amazon diversity in light of the paleoecological record *Quaternary Science Reviews* **6**.
- Colinvaux, P. A., P. E. De Oliveira and M. B. Bush (2000). Amazonian and neotropical plant communities on glacial time-scales: The failure of the aridity and refuge hypotheses. *Quaternary Science Reviews* **19**: 141-169.
- Colinvaux, P. A., P. E. De Oliveira, J. E. Moreno, M. C. Miller and M. B. Bush (1996). A long pollen record from lowland Amazonia: Forest and cooling in glacial times. *Science* **274**: 85-88.
- Cornuet, J.-M., F. Santos, M. A. Beaumont, C. P. Robert, J.-M. Marin, D. J. Balding, T. Guillemaud and A. Estoup (2008). Inferring population history with DIY ABC: a user-friendly approach to approximate Bayesian computation. *Bioinformatics* **24**: 2713-2719.
- Cox, P. M., R. A. Betts, M. Collins, P. P. Harris, C. Huntingford and C. D. Jones (2004). Amazonian forest dieback under climate-carbon cycle projections for the 21st century. *Theoretical and Applied Climatology* **78**: 137-156.
- Csilléry, K., M. G. B. Blum, O. E. Gaggiotti and O. François (2010). Approximate Bayesian Computation (ABC) in practice. *Trends in Ecology & Evolution* **25**: 410-418.
- Dauby, G., J. Jérôme Duminil, M. Heuertz and O. J. Hardy (2010). Chloroplast DNA polymorphism and phylogeography of a central african tree species widespread in mature rainforests: *Greenwayodendron suaveolens* (Annonaceae). *Tropical Plant Biology* **3**: 4-13.
- de Granville, J. J. (1982). Rain forest and xeric flora refuges in French Guiana. *Biological diversification in the Tropics*. G. T. Prance. New York, Columbia University Press: 137-158.
- de Souza, E. B., M. T. Kayano, J. Tota, L. Pezzi, G. Fisch and C. Nobre (2000). On the influence of the el Nino, la Nina and atlantic dipole pattern on the amazonian rainfall during 1960-1998. *Acta Amazonica* **30**: 305-318.
- Del Moral, P., A. Doucet and A. Jasra (2012). An adaptive sequential Monte Carlo method for approximate Bayesian computation. *Statistics and computing* **22**: 1009-1020.
- Di Rienzo, A., A. C. Peterson, J. C. Garza, A. M. Valdes, M. Slatkin and N. B. Freimer (1994). Mutational processes of simple-sequence repeat loci in human populations. *Proceedings of the National Academy of Sciences of the USA* **91**.
- Dick, C. W. and M. Heuertz (2008). The complex biogeographic history of a widespread tropical tree species. *Evolution* **62**: 2760-2774.
- Doksum, K. A. and A. Y. Lo (1990). Consistent and robust bayes procedures for location based on partial information. *The Annals of Statistics* **18**: 443-453.
- Dupouey, J. L., E. Dambrine, J. D. Laffite and C. Moares (2002). Irreversible impact of past land use on forest soils and biodiversity. *Ecology* **83**: 2978-2984.
- Durrieu de Madron, L. (1993). *Mortalité, châblis et rôle des trouées dans la sylvigénèse avant et après exploitation sur le dispositif d'étude sylvicole de Paracou - Guyane française*. Thèse de Doctorat, Ecole nationale du génie rural, des eaux et des forêts, Paris, France, 193p.
- Dutech, C., H. I. Joly and P. Jarne (2004). Gene flow, historical population dynamics and genetic diversity within French Guianan populations of a rainforest tree species, *Vouacapoua americana*. *Heredity* **92**: 69-77.

- Eden, M. J., W. Bray, L. Herrera and C. McEwan (1984). Terra preta soils and their archaeological context in the Caqueta Basin of Southeast Colombia. *American Antiquity* **49**: 125-140.
- EPICA (2004). Eight glacial cycles from an Antarctic ice core. *Nature* **429**: 623-628.
- Estoup, A., P. Jarne and J.-M. Cornuet (2002). Homoplasy and mutation model at microsatellite loci and their consequences for population genetics analysis. *Molecular Ecology* **11**: 1591-1604.
- Ewens, W. J. (1972). The sampling theory of selectively neutral alleles. *Theoretical population biology* **3**: 87-112.
- Excoffier, L., J. Novembre and S. Schneider (2000). SIMCOAL: a general coalescent program for the simulation of molecular data in interconnected populations with arbitrary demography. *The Journal of Heredity* **91**: 506-509.
- Fedorov, V. B. and N. C. Stenseth (2002). Multiple glacial refugia in the North American Arctic: inference from phylogeography of the collared lemming (*Dicrostonyx groenlandicus*). *Proceedings of the Royal Society of London B* **269**: 2071-2077.
- Fisher, R. A. (1930). The genetical theory of natural selection. Oxford, UK, Oxford University Press.
- Freycon, V., M. Krencker, D. Schwartz, R. Nasi and D. Bonal (2010). The impact of climate changes during the Holocene on vegetation in northern French Guiana. *Quaternary Research* **73**: 220-225.
- Fu, Y.-X. and W.-H. Li (1993). Statistical tests of neutrality of mutations. *Genetics* **133**: 693-709.
- Good, P., J. A. Lowe, M. Collins and W. Moufouma-Okia (2008). An objective tropical Atlantic sea surface temperature gradient index for studies of south Amazon dry-season climate variability and change *Philosophical Transactions of the Royal Society B* **363**: 1761-1766.
- Grenand, P., C. Moretti, H. Jacquemin and M. F. Prévost (2004). Pharmacopées traditionnelles en Guyane.
- Grivet, D., M.-F. Deguilloux, R. J. Petit and V. L. Sork (2006). Contrasting patterns of historical colonization in white oaks (*Quercus spp.*) in California and Europe. *Molecular Ecology* **15**: 4085-4093.
- Haffer, J. (1969). Speciation in amazonian forest birds. *Science* **165**: 131-137.
- Haffer, J. (1997). Alternative models of vertebrate speciation in Amazonia: an overview. *Biodiversity and Conservation* **6**: 451-476.
- Haffer, J. and G. T. Prance (2001). Climatic forcing of evolution in Amazonia during the Cenozoic: on the refuge theory of biotic differentiation *Amazoniana* **16**: 579-605.
- Hammond, D. S. (2005). Tropical forests of the Guiana shield. Ancient forests in a modern world. Wallingford, UK and Cambridge, MA, USA, CABI publishing.
- Harpending, H. C., M. A. Batzer, M. Gurven, L. B. Jorde, A. R. Rogers and S. T. Sherry (1998). Genetic traces of ancient demography. *Proceedings of the National Academy of Sciences of the USA* **95**: 1961-1967.
- Heuertz, M., E. De Paoli, T. Kallman, H. Larsson, I. Jurman, M. Morgante, M. Lascoux and N. Gyllenstrand (2006). Multilocus patterns of nucleotide diversity, linkage disequilibrium and demographic history of Norway Spruce [*Picea abies* (L.) Karst]. *Genetics* **174**: 2095-2105.
- Hewitt, G. M. (2000). The genetic legacy of the Quaternary ice ages. *Nature* **405**: 907-913.

- Holliday, J. A., M. Yuen, K. Ritland and S. N. Aitken (2010). Postglacial history of a widespread conifer produces inverse clines in selective neutrality tests. *Molecular Ecology* **19**: 3857-3864.
- Hoorn, C., F. P. Wesselingh, H. ter Steege, M. A. Bermudez, A. Mora, J. Sevink, I. Sanmartín, A. Sanchez-Meseguer, C. L. Anderson, J. P. Figueiredo, C. Jaramillo, D. Riff, F. R. Negri, H. Hooghiemstra, J. Lundberg, T. Stadler, T. Särkinen and A. Antonelli (2010). Amazonia through time: Andean uplift, climate change, landscape evolution, and biodiversity. *Science* **330**: 927-931.
- Ingvarsson, K. (2008). Multilocus patterns of nucleotide polymorphism and the demographic history of *Populus tremula*. *Genetics society of America* **180**: 329-340.
- IPCC (2007). IPCC Fourth assessment report: Climate Change 2007 (AR4). Cambridge, United Kingdom and New York, USA, Cambridge University Press.
- Jaramillo-Correa, J. P., J. Beaulieu and J. Bousquet (2004). Variation in mitochondrial DNA reveals multiple distant glacial refugia in black spruce (*Picea mariana*), a transcontinental North American conifer. *Molecular Ecology* **13**: 2735-2747.
- Kashi, Y., D. King and M. Soller (1997). Simple sequence repeats as a source of quantitative genetic variation. *Trends in Genetics* **13**: 74-78.
- Kimura, M. (1968). Evolutionary rate at the molecular level *Nature* **217**: 624-626.
- Kimura, M. (1969). The number of heterozygous nucleotide sites maintained in a finite population due to steady flux. *Genetics* **61**: 893-903.
- Kimura, M. (1985). The neutral theory of molecular evolution. Cambridge, UK, Cambridge University Press.
- Kimura, M. (1985). The role of compensatory neutral mutations in molecular evolution. *Journal of Genetics* **64**: 7-19.
- Kimura, M. (1986). DNA and the neutral theory. *Philosophical Transactions of the Royal Society of London B* **312**: 343-354.
- Kimura, M. and J. F. Crow (1964). The number of alleles that can be maintained in a finite population. *Genetics* **49**: 725-738.
- Kingman, J. F. C. (1982). On the genealogy of large populations. *Journal of Applied Probability* **19**: 27-43.
- Kuhner, M. K. (2008). Coalescent genealogy samplers: windows into population history. *Trends in Ecology and Evolution* **24**: 86-93.
- Laval, G., E. Patin, L. B. Barreiro and L. Quintana-Murci (2010). Formulating a historical and demographic model of recent human evolution based on resequencing data from noncoding regions. *PLoS One* **5**: e10284.
- Lavallée, D. (1995). Promesse d'Amérique. La préhistoire de l'Amérique du Sud. Paris, Hachette.
- Ledru, M.-P., J. Bertaux and A. Sifeddine (1998). Absence of last glacial maximum records in lowland tropical forests. *Quaternary Research* **49**: 233-237.
- Ledru, M.-P., R. C. Cordeiro, J. M. L. Dominguez, L. Martin, P. Mourguiart, A. Sifeddine and B. Turcq (2001). Late-glacial cooling in Amazonia inferred from pollen at Lagoa do Caço, northern Brazil. *Quaternary Research* **55**: 47-56.
- Ledru, M.-P., D.-D. Rousseau, F. W. Cruz Jr., C. Riccomini, I. Karmann and L. Martin (2005). Paleoclimate changes during the last 100,000 yr from a record in the Brazilian Atlantic rainforest region and interhemispheric comparison. *Quaternary Research* **64**: 444 - 450.

- Legates, D. R. and C. J. Wilmott (1990). Mean seasonal and spatial variability in global surface air temperature *Theoretical and Applied Climatology* **41**: 11-21.
- Lombardo, U., E. Canal-Beeby, S. Fehr and H. Veit (2011). Raised fields in the Bolivian Amazonia: a prehistoric green revolution or a flood risk mitigation strategy? *Journal of Archaeological Science* **38**: 502-512.
- Lopes, J. S., D. Balding and M. A. Beaumont (2009). PopABC: a program to infer historical demographic parameters. *Bioinformatics* **25**: 47-49.
- Lowe, A. J., D. Harris, E. Dormontt and I. K. Dawson (2010). Testing putative african tropical forest refugia using chloroplast and nuclear DNA phylogeography. *Tropical Plant Biology* **3**: 50-58.
- Maley, J. (1991). The african rain forest vegetation and palaeoenvironments during late Quaternary. *Climate Change* **19**: 79-98.
- Maley, J. (1996). The African rain forest. Main characteristics of changes in vegetation and climate from the Upper Cretaceous to the Quaternary *Proceedings of the Royal Society of Edinburgh* **104B**: 31-73.
- Marjoram, P., J. Molitor, V. Plagnol and S. Tavaré (2003). Markov chain Monte Carlo without likelihoods. *Proceedings of the National Academy of Sciences* **100**: 15324-15328.
- Markgraf, V. (1989). Palaeoclimates in central and south america since 18,000 BP based on pollen and lake-level records. *Quaternary Science Reviews* **8**: 1-24.
- Martin, L., J. Bertaux, T. Corrège, M.-P. Ledru, P. Mourguiart, A. Sifeddine, F. Soubiès, D. Wirrmann, K. Suguio and B. Turcq (1997). Astronomical forcing of contrasting rainfall changes in tropical South America between 12,400 and 8800 cal yr B.P. *Quaternary Research* **47**: 117-122.
- Mayle, F. E., D. J. Beerling, W. D. Gosling and M. B. Bush (2004). Responses of Amazonian ecosystems to climatic and atmospheric carbon dioxide changes since the last glacial maximum. *Philosophical transactions of the royal society Lond. B* **359**: 499-514.
- Mayle, F. E., R. Burbridge and T. J. Killeen (2000). Millennial-scale dynamics of southern Amazonian rain forests. *Science* **290**: 2291-2294.
- Mayle, F. E. and M. J. Power (2008). Impact of a drier early–mid-Holocene climate upon Amazonian forests. *Philosophical transactions of the royal society B* **363**: 1829-1838.
- Mazière, M. and G. Mazière (1994). L'archéologie amérindienne en Guyane. Etat actuel de la recherche. *Bulletin de la Société Préhistorique Française* **91**: 333-341.
- Mazières, S. (2006). *Le peuplement amérindien de la Guyane française: apports des marqueurs moléculaires* Thèse de Doctorat, Paul Sabatier, Toulouse III, France, 244.
- McKey, D., R. Rostain, J. Iriarte, B. Glaser, J. J. Birk, I. Holste and D. Renarda (2010). Pre-Columbian agricultural landscapes, ecosystem engineers, and self-organized patchiness in Amazonia. *Proceedings of the National Academy of Sciences* **107**: 7823–7828.
- Mestre, M. and S. Delpech (2008). Saint-Laurent du Maroni. Plateau des Mines. *Rapport de fouille INRAP*.
- Migeon, G. (2009). Eléments pour une esquisse de l'occupation amérindienne ancienne de la forêt guyanaise. *Amaz'Hommes. Sciences de l'Homme et sciences de la Nature en Amazonie*. E. Barone-Visigalli and A. Roosevelt: 109-131.

- Milligan, B. G., J. Leebens-Mack and A. E. Strand (1994). Conservation genetics: beyond the maintenance of marker diversity. *Molecular Ecology* **3**: 423-435.
- Molinier, M., J.-L. Guyot, E. de Oliveira and V. Guimaraes (1996). Les régimes hydrologiques de l'Amazonie et de ses affluents. L'hydrologie tropicale: géoscience et outil pour le développement Actes de la conférence de Paris, IAHS 209-222.
- Molinier, M., J. L. Guyot, G. Cochonneau, V. Guimarães and E. de Oliveira (1999). La variabilité hydrologique du bassin amazonien et la circulation atmosphérique océanique. Hydrological and Geochemical Processes in Large Scale River Basins. Manaus.
- Montoya, E., V. Rull, N. D. Stansell, B. W. Bird, S. Nogué, T. Vegas-Vilarrubia, M. B. Abbott and W. Diaz (2011). Vegetation changes in the Neotropical Gran Sabana (Venezuela) around the Younger Dryas chron. *Journal of Quaternary Science* **26**: 207-218.
- Nepstad, D. C., C. M. Stickler, B. Soares-Filho and F. Frank Merry (2008). Interactions among Amazon land use, forests and climate: prospects for a near-term forest tipping point. *Philosophical Transactions of the Royal Society B* **363**: 1737-1746.
- Nogué, S., V. Rull, E. Montoya, O. Huber and T. Vegas-Vilarrubia (2009). Paleocology of the Guayana Highlands (northern South America): Holocene pollen record from the Eruoda-tepui, in the Chimantá massif. *Palaeogeography, Palaeoclimatology, Palaeoecology* **281**: 165-173.
- Ohta, T. and M. Kimura (1973). A model of mutation appropriate to estimate the number of electrophoretically detectable alleles in a finite population. *Genetical Research, Cambridge* **22**: 201-204.
- Olsen, P. E. and D. V. Kent (1996). Milankovitch climate forcing in the tropics of Pangaea during the late Triassic. *Palaeogeography, Palaeoclimatology, Palaeoecology* **122**: 1-26.
- Pagán Jiménez, J. R. (2011). Early phytocultural processes in the pre-Colonial Antilles. *Essays in archaeology, ethnohistory & ethnography of the Amerindian Circum-Caribbean*. C. Hofman and A. van Duijvenbode. Leiden, Sidestone Press: 87-116.
- Petit, J. and G. Prudent (2008). Climate change and biodiversity in the European Union overseas entities, UICN.
- Petit, R. J., I. Aguinagalde, J.-L. de Beaulieu, C. Bittkau, S. Brewer, R. Cheddadi, R. Ennos, S. Fineschi, D. Grivet, M. Lascoux, A. Mohanty, G. Muller-Starck, B. Demesure-Musch, P. Palme, J. P. Martin, S. Rendell and G. G. Vendramin (2003). Glacial refugia: Hotspots but not melting pots of genetic diversity. *Science* **300**: 1563-1565.
- Petit, R. J., S. Brewer, S. Bordacs, K. Burg, R. Cheddadi, E. Coart, J. Cottrell, U. M. Csaikl, B. van Dam, J. D. Deans, S. Espinel, S. Fineschi, R. Finkeldey, I. Glaz, P. G. Goicoechea, J. S. Jensen, A. O. König, A. J. Lowe, S. F. Madsen, G. Matyas, R. C. Munro, F. Popescu, D. Slade, H. Tabbener, S. G. M. de Vries, B. Ziegenhagen, J.-L. de Beaulieu and A. Kremer (2002). Identification of refugia and post-glacial colonisation routes of European white oaks based on chloroplast DNA and fossil pollen evidence. *Forest Ecology and Management* **156**: 49-74.
- Petit, R. J., F. S. Hu and C. W. Dick (2008). Forests of the past: a window to future changes. *Science* **320**: 1450-1452.
- Petitjean Roget, H. (1991). 50 sites de montagne en Guyane française : contribution à l'inventaire archéologique d'Emile Abonnenc. Proceedings of the Twelfth Congress of

- the International Association of Caribbean Archaeology, Cayenne 1987, International Association Caribbean Archaeology, Martinique.
- Power, M. J., M. B. Bush, H. Behling, S. P. Horn, F. E. Mayle and D. H. Urrego (2010). Paleofire activity in tropical America during the last 21 ka: A regional synthesis based on sedimentary charcoal *Pages news* **18**: 73-75.
- Prance, G. T. (1973). Phytogeographic support for the theory of Pleistocene forest refuges in the Amazon Basin, based upon evidence from distribution pattern in Caryocaraceae, Chrysobalanaceae, Dichapetalaceae and Lecythydaceae. *Acta Amazon* **3**: 5-28.
- Pritchard, J. K., M. T. Seielstad, A. Perez-Lezaun and M. W. Feldman (1999). Population growth of Human Y chromosomes: A study of Y chromosome microsatellites. *Molecular Biology and Evolution* **16**: 1791-1798.
- Provan, J. and K. D. Bennett (2008). Phylogeographic insights into cryptic glacial refugia. *Trends in Ecology and Evolution* **23**: 564-571.
- Pyhäjärvi, T., M. R. Garcia-Gil, T. Knurr, M. Mikkonen, W. Wachowiak and O. Savolainen (2007). Demographic History Has Influenced Nucleotide Diversity in European *Pinus sylvestris* Populations. *Genetics* **177**: 1713-1724.
- Rial, J. A. (1999). Pacemaking the Ice Ages by frequency modulation of Earth's orbital eccentricity. *Science* **265**: 564-568.
- Ridley, M. (2004). *Evolution*. Oxford, Blackwell.
- Rosenberg, N. A. and M. Nordborg (2002). Genealogical trees, coalescent theory and the analysis of genetic polymorphisms. *Nature reviews Genetics* **3**: 380-390.
- Rostain, R. (1994). *L'occupation amérindienne ancienne du littoral de Guyane*. Thèse de Doctorat, Panthéon-Sorbonne, Paris I, 721p.
- Rostain, S. (2008). The archaeology of the Guianas: an overview. *Handbook of South American Archaeology*. H. Silverman and W. H. Isbell. New York, Springer: 279-302.
- Scarpato Cunha, A. A. and E. A. M. Koutsoukos (2001). Orbital cyclicity in a Turonian sequence of the Cotinguiba formation, Sergipe basin, NE Brazil. *Cretaceous Research* **22**: 529-548.
- Schönberg, A., C. Theunert, M. Li, M. Stoneking and I. Nasidze (2011). High-throughput sequencing of complete human mtDNA genomes from the Caucasus and West Asia: high diversity and demographic inferences. *European Journal of Human Genetics* **19**: 988-994.
- Scotti-Saintagne, C., C. W. Dick, H. Caron, V. G. G., V. Troispoux, P. Sire, M. Casalis, A. Buonamici, R. Valencia, M. R. Lemes, R. Gribel and I. Scotti (2012a). Amazon diversification and cross-Andean dispersal of the widespread Neotropical tree species *Jacaranda copaia* (Bignoniaceae). *Journal of Biogeography*: 10.1111/j.1365-2699.2012.02797.x.
- Scotti-Saintagne, C., C. W. Dick, H. Caron, G. G. Vendramin, E. Guichoux, A. Buonamici, C. Duret, P. Sire, R. Valencia, M. R. Lemes, R. Gribel and I. Scotti (2012b). Phylogeography of a species complex of lowland Neotropical rain forest trees (*Carapa*, Meliaceae). *Journal of Biogeography*: 10.1111/j.1365-2699.2011.02678.x.
- Servant, M., J. Maley, B. Turcq, M.-L. Absy, P. Brenac, M. Fournier and M.-P. Ledru (1993). Tropical forest changes during the Late Quaternary in African and South American lowlands *Global and Planetary Change* **7**: 25-40.

- Servant, M. and S. Servant-Vildary (1996). Dynamique à long terme des écosystèmes forestiers intertropicaux. Paris.
- Shafer, A. B., C. I. Cullingham, S. D. Côté and D. W. Coltman (2010). Of glaciers and refugia: a decade of study sheds new light on the phylogeography of northwestern North America. *Molecular Ecology* **19**: 4589-4621.
- Sisson, S. A., Y. Fan and M. M. Tanaka (2007). Sequential Monte Carlo without likelihoods. *Proceedings of the National Academy of Sciences* **104**: 1760-1765.
- Stephens, M. (2001). Inference under the coalescent. *Handbook of Statistical genetics*. D. J. Balding, M. Bishop and C. Cannings. England, Wiley.
- Sternberg, L. D. (2001). Savanna-forest hysteresis in the tropics. *Global Ecology and Biogeography* **10**: 369-378.
- Stirling, C. H., T. M. Esat, K. Lambeck, M. T. McCulloch, S. G. Blake, D.-C. Lee and A. N. Halliday (2001). Orbital forcing of the marine isotope stage 9 Interglacial. *Science* **291**: 290-293.
- Taberlet, P., L. Fumagalli, A.-G. Wust-Saucy and J.-F. Cosson (1998). Comparative phylogeography and postglacial colonization routes in Europe. *Molecular Ecology* **7**: 453-464.
- Tajima, F. (1989). Statistical method for testing the Neutral mutation hypothesis by DNA polymorphism *Genetics* **123**: 585-595.
- Tardy, C. (1998). *Paléoincendies naturels, feux anthropiques et environnements forestiers de Guyane française du tardiglaciaire à l'holocène récent. Approches chronologique et anthracologique*. Thèse de Doctorat, Sciences et Techniques, Montpellier II, France, 343.
- Tavaré, S., D. J. Balding, R. C. Griffiths and P. Donnelly (1997). Inferring coalescence times from DNA sequence data *Genetics* **145**: 505-518.
- Thalmann, O., D. Wegmann, M. Spitzner, M. Arandjelovic, K. Guschanski, C. Leuenberger, R. A. Bergl and L. Vigilant (2011). Historical sampling reveals dramatic demographic changes in western gorilla populations. *BMC Evolutionary Biology* **11**: 85.
- Thuiller, W., C. Albert, M. B. Araujo, P. M. Berry, M. Cabeza, A. Guisan, T. Hickler, G. F. Midgley, J. Paterson, F. M. Schurr, M. T. Sykes and N. E. Zimmermann (2008). Predicting global change impacts on plant species' distributions: Future challenges. *Perspectives in Plant Ecology, Evolution and Systematics* **9**: 137-152.
- Thuillet, A.-C., D. Bru, J. David, P. Roumet, S. Santoni, P. Sourdille and T. Bataillon (2002). Direct Estimation of Mutation Rate for 10 Microsatellite Loci in Durum Wheat, *Triticum turgidum* (L.) Thell. ssp *durum* desf. *Molecular Biology and Evolution* **19**: 122-125.
- Toth, G., Z. Gaspari and J. Jurka (2000). Microsatellites in different eukaryotic genomes: Survey and analysis. *Genome Research* **10**: 967-981.
- Vacher, S., S. Jérémie and J. Briand (1998). Amérindiens du Sinnamary (Guyane) : archéologie en forêt équatoriale. *Document d'Archéologie française* Paris, Maison des Sciences de l'homme. **70**.
- Valencia, R., R. B. Foster, G. Villa, R. Condit, J.-C. Svenning, C. Hernandez, K. Romoleroux, E. Losos, E. Magård and H. Balslev (2004). Tree species distributions and local habitat variation in the Amazon: large forest plot in eastern Ecuador. *Journal of Ecology* **92**: 214 -229.

- van den Bel, M., N. Serrand, L. Bruxelles, P. Verdin, J. R. Pagán Jiménez, J. Brancier, C. Cammas, J. Birk, B. Glazer and S. Knippenberg (2011). Chemin Saint-Louis. Guyane française, Rapport de fouille Inrap.
- Vargas-Ramírez, M., J. Maran and U. Fritz (2010). Red- and yellow-footed tortoises (*Chelonoidis carbonaria* and *C. denticulata*), in South American savannahs and forests: do their phylogeographies reflect distinct habitats? *Organisms Diversity and Evolution* **10**: 161-172.
- Veeramah, K. R., D. Wegmann, A. Woerner, F. L. Mendez, J. C. Watkins, G. Destro-Bisol, H. Soodyall, L. Louie and M. F. Hammer (2012). An early divergence of KhoeSan ancestors from those of other modern Humans is supported by an ABC-based analysis of autosomal resequencing data. *Molecular Biology and Evolution* **29**: 617-630.
- Vendramin, G. G., B. Fady, S. C. Gonzalez-Martinez, F. S. Hu, I. Scotti, F. Sebastiano, A. Sotto and R. J. Petit (2008). Genetically depauperate but widespread: the case of an emblematic mediterranean pine. *Evolution* **62**: 680-688.
- Wachowiak, W., M. J. Salmela, R. A. Ennos, G. Iason and S. Cavers (2011). High genetic diversity at the extreme range edge: nucleotide variation at nuclear loci in Scots pine (*Pinus sylvestris* L.) in Scotland. *Heredity* **106**: 775-787.
- Wagner, F. (2011). *La réponse des forêts tropicales humides aux variations climatiques. Evolution de la structure et de la dynamique des peuplements forestiers guyanais.* Thèse de Doctorat, Université des Antilles et de la Guyane, Kourou, Guyane française, 250.
- Waltari, E., R. J. Hijmans, A. T. Peterson, A. S. Nyari, S. L. Perkins and R. P. Guralnick (2007). Locating Pleistocene refugia: comparing phylogeographic and ecological niche model predictions. *PLoS ONE* **2**: e563.
- Watterson, G. A. (1978). The homozygosity test of neutrality. *Genetics* **88**: 405-417.
- Watterson, G. A. (1986). The homozygosity test after a change in population size. *Genetics* **122**: 899-907.
- Wegmann, D., C. Leuenberger, S. Neuenschwander and L. Excoffier (2010). ABCtoolbox: a versatile toolkit for approximate Bayesian computations. *BMC Bioinformatics* **2010**, : **11**: 116.
- Wolfe, K., W.-H. Li and P. M. Sharp (1987). Rates of nucleotide substitution vary greatly among plant mitochondrial, chloroplast, and nuclear DNAs. *Proceedings of the National Academy of Sciences of the USA* **84**: 9054-9058.
- Wright, S. (1929). Fisher's theory of dominance. *American Naturalist* **63**: 274-279.
- Wright, S. (1931). Evolution in mendelian populations. *Genetics* **16**: 97-159.
- Zhivotovsky, L. A., M. W. Feldman and S. A. Grishchkin (1997). Biased mutations and microsatellite variation. *Molecular Biology and Evolution* **14**: 926-933.

Annexes

Annexe 1

Always Look on Both Sides: Phylogenetic Information Conveyed by Simple Sequence Repeat Allele Sequences

Stéphanie Barthe¹, Felix Gugerli², Noelle A. Barkley³, Laurent Maggia⁴, Céline Cardi⁵, Ivan Scotti^{6*}

1 Unité Mixte de Recherche "Ecologie des forêts de Guyane", University of French West Indies and French Guiana, Kourou, French Guiana, **2** Biodiversity and Conservation Biology Research Unit, Swiss Federal Research Institute for Forest, Snow and Landscape Research, Birmensdorf, Switzerland, **3** Plant Genetic Resources Conservation Unit, United States Department of Agriculture – Agricultural Research Service, Griffin, Georgia, United States of America, **4** Unité Mixte de Recherche "Amélioration génétique et adaptation des plantes méditerranéennes et tropicales", Institut Agronomique néo-Calédonien, Nouméa, New Caledonia, **5** Unité Mixte de Recherche "Amélioration génétique et adaptation des plantes méditerranéennes et tropicales", Centre de coopération internationale en recherche agronomique pour le développement, Montpellier, France, **6** Unité Mixte de Recherche Ecologie des forêts de Guyane, Institut National de la Recherche Agronomique, Kourou, French Guiana

Abstract

Simple sequence repeat (SSR) markers are widely used tools for inferences about genetic diversity, phylogeography and spatial genetic structure. Their applications assume that variation among alleles is essentially caused by an expansion or contraction of the number of repeats and that, accessorially, mutations in the target sequences follow the stepwise mutation model (SMM). Generally speaking, PCR amplicon sizes are used as direct indicators of the number of SSR repeats composing an allele with the data analysis either ignoring the extent of allele size differences or assuming that there is a direct correlation between differences in amplicon size and evolutionary distance. However, without precisely knowing the kind and distribution of polymorphism within an allele (SSR and the associated flanking region (FR) sequences), it is hard to say what kind of evolutionary message is conveyed by such a synthetic descriptor of polymorphism as DNA amplicon size. In this study, we sequenced several SSR alleles in multiple populations of three divergent tree genera and disentangled the types of polymorphisms contained in each portion of the DNA amplicon containing an SSR. The patterns of diversity provided by amplicon size variation, SSR variation itself, insertions/deletions (indels), and single nucleotide polymorphisms (SNPs) observed in the FRs were compared. Amplicon size variation largely reflected SSR repeat number. The amount of variation was as large in FRs as in the SSR itself. The former contributed significantly to the phylogenetic information and sometimes was the main source of differentiation among individuals and populations contained by FR and SSR regions of SSR markers. The presence of mutations occurring at different rates within a marker's sequence offers the opportunity to analyse evolutionary events occurring on various timescales, but at the same time calls for caution in the interpretation of SSR marker data when the distribution of within-locus polymorphism is not known.

Citation: Barthe S, Gugerli F, Barkley NA, Maggia L, Cardi C, et al. (2012) Always Look on Both Sides: Phylogenetic Information Conveyed by Simple Sequence Repeat Allele Sequences. PLoS ONE 7(7): e40699. doi:10.1371/journal.pone.0040699

Editor: Shin-Han Shiu, Michigan State University, United States of America

Received: February 7, 2012; **Accepted:** June 12, 2012; **Published:** July 13, 2012

This is an open-access article, free of all copyright, and may be freely reproduced, distributed, transmitted, modified, built upon, or otherwise used by anyone for any lawful purpose. The work is made available under the Creative Commons CC0 public domain dedication.

Funding: Funding provided by PO-FEDER Guyane "ENERGIRAVI" http://www.europe-guyane.eu/index.php?option=com_content&task=view&id=14&Itemid=206. SB's Doctoral program has been partly funded by the European Social Fund. The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

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

* E-mail: ivan.scotti@ecofog.gf

Introduction

Simple sequence repeats (SSRs; see glossary, Table 1) can conveniently be used as genetic markers owing to their polymorphism and relative ease of interpretation [1]. SSRs are widely used for, e.g., reconstructing phylogenetic relationships [2], for analysing spatial genetic structure among and within populations [1], and for detecting and explaining patterns linked to habitat fragmentation and gene flow [3]. In many studies, SSR markers demonstrated recent expansions or bottlenecks in various plant [4,5], animal [6–8] and human populations [9,10].

SSRs are often expected to mutate following the stepwise mutation model (SMM) [11], whereby mutations alter the length of the repeat either by adding or by deleting a single repeat unit at a fixed rate [12]. More elaborate mutational processes, such as the two-phase model (TPM; [13]) and the generalized stepwise model (GSM; [14,15]), allow for multi-step mutations. Departures from these models can however occur [16–20]. Moreover, SSR alleles

are generally scored as the length (in base pairs) of PCR amplicons (see glossary, Table 1) containing the SSR, with differences in amplicon size taken to represent differences in repeat number in the SSR. Amplicon size includes of course SSR repeat number plus the length of the flanking region (FR; see glossary, Table 1). These chunks of sequence can be and actually often are polymorphic and may contain both single nucleotide polymorphisms (SNPs; see glossary, Table 1) and insertions/deletions (indels; see glossary, Table 1) [18,21–24]. Indels which have been demonstrated to occur in the FR sequences of some SSR alleles clearly contribute to total amplicon size, but they are interpreted under the SMM as (false) variation in SSR length. SNPs contribute to the total sequence variance of SSR loci (see glossary, Table 1) that goes largely unnoticed when only amplicon size is recorded (which is typical of most SSR studies). As a consequence, SSR amplicon size data are prone to a particular form of size homoplasy [15], i.e. equally sized alleles may have different sequences and therefore be evolutionarily different. This casts

Table 1. Glossary.

Amplicon	Product of a DNA amplification reaction.
SSR	Simple sequences repeat (or "microsatellite"). Tandem repeat of simple di- to hexa-nucleotide sequence motifs.
FR	Flanking region. DNA sequences appearing in an amplicon on either side of the SSR sequence.
Indel	A sequence gap in a DNA sequence alignment caused by an insertion or deletion mutation.
SNP	Single nucleotide polymorphism. DNA polymorphism that involves a change in a single base of a DNA sequence.
SSR locus	A specific genomic region consisting of SSR (microsatellite) DNA and its flanking regions (FRs).

doi:10.1371/journal.pone.0040699.t001

doubt on the evolutionary and population-genetic inferences that can be drawn from estimates of divergence obtained from SSR amplicon size data. It is therefore necessary to examine the kinds of evolutionary information conveyed in SSR alleles, to assess their consequences on fragment size-based evolutionary inferences, and to find ways to take advantage of this so far unexploited source of variability.

Figure 1 succinctly shows how information in FR sequence variation may change one's view of SSR marker evolutionary information content. Figure 1A shows a "plain" SSR locus, where a mutation (bar on the right branch) changes the number of repeats from n to $n+1$. The four observed alleles are then linked by a simple genealogy where alleles with the same number of repeats are evolutionarily closer. This is the classical view of SSR allele evolution under repeat-number mutational models. Figure 1B shows four alleles with the same numbers of repeats as in Figure 1A, but in this case information from a SNP in the FR (indicated by the "...A/...C" symbols) is added. Here, alleles with different numbers of repeats share the same SNP. Given that SNP mutation rates are generally lower than those of SSRs, the most-likely genealogy is one in which divergence occurs first at the SNP (mutation in the upper right branch) and then in SSRs within flanking-region haplotypes (mutations in the lower branches). In this case, three mutations have occurred, the genealogy is expected to be older than in 1A, and the alleles are grouped based on their FR sequences and not based on the number of repeats. A simulation study [25] revealed that SSR and SNP sources of variation provide independent evolutionary information, so it is reasonable to study the role of SNP variation in SSR flanking regions. We therefore asked the following questions: How much does amplicon size variation reflect variation in SSR repeat number? How important is the contribution of the various sources of sequence variation in total SSR diversity? Do these extra-SSR polymorphisms carry a different population and phylogenetic differentiation signal than the SSRs themselves? If so, how do the different signals combine? How does one best interpret SSR data, once these departures from the SMM have been acknowledged and assessed? To answer these questions, we sequenced SSR amplicons obtained from three divergent genera of long-lived angiosperm tree species to inspect the distribution of polymorphisms within and around SSRs and to evaluate the impact of each source of sequence variation on the detection of evolutionary and population genetic signals.

Materials and Methods

Choice of Data Sets

Three data sets were utilized in this study. The structures of the three data sets analysed here varied in the following ways: the *Citrus* data represent sequences of individual trees each belonging to a different species, variety or cultivar; the *Jacaranda* data were

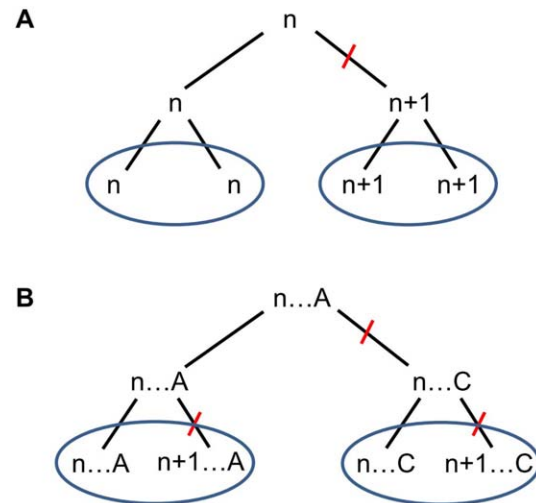


Figure 1. Two alternative genealogies for simple sequence repeat (SSR) alleles containing the same number of repeats. SSR alleles are indicated by their number of repeats (n , $n+1$). The A/C letters indicate a SNP in the flanking regions. Red bars correspond to mutational events in the flanking region sequence or in the number of SSR repeats. (A) No sequence information: deduced genealogy of observed data (third line) groups alleles together according to their number of repeats and involves a single SSR mutation. Genealogy is recent. (B) Consideration of sequence information: deduced genealogy will involve two identical and independent substitutions at the same nucleotide site and a unique SSR mutation, which is less likely). Genealogy is ancient and SSR alleles do not group according to their numbers of repeats.

doi:10.1371/journal.pone.0040699.g001

collected in four different sampling sites considered as four distinct populations; and the oak data came from a single mixed stand of sessile/pedunculate oaks (*Quercus robur*, *Q. petraea*). The *Citrus*, *Jacaranda*, and *Quercus* data sets were submitted to partially different types of analyses according to the structure of each data set.

Given the taxonomy and prevailing theory on many *Citrus* species being derived by natural hybridization from ancestral species, alleles were chosen to minimize known hybrids from the data set of Barkley et al. [26,27], who examined the genetic diversity of 370 *Citrus* accessions. The individual identity of each accession (as provided by the University of California, Riverside citrus database, <http://www.citrusvariety.ucr.edu/>) is given in the Supplementary Table 1 of Barkley et al. [26]. The *Jacaranda* data were obtained from three sites in French Guiana (Counami, Paracou and Saint-Laurent) and one in Brazil (Tapajos) (Table S1). The *Quercus* alleles were taken from genetic analyses of *Q. petraea* and *Q. robur* in a mixed oak stand in Switzerland [28,29].

Table 2. Characteristics of simple sequence repeat (SSR) data sets of the three taxa, *Citrus* (C), *Jacaranda* (J) and *Quercus* (Q).

Data set	Type of data	SSR locus	Size (bp) ^a	N ^b	Reference
Set C	Collection of provenances	cAGG9	123	34	[30]
		CCT01	167	34	[27]
		GT03	177	36	[27]
Set J	Four populations	Jc3A10	180	100	This study
		Jc3F4	144	101	This study
		Jc3H10	215	96	This study
Set Q	One mixed population	QrZAG30	255	47	[31]

^aTotal length of the amplicon consensus sequence in base pairs.

^bTotal number of analysed alleles.

doi:10.1371/journal.pone.0040699.t002

The SSR markers used for this study were: cAGG9, CCT01 and GT03 for *Citrus* [26,30]; Jc3H10, Jc3F4 and Jc3H10 for *Jacaranda*; QrZAG30 for *Quercus* [31]. Allele sequencing for the *Citrus* and *Quercus* data sets was as reported in Barkley et al. [26] and Gugerli et al. [29].

Molecular Methods

Jacaranda markers were obtained following the method developed for SSR markers in tropical crops by Billotte et al. [32]. Ten nanograms of DNA was digested with *RsaI*, followed by an enrichment step in (GT)_n and (GA)_n repeats by hybridization of cleaved fragments with biotin-labelled (CA)₈ and (CT)₈ oligoprobes and capture of the selected sequences with streptavidin-coated magnetic beads. Those fragments were ligated into a pGEM-TEasy vector (Promega, Madison, WI), and plasmids were then used to transform competent XL-1 Blue strain *Escherichia coli* (Stratagene, La Jolla, CA). Positive colonies were tested using colony polymerase chain reaction (PCR) to check the presence of inserts. Amplified fragments were then transferred from an agarose gel onto a N+ Hybond membrane for SSR presence screening by hybridization with labeled γ -³²P (GT)₁₅ and (GA)₁₅ oligoprobes. One hundred and thirteen bacterial clones containing plasmids with inserts that gave strong hybridization signals, with sizes ranging from 500 to 1000bp, were selected and cultured. Plasmids were extracted and sequenced using the universal T7 primer on an automated ABI 377 capillary sequencer (Applied Biosystems, Foster City, CA). After discarding duplicates, hybrid clones and clones with the SSR region too close to either end of the sequence, thirteen sequences were suitable for primer design and allowed a successful design using Oligo 3 software (Molecular Biology Insights Inc, USA). Upon screening for polymorphism and clarity of PCR patterns, three primer pairs were chosen for subsequent analyses. Primer sequences, annealing temperatures and GenBank/EMBL accession numbers are shown in Table S2. Dried cambium discs and leaf samples were flash-frozen in liquid nitrogen and later ground to a powder using a mortar and pestle. DNA extractions were performed by following Colpaert et al. [33]. PCRs for the detection of SSR polymorphisms were carried out in a 12 μ L volume containing 6 μ L 20-fold diluted DNA, 1 \times *Taq* buffer, 0.26 mM of dNTP, 0.03 U/ μ L *Taq* DNA polymerase (all products from Invitrogen, Carlsbad, CA) and 0.54 μ M of each primer (MWG Biotech, Ebersberg, Germany). For Jc3A10 primers, 0.3% BSA was added. An initial denaturation at 94°C

for 5 min was followed by 35 cycles of 94°C for 30 s, annealing temperature for 30 s and 72°C for 30 s, and a final extension at 72°C for 10 min. Genotyping was performed using fluorescently labeled primers (PET, 6FAM and NED) in the previous PCR protocol and fragments were separated on an ABI 3130XL capillary sequencer (Applied Biosystems, Foster City, CA) using ABI POP4 and Applied Biosystems LIZ-500 as internal standard, following the manufacturer's instructions. DNA was bidirectionally sequenced directly from PCR amplification products. When necessary, gametic phase was determined in heterozygotes by cloning amplicons and sequencing at least one allele per individual following the TA cloning kit protocol from Invitrogen (Carlsbad, CA). PCR products were ligated with the plasmid pCR^{2.1} and used to transform competent DH5 cells. After an overnight incubation at 37°C, white colonies were isolated. The presence of an SSR allele was checked by PCR with universal M13 primers. PCRs consisted of 1 μ L of cultivated colonies, 1 \times *Taq* buffer, 0.125 mM dNTPs, 0.025 U/ μ L *Taq* DNA polymerase (all products from Invitrogen, Carlsbad, CA) and 0.83 μ M forward and reverse primers. Cycling conditions consisted of 94°C for 5 min; 35 cycles of: 94°C for 1 min, 50°C for 1 min, 72°C for 2 min; and one cycle of 72°C for 5 min. PCR products were checked on a 4% agarose gel in 0.5 \times TAE. Plasmids were sequenced with the ABI BigDye[®] Terminator V3.1 kit (Applied Biosystems, Foster City, CA) following the manufacturer's protocol. Cloning products were diluted 1:40, purified with ExoSap-IT (USB Corporation, Cleveland, OH), and separated on an ABI 3130XL capillary sequencer (Applied Biosystems, Foster City, CA). All sequences were aligned and edited using CodonCode Aligner V1.6.3 (Codoncode Corporation, Dedham, MA).

Data Analyses

For each data set (*Citrus*, *Jacaranda*, and *Quercus*), each allele was characterised by the following information: (a) amplicon size variation, (b) SSR variation, (c) FR sequence variation and (d) amplicon sequence variation. Each of these portions of information conveyed by SSR marker data was analysed as a separate source of variation. Results were compared across data sets to assess their evolutionary information content. Indels were coded as SNP mutations to represent them as single mutational events.

Levels of polymorphism (number of alleles (*A*) or haplotypes (*h*), Nei's genetic diversity (*H_e*), number of SNPs and indels) were recorded for each source of variation. Linkage disequilibrium (LD) was computed for FR sequence variation.

Data were analysed in two ways: (a) for all data sets, at the level of individual alleles; (b) for *Jacaranda* only, at the population level.

- (a) At the individual-allele level, matrices of pairwise genetic distances were computed between individuals and between alleles based on each source of variation. For amplicon length, distance is represented by difference in length; for SSR length, distance is the difference in number of repeat units; for whole amplicon sequence, distance is the total number of differences, including SNPs, indels and number of repeat units; for flanking region sequences, distance is the total number of SNPs and indels. The correlation between matrices was tested by a Mantel test. LD between all pairs of polymorphic sites in FR sequences was tested with a Markov chain Monte Carlo (MCMC) procedure of 10,000 steps and a burn-in phase of 1,000 steps (default values). Pairwise genetic distances for amplicon and FR sequence variation were computed based on the number of mutational steps between sequences. Pairwise genetic distances for SSR

Table 3. Polymorphism at each simple sequence repeat (SSR) locus differentiated by segments of the DNA amplicon. SSR1, SSR2 and SSR3; respectively first, second and third SSR occurring in each amplicon (see figure S1 for details on each amplicon's sequence); FR: flanking regions; H_e : Nei's genetic diversity; A: number of alleles; h: number of haplotypes. Data sets: C: *Citrus*; J: *Jacaranda*; Q: *Quercus*.

SSR locus (data set)	Amplicon size variation			SSR1 variation			SSR2 variation			SSR3 variation			FR sequence variation					
	H_e	A	h	Repeat unit	H_e	A	Repeat unit	H_e	A	Repeat unit	H_e	A	Sequence length (bp)	Number of SNPs	SNPs per 100 bp	Number of indels	H_e	h
cAGG9 (C)	0.772	4	0.868	11	(GAG) _n	0.772	4	(GAA) _n	0.485	4	-	-	82	5	6.10	1	0.542	8
CCT01 (C)	0.758	4	0.966	20	(TCC) _n	-	-	-	-	-	-	-	140	17	12.14	0	0.861	14
GT03 (C)	0.783	5	0.900	12	(GT) _n	-	-	(AT) _n	-	-	-	-	138	12	8.70	1	0.808	9
Jc3A10 (J)	0.856	13	0.937	43	(CT) _n	0.769	12	(CA) _n	0.700	6	-	-	116	2	1.72	1	0.306	5
Jc3F4 (J)	0.912	20	0.920	32	(GA) _n	0.853	15	-	-	-	-	-	104	3	2.88	4	0.662	11
Jc3H10 (J)	0.856	13	0.922	40	(CT) _n	0.748	19	(CACG) _n	0.476	3	(CGCAC) _n	0.527	4	1	0.90	2	0.409	2
QrZAG30 (Q)	0.966	23	0.983	31	(GA) _n	0.915	16	-	-	-	-	-	185	14	7.57	15	0.922	18

doi:10.1371/journal.pone.0040699.t003

variation and amplicon size variation data were computed as Euclidian distances (genetic distances based on the SSR variation were not computed for *Citrus* loci CCT01 and GT03, which displayed extremely complex repeat sequences; see Figure S1). Most of the above calculations were performed with ARLEQUIN versions 3.11 and 3.5.1.2 [34]. Genetic distance matrices for SSR variation and amplicon size variation data were computed with an ad-hoc routine in R [35], and correlations between genetic distance matrices were computed with the mantel.test function of the R package "ape".

- (b) At the population level, for *Jacaranda*, global and pairwise F_{ST} values [36] were computed for all sources of variation; R_{ST} values [37] were computed on sources of variability comprising amplicon size variation and SSR variation; N_{ST} values [38] were computed on amplicon and FR sequence variations. In N_{ST} calculations on amplicon sequence variation, mutations in repeat number were weighted less than substitutions (weight = 0.05) to account for their faster mutation rate. Weight choice was based on the observation that SSRs have, in our *Jacaranda* sample, approximately ten times more alleles than each SNP (see Results section). Considering that rare alleles might be missing from our sample, we estimated a ratio of 1:20 in allele richness between SNPs and SSRs and used this ratio to establish the relative weight. A hierarchical analysis of molecular variance (AMOVA) was performed to partition genetic variance into within- and among-population components. The significance of variance components was tested by 1,000 permutations [39]. These analyses were performed with ARLEQUIN 3.11 and 3.5.3.1. Locus-by-locus pairwise F_{ST} , R_{ST} , and N_{ST} matrices were used to construct population-level consensus UPGMA (Unweighted Pair Group Method with Arithmetic Mean) cladograms by averaging genetic distance information of three loci with SplitsTree4 v4.6 [40,41].

Results

Distribution of Polymorphism

In order to evaluate homoplasy and the variation in phylogenetic signal from different regions of SSR alleles (repeat region and FR), sequence data produced from three divergent tree genera were evaluated. Forty-six, 101, and 47 samples constituted the *Citrus*, *Jacaranda*, and *Quercus* data sets, respectively. Three, three and one SSRs were genotyped and sequenced for the *Citrus*, *Jacaranda* and *Quercus* data sets, respectively. The lengths of the sequenced fragments varied between 123 and 255 bp (Table 2). Whenever possible, multiple alleles of the same fragment size were genotyped and sequenced to estimate the degree of sequence divergence among equally sized alleles ("size homoplasy"). This includes both variation in SSR repeat number (that is taken into account by the SMM) and variation in FR sequences, which is the focus of our study.

Levels of polymorphism (number of alleles/haplotypes, Nei's genetic diversity, number of SNPs and indels) for each amplicon and each source of variation are displayed in Table 3. Large amounts of variability were observed, as expected, in the amplicon sequences. SNPs were found to interrupt the repeat in all three taxa (imperfect SSRs: 33 sequences, data not shown). Sequences with SNPs within the repeat were excluded from subsequent analyses owing to the complexity of the mutation model of SSR repeats containing SNPs. Four loci had two or more repeats

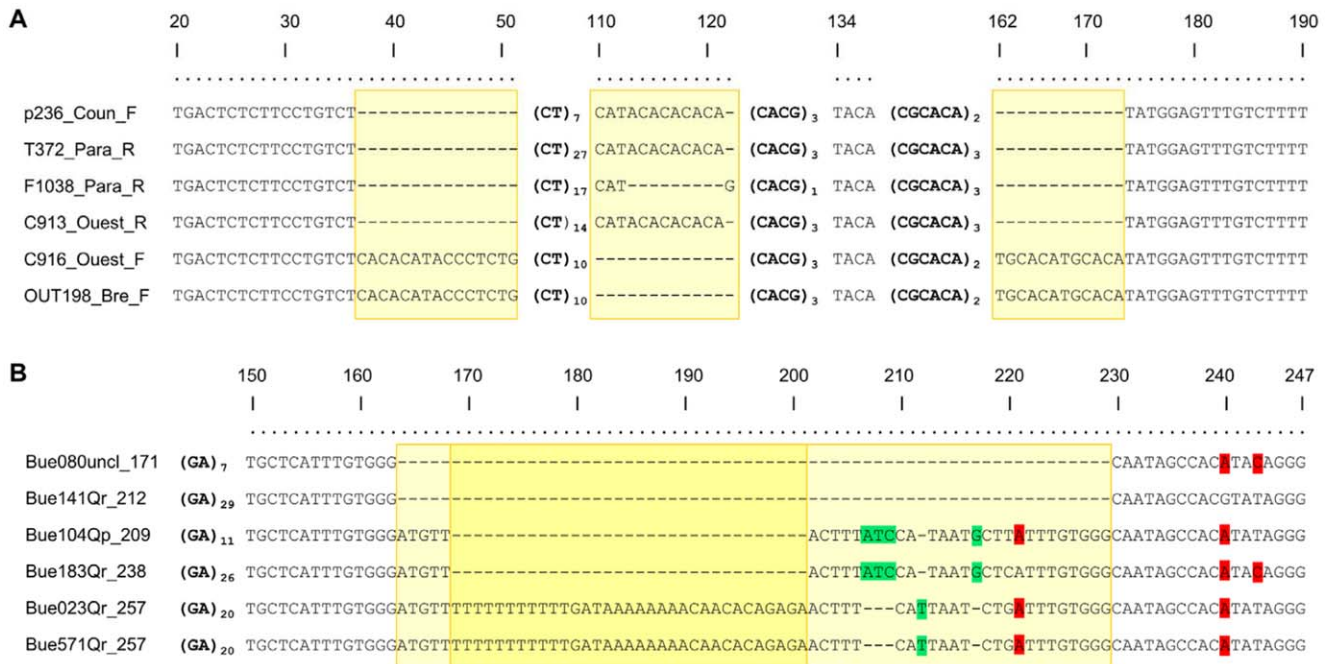


Figure 2. Alignment of a subset of DNA fragments at simple sequence repeat (SSR) loci Jc3H10 (*Jacaranda copaia*) (A) and QrZAG30 (*Quercus robur*) (B). Numbers in the top line indicate positions relative to the consensus sequence of (A) 190 bp and (B) 247 bp. Bold nucleotides in brackets indicate SSR motifs and their number of repeats. Dashes indicate gaps, highlighted nucleotides in green (light grey*) indicate indels of one to three bases, highlighted nucleotides in red (dark grey*) mark mutations from one base to another, and yellow (light grey*) boxes indicate groups of insertion/deletions longer than three bases and considered as a single mutational event. * In shades-of-grey printouts. doi:10.1371/journal.pone.0040699.g002

(compound SSRs). The highest number of alleles and Nei's genetic diversity were detected for amplicon sequence variation, then for amplicon size variation and lowest was found for SSR variation (Table 3). FR sequence variation showed the lowest diversity. Indels were observed in six loci out of seven, while SNP density varied between 0.90 and 12.14 SNPs per 100 bp.

An example of the complexity of the polymorphisms observed is shown for markers Jc3H10 and QrZAG30 (*Jacaranda* and *Quercus* data set, Figure 2; the complete alignments for all sequenced amplicons are given in Figure S1). Three indels of several nucleotides (between 12 and 15 bases) were detected in the FR sequences of the compound marker Jc3H10. One was upstream of the first SSR, the second was between the repeat motifs and the last was detected after the SSR repeats. In the oak marker QrZAG30, in addition to SNPs, a large 66-nucleotide indel was observed, with SNPs occurring among alleles carrying the DNA fragment involved in the indel. Several cases of size homoplasy were detected, such as: (i) indels in the FRs compensating differences in number of repeats, (ii) compound SSRs with the same number of repeats, but composed of different numbers of repeats in different motifs, (iii) SSRs with the same amplicon size and number of repeats but with SNPs in the FR sequences or interrupting the repeat (Figure S1). For the three data sets, 50 alleles with the same amplicon sizes out of 68 (74%) showed size homoplasy (Table S3).

Linkage Disequilibrium

LD was tested for all pairs of SNP and indel sites within the FR sequence variation of each marker (Figure 3). For the *Citrus* data set, only 19 pairs out of 217 (8.8%) showed significant LD. For the three SSR loci, LD was very irregular and there was no relationship between LD and sequence distance (Mantel tests,

$P > 0.05$ for all loci). For the *Jacaranda* data set, ten pairs out of 27 (37%) showed significant LD. For marker Jc3F4, the SNP at position 81 was in disequilibrium with two indels and one SNP, all located upstream of the SSR. LD was strongest for Jc3H10, where all three pairs were significant. The relationship between LD and sequence distance was significant (Mantel test, $P = 0.023$). Tests of correlation between LD and distance in bp could not be performed for Jc3A10 and Jc3H10 because only three nucleotide sites were polymorphic (three LD values). For the *Quercus* data set (marker QrZAG30), 146 pairs of SNPs and indels out of 325 (44.9%) showed significant LD. In particular, 13 or 14 contiguous SNPs and indels located downstream of the SSR at positions 84–107 formed a disequilibrium block, itself in weaker disequilibrium with upstream polymorphic sites at positions 23, 45 and 46 and another SNP downstream the SSR (position 60).

Correlation between Genetic Distance Matrices

We tested whether different sources of variation conveyed the same information on relatedness among alleles. To do so, genetic distances were computed for each pair of individuals with each source of variation (SSR variation, amplicon size, FR sequence, and amplicon sequence variation) within each marker, and the correlation between genetic distance matrices was tested by a Mantel test. Among the 36 pairs of matrices, 27 (75%) showed a significant correlation (Figure S2). Genetic distance matrices based on SSR variation and on FR sequence variation were the least correlated (one significant pair out of five). Amplicon sequence variation and amplicon size variation, as well as SSR variation and amplicon size variation, were correlated for all markers. The matrices computed on amplicon sequence variation were correlated to the remaining matrices in 95% of the cases (18 out of 19).

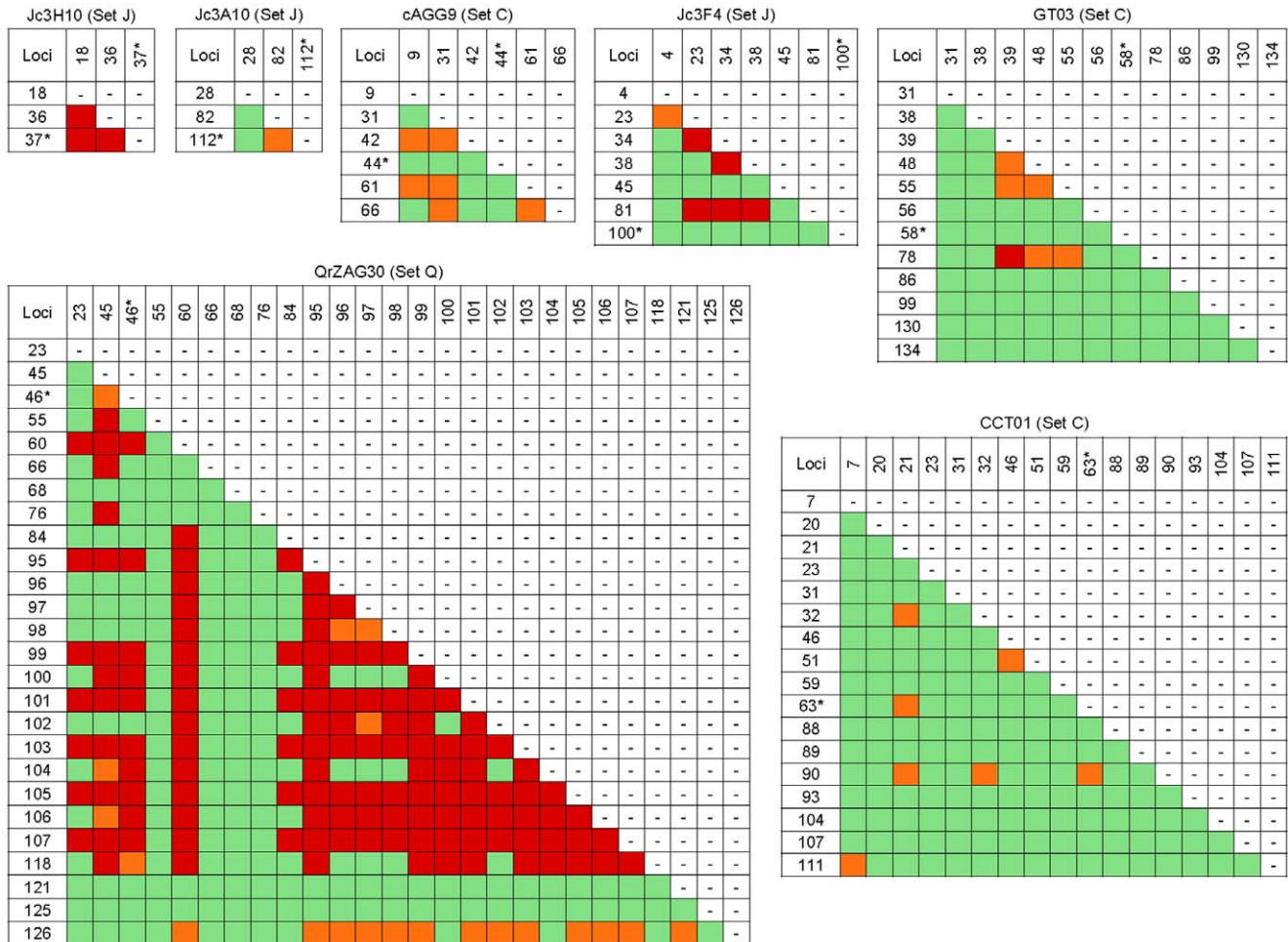


Figure 3. Linkage Disequilibrium (LD) for pairs of polymorphisms in the flanking regions. Dark cells contain LD values significantly different from zero. An asterisk next to a locus name indicates that the SSR repeat(s) is (are) located between that locus and the next. Colouring (shading) indicates the degree of significance of the test: green (light grey*), $P > 0.05$; orange (grey*), $0.05 < P < 0.001$; red (dark grey*), $P < 0.001$. * In shades-of-gray printouts.

doi:10.1371/journal.pone.0040699.g003

Conversely, FR sequence variation alone predicted other matrices in only 53% of the cases.

Population Differentiation

For the *Jacaranda* data set, we tested population genetic differentiation for each source of variation by computing F_{ST} , R_{ST} and N_{ST} as appropriate (see Methods; Figure 4). All F_{ST} values were significant. FR sequence variation showed the largest value (0.42) and SSR variation the smallest (0.05). R_{ST} values were very small (< 0.02) and non-significant when computed on both amplicon size variation and SSR variation. On the contrary, N_{ST} values were large and significant, with values of 0.50 for FR sequence variation and 0.36 for amplicon sequence variation. In the phylogenetic trees obtained from each combination of source of variation/genetic distance measures, F_{ST} and N_{ST} provided the same topology for all cases they were applied to, indicating a more or less close link between pairs of populations Tapajos/Saint-Laurent and Counami/Paracou (Figure 4). R_{ST} grouped Paracou with Tapajos and Counami with Saint-Laurent (for amplicon size variation) or separated the three Guiana shield populations from Tapajos (for SSR variation; Figure 4).

Discussion

Distribution of Polymorphism

SSR polymorphisms are widely used to infer population history and biogeographic patterns. These inferences rely on assumptions about SSR mutational models, and departures from such models are likely to bias divergence and diversity estimates. SSR alleles have been demonstrated to contain other sources of polymorphism in addition to variation in repeat number. In this study we have characterised the contribution of molecular variation occurring outside SSR repeats to SSR marker variability. Whole allele sequencing from SSR markers was analyzed from three tree genera (*Citrus*, *Jacaranda* and *Quercus*). A total of seven (dinucleotide and trinucleotide) SSR markers were targeted, including compound and imperfect markers. More genetic diversity was found for amplicon sequence variation than for amplicon size variation, suggesting that much of data polymorphism is neglected when SSRs are described through amplicon size alone. On average, 7.7 SNPs and 3.4 indels were detected in the flanking sequences of each marker. Genetic diversity (H_e) ranged between 0.306 and 0.922 for the FRs. LD among SNPs and indels in the FR was generally significant. These results demonstrate that SSR alleles are very often riddled with abundant non-SSR indels, plus vast

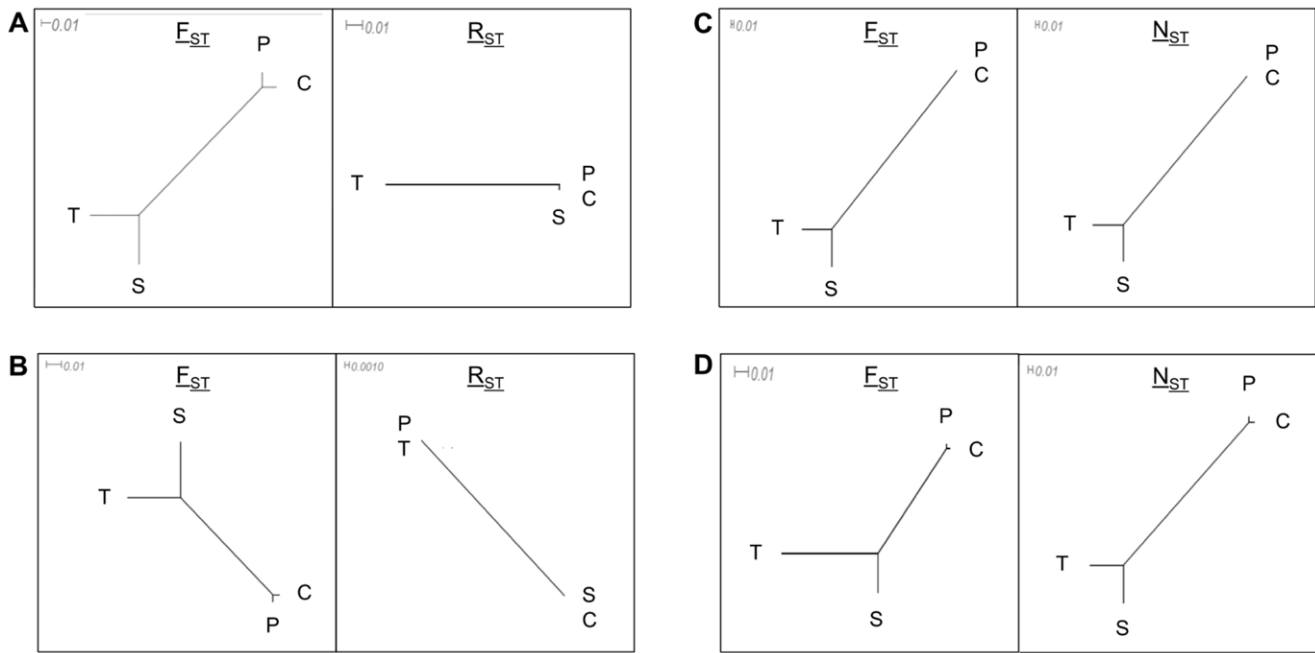


Figure 4. Phylogenetic trees of *Jacaranda* populations based on different components of simple sequence repeat (SSR) data. SSR variation (A), amplicon size variation (B), flanking region (FR) sequence variation (C) and amplicon sequence variation (D). Each type of data was analysed according to a pair of suitable genetic distance estimators: F_{ST} , suitable for loci following the infinite allele model (IAM); R_{ST} , for loci following the stepwise mutation model (SMM); and N_{ST} , for loci following the infinite site model (ISM) model. There are four geographic populations: Counami (C), Paracou (P) and Saint-Laurent (S) in French Guiana; Tapajos (T) in Brazil. Note that scales are not the same among trees. doi:10.1371/journal.pone.0040699.g004

amounts of SNPs both within and near the repeat itself. Similar genetic variation was previously observed at intra- or inter-specific levels [18,42–44] and does not seem to be exceptional in plants. SSR allele mutations are clearly not restricted to the hypervariable tandem repeat region, and polymorphisms outside SSR motifs are also present in animals [22,45,46] and humans [21]. Hidden indels can directly confound the estimation of repeat number from amplicon size (“allele size homoplasy”). While undetected SNPs do not directly influencing repeat number estimates, they do cause haplotypes with different evolutionary histories to be treated as if they were identical. Both forms of homoplasy can have profound consequences on the way data should be interpreted.

Correlation of Polymorphism at the Individual-allele Level

Our results show that pieces of genetic information carried by the different portions of an amplicon are only partially correlated. In particular, blocks of linkage disequilibrium never cover the whole amplicon, and matrices of allele-level pairwise genetic distances built on each source of variation are not always well correlated. In other terms, the different sources of variation are not all equally capable to represent the “true” allele genealogy. Not surprisingly, whole-amplicon sequences were the most parsimonious predictors of relatedness described by the different parts of the amplicons in all our data sets; conversely, FR sequences alone did not adequately synthesize the information from the whole amplicon sequence. Correlation between genetic distances as obtained from FR sequence variation and from SSR variation was globally weak (Figure S2). This recalls the finding of Payseur and Cutter [25] which stated that coalescence times for SSRs and linked SNPs were uncorrelated. Thus, analysing all parts of SSR amplicons increases the amount of available, phylogenetically independent information, with the additional advantage of providing data from DNA regions with different mutation rates.

Nevertheless, the correlation analysis shows that the best predictor of amplicon size is SSR length. This indicates that amplicon size is a reliable first-approximation proxy for SSR repeat length, even though extensive homoplasy blurs the correlation.

Correlation of Polymorphism at the Population Level

The consequences – and usefulness – of hidden sequence variation are shown when we compare how sources of variation perform in detecting population divergence. This analysis was performed in the *Jacaranda* data set comprising four populations. In this example, two alternative patterns were expected for population relatedness: (a) a strictly geographical clustering (the three Guiana shield populations cluster and are separated from the Amazonian population), and (b) a pattern derived from independent results on chloroplast DNA divergence (Counami and Paracou populations form a group, Tapajos and Saint-Laurent form another group (Caroline Scotti-Saintagne, INRA, UMR « Ecologie des forêts de Guyane », article under second round of revision for *Journal of Biogeography*). Our results support the latter hypothesis, as proven by the convergent topology of trees obtained with F_{ST} and N_{ST} independently of the source of variation. The purely geographical hypothesis, on the other hand, is only supported by SSR variation in combination with R_{ST} (although with a non-significant global R_{ST} value). R_{ST} estimates have larger variance than F_{ST} and N_{ST} , particularly when based on small numbers of loci, and R_{ST} values obtained here may be unreliable. Nevertheless, the pattern of R_{ST} differentiation closely follows geographical distance. This may indicate that repeat number variation has arisen locally from small numbers of founding alleles and reflects recent population divergence, while variation in flanking regions may follow larger phylogeographic patterns and reflect deeper divergence. If R_{ST} estimated can be trusted, all sources of variation converge with independent, chloroplast-based

results, except SSR variation itself. Therefore, information derived from the “true” nature of SSR data (SSR variation) is actually not representative of the information conveyed by alleles as a whole. This finding strongly hints that other sources of variability, more correctly described by mutation models that apply to DNA sequences, actually provide the bulk of information carried by SSRs. In other words, phylogeographic information carried by amplicon length is not provided by the SSR (which amplicon length is intended to represent) but by variation in flanking regions. Note that this contrasts with analyses carried out at the individual level (see above), where information on genetic distance match when measured as amplicon size and as SSR size. This issue raises the question of the application of the SMM, or other step-based mutational models, to SSRs, as already discussed by Colson and Goldstein [47] and Ellegren [12]. At the same time, these results substantiate the hypothesis of Cornuet *et al.* [48], who postulated that not all SSRs would evolve under the same mutational model; additional mutational models are required to appropriately handle variation in SSR data produced from different genera. Moreover, variation in the FRs contributed significantly to the phylogenetic signal and sometimes represented the main source of differentiation among individuals and populations, as shown by *Jacaranda* SSR markers and by the *Citrus* SSR marker GT03 [27].

Conclusion

Given the complexity of amplicon sequences described here, one may wonder how to properly exploit SSR data. The weight of polymorphisms other than variation in repeat number cannot and should not be overlooked: identity in allele size does not necessarily indicate identity in sequence content or the number of repeats within same sized alleles. When considering individual alleles, amplicon size correlates well with SSR repeat number. Nevertheless, even if repeat number could be obtained directly, it may provide only weak and inconsistent signals of population genetic differentiation, as shown by the *Jacaranda* data set (but larger numbers of markers may reduce estimation variance and mitigate this problem). Taking sequence variation into account actually adds a significant piece of information to phylogenetic or phylogeographic reconstruction. Population structuring emerges more clearly on the basis of sequence data, and the combination of sequences and SSR variation provides higher resolution. Consequently, genetic distance measures assuming the SMM or related models should be restricted to perfect SSRs with invariable FRs. On the other hand, measures such as F_{ST} (for amplicon size data) and N_{ST} (for amplicon sequences) seem to be the tool of choice for

the analysis of the average, not-so-ideal SSR markers that a population geneticist meets in his or her everyday work.

Supporting Information

Figure S1 Complete alignments of DNA fragments for all SSR markers studied. Marker names are shown above each alignment.

(PDF)

Figure S2 Z-statistics and summary of Mantel tests for the correlations of pairs of genetic distance matrices within species. Colouring (shading) indicates the level of significance of the test: green (light grey*), $P > 0.05$; orange (grey*), $0.05 < P < 0.001$; red (dark grey*), $P < 0.001$. Identity of DNA fragment sections is reported: simple sequence repeat (SSR) variation; amplicon size variation, flanking region (FR) sequence variation, and amplicon sequence variation. Marker names are displayed above each panel and data sets refer to *Citrus* (C), *Jacaranda* (J) and *Quercus* (Q). Sequences with single nucleotide polymorphisms (SNPs) within the repeat were excluded from subsequent analyses because of the complexity to describe the mutation model of SSR repeats bracketing SNPs.

(PDF)

Table S1 List and details of the sampled sites for the *Jacaranda* (J) data set.

(DOC)

Table S2 Description, amplification conditions and polymorphism of the nuclear simple sequence repeats (SSRs) analysed in *Jacaranda copaia*.

(DOC)

Table S3 List and details of alleles for the three data sets. Amplicon sizes, frequencies and number of associated haplotypes are indicated for each allele.

(DOC)

Acknowledgments

We wish to thank Maristerra Lemes for providing a part of the samples, Raphaël Leblois and Caroline Scotti-Saintagne for fruitful discussions and for help with preliminary data analyses, and Valérie Troispoux and Saint-Omer Cazal for technical support.

Author Contributions

Conceived and designed the experiments: IS. Performed the experiments: SB FG NB. Analyzed the data: SB IS. Contributed reagents/materials/analysis tools: LM CC. Wrote the paper: SB FG NB IS.

References

1. Tautz D (1989) Hypervariability of simple sequences as a general source for polymorphic DNA markers. *Nucleic Acids Res* 17: 6465–6471.
2. Takezaki N, Nei M (1996) Genetic distances and reconstruction of phylogenetic trees from microsatellite DNA. *Genetics* 144: 389–399.
3. Rodriguez H, Geistlinger J, Berlyn G, Kahl G, Weising K (2000) Characterization of novel microsatellite loci isolated from the tropical dioecious tree *Simarouba amara*. *Mol Ecol* 9: 489–504.
4. England PR, Usher AV, Whelan RJ, Ayre DJ (2002) Microsatellite diversity and genetic structure of fragmented populations of the rare, fire-dependent shrub *Grevillea macleayana*. *Mol Ecol* 11: 967–977.
5. Taylor DR, Keller SR (2007) Historical range expansion determines the phylogenetic diversity introduced during contemporary species invasion. *Evolution* 61: 335–345.
6. Estoup A, J Wilson IJ, Sullivan C, Cornuet J-M, Moritz C (2001) Inferring population history from microsatellite and enzyme data in serially introduced cane toads, *Bufo marinus*. *Genetics* 159: 1671–1687.
7. Randi E, Tabarroni C, Rimondi S, Lucchini V, Sfougaris A (2003) Phylogeography of the rock partridge (*Alectoris graeca*). *Mol Ecol* 12: 2201–2214.
8. Bonhomme M, Blancher A, Cuartero S, Chikhi L, Crouau-Roy B (2008) Origin and number of founders in an introduced insular primate: estimation from nuclear genetic data. *Mol Ecol* 17: 1009–1019.
9. Reich DE, Goldstein DB (1998) Genetic evidence for a Paleolithic human population expansion in Africa. *Proc Nat Acad Sci USA* 95: 8119–8123.
10. Liu H, Prugnolle F, Manica A, Balloux F (2006) A geographically explicit genetic model of worldwide human-settlement history. *Am J Hum Genet* 79: 230–237.
11. Kimura M, Ohta T (1978) Stepwise mutation model and distribution of allelic frequencies in a finite population. *Proc Nat Acad Sci USA* 75: 2868–2872.
12. Ellegren H (2004) Microsatellites: simple sequences with complex evolution. *Nat Rev Genet* 5: 435–445.
13. Di Rienzo A, Peterson AC, Garza JC, Valdes AM, Slatkin M, et al. (1994) Mutational processes of simple-sequence repeat loci in human populations. *Proc Nat Acad Sci USA* 91: 3166–3170.
14. Zhivotovsky LA, Feldman MW, Grishchkin SA (1997) Biased mutations and microsatellite variation. *Mol Biol Evol* 14: 926–933.
15. Estoup A, Jarne P, Cornuet J-M (2002) Homoplasy and mutation model at microsatellite loci and their consequences for population genetics analysis. *Mol Ecol* 11: 1591–1604.

16. Primmer CR, Ellegren H (1998) Patterns of molecular evolution in avian microsatellites. *Mol Biol Evol* 15: 997–1008.
17. Symonds VV, Lloyd AM (2003) An analysis of microsatellite loci in *Arabidopsis thaliana*: Mutational dynamics and application. *Genetics* 165: 1475–1488.
18. Curtu A-L, Finkeldey R, Gailing O (2004) Comparative sequencing of a microsatellite locus reveals size homoplasy within and between European oak species (*Quercus* spp.). *Plant Mol Biol Rep* 339–346: 22.
19. Stajner N, Jakse J, Kozjak P, Javornik B (2005) The isolation and characterisation of microsatellites in hop (*Humulus lupulus* L.). *Plant Sci* 168: 213–221.
20. Xie H, Sui Y, Chang F-Q, Xu Y, Ma R-C (2006). SSR allelic variation in almond (*Prunus dulcis* Mill.). *Theor Appl Genet* 112: 366–372.
21. Grimaldi M-C, Crouau-Roy B (1997) Microsatellite allelic homoplasy due to variable flanking sequences. *J Mol Evol* 44: 336–340.
22. Orti G, Pearse DE, Avise JC (1997) Phylogenetic assessment of length variation at a microsatellite locus. *Proc Nat Acad Sci USA* 94: 10745–10749.
23. Mogg R, Batley J, Hanley S, Edwards D, O'Sullivan H, et al. (2002) Characterization of the flanking regions of *Zea mays* microsatellites reveals a large number of useful sequence polymorphisms. *Theor Appl Genet* 105: 532–543.
24. Lia VV, Bracco M, Gottlieb AM, Poggio L, Confalonieri VA (2007) Complex mutational patterns and size homoplasy at maize microsatellite loci. *Theor Appl Genet* 115: 981–991.
25. Payseur BA, Cutter AD (2006) Intergrating patterns of polymorphism at SNPs and STRs. *Trends Genet* 22: 424–429.
26. Barkley NA, Roose ML, Krueger RR, Federici C'T (2006) Assessing genetic diversity and population structure in a citrus germplasm collection utilizing simple sequence repeat markers (SSRs). *Theor Appl Genet* 112: 1519–1531.
27. Barkley NA, Krueger RR, Federici C'T, Roose ML (2009) What phylogeny and gene genealogy analyses reveal about homoplasy in citrus microsatellite alleles. *Plant Syst Evol* 282: 71–86.
28. Gugerli F, Walser J-C, Douvani K, Holderegger R, Finkeldey R (2007) Coincidence of small-scale spatial discontinuities in leaf morphology and nuclear microsatellite variation of *Quercus petraea* and *Q. robur* in a mixed forest. *Ann Bot* 99: 713–722.
29. Gugerli F, Brodbeck S, Holderegger R (2008) Insertions–deletions in a microsatellite flanking region may be resolved by variation in stuttering patterns. *Plant Mol Biol Rep* 26: 255–262.
30. Kijas JMH, Thomas MR, Fowler JCS, Roose ML (1997) Integration of trinucleotide microsatellites into a linkage map of *Citrus*. *Theor Appl Genet* 94: 701–706.
31. Kampfer S, Lexer C, Glössl J, Steinkellner H (1998) Characterization of (GA)_n microsatellite loci from *Quercus robur*. *Hereditas* 129: 183–186.
32. Billotte N, Lagoda PJJ, Risterucci A-M, Baurens F-C (1999) Microsatellite-enriched libraries: applied methodology for the development of SSR markers in tropical crops. *Fruits* 54: 277–288.
33. Colpaert N, Cavers S, Bandou E, Caron H, Gheysen G, et al. (2005) Sampling tissue for DNA analysis of trees: Trunk cambium as an alternative to canopy leaves. *Silvae Genet* 54: 265–269.
34. Excoffier L, Lischer HEL (2010) Arlequin suite ver 3.5: a new series of programs to perform population genetics analyses under Linux and Windows. *Mol Ecol Res* 10: 564–567.
35. R Development Core Team (2011) R: A language and environment for statistical computing. v. 2.14.1 R Foundation for Statistical Computing, Vienna, Austria. Available: <http://www.R-project.org>. Accessed 2012 June 18.
36. Weir BS, Cockerham CC (1984) Estimating F-statistics for the analysis of population structure. *Evolution* 38: 1358–1370.
37. Michalakis Y, Excoffier L (1996) A generic estimation of population subdivision using distances between alleles with special reference for microsatellite loci. *Genetics* 142: 1061–1064.
38. Pons O, Petit RJ (1996) Measuring and testing genetic differentiation with ordered versus unordered alleles. *Genetics* 144: 1237–1245.
39. Excoffier L, Smouse PE, Quattro JM (1992) Analysis of molecular variance inferred from metric distances among DNA haplotypes: Application to human mitochondrial DNA restriction data. *Genetics* 131: 479–491.
40. Huson DH (1998) SplitsTree: analysing and visualizing evolutionary data. *Bioinformatics* 14: 68–73.
41. Huson DH, Bryant D (2006) Application of phylogenetic networks in evolutionary studies. *Mol Bio Evol* 23: 254–267.
42. Peakall R, Gilmore S, Keys W, Morgante M, Rafalski A (1998) Cross-species amplification of Soybean (*Glycine max*) simple sequence repeats (SSRs) within the genus and other legume genera: Implications for the transferability of SSRs in plants. *Mol Biol Evol* 15: 1275–1287.
43. Saha MC, Mian MAR, Eujayl I, Zwonitzer JC, Wang L, et al. (2004) Tall fescue EST-SSR markers with transferability across several grass species. *Theor Appl Genet* 109: 783–791.
44. McDonald MJ, Wang W-C, Huang H-D, Leu J-Y (2011) Clusters of nucleotide substitutions and insertion/deletion mutations are associated with repeat sequences. *PLoS Biol* 9: e1000622. doi: 10.1371/journal.pbio.1000622.
45. Brohede J, Ellegren H (1999) Microsatellite evolution: polarity of substitutions within repeats and neutrality of flanking sequences. *Proc R Soc Lond B*, 266: 825–833.
46. Shepherd LD, Lambert DM (2005) Mutational bias in penguin microsatellite DNA. *J Hered* 96: 566–571.
47. Colson I, Goldstein DB (1999) Evidence for complex mutations at microsatellite loci in drosophila. *Genetics* 152: 617–627.
48. Cornuet J-M, Beaumont MA, Estoup A, Solignac M, Carey PD (2006) Inference on microsatellite mutation processes in the invasive mite, *Varoa destructor*, using reversible jump Markov chain Monte Carlo. *Theor Popul Biol* 69: 129–144.

Annexe 2

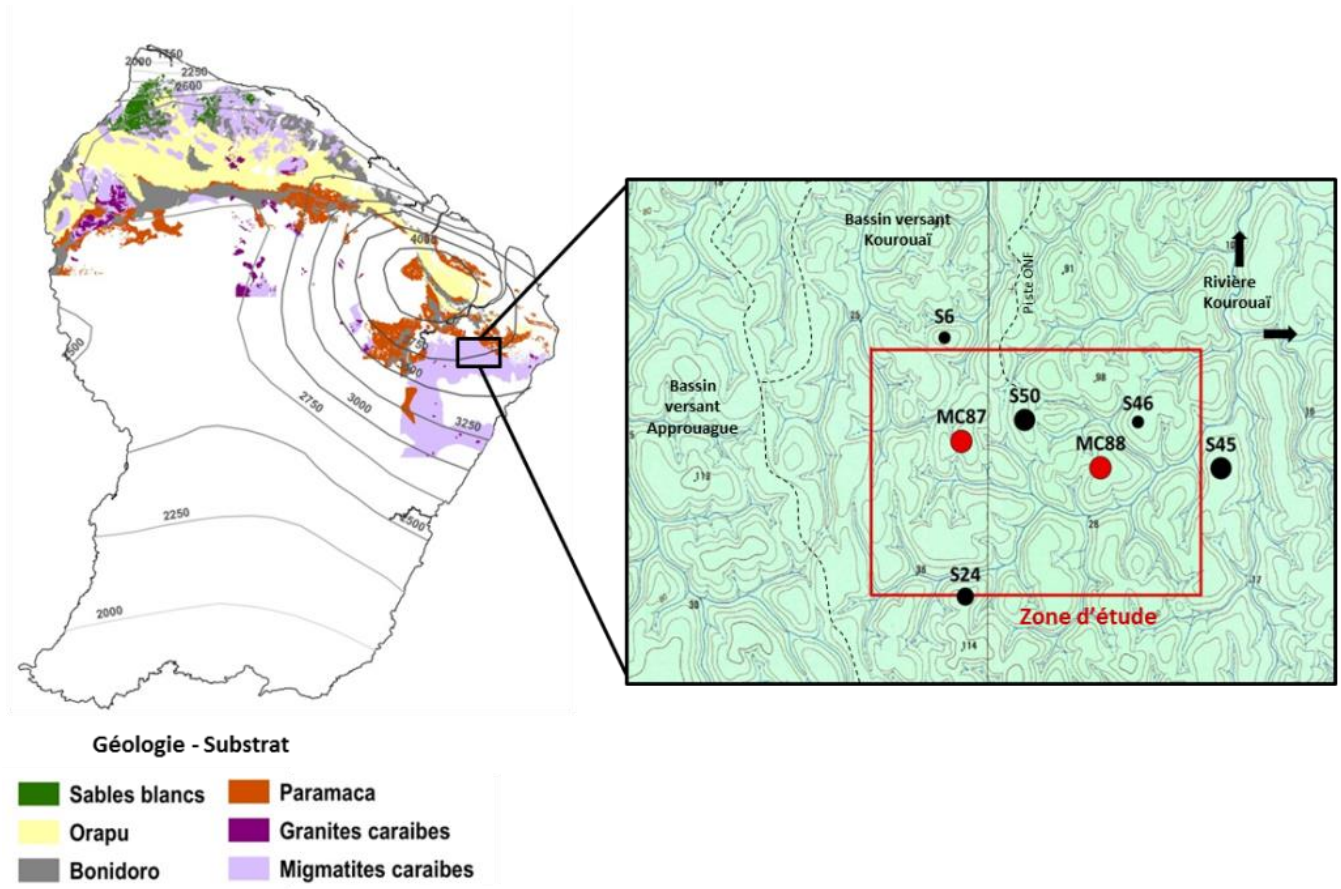


Figure A.1. Localisation de la zone d'étude en Guyane française

Impact des occupations amérindiennes anciennes à l'échelle du paysage forestier amazonien : études préliminaires

La présence de l'homme en Guyane à l'Holocène laisse supposer qu'il s'est installé et a utilisé les ressources à sa disposition pour survivre mais il est encore difficile de quantifier l'impact de ces populations précolombiennes sur leur environnement. Même si les fouilles archéologiques d'un site révèlent des artefacts indicateurs de présence humaine, l'interprétation de la structure d'un village, la densité de son occupation et son impact sur la couverture forestière restent hypothétiques (Sylvie Jérémie, communication personnelle). De plus, la chronologie des occupations sur un même site est souvent délicate à déterminer car si les charbons utilisés pour les datations peuvent migrer d'un horizon de sol à l'autre dans des sols remaniés, ils peuvent également avoir de multiples origines telles que des paléo-incendies naturels. L'occupation précolombienne ne se résume donc pas à l'occupation d'un lieu forestier à un instant donné mais a probablement été continue dans le temps. Le but de cette étude préliminaire a été de savoir si, à l'échelle du paysage, les perturbations liées à une occupation humaine ancienne avaient influencé la distribution actuelle des espèces forestières.

Cette étude a été menée à l'Est de la Guyane française, dans une région où deux sites archéologiques de montagnes couronnées avaient été au préalable découverts (Figure A.1). Une grille d'échantillonnage de 3,5 km sur 2,5 km, couvrant une surface d'environ 9 km² et incluant un maillage de 500m de côté, a été définie. Les coordonnées spatiales, la topographie, l'altitude, la couleur et la texture du sol, la présence de charbons et de poteries et l'abondance des espèces ont été relevés sur un total de 90 points, comprenant les intersections du maillage d'échantillonnage et tous les sommets intermédiaires. Les individus de 4 espèces d'arbres (*Dicorynia guianensis*, *Carapa sp.*, *Virola michelii* et *Vouacapoua americana*) avec un diamètre (DBH) supérieur à 10 cm et de 5 palmiers forestiers (*Astrocaryum paramaca*, *Astrocaryum sciophilum*, *Euterpe oleracea*, *Oenocarpus bacaba* et *Oenocarpus bataua*) ont été comptés à vue dans un rayon de 25 m à chaque point d'échantillonnage.

Tableau A.1. Effets de l'environnement et des perturbations sur l'abondance de 9 espèces de la forêt tropicale humide en Guyane française. Les coefficients de détermination et la significativité des paramètres du modèle retenu sont indiqués ainsi que le signe des paramètres pour les variables qualitatives.

		Variables spatiales			Variables environnementales				Variables archéologiques			r ²
		X	Y	X:Y	Topographie	Altitude	Couleur sol	Texture sol	Tessons	Charbons	Éloignement au site	
Arbres Forestiers	<i>D. guianensis</i>	(+)	++		0		X		(-)			0,26
	<i>Carapa sp.</i>				XX	+++						0,12
	<i>V. michelii</i>	(+)	(+)	(-)		0			+			0,14
	<i>V. americana</i>	+	+	-	0	+++	X			++		0,25
Palmyers Forestiers	<i>A. paramaca</i>	+++	(-)			---		X	(+)	+++		0,3
	<i>A. sciophilum</i>	(-)	---		X		X			(-)		0,32
	<i>E. oleraceae</i>		+		X	+++	X		(-)	---		0,55
	<i>O. bacaba</i>	(+)										0,01
	<i>O. bataua</i>				0			X		+		0,14

P-value : (>)>0,05; X<0,05; XX<0,01 ; XXX<0,001.

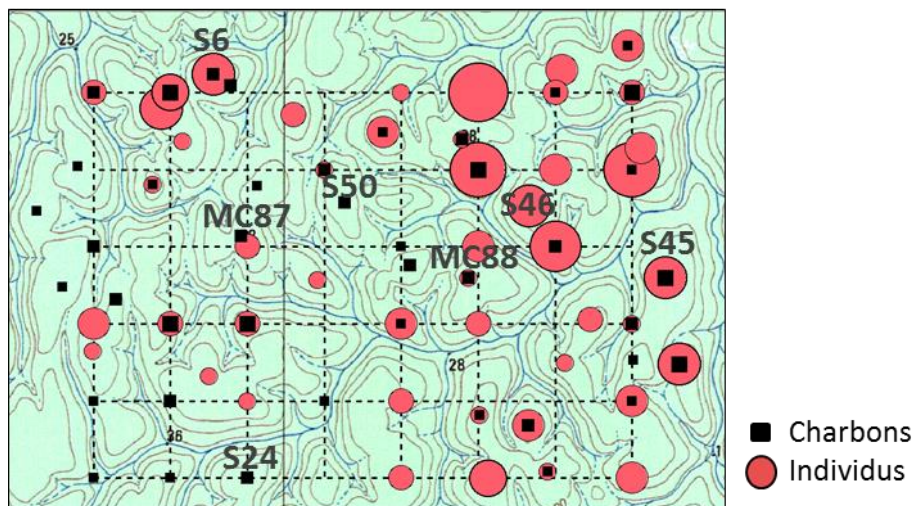


Figure A.2. Abondance de *Vouacapoua americana* en fonction de la quantité de charbons et de la présence de tessons

D'autres informations ont également été notées et spatialisées telles que la présence de quartz ou d'espèces à vocation alimentaire. Comme le nombre d'individus par espèce est une variable de comptage, la loi de Poisson est appropriée dans le cadre du modèle linéaire généralisé suivant :

$$\begin{aligned} \log(E(\text{Abondance})) &= \theta_{s1}X + \theta_{s2}Y + \theta_{s3}XY + \theta_{a1}\text{Tessons} + \theta_{a2}\text{Charbons} + \theta_{a3}\text{Distance} \\ &+ \theta_{e1}\text{Topographie} + \theta_{e2}\text{Altitude} + \theta_{e3}\text{Sol} + \theta_{e4}\text{Texture} \end{aligned}$$

Les variables X , Y et XY permettent de prendre en compte d'éventuels gradients de distribution spatiale liés à des paramètres non locaux et le critère d'information d'Akaike (AIC) a été utilisé pour choisir le modèle le plus parcimonieux.

Le tableau 1 présente les modèles retenus pour chaque espèce. Les variables environnementales les plus explicatives reflètent les préférences écologiques des espèces. Dans le cas des variables archéologiques, qui sont des estimateurs directs de perturbations anciennes, on constate que la quantité de charbons est corrélée positivement avec l'abondance de *V. americana* (Figure A.2) alors que *E. oleraceae* est présent dans des milieux sans charbons et tessons. *V. michelii* est, lui, plus abondant en présence de tessons et dans le cas de *A. paramaca* plus on s'éloigne d'un site perturbé plus les individus sont nombreux. Même si certaines variables peuvent présenter des biais, tels que l'éloignement au site qui ne prend pas en compte la distance avec des sites pouvant exister hors de la zone prospectée, cette étude encore très exploratoire laisse suggérer que les milieux perturbés ont une influence sur la présence de certaines espèces. Si le *Carapa sp.* paraît insensible aux perturbations, d'autres espèces semblent être favorisées telles que *V. michelii* et *V. americana*. La présence d'*E. oleraceae* est, elle, plus liées à l'altitude et à la topographie qu'à la présence de charbons et/ou de fragments de poterie.

L'utilisation d'un système d'information géographique (SIG) permettrait également de mieux comprendre et représenter l'impact des populations précolombiennes à l'échelle du paysage. Cette approche permet notamment de déterminer des couloirs de déplacement d'un point A à un point B en prenant en compte le coût pour l'homme engendré par la topographie du milieu ou encore la présence de rivières. Un tel modèle a été esquissé dans la Figure A.3 entre les deux sites de montagnes couronnées MC87 et MC88, étudiés précédemment. Nous pouvons émettre l'hypothèse que les populations précolombiennes préféreraient se déplacer de façon la plus directe et la plus rapide possible, en empruntant les lignes de crêtes plutôt que les bas-fonds, pour aller d'un site à l'autre, en supposant que

ceux-ci aient été contemporains. Si les hommes se déplaçaient probablement entre les villages, on peut également imaginer qu'ils utilisaient quotidiennement les ressources environnantes telles que l'eau, les animaux, les végétaux, les roches, l'argile pour se nourrir, construire des outils, des habitations, fabriquer des poteries. L'analyse des données récoltées dans cette étude par les méthodes SIG pourrait permettre de reconstruire des aires de circulation liés à l'accès aux ressources et au déplacement entre habitats afin de mieux visualiser l'impact général des populations précolombiennes sur la forêt. Les travaux préliminaires décrits ici ne sont encore que des pistes qui permettront peut-être un jour de quantifier et de délimiter les zones forestières qui ont été perturbées par les occupations précolombiennes.

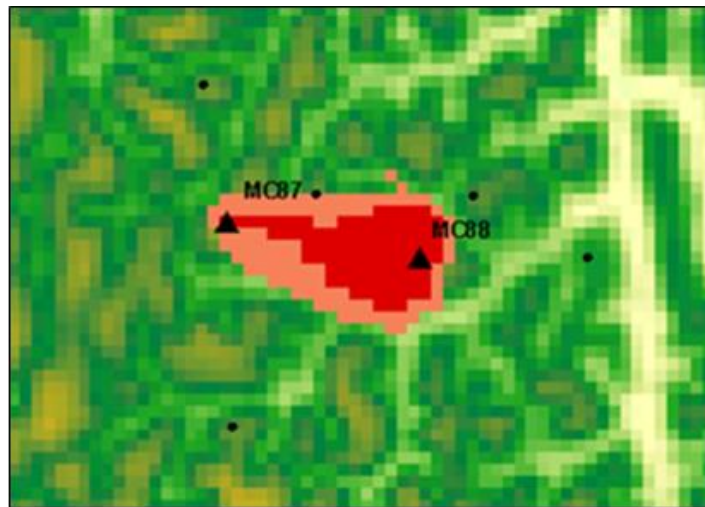


Figure A.4. Aire de circulation du site MC87 vers le site MC88 réalisée avec une méthode SIG. La construction du couloir de coût entre les deux sites nécessite un modèle numérique de terrain (MNT) pour représenter la pente et construire les couches de direction, d'allocation et de distance. Les points noirs représentent les autres sites sur lesquels des fragments de poterie ont été retrouvés.