1 # How ancient forest fragmentation and riparian connectivity

2 # generate high levels of genetic diversity in a micro-endemic

3 # Malagasy tree

4

5 Jordi Salmona[1*], Axel Dresen[1], Anicet E. Ranaivoson[1,2], Sophie Manzi[1], Barbara Le Pors[3],

6 Cynthia Hong-Wa[4], Jacqueline Razanatsoa[5], Nicole V. Andriaholinirina[2], Solofonirina

7 Rasoloharijaona[2], Marie-Elodie Vavitsara[2], Guillaume Besnard[1*]

8

9 [1] CNRS-UPS-IRD, UMR5174, Laboratoire Évolution & Diversité Biologique, Université Paul

10 Sabatier, 118 route de Narbonne, 31062 Toulouse, France

11 [2] Faculté des Sciences, Université de Mahajanga, BP 652 401, Mahajanga, Madagascar

12 [3] Instituto Gulbenkian de Ciênca, Rua da Quinta Grande, 6, P-2780-156 Oeiras, Portugal

13 [4] Claude E. Phillips Herbarium, Delaware State University, 1200 N. Dupont Hwy, Dover, DE

14 19901-2277, USA

15 [5] Herbier, Département Flore, Parc Botanique et Zoologique de Tsimbazaza, BP 4096,

16 Antananarivo - 101, Madagascar

17

18 *Corresponding authors:

19 Jordi Salmona: jordi.salmona@gmail.com

20 Guillaume Besnard: guillaume.besnard@univ-tlse3.fr

21

22 Total word count: 5932 (Introduction: 1104, Materials and Methods: 1732, Results: 1031,

23 Discussion: 2065),

24 Number of figures: 4, all figures should be published in color.

25 Number of tables: 1

26 The Supporting Information comprises 10 method and 3 result supporting paragraphs, 10 tables,

27 25 figures and associated references.

28

Abstract

- Understanding landscape changes is central to predicting evolutionary trajectories and defining conservation practices. While human-driven deforestation is intense throughout Madagascar, exception in areas like the Loky-Manambato region (North) raises questions. This region also harbors a rich and endemic flora, whose evolutionary origin remains poorly understood.

- We assessed the genetic diversity of an endangered micro-endemic Malagasy olive species (*Noronhia spinifolia*) to better understand the vegetation dynamic in the Loky-Manambato region and its influence on past evolutionary processes. We characterized 72 individuals sampled across eight forests through nuclear and mitochondrial restriction associated sequencing data (RADseq) and chloroplast microsatellites (cpSSR).

- Extremely high genetic diversity was revealed in the three genomic compartments (chloroplast $h = 0.99$, mitochondrial $h = 0.85$, and nuclear $H_O = 0.07\text{-}0.20$). Combined population and landscape genetics analyses indicate that *N. spinifolia* diversity is best explained by the current forest cover ($R^2 = 0.90$), highlighting a long-standing forest fragmentation in the region. Our results further suggest a predominant role of forest-dwelling organisms in mediating pollen and seed dispersals.

- This sustains a major and long-term role of riparian corridors in maintaining connectivity across those antique mosaic-habitats, calling for the study of organismal interactions that promote gene flow.

Key words: Habitat loss and fragmentation, Landscape genetics, Malagasy olive, Mitochondrial DNA, gene flow, connectivity, cpSSR, RADseq, Madagascar.

Introduction

Offsetting rapid anthropogenic habitat destruction and fragmentation, the primary causes of declines in global biodiversity (Fahrig, 2003; Lindenmayer & Fischer, 2013; Goudie, 2018), requires, among others, to urgently preserving connectivity (Haddad *et al.*, 2015). Although defining appropriate conservation programs largely depends on knowledge of species dispersal strategies (Sutherland *et al.*, 2004; LeBuhn *et al.*, 2015; Gardner *et al.*, 2018), these remain poorly understood, in particular in tropical hotspots. This typically requires understanding species diversity, their dynamic, behavior and interactions across the landscape (Pressey *et al.*, 2007), which can be efficiently inferred from genetic data (Frankham, 2010; Salmona *et al.*, 2017a).

Madagascar's unique biodiversity (Goodman & Benstead, 2003; Myers *et al.*, 2000), constitutes an ideal model to study evolutionary processes of diversification (Vences, 2005; Wilmé *et al.*, 2006; Vences *et al.*, 2009). Drivers of evolution, such as riverine barriers (Craul *et al.*, 2008), refugia interconnection (Wilmé *et al.*, 2006), and habitat loss and fragmentation (Yoder *et al.*, 2016; Salmona *et al.*, 2017b), have been identified from taxonomic diversity and the genetic makeup of the Malagasy biota. However, assessing the relative and confounding effects of complex landscape dynamics (forest loss, fragmentation, barriers emergence, etc.) on population dynamics, is notoriously challenging (Nater *et al.*, 2015; Salmona *et al.*, 2017a,b; Beichman *et al.*, 2018).

Although deforestation has greatly intensified across Madagascar (~40-50%) since the 1950's (Harper *et al.*, 2007; Vieilledent *et al.*, 2018), it has also remained surprisingly mild in some regions like the Loky-Manambato (northern-Madagascar) (Quéméré *et al.*, 2012). Furthermore, the Miocene expansion of the Malagasy savannas (Vorontsova *et al.*, 2016; Hackel *et al.*, 2018; Solofondranohatra *et al.*, 2018; Salmona *et al.*, 2020) evidenced the antiquity of open-canopy environments. In some regions, it therefore remains crucial to confirm the questionable antiquity of landscapes and to identify their related change-inducing factors [e.g. (Quéméré *et al.*, 2010; Yoder *et al.*, 2016; Salmona *et al.*, 2017b, 2020)].

The Loky-Manambato (LM) region rose as a small-scale model-region to study habitat loss and fragmentation, thanks to its well-characterized matrix of forests and open-habitats, the

3

82   diversity of its putative barriers to gene flow, as well as its high levels of endemicity across

83   living kingdoms (Goodman & Wilmé, 2006; Goodman *et al.*, 2018). For instance, the forest-

84   matrix was identified as the landscape feature shaping genetic diversity across all species studied

85   in the LM region, while the Manankolana River, showed a strong effect on *Propithecus*

86   *tattersalli*, not consistently recovered in other species (Quéméré *et al.*, 2010; Rakotoarisoa *et al.*,

87   2013a; Sgarlata *et al.*, 2018; Aleixo-Pais *et al.*, 2019; Tang *et al.*, 2020). Although multiple

88   studies on mammals attempted to describe and understand the processes that shaped its

89   landscape and generated its diversity (Quéméré *et al.*, 2012; Rakotoarisoa *et al.*, 2013b; Salmona

90   *et al.*, 2017b; Sgarlata *et al.*, 2018, 2019), contributions on other taxa, such as plants, are crucial

91   to draw taxonomically-broad generalities regarding the antiquity of its landscape, its connectivity

92   and conservation.

93   Native tree species are putatively good models for landscape genetics studies in

94   fragmented habitats, being the primary and immediate victims of deforestation and landscape

95   changes. However, few studies have used the genetic diversity of Malagasy plant populations

96   (Andrianoelina *et al.*, 2009; Gardiner *et al.*, 2017; Salmona *et al.*, 2020) to infer landscape

97   dynamics and inform conservation. The Malagasy olives (genus *Noronhia*), with a high number

98   of taxa and a high micro-endemism rate, are among the major components of Madagascar forests

99   and of the LM region in particular (Hong-Wa & Besnard, 2014; Hong-Wa, 2016). Among them,

100 the Malagasy spiny olive (*Noronhia spinifolia* Hong-Wa) is mostly endemic to the dry to sub-

101 humid forests of the LM region; and although it is relatively frequent there, it is of high

102 conservation concern due to its narrow range. With such a distribution, *N. spinifolia*'s genetic

103 diversity holds the potential to have retained information about the macro- and micro-

104 evolutionary processes that have shaped the genus and species-level diversity in the region.

105 Furthermore, being narrowly distributed, it may hold relatively low genetic diversity (Kimura,

106 1983) and suffer from inbreeding depression due to recent population collapse. *Noronhia*

107 *spinifolia* therefore represents an excellent model to better understand Malagasy olives' ecology

108 and offers a case study to define appropriate action for dry-forests plant conservation in northern

109 Madagascar.

110   In such sexually-reproducing plants, dispersal occurs by two means: via haploid male

111 gametes in pollen, and via diploid embryos in seeds. Without field data, population and

112 landscape genetics offer an alternative way to estimate effective dispersal (Holderegger *et al.*,

4

113   2010; Balkenhol *et al.*, 2016). In particular, the combined use of complementary maternally and

114   biparentally inherited genetic data [respectively from chloroplast or mitochondrial genomes

115   (cpDNA or mtDNA) and the nuclear genome (nDNA)] allows disentangling, to a certain level,

116   the relative contribution of seed and pollen dispersals in gene flow. For instance, the congeneric

117   *N. lowryi* exhibited contrasting strong chloroplast and near-panmixia nuclear genetic structure

118   suggesting a long and short distance dispersal of pollen and seed, respectively (Salmona *et al.*,

119   2020). While progresses in sequencing technologies facilitated the generation of such genetic

120   data for non-model organisms (Allendorf *et al.*, 2010), recent advances in spatially explicit

121   analyses also unlocked our ability to estimate the effect of numerous collinear landscape features

122   on genetic diversity (Balkenhol *et al.*, 2016; Prunier *et al.*, 2017). Furthermore, although the

123   limited number of tested alternative landscape hypotheses long relied on prior knowledge or

124   expert opinions, recent approaches iterating around a large panel of resistance values (Graves *et*

125   *al.*, 2013) or searching for Bayesian optima (Peterman, 2018), widened the potential to

126   identifying relevant landscape components while optimizing their cost values from the genetic

127   data itself.

128   Here, we used genomic data from recently collected specimens of *N. spinifolia* across

129   most of its range, the LM region. We first tested whether its restricted geographic distribution

130   resulted in a low genetic diversity, as expected under a neutral model (Kimura, 1983), or

131   remained relatively high as for co-distributed primates [*P. tattersalli* and *Microcebus tavaratra*

132   (Quéméré *et al.*, 2010; Aleixo-Pais *et al.*, 2019)]. We then inferred the effect of landscape

133   components on maternally and biparentally inherited genetic diversity, to investigate patterns of

134   seed and pollen dispersals, and assessed their congruence with those of a congeneric species

135   from the High Plateau [*N. lowryi* (Salmona *et al.*, 2020)], and of co-distributed taxa

136   (abovementioned). We also examined whether the relative stability of the forest cover in the past

137   70 years (Quéméré *et al.*, 2012; Salmona *et al.*, 2017b) is reflected in *N. spinifolia* genetic

138   makeup, comparing the effect of recent and historical forest covers on gene flow, as a proxy for

139   the temporality of its habitat fragmentation. Finally, we present the application of our work to the

140   conservation of the LM region forest network.

Material and methods

Study region

The Loky-Manambato (LM) region (Daraina; Fig. 1) is a biogeographical transition zone between dry deciduous and humid forests (Goodman & Wilmé, 2006), which is delimited by the Loky and Manambato Rivers. This region is crossed by the relatively shallow Manankolana River, bordered by riparian forests along most of its course, and by a national dirt road (Fig. 1). It consists of an area of ~2,500 km² covered by ~360 km² of forests (Goodman *et al.*, 2018), fragmented into a dozen major forest patches surrounded by human-altered grasslands, dry scrub and agricultural lands. Most forests are situated at low- to mid-elevations and mostly consist of dry deciduous vegetation. In contrast, some mountain forests (Binara and Antsahabe, plus Bobankora to a lower extent) are covered by a gradient of dry deciduous, transition, humid and ericoid vegetation (Gautier *et al.*, 2006). Despite sustained grassland fires, slash-and-burn agriculture and charcoal production, as well as exploitation of wood, gold and sapphires (Fanamby, 2010; Goodman *et al.*, 2018), deforestation rate in the LM region is relatively low (Quéméré *et al.*, 2012) compared with those of eastern and southwestern Madagascar (Vieilledent *et al.*, 2018), likely stemming from its remoteness, difficult accessibility and climate. However, to mitigate the threats, the LM region progressively became managed as a protected area by the Malagasy NGO "Fanamby" since 2005 (Fanamby, 2010; Goodman *et al.*, 2018).

Study species

Noronhia spinifolia (Oleaceae) is a small-sized, hermaphroditic, understory tree that is easily distinguishable from other *Noronhia* species by its narrow linear leaves with a spiny tip. It is micro-endemic to northern Madagascar, mainly found in the LM region except for one record from further north in Montagne des Français, and is reported mainly in semi-deciduous forests of low altitude. *Noronhia spinifolia* has been assigned a preliminary conservation status of "Endangered" due to threats to its habitat (Hong-Wa, 2016).
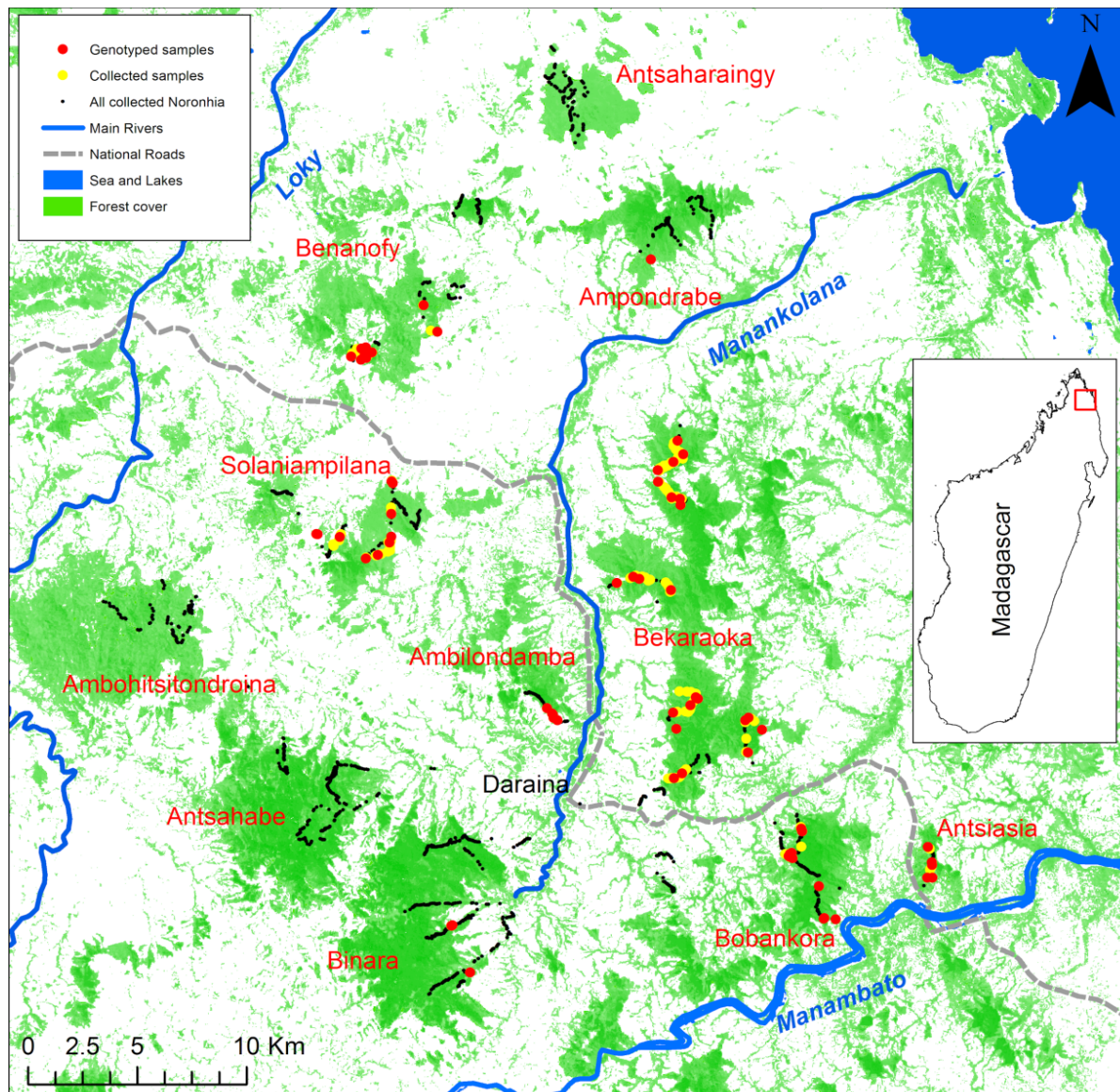
Figure 1: Map of *Noronhia spinifolia* sampling in the Loky-Manambato (LM) region.

The small black points represent samples collected for all *Noronhia* species (ca. 30 distinct taxa) and illustrate the survey effort conducted in the region. The yellow and red dots represent *N. spinifolia* samples, with the red dots corresponding to samples included in our genomic analyses. The forest cover is adapted from Hansen *et al.* (2013). Pixels with less than 30% tree cover are represented in white. The remaining tree cover percentage values are represented from light green (30%) to dark green (100%). This forest cover representation also illustrates the presence of riparian forests along streams of the LM region.

Plant sampling

To sample *N. spinifolia* populations, we surveyed all major forests of the LM region (Fig. 1) in 2017 and 2018, during the dry season (July-September), and used topography (altitude and shape) as a sampling guide to maximize the representation of all landscape features. Most surveys started from the forest edge at low altitude towards the forest core and higher elevation. We identified *Noronhia* species based on tree characteristics, leaf morphology and tissue structure, and collected 220 leaf samples, preserved in silica gel for DNA conservation. For each tree, we systematically recorded its height, diameter and reproductive state, as well as its geographical coordinates (GPS), elevation, and habitat type. For all forests, at least one specimen voucher was prepared and deposited at the herbarium of the Parc Botanique et Zoologique de Tsimbazaza (TAN).

Laboratory procedures

DNA extraction, organellar and nuclear genotyping

We extracted DNA from 137 samples of *N. spinifolia*, selected to maximize geographic and altitudinal representation and prioritizing mature and fully-grown trees. DNA extraction used a commercial protocol adapted to plants, followed by quality control procedures ensuring high quality genomic DNA for 72 samples selected for genotyping (Fig. 1, Methods S1). Using a two-pronged approach, we genotyped 15 chloroplast microsatellites (cpSSR) and one mitochondrial microsatellite (mtSSR), originally developed on *Olea europaea* (Table S1, Methods S2, S3; Besnard *et al.*, 2011), and also used restriction associated DNA sequencing (RADseq; generating data from the biparentally inherited nuclear genome and the mitogenome; Methods S4). RADseq consists in sequencing regions neighboring restriction sites, to obtain homologous sequences across individuals, spread across the genome, at a decent coverage and a reasonable cost (Baird *et al.*, 2008; Andrews *et al.*, 2016).

Data processing

Organellar RADseq loci and *de-novo* assembly of the nuclear loci catalog

After ad-hoc demultiplexing and cleaning of reads (Methods S4), we screened the organellar genomes using bwa-mem sequence alignment (Li, 2013) to the *N. clarinerva* mitogenome and *N. spinifolia* plastome (MW202230 and MT081057, respectively; Methods S5). We recovered ten mitochondrial *Sbf*I RAD loci identified *in silico* and called haplotypes using ANGSD v0.92 (Nielsen *et al.*, 2012; Korneliussen *et al.*, 2014), based on their highest effective base depth (Wang *et al.*, 2013). Conversely, no cpDNA RAD locus was recovered, confirming *in silico* analyses (Methods S5).

A catalog of nuclear tags (loci) was *de-novo* optimized (Methods S6) by iterating around the core parameters of Stacks (Rochette *et al.*, 2019) to maximize the amount of available biological information (Paris *et al.*, 2017). The final catalog was further cleaned (Methods S6) for exogenous contaminants using DeconSeq (Schmieder & Edwards, 2011) and endogenous orthologs using MUMmer (Kurtz *et al.*, 2004).

RADseq genotyping

We used two fundamentally distinct genotyping approaches to ensure the robustness of our results: single nucleotide polymorphism (SNPs) called in Stacks, and genotype likelihoods (GLs) estimated with ANGSD (Methods S7). GLs retain information about uncertainty in base calls, which alleviates some issues associated with RADseq data such as unevenness in sequencing depth and allele drop-outs (Heller *et al.*, accepted; Pedersen *et al.*, 2018; Warmuth & Ellegren, 2019).

Landscape genetics

We conducted complementary analyses to assess the effect of landscape components on the genetic diversity of *N. spinifolia*. We first investigated the raw patterns of genetic diversity and structure without priors to describe the major trends and build hypotheses. Then, using univariate approaches under an isolation-by-resistance model (IBR; McRae, 2006), we assessed the effect of each landscape component, iterating through their cost and resolution. Finally, using a

227 multivariate model considering spatial autocorrelation and multicollinearity, we assessed the
228 contribution of selected landscape components.

Genetic diversity

230 We assessed the proportion of heterozygous genotypes (H_E) from nuclear genotype likelihoods
231 (GL) based on folded site frequency estimated in ANGSD. We further estimated organellar
232 diversity (h), the probability that two haplotypes are different (Nei, 1987).

Population structure

234 We assessed the level of genetic differentiation among localities with Reynolds' weighted F_{ST}
235 (Reynolds *et al.*, 1983) from GL inferred in ANGSD. We explored the genetic structure of our
236 study system through naive clustering analyses (Methods S8), based on ANGSD GLs using
237 NgsAdmix v32 (Skotte *et al.*, 2013) and on Stacks called genotypes using ADMIXTURE v1.3.0
238 (Alexander *et al.*, 2009), and with a principal component analysis (PCA) from GLs with
239 PCAngsd. We estimated the level of organellar genetic differentiation among forests with Nei's
240 weighted F_{ST} (Nei, 1973) using the R package *hierfstat*. We also investigated the phylogenetic
241 structure of organellar DNA data using minimum spanning networks of genetic distances (see
242 below) constructed with the R package *poppr* (Kamvar *et al.*, 2015).

Genetic distances

244 We assessed the power of several measures of among-individuals genetic relationships (distances
245 or relatedness) from chloroplast, mitochondrial and nuclear data. For cpSSR data, we used the
246 Bruvo's and Prevosti's genetic distances (Prevosti *et al.*, 1975; Bruvo *et al.*, 2004). From
247 mtRAD SNPs, we inferred Euclidian and Manhattan distances. We estimated an overall genetic
248 distance for organellar genomes by combining weighted Manhattan mtDNA and Bruvo's
249 cpDNA distances (Methods S3).
250 We estimated the covariance of nuclear RADseq GLs (Meisner & Albrechtsen, 2018), as
251 well as Hall's and Vieira's metrics (Hall *et al.*, 2012; Vieira *et al.*, 2013) in PCAnsgd. Using
252 nuclear SNP data, we also computed the Nei's genetic distance (Nei, 1972) and the Yang's
253 relatedness (Yang *et al.*, 2010) in the *StAMPP* R package (Pembleton *et al.*, 2013).

Isolation by distance

We investigated patterns of isolation by distance (IBD) to assess how the geographic distance alone explains the genetic diversity (Wright, 1943; Slatkin, 1993). We used Mantel tests (Mantel, 1967) between individual geographic and genetic distances (Methods S9). Since IBD may be limited to a certain scale (e.g. Keller & Holderegger, 2013; Van Strien *et al.*, 2015; Cayuela *et al.*, 2019), we compared subsets of pairwise data defined by a maximum geographic distance (S) between samples (Methods S9).

Isolation by resistance

Landscapes are rarely homogeneous, and gene flow may be limited or facilitated by its components. We used an IBR approach (McRae, 2006) to assess the cost associated with effective dispersal through each landscape feature.

Landscape variables, cost and resolution

As *N. spinifolia* was recently described and occurs in a remote area (Hong-Wa, 2016), we had little prior knowledge on the landscape variables that may affect pollen and seed dispersals. We therefore assessed the effect of most available landscape variables (e.g. roads, rivers, vegetation cover; Methods S10). To test if the genetic diversity of old trees may be better explained by past forest cover, we used forest cover data from 1953, 1973, and 2000s (Hansen *et al.*, 2013; Vieilledent *et al.*, 2018).

Although strong priors associating a landscape component to a particular cost may be available for well-studied species (e.g. Dellicour *et al.*, 2019; Quéméré *et al.*, 2010), landscape variables and their associated cost are often chosen almost arbitrarily when little or no data are available (Beier *et al.*, 2008, 2011). To identify the variable-cost associations that matter for our study system, we iteratively tested 14 conductance-resistance values (Methods S10). Similarly, organisms do not necessarily perceive each environmental component at the same resolution (or granularity: Baguette & Van Dyck, 2007; Everson & Boucher, 1998; Laurance *et al.*, 2007; Murcia, 1995). To identify the variable-cost-granularity relevant for *N. spinifolia*, we tested four pixel resolutions (Methods S10).

281 *Movement models*

282 To determine which dispersal model best applies to *N. spinifolia*, we used both the Least Cost

283 Path (LCP) and the Circuit Theory (CT). These two approaches, respectively, consider the least

284 cost trajectory and the cost of all possible trajectories (McRae & Beier, 2007). We computed

285 landscape distances using the R package *gdistance* (Van Etten, 2012).

286 *Statistical procedures*

287 We used a two-step procedure to first select landscape components, as well as their best fitting

288 cost, resolution, and movement model, and then, to assess their unique and common

289 contributions to *N. spinifolia*'s genetic diversity.

290 We estimated the correlation between geographic or landscape distance and genetic

291 matrices using Mantel tests (Mantel, 1967) in the R Package *vegan* (Dixon, 2003). We retained

292 variables showing a better fit (R^2) than IBD, exhibiting sensitivity to cost values (i.e. variables

293 with a fixed fit across all cost values were discarded), and selected their best fitting cost,

294 movement model, and resolution. We modeled the contribution of the retained landscape

295 variables using logistic regressions on distance matrices (LRDM; Prunier *et al.*, 2015; Smouse *et*

296 *al.*, 1986), a statistical procedure that is similar to classical multiple ordinary least-square

297 regressions, except that the significance of model fit (multiple R^2) is assessed through

298 permutations of the dependent matrix (Legendre *et al.*, 1994). We finally disentangled

299 multicollinearity among variables and decomposed their unique and common contributions using

300 commonality analyses (CA; Prunier *et al.*, 2015).

301

Results

Species occurrence

We sampled *N. spinifolia* in eight of the 11 surveyed major forests of the LM region (Fig. 1). The species occurs from low to medium elevation, between 87 and 505 m, but with strong discrepancies among forests (Fig. S1). While it was mainly recorded in dry forests, it was surprisingly found in dry to wet transition forests at medium elevation (451-505 m) in Binara. Furthermore, the species was not found in three major forest patches of the LM region - namely Antsahabe, Ambohitsitondroina and Antsaharaingy - despite (*i*) large prospection efforts in these forests, and (*ii*) apparently similar habitat as the neighboring forests harboring the species (Fig. 1).

Organellar DNA genotyping and nuclear catalog construction

Of the 15 chloroplast microsatellites, 14 showed polymorphism (Table S2), and allowed distinguishing 55 chlorotype profiles among 72 trees (Results S1). The ten mitochondrial RAD loci (mtRAD) allowed identifying 11 SNPs (Results S1; Table S3). The combination of mtRADs and the mtSSR locus permits the identification of 15 mitotypes among 72 trees (Table 1). The cpSSR markers showed low to moderate linkage disequilibrium (LD; Fig. S2), a likely consequence of microsatellite-repeat-length homoplasy. Meanwhile, the mtDNA markers showed either high (among seven loci) or null LD (Fig. S3). Because SNPs are expected to be more stable (unlikely homoplasy) than SSRs, null LD between SNP loci was not expected, and could indicate recombination in the mitogenome. Finally, the overall LD among mtDNA and cpDNA markers (Fig. S4) suggests that they are both maternally inherited, although paternal leaks may occur occasionally.

The nuclear catalog parameter space exploration allowed selecting values ($m = 4$, $M = 5$, $N = 8$) that offer a trade-off between the coverage, loci number, and SNP number, while limiting the number of paralogs and the presence of contaminants (Figs S5-7; Results S2). The SNP-calling procedure showing low ability to recover the genetic makeup of *N. spinifolia* (when compared to the GL-based procedure; Figs S8-13), we therefore limited its use to preliminary

13

329 analyses (ADMIXTURE & genetic distances) and proceeded with the GL-based procedure for

330 downstream analyses.

Genetic diversity

332 Chloroplast microsatellites revealed a relatively high genetic diversity with only two chlorotypes

333 shared by individuals from more than one forest, resulting in a high probability that two

334 randomly sampled haplotypes are different ($h = 0.99$) and a mean allelic richness (A_r; estimated

335 for five individuals) of 2.41 (Table 1). Consequently, most forests showed an extremely high

336 cpSSR genetic diversity ($h > 0.92$) with the exception of Binara that appeared slightly less

337 diverse ($h = 0.73$; Table 1). A relatively high mitotype diversity was also revealed [$h = 0.85$

338 (ranging from 0.66 to 0.97 per forest), $A_r = 2.12$]. Similarly, most sampled individuals exhibit

339 relatively high levels of nuclear diversity with ~7 to ~20% of polymorphic sites and large

340 discrepancies within and among forests (Table S1; Fig. S14). This diversity is not

341 homogeneously distributed in space, and higher levels of genetic diversity seem to occur in the

342 area from Solaniampilana to southern Bekaraoka (Fig. S15). Furthermore, genetic diversity does

343 not seem influenced by altitude (Fig. S16).

344 **Table 1: Chloroplast and mitochondrial summary statistics.**

Forests	cpSSR				mtRAD			
	N	n_h	h	A_r	N	n_h	h	A_r
Ambilondamba	6	5	0.98	2.22	6	4	0.97	2.16
Ampondrabe	1	1	-	-	1	1	-	-
Antsiasia	6	4	0.92	2.67	6	3	0.81	2.14
Bekaraoka	25	19	0.99	2.38	22	5	0.66	2.04
Benanofy	11	8	0.94	2.39	11	4	0.78	2.26
Binara	5	2	0.73	2.36	5	2	0.73	2.05
Bobankora	11	10	0.99	2.45	11	3	0.73	2.04
Solaniampilana	10	8	0.97	2.37	10	5	0.87	2.17
Total / Mean	75	55	0.99	2.41	72	15	0.85	2.12

345 N = number of analyzed individuals; n_h = number of haplotypes; h = haplotype diversity; A_r:
346 allelic richness (estimated for five individuals).

14

Population structure

348 The chloroplast and mitochondrial data both revealed substantial differentiation among forests

349 (F_{ST} estimates ranging from 0.040 to 0.393 for cpSSRs; and 0.005 to 0.661 for mtRADs). As

350 expected, a strong differentiation was also observed when combining cpDNA and mtDNA data

351 (F_{ST} estimates ranging from 0.101 to 0.401; Table S4). The Solaniampilana-Benanofy forest

352 cluster was clearly distinguished from other forests for both mtDNA and cpDNA (Figs S17-18),

353 while Bekaraoka and Bobankora showed limited divergence with their neighboring forests.

354 Haplotype networks based on cpSSR and/or mtRAD data also revealed that one maternal lineage

355 is unique to Solaniampilana and Benanofy (Fig. 2). Furthermore, the geographic Euclidean

356 distances showed low, but highly significant, power at explaining genetic distances among

357 individuals (R^2 [cpSSR]: 11.7%; R^2 [mtRAD]: 20.7%; and R^2 [cpSSR + mtRAD]: 21.3%; Figs

358 S13, S19; Results S3).

359
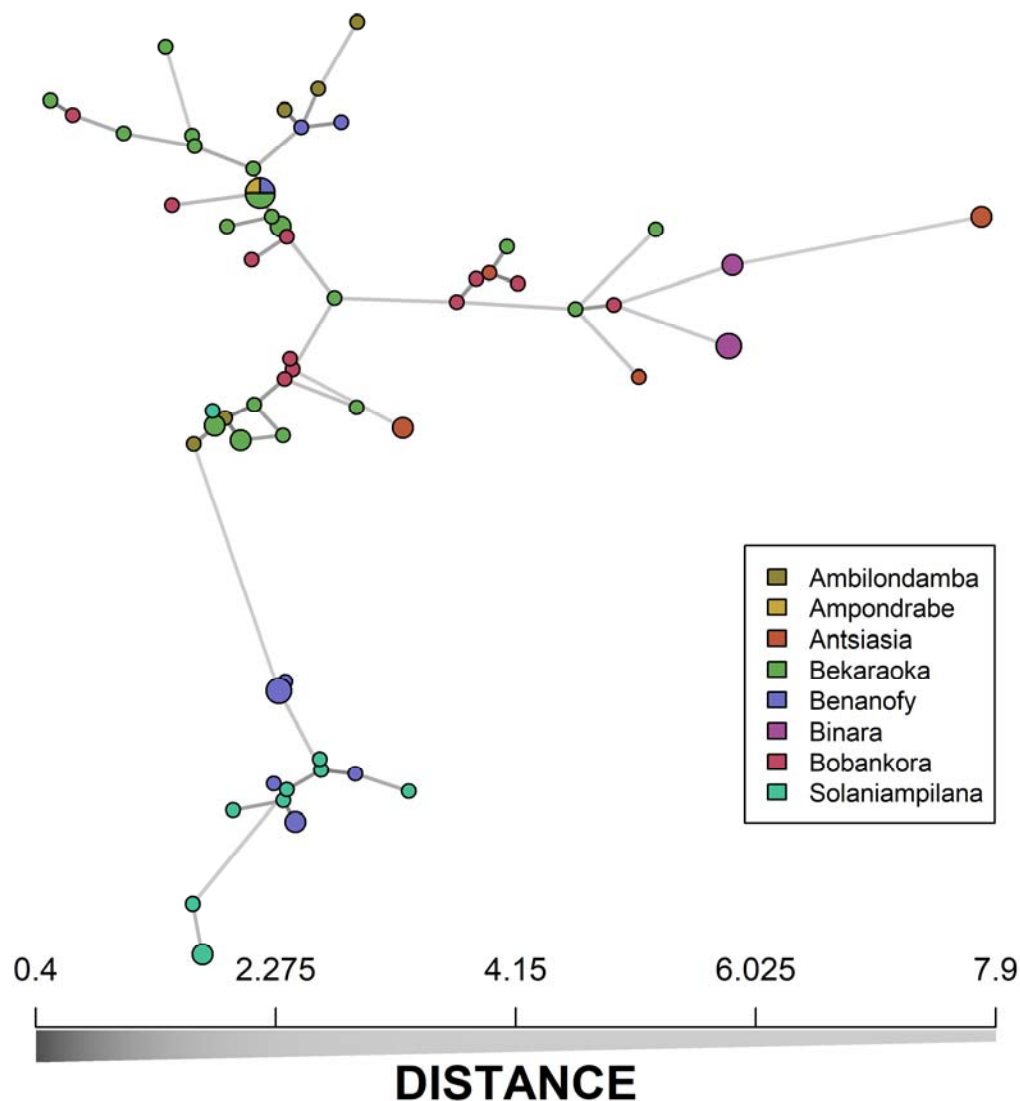
Figure 2: Organellar DNA haplotype network of *Noronhia spinifolia*.

Line length and grey scale are proportional to the Bruvo's cpDNA + Manhattan mtDNA combined genetic distances between distinct organellar haplotypes. Pie chart size is proportional to the occurrence number of a given haplotype. All edges of equal weight are represented. Distances among haplotypes are represented both through longer edges and the grey scale. The network highlights the huge organellar DNA diversity in *N. spinifolia*, with only one haplotype shared by individuals from at least two forests. It further shows a limited spatial structure, with, for instance, haplotypes from Solaniampilana and Benanofy grouping together at the bottom of the network.

F_{ST} estimates based on nuclear markers (Table S5) ranged from 0.089 to 0.210, indicating that most forests are differentiated from each other. However, we found no strong structure in sub-populations, with no particular support for number of clusters >1, both for GL- and SNP-

16

373   based analyses (Figs S8-9). Instead, we found a clear northeast-southwest signal of continuous
374   genetic differentiation across space, through GL-based PCA (First axis, ~15% of the variance
375   explained; Fig. S20), clustering (Figs 3, S10-11), and IBD analyses (Figs S13, S19). The
376   observed continuous structure is well illustrated by the clustering structure for $K = 3$ that shows
377   an admixed pattern from and to most sampling sites (Fig. 3). We found a clear IBD signal
378   explaining up to 56.6% of the among-individuals nuclear GL covariance (Fig. S19).

379
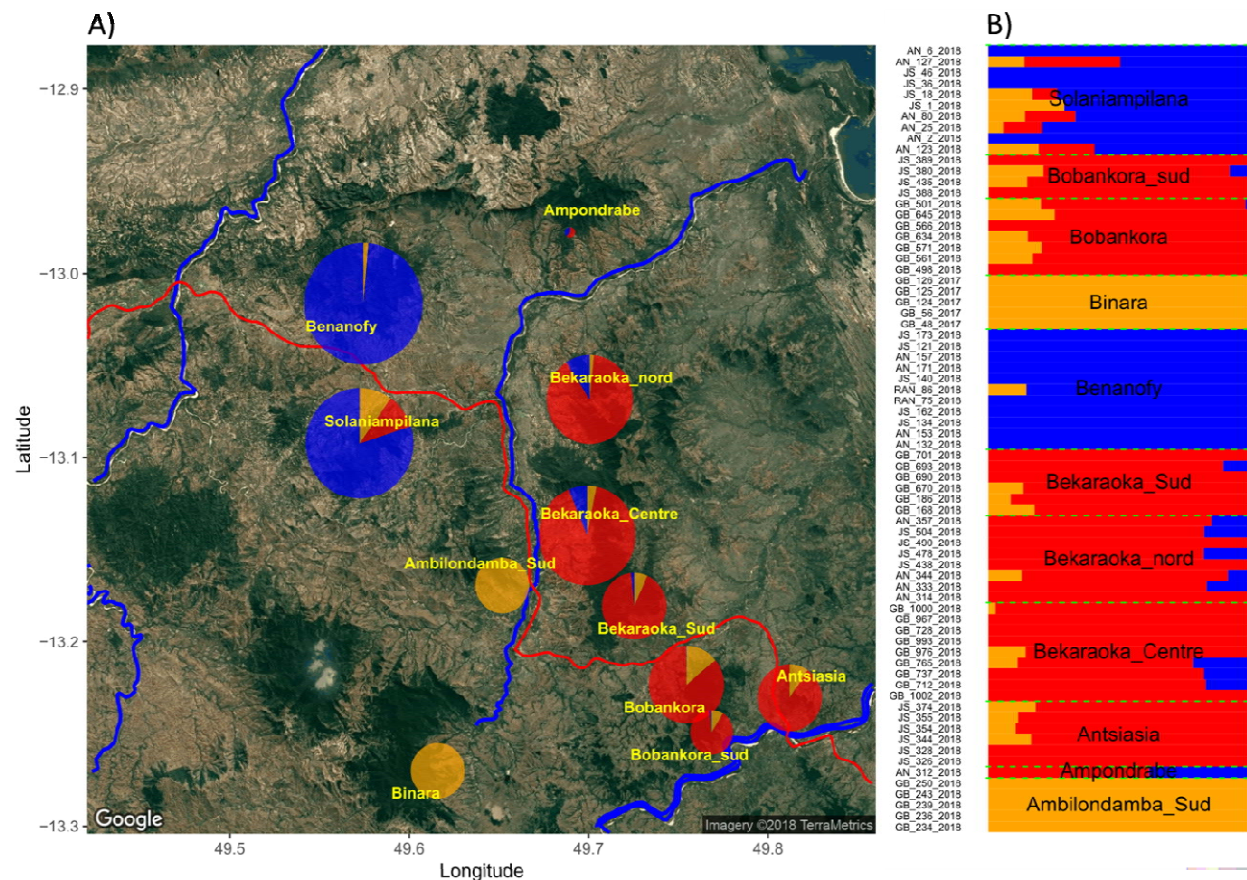
Figure 3: Spatial genetic structure of *Noronhia spinifolia* in the Loky-Manambato region.
NgsAdmix ancestry proportions (for $K = 3$ genetic clusters) represented either (**A**) spatially by
sampling site, or (**B**) per individual. Size of pie charts (in A) is proportional to the number of
samples per site. Pie shares represent the sums of individual ancestry proportions that are showed
in B. Results are arbitrarily represented for $K = 3$, according to the likelihood and delta$K$ results
in Fig. S8, because this K value best illustrates the continuous pattern of structure inferred using
ngsAdmix and other approaches.

387

Landscape genetics

The optimization of resistance surfaces through univariate comparison of genetic and landscape distances (IBR) showed lower fit for cpDNA (R^2 max ~0.14) than for mtDNA (R^2 max ~0.38) and nDNA (R^2 max ~0.90). We discarded landscape variables showing uniform response at varying costs (e.g. wind) and those exhibiting lower fit than the null IBD layer. Among the four vegetation layers, the continuous and discrete percent tree cover layer always exhibited the highest fit for conductance values at high resolution with cpDNA, mtDNA and nDNA (R^2 = 0.14; 0.38 and 0.90, respectively; Figs S21- 24). In other words, the percent tree cover data alone shows a strong conducting effect on gene flow and explains a very large portion of the genetic variation (R^2 = 0.90). Altogether the parameter space exploration reveals a strong effect of all forest cover layers, whereas some other variables (i.e., rivers, roads and slope) may have subtle lower effects too. Our results combining LRDM and CA confirmed that forest cover was the best landscape predictor of genetic differentiation, releasing other landscape components and IBD to account mostly for collinearity with the forest cover (Fig. 4; Table S6). This pattern was consistent across organellar and nuclear DNA (Table S6), and the high quality percent tree cover from Hansen *et al.* (2013) was always the best forest cover predictor (Table S6). The 2000's forest covers all better fit genetic diversity than the 1953 and 1973 forest covers, meaning we did not recover particular effect of the documented forest-cover changes on the genetic diversity of *N. spinifolia*.
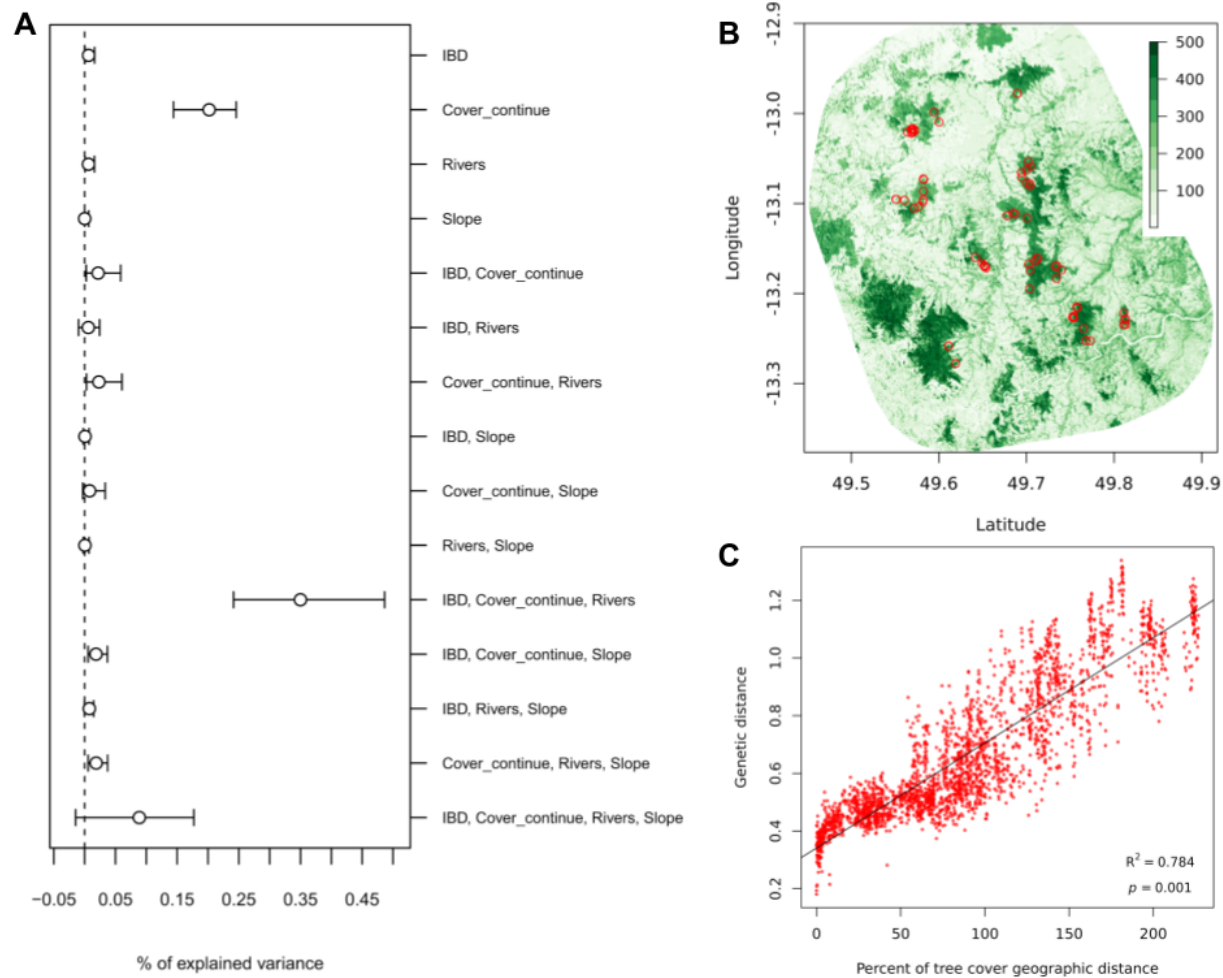
Figure 4: Landscape contribution to nuclear gene flow in *Noronhia spinifolia*.

A) Unique and common contributions of four selected landscape variables to nuclear gene flow, estimated using commonality analysis. **B**) Geographic representation of the percent tree cover conductance (inverse of cost), which illustrates the landscape conductance. **C**) Graphic representation of the relationship between percent tree cover geographic distances and genetic distances (isolation by resistance). This figure illustrates a strong conducting effect of forest cover (percent tree cover) on the connectivity of *N. spinifolia*, and it further shows that Euclidean geographic distance (IBD), the Manankolana River (Rivers) and the topology (Slope) have very low unique contribution, if any, to *N. spinifolia* nuclear gene flow. Cover_continue: Percent tree cover, conductance = 5; IBD: Isolation by distance, resistance = 1; Rivers: resistance = 5; Slope: conductance = 5.

Discussion

420

421 From a comprehensive and extensive sampling of *Noronhia spinifolia* in its core distribution
422 area, and leveraging the rare combination of nuclear and mitochondrial RADseq data with
423 cpDNA microsatellites, this study allowed us to reveal a strong effect of forest cover on gene
424 flow in a fragmented habitat in northern Madagascar. We not only report a surprisingly high
425 organellar and nuclear genetic diversity unevenly distributed in space, but also found that GL-
426 based approaches were able to recover much more information than SNP-calling approaches in
427 our model species. Moreover, the iterative optimization of resistance surface allowed identifying
428 outstanding landscape variables with a strong effect on the connectivity of *N. spinifolia*. Finally,
429 we show that recent forest cover better explains the genetic diversity of *N. spinifolia* than more
430 ancient ones.

Noronhia spinifolia, a highly diverse Malagasy micro-endemic

431

432 Our analyses exhibit unexpectedly high chloroplast ($h = 0.99$; 55 chlorotypes for 72 individuals),
433 mitochondrial ($h = 0.85$; 15 mitotypes), and nuclear (~7-20% polymorphic sites) genetic
434 diversity in a micro-endemic Malagasy tree species.

435 Firstly, the cpDNA diversity is tremendously higher than that of another micro-endemic
436 congener of the High Plateau (*N. lowryi*) when using the same 15 cpSSR loci [6 haplotypes in 77
437 individuals; $h = 0.58$ (Salmona *et al.*, 2020)]. More surprisingly, more cpDNA haplotypes and
438 diversity were revealed in 72 *N. spinifolia* individuals than in 1263 wild olive trees from the
439 whole Mediterranean basin [47 chlorotypes; $h_{cp} = 0.35$ (Besnard *et al.*, 2013)] and thus across
440 very different geographic scales (LM region = 900 km² *vs* Mediterranean basin = ~2.5 Million
441 km²) and despite the use of more polymorphic cpSSRs (35) in olive. Similarly, the *N. spinifolia*
442 mtDNA diversity is also higher than in the Mediterranean olive [4 mitotypes; $h_{mt} = 0.58$;
443 (Besnard *et al.*, 2002)], although comparable diversity level has been revealed in other plant
444 groups exhibiting large mitogenomes with high mutation rates as *Silene vulgaris* in Central
445 Europe [30 mitotypes; $h = 0.94$; (Štorchová & Olson, 2004)]. Finally, the nuclear genomic
446 diversity is ~20-40 times higher than that estimated in poplar populations across all Eurasia (Ma
447 *et al.*, 2018). The diversity is also twice as large as in five eastern-Madagascar mouse lemurs and

448   two orders of magnitude higher than in African plains zebra both estimated using similar data

449   and procedure (Pedersen *et al.*, 2018; Poelstra *et al.*, 2020). This high genetic diversity is

450   particularly unexpected for a narrowly distributed micro-endemic, and thus threatened, species.

451   Several non-exclusive evolutionary mechanisms may explain such an exceptionally high

452   intraspecific genetic diversity in *N. spinifolia*. Firstly, it suggests that a long-term maintenance of

453   a large effective population size precluded significant genetic drift. Persistent connectivity

454   between forest patches may have been key in this process, particularly during climatic

455   fluctuations of the Late Quaternary that may have contributed to fragmenting habitat, as

456   suggested for other species of the LM region (Quéméré *et al.*, 2012; Salmona *et al.*, 2017b).

457   Secondly, the genus *Noronhia* has extremely diversified in northern Madagascar (Hong-Wa,

458   2016), and about 30 taxa have been recently recorded and sampled in the LM region (JS & GB,

459   unpublished data). What caused such diversification remains unknown, but the co-occurrence of

460   closely related taxa may offer opportunities for multiple hybridization events, which could have

461   thus contributed to the increased genetic diversity in *N. spinifolia*. However, the cpSSR

462   characterization of four sympatric/parapatric LM *Noronhia* (i.e. *N. candicans*, *N. clarinerva*, *N.*

463   *crassinodis* and *N. intermedia*; > 200 individuals), closely related to *N. spinifolia* (according to

464   cpDNA and nrDNA data; Salmona *et al.*, 2020), shows that these species have no shared

465   chlorotype with our study model (GB, unpubl. data), thus suggesting that maternal introgression

466   events to *N. spinifolia*, if any, may be prehistoric. Lastly, high mutation rate may also contribute

467   to the high genetic diversity in *N. spinifolia*. An obvious acceleration of the mitogenome

468   evolutionary rate has been recently documented in the closely related species *N. candicans*, *N.*

469   *clarinerva*, *N. intermedia* and *N. spinifolia,* with a high number of di- or tri-nucleotide mutations

470   possibly reflecting frequent mtDNA recombination in this clade (Van de Paer, 2017), as also

471   suggested by a lack of LD between some SNPs. While accelerated mutation rate was missing on

472   the plastome (Salmona *et al.*, 2020), we are still lacking any evidence for the nuclear genome.

473   Such accelerating evolutionary rate could result from relatively frequent and recurrent

474   hybridization events in this group, promoting genomic instability (Fontdevila, 2005; Payseur &

475   Rieseberg, 2016). In conclusion, the surprisingly high genetic diversity calls for the identification

476   of the evolutionary, ecological and/or molecular mechanisms underlying this peculiar pattern.

477   **Landscape effects on the genetic diversity of *Noronhia spinifolia***

A strong continuous spatial structure

478
479 Beyond revealing surprisingly high levels of diversity, our results also show complementary
480 signals of a strong continuous structure in space (PCA, clustering and IBD), from both organelles
481 and the nucleus, in contrast to generally expected incongruent patterns among genomes
482 (Olofsson *et al.*, 2019; Bianconi *et al.*, 2020). While the Northwest-Southeast differentiation
483 cline represented as much as ~15% of the variance of the PCA, the geographic Euclidean
484 distance alone explained up to ~55% of the nuclear genetic variance using IBD tests. This strong
485 pattern of nuclear genetic structure sharply contrasts with the absence of nuclear spatial structure
486 in the savanna olive tree, *N. lowryi* (Salmona *et al.*, 2020). However, reported IBD patterns in
487 trees show a wide range from low values in *Dalbergia monticola* across eastern Madagascar
488 humid forests [$R^2 = 0.18$; (Andrianoelina *et al.*, 2009)], or *Coffea mauritiana* in the Reunion
489 Island [$R^2 = 0.21$; (Garot *et al.*, 2019)], to high values in *Swietenia macrophylla* in Central
490 America [$R^2 = 0.62$; (McRae & Beier, 2007)]. Unexpectedly, this genetic structure was here
491 extremely well explained by the vegetation cover (percent tree cover; mtDNA $R^2 = 0.38$; nDNA
492 $R^2 = 0.90$), releasing IBD to account mostly for collinearity with the forest cover. Although
493 strong landscape effects were also found in *S. macrophylla* (McRae & Beier, 2007), we report a
494 unique evidence of a strong habitat fragmentation effect explained mostly by one landscape
495 variable.

On seed-mediated gene flow: the organellar DNA testimony

496
497 Our results evidence striking differences between seed- and pollen-mediated gene flows.
498 Organellar IBR patterns (Figs S19, S21-24) suggest that seed-mediated gene flow is driven by
499 forest cover. However, the mtDNA and cpDNA haplotype networks (Figs 2, S17-18) show
500 complex histories with multiple potential fluxes among forests. In brief, we could not recover
501 landscape variables with striking effect on seed dispersal. Two non-exclusive interpretations can
502 be invoked for explaining this pattern: (*i*) relevant landscape variables are not available or of low
503 resolution (e.g. forest type and climatic variables); or (*ii*) the cpDNA and mtDNA diversities are
504 confounded by homoplasy, recombination, strong drift, long-term phylogenetic or demographic
505 history. Nevertheless, the prevailing effects of forest cover suggest that seed dispersal may be
506 primarily performed by forest-dwelling animals (zoochory), especially those with limited and/or

22

507 rare across-forest movements, such as lemurs, rodents and territorial birds (Quéméré *et al.*, 2010;

508 Rakotoarisoa *et al.*, 2013a; Sgarlata *et al.*, 2018; Aleixo-Pais *et al.*, 2019). However, such a

509 pattern could also result from infrequent seed ingestion by birds with broad-movement behavior.

510 A deep forest cover effect on gene flow

511 Unlike organellar DNAs, nDNA diversity is deeply explained by the LM region forest cover

512 (Fig. 4). While this partially confirms the effect of forest cover on seed dispersal since nDNA

513 diversity is influenced by both seed and pollen movement, wind-mediated pollen dispersal

514 favored in open-canopy environments is not supported here. It thus further sustains that pollen

515 dispersal is mediated by forest-dwelling organisms with movements limited by open-canopy

516 environments. Insect-mediated pollen dispersal in *N. spinifolia* is also strongly suggested by its

517 flower morphology and color (Hong-Wa, 2016). However, the currently limited knowledge of

518 the Malagasy entomofauna and plant-pollinator networks prevents us from clearly identifying

519 this species' forest-dwelling pollinators.

The antiquity of forest fragmentation in northern Madagascar

521 Our results further support a long-standing forest fragmentation in the LM region. First, a better

522 fit of all recent forest cover [2000's; (Hansen *et al.*, 2013; Vieilledent *et al.*, 2018)], compared to

523 older vegetation cover (1953, 1973), suggests that the small forest changes that have occurred

524 through this period (Quéméré *et al.*, 2012) are unable to explain the genetic diversity of *N.*

525 *spinifolia*. Second, because we sampled and genotyped mostly fully-grown and mature trees, and

526 since the generation time of *Noronhia* is potentially long [>20-50 years; (Salmona *et al.*, 2020)],

527 its estimated genetic diversity is expected to reflect ancient (e.g. ~50-100 years old) forest cover.

528 Although, the better resolution of the most recent forest layer could positively bias its fit, its

529 strong genetic correlation with aging trees' genetic diversity is sound evidence that the landscape

530 of the LM region was relatively stable for the last century. Our study does not infer demography

531 over time, nor does it try to assess the effect of forest cover changes over larger timeframe, e.g.

532 the last ten or so millennia, although ancient forest dynamics may have only induced limited drift

533 on *N. spinifolia*. Furthermore, its predominant distribution in low-elevation dry forest suggests

534 that this habitat type may have been spatially, topographically, and temporally extensive in

535 northern Madagascar, albeit frequently fragmented, as seemingly evidenced by a rare and likely

23

536 relictual occurrence of the species in contemporary high-elevation humid forest (e.g. Binara) and
537 similarly peculiar presence further north (e.g. Montagne des Français). Nevertheless, our result
538 concurs with those of several recent studies (Bond *et al.*, 2008; Quéméré *et al.*, 2012;
539 Vorontsova *et al.*, 2016; Yoder *et al.*, 2016; Salmona *et al.*, 2017b, 2020) supporting the
540 antiquity of open habitats in Madagascar and thus corroding the narrative that human arrival
541 alone changed the island's landscapes.

Further prospects and conservation implications

543 The power of coupling genomic data to landscape genetics allowed not only identifying major
544 landscape components influencing effective dispersal, but also their respective effects on seed
545 and pollen dispersals. This surprising result warrants further investigation using higher resolution
546 landscape and environmental layers, not available to our study. In particular, it would benefit
547 from the use of forest type, land use, and climate data of better resolution. Furthermore, while
548 our study clearly identifies that seed and pollen are dispersed by forest-dwelling organisms, it
549 neither identifies these organisms nor does it clearly show that seed and pollen do still effectively
550 disperse among forests. These questions could be tackled by inferring pedigree data from high
551 density population sampling, using field survey of potential dispersers during flowering and
552 fructification (e.g. camera tracking), and/or using metabarcoding approaches to assess the
553 interaction network within the LM forests.

554 While our study confirms the biological importance of the LM region, which is known
555 for its species richness and endemism across taxa (Goodman & Wilmé, 2006; Rakotondravony,
556 2006, 2009; Sgarlata *et al.*, 2019), and more specifically for the genus *Noronhia* (Hong-Wa,
557 2016), our results also have several implications for biodiversity conservation in the region:
558 - First, they underscore the conservation value of the often-overlooked intraspecific genetic
559 diversity, which is unexpectedly high in *N. spinifolia*.
560 - Second, this study highlights the importance of riparian forests of the LM region for their major
561 role both as corridors connecting forest fragments, which is supported by the fact that genetic
562 diversity in *N. spinifolia* is explained by forest cover rather than Euclidian distance, and as
563 vectors promoting the roles of vertebrates and insects on seed and pollen dispersal. Therefore,
564 actively maintaining, protecting, and reforesting riparian and corridor forests, which are likely

24

565   pivotal for the functional connectivity of *N. spinifolia* but also most native and endemic species

566   of the LM region (Quéméré *et al.*, 2010; Rakotoarisoa *et al.*, 2013a; Sgarlata *et al.*, 2018; Aleixo-

567   Pais *et al.*, 2019), remain critical conservation actions.

568   - Third, our study identifies the Binara forest as unique among the major forests of the LM

569   region and in urgent need of deeper conservation focus. Indeed, our extensive forest survey

570   allowed us to find and collect just a few samples in this forest, where they were found only at

571   unexpectedly higher altitude and wetter habitat (Fig. S1). Similarly, several other Malagasy olive

572   species (e.g. *N. ankaranensis*, *N. candicans*, *N. christenseniana* and *N. oblanceolata*; GB and JS

573   unpublished data) were found to occur only at higher altitude in the mountain evergreen forests

574   of this region (e.g. Binara and Antsahabe). Altogether, this pattern, though unclear, echoes the

575   peculiarities of these forests, that likely acted as refugia for numerous taxa during drier periods

576   (Raxworthy & Nussbaum, 1995; Goodman & Wilmé, 2006; Rakotoarisoa *et al.*, 2013b; Sgarlata

577   *et al.*, 2019).

Data availability

579   Raw RADseq data and RADseq mtDNA alignments have been deposited to the Short Read

580   Archive (SRA) NCBI database under the reference PRJNA632767. Organellar microsatellite

581   genotypes and mtRAD variants are available in Tables S7 and S8, respectively. All additional

582   data, scripts and materials are available to readers at 10.5281/zenodo.4290421.

Conflict of interest disclosure

584   The authors of this article declare that they have no financial conflict of interest with the content

585   of this article.

25

Acknowledgments

Author Contribution

JS and GB designed the experiment. JS, AER, BLP, RJ, CHW and GB were pivotal to field material collection and herbarium composition. JS, SM, and GB generated the genetic data. JS conducted bioinformatics and population genetic analyses. JS and AD conducted IBR analyses. JS and GB drafted a first version of the manuscript with a significant input from CHW. All co-authors agreed with the last version of the manuscript.

Supporting Information

References

663 Aleixo-Pais I, Salmona J, Sgarlata GM, Rakotonanahary A, Sousa AP, Parreira B, Kun-
664 Rodrigues C, Ralantoharijaona T, Jan F, Rasolondraibe E, *et al.* 2019. The genetic structure of a
665 mouse lemur living in a fragmented habitat in northern Madagascar. *Conservation Genetics* 20:
666 229–243.

667 Alexander DH, Novembre J, Lange K. 2009. Fast model-based estimation of ancestry in
668 unrelated individuals. *Genome Research* 19: 1655–1664.

669 Allendorf FW, Hohenlohe PA, Luikart G. 2010. Genomics and the future of conservation
670 genetics. *Nature Reviews Genetics* 11: 697–709.

671 Andrews KR, Good JM, Miller MR, Luikart G, Hohenlohe PA. 2016. Harnessing the power of
672 RADseq for ecological and evolutionary genomics. *Nature Reviews Genetics* 17: 81–92.

673 Andrianoelina O, Favreau B, Ramamonjisoa L, Bouvet J-M. 2009. Small effect of fragmentation
674 on the genetic diversity of *Dalbergia monticola*, an endangered tree species of the eastern forest
675 of Madagascar, detected by chloroplast and nuclear microsatellites. *Annals of Botany* 104: 1231–
676 1242.

677 Baguette M, Van Dyck H. 2007. Landscape connectivity and animal behavior: functional grain
678 as a key determinant for dispersal. *Landscape Ecology* 22: 1117–1129.

679 Baird NA, Etter PD, Atwood TS, Currey MC, Shiver AL, Lewis ZA, Selker EU, Cresko WA,
680 Johnson EA. 2008. Rapid SNP discovery and genetic mapping using sequenced RAD markers.
681 *PLoS One* 3: e3376.

682 Balkenhol N, Cushman S, Storfer A, Waits L (Eds.). 2016. *Landscape genetics: concepts,*
683 *methods, applications*. Oxford: John Wiley & Sons.

684 Beichman AC, Huerta-Sanchez E, Lohmueller KE. 2018. Using genomic data to infer historic
685 population dynamics of nonmodel organisms. *Annual Review of Ecology, Evolution, and*
686 *Systematics* 49: 433–456.

687 Beier P, Majka DR, Spencer WD. 2008. Forks in the road: choices in procedures for designing
688 wildland linkages. *Conservation Biology* 22: 836–851.

689 Beier P, Spencer W, Baldwin RF, McRAE BH. 2011. Toward best practices for developing
690 regional connectivity maps. *Conservation Biology* 25: 879–892.

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

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

696 Besnard G, Khadari B, Navascués M, Fernández-Mazuecos M, El Bakkali A, Arrigo N, Baali-
697 Cherif D, Brunini-Bronzini de Caraffa V, Santoni S, Vargas P. 2013. The complex history of the
698 olive tree: from Late Quaternary diversification of Mediterranean lineages to primary
699 domestication in the northern Levant. *Proceedings of the Royal Society B: Biological Sciences*
700 280: 20122833.

701 Bianconi ME, Dunning LT, Curran EV, Hidalgo O, Powell RF, Mian S, Leitch IJ, Lundgren
702 MR, Manzi S, Vorontsova MS, *et al.* 2020. Contrasted histories of organelle and nuclear
703 genomes underlying physiological diversification in a grass species. *Proceedings of the Royal*
704 *Society, London, Series B* 287: 20201960.

705 Bond WJ, Silander Jr JA, Ranaivonasy J, Ratsirarson J. 2008. The antiquity of Madagascar's
706 grasslands and the rise of C4 grassy biomes. *Journal of Biogeography* 35: 1743–1758.

707 Bruvo R, Michiels NK, D'Souza TG, Schulenburg H. 2004. A simple method for the calculation
708 of microsatellite genotype distances irrespective of ploidy level. *Molecular Ecology* 13: 2101–
709 2106.

710 Cayuela H, Boualit L, Laporte M, Prunier JG, Preiss F, Laurent A, Foletti F, Clobert J, Jacob G.
711 2019. Kin-dependent dispersal influences relatedness and genetic structuring in a lek system.
712 *Oecologia* 191: 97–112.

713 Craul M, Radespiel U, Rasolofoson DW, Rakotondratsimba G, Rakotonirainy O,
714 Rasoloharijaona S, Randrianambinina B, Ratsimbazafy J, Ratelolahy F, Randrianamboavaonjy
715 T. 2008. Large rivers do not always act as species barriers for *Lepilemur* sp. *Primates* 49: 211–
716 218.

717 Dellicour S, Prunier JG, Piry S, Eloy M-C, Bertouille S, Licoppe A, Frantz AC, Flamand M-C.
718 2019. Landscape genetic analyses of *Cervus elaphus* and *Sus scrofa*: comparative study and
719 analytical developments. *Heredity* 123: 228–241.

720 Dixon P. 2003. VEGAN, a package of R functions for community ecology. *Journal of*
721 *Vegetation Science* 14: 927–930.

722 Everson DA, Boucher DH. 1998. Tree species-richness and topographic complexity along the
723 riparian edge of the Potomac River. *Forest Ecology and Management* 109: 305–314.

724 Fahrig L. 2003. Effects of habitat fragmentation on biodiversity. *Annual Review of Ecology,*
725 *Evolution, and Systematics* 34: 487–515.

726 Fanamby. 2010. Plan de gestion environnementale et de sauvegarde sociale (PGESS). Etude
727 d'impact environnemental et social (EIES) de la Nouvelle Aire Protégée Loky-Manambato.

728 Fontdevila A. 2005. Hybrid genome evolution by transposition. *Cytogenetic and Genome*
729 *Research* 110: 49–55.

730 Frankham R. 2010. Challenges and opportunities of genetic approaches to biological
731 conservation. *Biological Conservation* 143: 1919–1927.

732 Gardiner LM, Rakotoarinivo M, Rajaovelona LR, Clubbe C. 2017. Population genetics data help
733 to guide the conservation of palm species with small population sizes and fragmented habitats in
734 Madagascar. *PeerJ* 5: e3248.

735 Gardner CJ, Waeber PO, Razafindratsima OH, Wilmé L. 2018. Decision complacency and
736 conservation planning. *Conservation Biology* 32: 1469–1472.

737 Garot E, Joët T, Combes M-C, Lashermes P. 2019. Genetic diversity and population divergences
738 of an indigenous tree (*Coffea mauritiana*) in Reunion Island: role of climatic and geographical
739 factors. *Heredity* 122: 833–847.

740 Gautier L, Ranirison P, Nusbaumer L, Wohlhauser S. 2006. Aperçu des massifs forestiers de la
741 région Loky-Manambato. *Inventaires de la faune et de la flore du nord de Madagascar dans la*
742 *région Loky-Manambato, Analamerana et Andavakoera. Série Sciences Biologiques* 23: 81–99.

743 Goodman SM, Benstead JP. 2003. *Natural history of Madagascar*. Chicago and London:
744 University Chicago Press.

745 Goodman SM, Raherilalao MJ, Wohlhauser S. 2018. *The terrestrial protected areas of*
746 *Madagascar: their history, description, and biota*. Chicago: University of Chicago Press.

747 Goodman SM, Wilmé L. 2006. Inventaires de la faune et de la flore du nord de Madagascar dans
748 la région Loky-Manambato, Analamerana et Andavakoera. *Recherche pour le Dévelopement* 23:
749 1–238.

750 Goudie AS. 2018. *Human impact on the natural environment: Past, present and future*.
751 Hoboken, NJ: Wiley-Blackwell.

752 Graves TA, Beier P, Royle JA. 2013. Current approaches using genetic distances produce poor
753 estimates of landscape resistance to interindividual dispersal. *Molecular Ecology* 22: 3888–3903.

754 Hackel J, Vorontsova MS, Nanjarisoa OP, Hall RC, Razanatsoa J, Malakasi P, Besnard G. 2018.
755 Grass diversification in Madagascar: in situ radiation of two large C3 shade clades and support
756 for a Miocene to Pliocene origin of C4 grassy biomes. *Journal of Biogeography* 45: 750–761.

757 Haddad NM, Brudvig LA, Clobert J, Davies KF, Gonzalez A, Holt RD, Lovejoy TE, Sexton JO,
758 Austin MP, Collins CD. 2015. Habitat fragmentation and its lasting impact on Earth's
759 ecosystems. *Science Advances* 1: e1500052.

760 Hall N, Mercer L, Phillips D, Shaw J, Anderson AD. 2012. Maximum likelihood estimation of
761 individual inbreeding coefficients and null allele frequencies. *Genetics Research* 94: 151–161.

762 Hansen MC, Potapov PV, Moore R, Hancher M, Turubanova SA, Tyukavina A, Thau D,
763 Stehman SV, Goetz SJ, Loveland TR. 2013. High-resolution global maps of 21st-century forest
764 cover change. *Science* 342: 850–853.

765 Harper GJ, Steininger MK, Tucker CJ, Juhn D, Hawkins F. 2007. Fifty years of deforestation
766 and forest fragmentation in Madagascar. *Environmental Conservation* 34: 325–333.

767 Heller R, Nursyifa C, Garcia Erill G, Salmona J, Chikhi L, Meisner J, Korneliussen TS,
768 Albrechtsen A. accepted. A reference-free approach to analyze non-model RADseq data using
769 standard Next Generation Sequencing toolkits. *Molecular Ecology Resources*: Accepted.

770 Holderegger R, Buehler D, Gugerli F, Manel S. 2010. Landscape genetics of plants. *Trends in*
771 *Plant Science* 15: 675–683.

772 Hong-Wa C. 2016. A taxonomic revision of the genus *Noronhia* Stadtm. ex Thouars (Oleaceae)
773 in Madagascar and the Comoro Islands. *Boissiera* 70: 1–291.

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

776 Kamvar ZN, Brooks JC, Grünwald NJ. 2015. Novel R tools for analysis of genome-wide
777 population genetic data with emphasis on clonality. *Frontiers in Genetics* 6: 208.

778 Keller D, Holderegger R. 2013. Damselflies use different movement strategies for short-and
779 long-distance dispersal. *Insect Conservation and Diversity* 6: 590–597.

780 Kimura M. 1983. *The neutral theory of molecular evolution*. Cambridge University Press.

781 Korneliussen TS, Albrechtsen A, Nielsen R. 2014. ANGSD: analysis of next generation
782 sequencing data. *BMC Bioinformatics* 15: 356.

783 Kurtz S, Phillippy A, Delcher AL, Smoot M, Shumway M, Antonescu C, Salzberg SL. 2004.
784 Versatile and open software for comparing large genomes. *Genome Biology* 5: R12.

785 Laurance WF, Nascimento HE, Laurance SG, Andrade A, Ewers RM, Harms KE, Luizao RC,
786 Ribeiro JE. 2007. Habitat fragmentation, variable edge effects, and the landscape-divergence
787 hypothesis. *PLoS One* 2: e1017.

788 LeBuhn G, Droege S, Connor EF, Gemmill-Herren B, Potts SG, Minckley RL, Jean RP, Kula E,
789 Roubik DW, Wright KW. 2015. Evidence-based conservation: reply to Tepedino et al.
790 *Conservation Biology* 29: 283–285.

791 Legendre P, Lapointe F-J, Casgrain P. 1994. Modeling brain evolution from behavior: a
792 permutational regression approach. *Evolution* 48: 1487–1499.

793 Li H. 2013. Aligning sequence reads, clone sequences and assembly contigs with BWA-MEM.
794 *arXiv preprint arXiv:1303.3997*.

795 Lindenmayer DB, Fischer J. 2013. *Habitat fragmentation and landscape change: an ecological*
796 *and conservation synthesis*. Washington, DC: Island Press.

797 Ma T, Wang K, Hu Q, Xi Z, Wan D, Wang Q, Feng J, Jiang D, Ahani H, Abbott RJ. 2018.
798 Ancient polymorphisms and divergence hitchhiking contribute to genomic islands of divergence
799 within a poplar species complex. *Proceedings of the National Academy of Sciences of the United*
800 *States of America* 115: E236–E243.

801 Mantel N. 1967. The detection of disease clustering and a generalized regression approach.
802 *Cancer Research* 27: 209.

803 McRae BH. 2006. Isolation by resistance. *Evolution* 60: 1551–1561.

804 McRae BH, Beier P. 2007. Circuit theory predicts gene flow in plant and animal populations.
805 *Proceedings of the National Academy of Sciences of the United States of America* 104: 19885–
806 19890.

807 Meisner J, Albrechtsen A. 2018. Inferring population structure and admixture proportions in
808 low-depth NGS data. *Genetics* 210: 719–731.

809 Murcia C. 1995. Edge effects in fragmented forests: implications for conservation. *Trends in*
810 *ecology & evolution* 10: 58–62.

811 Myers N, Mittermeier RA, Mittermeier CG, Da Fonseca GA, Kent J. 2000. Biodiversity hotspots
812 for conservation priorities. *Nature* 403: 853–858.

813 Nater A, Greminger MP, Arora N, Schaik CP, Goossens B, Singleton I, Verschoor EJ, Warren
814 KS, Krützen M. 2015. Reconstructing the demographic history of orang-utans using
815 Approximate Bayesian Computation. *Molecular Ecology* 24: 310–327.

816 Nei M. 1972. Genetic distance between populations. *The American Naturalist* 106: 283–292.

817 Nei M. 1973. Analysis of gene diversity in subdivided populations. *Proceedings of the National*
818 *Academy of Sciences of the United States of America* 70: 3321–3323.

819 Nei M. 1987. *Molecular evolutionary genetics*. Columbia University Press.

820 Nielsen R, Korneliussen T, Albrechtsen A, Li Y, Wang J. 2012. SNP calling, genotype calling,
821 and sample allele frequency estimation from new-generation sequencing data. *PLoS One* 7:
822 e37558.

823 Olofsson JK, Dunning LT, Lundgren MR, Barton HJ, Thompson J, Cuff N, Ariyarathne M,
824 Yakandawala D, Sotelo G, Zeng K. 2019. Population-specific selection on standing variation
825 generated by lateral gene transfers in a grass. *Current Biology* 29: 3921–3927.

826 Paris JR, Stevens JR, Catchen JM. 2017. Lost in parameter space: a road map for stacks.
827 *Methods in Ecology and Evolution* 8: 1360–1373.

828 Payseur BA, Rieseberg LH. 2016. A genomic perspective on hybridization and speciation.
829 *Molecular ecology* 25: 2337–2360.

830 Pedersen C-ET, Albrechtsen A, Etter PD, Johnson EA, Orlando L, Chikhi L, Siegismund HR,
831 Heller R. 2018. A southern African origin and cryptic structure in the highly mobile plains zebra.
832 *Nature Ecology & Evolution* 1: 491–498.

833   Pembleton LW, Cogan NO, Forster JW. 2013. St AMPP: An R package for calculation of
834   genetic differentiation and structure of mixed-ploidy level populations. *Molecular Ecology*
835   *Resources* 13: 946–952.

836   Peterman WE. 2018. ResistanceGA: An R package for the optimization of resistance surfaces
837   using genetic algorithms. *Methods in Ecology and Evolution* 9: 1638–1647.

838   Poelstra J, Salmona J, Tiley GP, Schüßler D, Blanco MB, Andriambeloson JB, Bouchez O,
839   Campbell CR, Etter PD, Hohenlohe PA, *et al.* 2020. Cryptic Patterns of Speciation in Cryptic
840   Primates: Microendemic Mouse Lemurs and the Multispecies Coalescent. *Systematic Biology*:
841   10.1093/sysbio/syaa053.

842   Pressey RL, Cabeza M, Watts ME, Cowling RM, Wilson KA. 2007. Conservation planning in a
843   changing world. *Trends in Ecology & Evolution* 22: 583–592.

844   Prevosti A, Ocana J, Alonso G. 1975. Distances between populations of *Drosophila subobscura*,
845   based on chromosome arrangement frequencies. *Theoretical and Applied Genetics* 45: 231–241.

846   Prunier JG, Colyn M, Legendre X, Nimon KF, Flamand M-C. 2015. Multicollinearity in spatial
847   genetics: separating the wheat from the chaff using commonality analyses. *Molecular Ecology*
848   24: 263–283.

849   Prunier JG, Colyn M, Legendre X, Nimon KF, Flamand M-C. 2017. Multicollinearity in spatial
850   genetics: Separating the wheat from the chaff using commonality analyses (vol 24, pg 263,
851   2015). *Molecular Ecology* 26: 6183–6183.

852   Quéméré E, Amelot X, Pierson J, Crouau-Roy B, Chikhi L. 2012. Genetic data suggest a natural
853   prehuman origin of open habitats in northern Madagascar and question the deforestation
854   narrative in this region. *Proceedings of the National Academy of Sciences of the United States of*
855   *America* 109: 13028–13033.

856   Quéméré E, Crouau-Roy B, Rabarivola C, Louis EE, Chikhi L. 2010. Landscape genetics of an
857   endangered lemur (*Propithecus tattersalli*) within its entire fragmented range. *Molecular*
858   *Ecology* 19: 1606–1621.

859   Rakotoarisoa J-E, Raheriarisena M, Goodman SM. 2013a. A phylogeographic study of the
860   endemic rodent *Eliurus carletoni* (Rodentia: Nesomyinae) in an ecological transition zone of
861   northern Madagascar. *Journal of Heredity* 104: 23–35.

862   Rakotoarisoa J-E, Raheriarisena M, Goodman SM. 2013b. Late Quaternary climatic vegetational
863   shifts in an ecological transition zone of northern Madagascar: insights from genetic analyses of
864   two endemic rodent species. *Journal of Evolutionary Biology* 26: 1019–1034.

865   Rakotondravony HA. 2006. Patterns de la diversité des reptiles et amphibiens de la région de
866   Loky-Manambato. *Inventaires de la faune et de la flore du nord de Madagascar dans la région*
867   *Loky-Manambato, Analamerana et Andavakoera. série Sciences biologiques* 23: 101–148.

868 Rakotondravony HA. 2009. Aspects de la conservation des reptiles et des amphibiens dans la
869 région de Daraina. *Madagascar Conservation & Development* 1: 15–18.

870 Raxworthy CJ, Nussbaum RA. 1995. Systematics, speciation and biogeography of the dwarf
871 chameleons (*Brookesia*; Reptilia, Squamata, Chamaeleontidae) of northern Madagascar. *Journal*
872 *of Zoology* 235: 525–558.

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

875 Rochette NC, Rivera-Colón AG, Catchen JM. 2019. Stacks 2: Analytical methods for paired-end
876 sequencing improve RADseq-based population genomics. *Molecular Ecology* 28: 4737–4754.

877 Salmona J, Heller R, Lascoux M, Shafer A. 2017a. Inferring demographic history using genomic
878 data. In: Rajora O, ed. *Population Genomics*. Cham: Springer, 511–537.

879 Salmona J, Heller R, Quéméré E, Chikhi L. 2017b. Climate change and human colonization
880 triggered habitat loss and fragmentation in Madagascar. *Molecular Ecology* 26: 5203–5222.

881 Salmona J, Olofsson JK, Hong-Wa C, Razanatsoa J, Rakotonasolo F, Ralimanana H,
882 Randriamboavonjy T, Suescun U, Vorontsova MS, Besnard G. 2020. Late Miocene origin and
883 recent population collapse of the Malagasy savanna olive tree (*Noronhia lowryi*). *Biological*
884 *Journal of the Linnean Society* 129: 227–243.

885 Schmieder R, Edwards R. 2011. Fast identification and removal of sequence contamination from
886 genomic and metagenomic datasets. *PLoS One* 6: e17288.

887 Sgarlata GM, Salmona J, Aleixo-Pais I, Rakotonanahary A, Sousa AP, Kun-Rodrigues C,
888 Ralantoharijaona T, Jan F, Zaranaina R, Rasolondraibe E, *et al.* 2018. Genetic differentiation and
889 demographic history of the northern rufous mouse Lemur (*Microcebus tavaratra*) across a
890 fragmented landscape in northern Madagascar. *International Journal of Primatology* 39: 65–89.

891 Sgarlata GM, Salmona J, Le Pors B, Rasolondraibe E, Jan F, Ralantoharijaona T,
892 Rakotonanahary A, Randriamaroson J, Marques AJ, Aleixo-Pais I, *et al.* 2019. Genetic and
893 morphological diversity of mouse lemurs (*Microcebus* spp.) in northern Madagascar: The
894 discovery of a putative new species? *American Journal of Primatology* 81: e23070.

895 Skotte L, Korneliussen TS, Albrechtsen A. 2013. Estimating individual admixture proportions
896 from next generation sequencing data. *Genetics* 195: 693–702.

897 Slatkin M. 1993. Isolation by distance in equilibrium and non-equilibrium populations. *Evolution*
898 47: 264–279.

899 Smouse PE, Long JC, Sokal RR. 1986. Multiple regression and correlation extensions of the
900 Mantel test of matrix correspondence. *Systematic Zoology* 35: 627–632.

901 Solofondranohatra CL, Vorontsova MS, Hackel J, Besnard G, Cable S, Williams J, Jeannoda V,
902 Lehmann CE. 2018. Grass functional traits differentiate forest and savanna in the Madagascar
903 central highlands. *Frontiers in Ecology and Evolution* 6: 184.

904 Štorchová H, Olson MS. 2004. Comparison between mitochondrial and chloroplast DNA
905 variation in the native range of *Silene vulgaris*. *Molecular Ecology* 13: 2909–2919.

906 Sutherland WJ, Pullin AS, Dolman PM, Knight TM. 2004. The need for evidence-based
907 conservation. *Trends in Ecology & Evolution* 19: 305–308.

908 Tang Q, Fung T, Rheindt FE. 2020. ResDisMapper: An r package for fine-scale mapping of
909 resistance to dispersal. *Molecular Ecology Resources* 20: 819–831.

910 Van de Paer C. 2017. Structural diversity and contrasted evolution of cytoplasmic genomes in
911 flowering plants: a phylogenomic approach in Oleaceae. PhD thesis, University of Toulouse III-
912 Paul Sabatier.

913 Van Etten J. 2012. R package gdistance: distances and routes on geographical grids (version 1.1-
914 4). *Journal of Statistical Software* 76: 13.

915 Van Strien MJ, Holderegger R, Van Heck HJ. 2015. Isolation-by-distance in landscapes:
916 considerations for landscape genetics. *Heredity* 114: 27–37.

917 Vences M. 2005. Madagascar as a model region for the study of tempo and pattern in adaptive
918 radiations. In: Huber BA, Sinclair BJ, Lampe KH, eds. *Molecules, Organisms, Ecosystems*.
919 *African Biodiversity*. Boston, MA: Springer, 69–84.

920 Vences M, Wollenberg KC, Vieites DR, Lees DC. 2009. Madagascar as a model region of
921 species diversification. *Trends in Ecology & Evolution* 24: 456–465.

922 Vieilledent G, Grinand C, Rakotomalala FA, Ranaivosoa R, Rakotoarijaona J-R, Allnutt TF,
923 Achard F. 2018. Combining global tree cover loss data with historical national forest cover maps
924 to look at six decades of deforestation and forest fragmentation in Madagascar. *Biological
925 Conservation* 222: 189–197.

926 Vieira FG, Fumagalli M, Albrechtsen A, Nielsen R. 2013. Estimating inbreeding coefficients
927 from NGS data: impact on genotype calling and allele frequency estimation. *Genome Research*
928 23: 1852–1861.

929 Vorontsova MS, Besnard G, Forest F, Malakasi P, Moat J, Clayton WD, Ficinski P, Savva GM,
930 Nanjarisoa OP, Razanatsoa J. 2016. Madagascar's grasses and grasslands: anthropogenic or
931 natural? *Proceedings of the Royal Society, London, Series B* 283: 20152262.

932 Wang Y, Lu J, Yu J, Gibbs RA, Yu F. 2013. An integrative variant analysis pipeline for accurate
933 genotype/haplotype inference in population NGS data. *Genome research* 23: 833–842.

934 Warmuth VM, Ellegren H. 2019. Genotype-free estimation of allele frequencies reduces bias and
935 improves demographic inference from RADSeq data. *Molecular Ecology Resources* 19: 586–
936 596.

937 Wilmé L, Goodman SM, Ganzhorn JU. 2006. Biogeographic evolution of Madagascar's
938 microendemic biota. *Science* 312: 1063–1065.

939 Wright S. 1943. Isolation by distance. *Genetics* 28: 114–138.

940 Yang J, Benyamin B, McEvoy BP, Gordon S, Henders AK, Nyholt DR, Madden PA, Heath AC,
941 Martin NG, Montgomery GW. 2010. Common SNPs explain a large proportion of the
942 heritability for human height. *Nature Genetics* 42: 565.

943 Yoder AD, Campbell CR, Blanco MB, Dos Reis M, Ganzhorn JU, Goodman SM, Hunnicutt KE,
944 Larsen PA, Kappeler PM, Rasoloarison RM. 2016. Geogenetic patterns in mouse lemurs (genus
945 *Microcebus*) reveal the ghosts of Madagascar's forests past. *Proceedings of the National*
946 *Academy of Sciences of the United States of America* 113: 8049–8056.

947