Original Article

Genomic islands of divergence between *Drosophila yakuba* subspecies predominantly overlap with chromosomal inversions

Erina A. Ferreira[1,2], Cathy C. Moore[3], David Ogereau[1], Arnaud Suwalski[2], Stéphane R. Prigent[2], Rebekah L. Rogers[3], Amir Yassin[1,2,*]

[1]Laboratoire Évolution, Génomes, Comportement et Écologie, CNRS, IRD, Université Paris-Saclay – Institut Diversité, Ecologie et Evolution du Vivant (IDEEV), Gif-sur-Yvette, France. [2]Institut Systématique, Evolution, Biodiversité (ISYEB), CNRS, MNHN, Sorbonne Université, EPHE, Université des Antilles, 57 rue Cuvier, CP 50, 75005 Paris, France. [3]Department of Bioinformatics and Genomics, University of North Carolina, Charlotte NC, USA.

*** Corresponding author:** amir.yassin@universite-paris-saclay.fr

Short title: Inversions in noni-specialist *Drosophila*

Authors contributions

E.A.F. sequenced and assembled new genome, conducted population genomics and structural variation analyses, wrote the primary draft of the paper. C.C.M. sequenced individual lines. D.O. sequenced and assembled new genome. A.S. collected flies. S.R.P. taxonomically sorted flies. R.L.R. sequenced individual lines, obtained funds. A.Y. conceptualized the work, collected flies, taxonomically sorted flies, supervised population genomics and structural variation analyses, obtained funds, wrote the final version of the paper with inputs from all authors.

Acknowledgments

The authors are grateful to Claire Golléty for help identifying noni plantation locations in Mayotte. Our late colleague, Jean R. David (1931-2021), who initially discovered *D. y. mayottensis* in 2013, participated to the 2017 expedition. This work was funded by the Agence Nationale de la Recherche grant no. ANR-18-CE2-0008, a Richard Lounsbery Foundation grant and an ATM bursary from the Muséum National d'Histoire Naturelle to AY, and a National Institutes of Health grant NIH NIGMS R35 GM133376 to RLR.

Conflict of interest

The authors declare no conflict of interest.

Data availability statement

Scripts used in this study are publicly available on GitHub: https://github.com/AmirYassinLab/Population_genomics_scripts. Genome sequences generated for this study are available on NCBI BioProject: PRJNA972991, PRJNA973002 and PRJNA973271.

Abstract

During the early stages of local adaptation and speciation, genetic differences tend to accumulate at certain regions of the genome leading to the formation of genomic islands of divergence (GIDs). This pattern may be due to selection and/or difference in the rate of recombination. Here, we investigate the possible causes of GIDs in *Drosophila yakuba mayottensis*, and reconfirm using field collection its association with toxic noni (*Morinda citrifolia*) fruits on the Mayotte island. Population genomics revealed lack of genetic structure on the island and identified 20 GIDs distinguishing *D. y. mayottensis* from generalist mainland populations of *D. y. yakuba*. The GIDs were enriched with gene families involved in the metabolism of lipids, sugars, peptides and xenobiotics, suggesting a role in host shift. We assembled a new genome for *D. y. mayottensis* and identified five novel chromosomal inversions. Twelve GIDs (~72% of outlier windows) fell close to or within subspecies-specific inversions. However, three GIDs were in collinear, high recombining regions indicating strong signal of hard selective sweeps. Unlike *D. y. mayottensis*, *D. sechellia*, the only other noni-specialist, is homosequential with its generalist relatives. Thus, both selection and rearrangements shape GIDs and striking convergences can occur between species with distinct genomic architectures.

Keywords: ecological speciation; structural variation; population genomics; genetic parallelism; insect-plant interactions.

Introduction

The genomic landscape of inter-population differentiation is often heterogeneous, with certain genomic regions, known as genomic islands of divergence (GIDs), emerging above the average differentiation level (Feder and Nosil 2010; Wolf and Ellegren 2017). GIDs were originally hailed as a result of divergent selection acting on genes that underlie adaptation to distinct environment and/or reproductive isolation (Turner et al. 2005). However, a plethora of mechanisms may also lead to the formation of such islands even in the absence of divergent selection. Most importantly, variation in recombination rate, due to heterochromatin content or structural variation, strongly affects the rate of admixture and the accumulation of genetic differences in natural populations (Feder and Nosil 2010; Cruickshank and Hahn 2014; Quilodrán et al. 2020). Chromosomal inversions are among the most common forms of structural variation, and a number of studies has unraveled an overlap between inversions and GIDs (Feder and Nosil 2009; Michel et al. 2010; Ellegren et al. 2012; Sodeland et al. 2016; Huang et al. 2020). Inversions often suppress recombination, reducing the rate at which slightly-deleterious mutations are purged from the population (Berdan et al. 2021; Jay et al. 2021, 2022). Conversely, the reduction of recombination could conserve one or more favorable allelic combinations for survival in contrasting environments, hence becoming targets of balancing or diverging selection depending on the heterogeneity of the environment (Faria et al. 2019; Kapun and Flatt 2019; Mérot et al. 2021; Hager et al. 2022; Harringmeyer and Hoekstra 2022; Sanchez-Donoso et al. 2022; Tepolt et al. 2022). Untangling the complex interactions of selection and inversions in driving GIDs and ultimately speciation therefore requires proper understanding of the genomic architecture of the study model and its natural variation.

Host shifts or changes in the width of the ecological niche, e.g., specialization, are a major source of diversification in herbivorous insects (Wiens et al. 2015; Forbes et al. 2017). The complexity of this phenomenon, which should involve the simultaneous evolution of genes

underlying multiple behavioral and physiological phenotypes with an ultimate impact on reproductive isolation, has long been a major area of research (Futuyma and Moreno 1988; Vertacnik and Linnen 2017) . One of the best well-characterized systems at the genomic level is the specialization of the fly *Drosophila sechellia* on toxic fruits of noni (*Morinda citrifolia*) on the Seychelles islands in the Indian Ocean (Louis and David 1986; Auer et al. 2021). *Drosophila sechellia* is a member of the *simulans* species complex, which comprises two additional species, namely the cosmopolitan *Drosophila simulans* and *Drosophila mauritiana*, an endemic to Mauritius. Unlike *D. sechellia*, both species are generalists and are repulsed by the odor of and intolerant to noni main toxins, octanoic and hexanoic acids (R'Kha et al. 1991; Amlou et al. 1997, 1998; Drum et al. 2022). The three species represent a sister clade to *Drosophila melanogaster*. This facilitated the extension of the unique genetic and genomic tools from this paradigmatic species to the study of *D. sechellia* specialization (Matsuo et al. 2007; Lavista-Llanos et al. 2014; Prieto-Godino et al. 2017; Auer et al. 2020). Besides, the three species show partial reproductive isolation and are karyologically homosequential, making them ideal for quantitative trait loci (QTLs) mapping through interspecific crosses (Earley and Jones 2011; Hungate et al. 2013; Huang and Erezyilmaz 2015; Chakraborty et al. 2021). Despite these numerous advantages, the divergence of *D. sechellia* from its generalist sister species *D. simulans* is estimated at around 89.000-242,000 years ago (Garrigan et al. 2012; Schrider et al. 2018); its entire genome has accumulated so many differences that traces of possible GIDs during the early stages of divergence may be difficult to observe. Besides, a particularly low level of genetic diversity has been detected in this species (Legrand et al. 2009; Schrider et al. 2018), further complicating population genomics analyses.

In January 2013, a second case of *Drosophila* species associated with toxic noni was discovered in Mayotte, another island in the Indian Ocean. This was a population of *Drosophila yakuba*, which had higher preference for and survival on noni unlike its continental conspecifics

(Yassin et al. 2016). The noni-specialist population likely diverged from generalist continental populations nearly 28,000 years ago and specialization was accompanied by precopulatory isolation between Mayotte and continental flies, earning them subspecific status as *D. y. mayottensis* and *D. y. yakuba*, respectively (Yassin et al. 2016; Yassin 2017). Genomic comparisons between this single population of *D. y. mayottensis* and two *D. y. yakuba* continental populations identified multiple GIDs, widespread across the genome, consistent with theoretical expectations (Yassin et al. 2016). Many *D. y. mayottensis* GIDs overlap with large noni survival QTLs in *D. sechellia* (Yassin et al. 2016). Despite those similarities, *D. yakuba* is a species where, unlike the *simulans* species complex, multiple polymorphic chromosomal inversions have been identified in continental populations (Lemeunier and Ashburner 1976). Several GIDs in *D. y. mayottensis* may therefore be the result of such inversions.

In this paper, we report the population genomics analysis of three additional populations of *D. y. mayottensis* all associated with noni in Mayotte. We identified 20 GIDs distinguishing the three populations from continental *D. y. yakuba*. We assembled a genome for *D. y. mayottensis* and identified multiple inversions that predominantly overlap with the GIDs, hence allowing the discrimination between GIDs presenting genuine signal of hard selective sweep from those which may be mere results of structural variation.

Materials and methods

Fly collection and establishment of isofemale lines from Mayotte island

Flies were prospected in April 2017 from 14 localities on the Grande Terre and Petite Terre islands of Mayotte. We particularly searched fruiting noni trees but also collected drosophilids in places with no noni trees. Flies were collected by net sweeping and direct aspiration over fallen fruits or using standard baits with fermented banana hung from tree

branches. Flies were also bred from ripe fruits brought back to the laboratory. *D. y. mayottensis* isofemale lines were established by placing one female and one male in a small vial with instant Carolina medium and the F_1 progeny was transferred to an axenic medium in the laboratory. Fly species that were or could not be bred in the laboratory were killed in absolute ethanol and taxonomically sorted in the laboratory. Data on *D. yakuba* seasonal abundance and attraction to banana baits in West Africa was obtained from Vouidibio (1985).

Pooled and individual lines genome sequencing

Drosophila yakuba mayottensis were sequenced both as pooled (Pool-Seq) as well as individual lines (Ind-Seq). For Pool-Seq analyses, we pooled single F_1 female per line in three pools according to the populations, namely BE, MTS and SOU. Pools were sent in ethanol to End2end Genomics LLC, Davis, California where DNA was extracted, PE library prepared using NEBNext Protocol and 75x coverage double-paired Illumina NovoSeq sequencing was conducted. For Ind-Seq analyses, DNA was extracted from single flies flash frozen in liquid nitrogen following QIAamp Mini Kit (Qiagen) protocol without using RNase A. The resulting DNA samples were quantified (Qubit dsDNA HS assay kit, ThermoFisher Scientific), assessed for quality (Nanodrop ND-2000), and stored at -20°C. Illumina TruSeq Nano DNA libraries were prepared manually following the manufacturer's protocol (TruSeq Nano DNA, RevD; Illumina). Briefly, samples were normalized to 100ng DNA and sheared by sonication with Covaris M220 (microTUBE 50; Sage Science). The samples were end repaired, purified with Ampure XP beads (Agencourt; Beckman Coulter), adaptors adenylated, and Unique Dual Indices ligated. Adaptor enrichment was performed using eight cycles of PCR. Following Ampure XP bead cleanup, fragment sizes for all libraries were measured using Agilent Bioanalyzer 2100 (HS DNA Assay; Applied Biosystems). The libraries were diluted 1:10 000 and 1:20 000 and quantified in triplicate using the KAPA Library Quantification Kit (Kapa

Biosystems). Equimolar samples were pooled, and the libraries were size selected targeting 400-700bp range to remove adaptor monomers and dimers using Pippen Prep DNA Size Selection system (1.5% Agarose Gel Cassette #CDF1510; Sage Sciences). Library pools were run on an Illumina HiSeq 4000 platform using the 150bp paired end (PE) Cluster Kit.

Population genomics analyses and delimitation of the genomic islands of divergence (GIDs) in Drosophila yakuba

As in Yassin et al. (2016) and Ferreira et al. (2021), we used previously published genomic sequences from two continental *D. y. yakuba* populations, namely Cameroon (CY) and Kenya (NY) (Rogers et al. 2014), for comparisons with *D. y. mayottensis*. We also used previous Pool-Seq reads from a single *D. y. mayottensis* population collected from Soulou in 2013 for temporal comparisons within the subspecies (Yassin et al. 2016). For all comparisons, reads were mapped to the *D. yakuba* v.1.05 reference genome obtained from Flybase (https://flybase.org/, Thurmond et al. 2019) using Minimap2 software package (Li 2018). Minimap2-generated SAM files were converted to BAM format using samtools 1.9 software (Li et al. 2009). The BAM files were then cleaned and sorted using Picard v.2.0.1 (http://broadinstitute.github.io/picard/). Popoolation 2 software package v.1.20162 (Kofler et al. 2011) was used to generate a synchronized mpileup file, from which intra-population polymorphism (π) and between-populations genetic differentiation (F_{ST}) were estimated using customized perl scripts.

Delimitation of the genomic islands of divergence (GIDs) in D. y. mayottensis

We conducted Population Branch Excess (*PBE*) analysis (Yassin et al. 2016) to identify GIDs specifically distinguishing *D. y. mayottensis* from the two continental populations CY and NY. *PBE* is a standardized version of the Population Branch Statistic (*PBS*), which was

proposed to identify population-specific allelic changes from F_{ST} pairwise comparisons between three populations (Yi et al. 2010). The standardization measures how *PBS* at a locus deviates from the chromosome-wide median *PBS* divided by the median F_{ST} estimate from the two non-focal populations. Therefore, *PBE* is less sensitive than *PBS* to genome-wide variation in polymorphism levels. Due to the lack of genetic structure within *D. y. mayottensis* (see below), Pool-Seq reads of the three populations were pooled into a single Pool of Pools (PoP). *PBS* and *PBE* were made assuming an (CY,(NY,PoP)) tree. Genes with *PBE* values falling ≥ 97.5 quantiles were considered outliers, and the width of a genomic island of divergence was considered if two outlier windows did not fall more than 500-kb apart. Gene enrichment tests were conducted for each GID as well as using the whole list of genes from all GIDs using the online Database for Annotation, Visualization and Integrated Discovery (DAVID) (Sherman et al. 2022).

De novo assembly of D. y. mayottensis *genome and identification of chromosomal rearrangements*

A mass culture of *D. y. mayottensis* was used to assemble a reference genome for this subspecies. Genomic DNA was extracted using the Nucleobond AXG20 kit and buffer set IV from Macherey-Nagel (ref. 740544 and 740604, https://www.mn-net.com, Düren, Germany), and size selection was conducted using the SRE XS kit from Circulomics (https://www.circulomics.com/, Baltimore, Maryland, USA). Fragments less than 10 Kb were eliminated. The SQK-LSK109 kit from Oxford nanopore technology (Lu et al. 2016), https://nanoporetech.com/) was then used to prepare the samples for nanopore sequencing following manufacturer's protocol. The library was then loaded on a R9.4 flow cell for sequencing on MinION. Raw data were basecalled with Guppy v4.0.11.

Illumina short-read sequences have also been produced to reconstruct our genome, using a hybrid assembly strategy, combining those short accurate Illumina reads to our long Nanopore reads. Illumina sequencing was realized by Novogene Company Limited (https://en.novogene.com, Cambridge, UK) on our previously extracted DNA. DNA sequences, from Nanopore and Illumina technologies, were then assembled using MaSuRCA genome assembler software v3.4.0 (Zimin et al. 2013). We aligned our newly constructed *D. y. mayottensis* genome using minimap2 software package to the v.1.05 reference genome as well as to a newly assembled genome of the reference *D. yakuba* strain (David et al. 2022) to correct for early missassemblies in the v.1.05 reference genome (Miller et al. 2018). *D. y. mayottensis* scaffolds were then ordered and oriented according to the reference genome, and chromosomal rearrangements were visually detected from a dotplot constructed by R. We used the cytological position of *D. melanogaster* orthologs of *D. yakuba* genes to delimit *D. y. yakuba* rearrangements described by Lemeunier and Ashburner (1976).

Results

Drosophila yakuba mayottensis *is strictly associated with noni in Mayotte*

In April 2017, we investigated the association between *D. y. mayottensis* and noni across the two main islands of Mayotte, *i.e* Grande Terre and Petite Terre. Out of 14 collection sites, we identified noni trees in 9 locations, mostly near the coastline. However, the trees were not always at the same fruiting stage, with fallen ripe fruits found at only three locations, namely Soulou (the same locality for the 2013 sample, hereafter SOU), M'tsangamouji (MTS) in the humid northwest and Bambo Est (BE) in the arid southeast (Figure 1A). *D. y. mayottensis* flies were collected through net sweeping or direct aspiration over fallen, rotting noni fruits only at those three locations. No single *D. y. mayottensis* fly was collected using fermenting banana

239 traps at those sites or at any other site on the island, although the noni trees at Bambo Est were
240 present amid a banana plantation (Figure 1B). In West Africa, such traps often collect tens or
241 hundreds of *D. y. yakuba* (Rio et al. 1983; Vouidibio 1985; Llopart et al. 2005; Prigent et al.
242 2013; Turissini and Matute 2017) (Figure 1B; χ^2 $P < 0.001$ for comparison with data from
243 Vouidibio (1985)). The total number of collected *D. y. mayottensis* was low, ~200 flies
244 throughout the 10-days duration of the expedition, compared to ~600 flies that were collected
245 during the last 2 days of the January 2013 at a single location (David et al. 2014). This indicates
246 a strong seasonality in the abundance of the species, consistent with other observations on
247 *D. y. yakuba* on the continent (Vouidibio 1985; Prigent et al. 2013), and highlights the necessity
248 of year-round investigations. We also observed that flies were more abundant on noni fruits that
249 have been nibbled and thrown to the ground by the brown lemur (*Eulemur fulvus*) than on intact
250 fallen fruits (Figure 1C).

251

252 *Pooled and isofemale line sequencing of the three Mayotte populations reveal low level of*
253 *differentiation on the island*

254 We established 26, 22 and 29 isofemale lines from wild-caught flies from the three
255 collection sites, *i.e.,* BE, SOU and MTS, respectively. As in Yassin et al. (2016), we sequenced
256 a single F_1 fly per line in three pools according to the populations. The level of genetic
257 differentiation estimated in 10-kb windows in SOU between the 2013 and 2017 samples was
258 low ($F_{ST} = 0.0262 \pm 0.0001$; Figure 2A), as well as the level of differentiation between the three
259 2017 geographical samples ($F_{ST} = 0.0283 \pm 0.0001$; Figure 2B; Supplementary Table S1).
260 Because F_{ST} estimates from Pool-Seq data could be slightly biased (Hivert et al. 2018), we also
261 sequenced 7, 6 and 6 individual isofemale lines from BE, SOU and MTS, respectively. We
262 found a strong correspondence between F_{ST} estimates from both approaches, with Ind-Seq data
263 producing higher estimates probably due to their smaller sample size and higher ascertainment

bias (F_{ST} = 0.0352 ± 0.0001 between the 2013 and 2017 of SOU, and 0.0426 ± 0.0001 between the three Mayotte populations; Supplementary Table S1). Nonetheless, whether Pool- or Ind-Seq analyses were conducted, the genetic differentiation among *D. y. mayottensis* populations was significantly lower than that between the two *D. y. yakuba* populations from Cameroon (CY) and Kenya (NY) (F_{ST} = 0.0671±0.0002; Student's *t*-test $P < 2.2 \times 10^{-16}$ for both analyses).

Differentiation between the two D. yakuba *subspecies differed among chromosomal arms*

To quantify the level of genetic differentiation between Mayotte and mainland populations, we estimated F_{ST} in 10-kb windows across the genome. Because of the lack of genetic structure among the three populations of *D. y. mayottensis*, we pooled either the Ind-Seq or the Pool-Seq sequences of the three populations in a single population, referring to them as Pool of Individuals (PoI) and Pool of Pools (PoP), respectively (Supplementary Table S2). Results did not greatly differ between the two methods, and we will show here analyses on the PoP data due to its larger sampling size and consequently reduced ascertainment bias. Surprisingly, the average genome-wide differentiation between the subspecies differed whether we used CY and NY, being higher for the former case (0.0731 ± 0.0003 and 0.0638 ± 0.0001, respectively; Student's *t*-test $P < 2.2 \times 10^{-16}$). However, differentiation greatly depended on chromosomal arms (Figure 3A). Whereas for most chromosomal arms, differentiation between the two mainland populations CY and NY were significantly lower than between each from *D. y. mayottensis*, only for chromosomal arm 2R differentiation between CY and the island was the highest (Figure 3A). Differentiation among the three *D. y. mayottensis* populations did not show any chromosomal effect and for all chromosomal arms was significantly lower than differentiation between the two subspecies or within *D. y. yakuba* (Figure 3A; Supplementary Table S1).

289 *Twenty Genomic Islands of Divergence (GIDs) differentiate* D. y. mayottensis

290 The *PBE* analysis showed 20 GIDs falling above the 97.5% quantiles in *D. y.*
291 *mayottensis* (Figure 3B; Table 1). As in Yassin et al. (2016), we used the number of
292 polymorphic sites found in 10-kb windows to delimit low recombining regions that are close to
293 chromosomal arms extremities at the telomeres and centromeres (shaded areas in Figure 3B).
294 These regions strongly concorded with fine-scale recombination rate variation in *D. yakuba*
295 (Pettie et al. 2022). Half of the identified GIDs fell within these regions. Of the remaining 10
296 GIDs falling in high recombining regions, seven were on chromosome 2.

297 Eight GIDs were enriched for specific gene families (Table 1). Gene families that were
298 enriched for the whole list of genes from all GIDs included LITAF (LPS-induced TNF-
299 activating factor) endosome-associated membrane proteins, glucose-methanol-choline (GMC)
300 oxidoreductases, cytochrome P450, UDP-glycosyltransferases, and maltose-alpha-glucosidases
301 at $P < 0.05$, and serine-type endopeptidases and glutathione-S-transferases at $P < 0.10$
302 (Supplementary Table S3). All these gene families have major roles in the metabolism of
303 ecdysteroids, lipids, sugars, proteins and xenobiotics, as it would be expected in cases of major
304 dietary shifts.

305

306 *GIDs in* D. y. mayottensis *predominantly overlap with chromosomal inversions*

307 Our hybrid assembly of the *D. y. mayottensis* genome yielded 563 fragments with an
308 N50 of 6.9Mb, and a BUSCOv5 score of 99.5%. Comparing the assembly to the v.1.05
309 reference genome assembly v.1.05 of *D. yakuba* identified 9 large structural variations with the
310 characteristic V-shape of inversion breakpoints (Figure 4A). The first breakpoint was found on
311 the X chromosome and indicated a ~1 Mb-long inversion between X:8,648,348..9,652,060.
312 Lemeunier and Ashburner (1976) did not detect any inversion on chromosome X in *D. y.*
313 *yakuba*. Following their nomenclature, we call this new subspecies-specific inversion *Xto*, and

note that it overlaps with GID2. We also noted a collinear connection between X: 13,477,771..17,070,823 on a single scaffold (grey) that could have resulted from double inversions, but do direct evidence for breakpoints was observed at this region (Figure 4A).

The second breakpoint was found on chromosomal arm 2L and corresponded to a large, ~12 Mb-long inversion between 2L: 8423810..20352247. This large inversion encompassed the misassembled portion of this chromosomal arm in the reference genome between 8,586,415 and 11,101,068 (Figure 4A,B; cf. Supplementary Figure S1 for alignment with a correct assembly). Lemeunier and Ashburner (1976) found four polymorphic inversions in mainland *D. y. yakuba*, namely *2Lh*, *2Ll*, *2Lm* and *2Ln*, with *2Lh* being ancestral to *2Ll*, and *2Lm* and *2Ln* independently derived from *2Ll*. We found the large inversion in *D. y. mayottensis* to be independently derived from *2Ll* too, and called it *2Lo* (Figure 4B). Remarkably, the proximal end of this inversion falls within the large subcentromeric region of 2L. Compared to the two most common inversions of the mainland, *i.e. 2Lh* and *2Ll* (Lemeunier and Ashburner 1976), four GIDs (5 to 8) fall in an inverted order in *2Lo* and one was close to its proximal, subcentromeric breakpoint (GID9) (Table 1).

The third breakpoint fell in the subcentromeric region of chromosomal arm 2R at 2R:182,816..436,892 but did not overlap with a major GID. However, the fourth to the seventh breakpoints all fell within the high recombining region of 2R. Their cytogenetic positions corresponded to the breakpoints separating chromosomal inversions *2Rn* and *2Rj* (Lemeunier and Ashburner 1976). The reference genome has the *2Rn* order indicating that *2Rj*, the most common inversion in *D. yakuba* (Lemeunier and Ashburner 1976), is fixed in *D. y. mayottensis*. Both inversions independently derived from a hypothetical *2Ri* order that has never been observed in natural populations (Figure 4B; Lemeunier and Ashburner 1976). Two GIDs, 11 and 14, fell in the flanking region separating *2Rj* from *2Ri*, whereas GID12 and 13 overlapped with the breakpoints separating *2Rn* from *2Ri*. Remarkably, the large GID13 is inverted in

339 inversion *2Rk*, another common 2R inversion in *D. y. yakuba* that was not detected in *D. y.*
340 *mayottensis*.

341 The eighth and ninth breakpoints were found in the subcentromeric regions of
342 chromosomal arms 3L and 3R, respectively. Each corresponded to ~0.2 and 0.8 Mb-long
343 inversions and overlapped with GIDs 15 and 16, respectively. These small subcentromeric
344 inversions were not detected by Lemeunier and Ashburner (1976) and we denoted them *3Lk*
345 and *3Ri*, respectively.

346

347 **Discussion**

348 By combining field observation with population and structural genomics, we were able
349 to draw a clearer picture of the mechanisms underlying the evolution of genomic islands of
350 divergence (GIDs) in the noni-specialist subspecies *Drosophila yakuba mayottensis*. First, we
351 showed that *D. y. mayottensis* GIDs were subspecies-specific. There was no different pattern
352 of divergence whether F_{ST} was inferred by comparing mainland populations to the Soulou
353 population collected in January 2013 or April 2017, or to any other noni-associated *D. y.*
354 *mayottensis* population collected in 2017 across the island. Second, we found that the two
355 subspecies had distinct inversions. Of the 20 GIDs distinguishing *D. y. mayottensis* from two
356 mainland populations, 12 fell close to or within subspecies-specific chromosomal inversions,
357 corresponding to ~72% of the outlier windows in the *PBE* analysis.

358 Chromosomal inversions can play an important role in promoting local adaptation,
359 either through reducing gene flow at selected genomic regions or through bringing to a new
360 habitat an advantageous allelic combination segregating at the ancestral range (Faria et al. 2019;
361 Mérot et al. 2020; Schaal et al. 2022). The two scenarios could have contributed to *D. y.*
362 *mayottensis* specialization on noni in Mayotte. For example, GIDs falling in the subspecies-
363 specific inversions *Xtb* and *2Lo* include several ecologically-relevant candidates that are

involved in the metabolism of ecdysone, lipids and carbohydrates (GMC oxidoreductases (Iida et al. 2007; Glaser-Schmitt and Parsch 2018); maltases (Inomata et al. 2019)), ethanol tolerance (*Adh* (Siddiq and Thornton 2019)), and hexanoic acid sensing (*pdm3* (Arguello et al. 2021)). The *2Rj* inversion that is fixed in *D. y. mayottensis* is, on the other hand, widespread in *D. y. yakuba*, and it is ancestral to three 2R inversions that are specific to the mainland subspecies (Lemeunier and Ashburner 1976). GIDs near the breakpoints of this ancestral inversion contain important clusters of detoxification genes of cytochrome P450 oxidases and glutathione-S-transferases, which are involved in basal resistance to noni toxins in *D. sechellia* and its close relatives (Peyser et al. 2017; Lanno and Coolon 2019). Alternatively, both derived and shared inversions might have been fixed by drift, although demographic models do not suggest major bottlenecks during *D. y. mayottensis* colonization of Mayotte (Yassin et al. 2016). Testing the potential role of derived and shared inversions would require experimental isolation of each karyotype and comparison of the effect of each on different noni use traits, but the large breadth of these inversions would complicate fine-scale QTL mapping or genome-wide association studies (GWAS) in *D. yakuba*.

High levels of differentiation in collinear, high recombining regions constitute strong signals of hard selective sweeps. Only 3 GIDs (4, 19 and 20) satisfied these conditions. Most notably, GID20 consisted of syntenic members of the *Osiris* gene family. The role of those genes in conferring resistance against noni's main toxin, octanoic acid, has been demonstrated through genetic mapping (Hungate et al. 2013) and transcriptomic analyses in *D. sechellia* (Lanno et al. 2017), as well as functional gene silencing in *D. melanogaster* (Andrade López et al. 2017; Lanno et al. 2019). The two other GIDs also involve some genes with a potential role in resistance to plant chemical defenses in other insects, such as *acetylcholine esterase* (*Ace*) or serine-type endopeptidases (Hoang et al. 2015; Alyokhin and Chen 2017; Lanno and Coolon 2019). Those genes therefore are best candidates for future functional genetics analyses.

389 However, hard selective sweeps indicate a particular mode of selection wherein the selected
390 alleles were absent or had low frequency at the onset of selection. Alternatively, incomplete
391 selective sweeps wherein the selected alleles may have not reached fixation yet or soft selective
392 sweeps on alleles segregating at intermediate frequencies in the ancestral range could also be
393 important factors, although loci under these patterns of sweeps will hardly be detected by GID
394 analyses. A signal of soft selective sweep on the odorant receptor *Or22a*, a gene that controls
395 *D. sechellia* attraction to methyl hexanoate, the main noni volatile (Auer et al. 2020), was
396 recently demonstrated in *D. y. mayottensis* (Ferreira et al. 2021). However, the efficiency of
397 this approach will be limited to collinear, high recombining regions, which constitute a
398 relatively small proportion of the *D. yakuba* genome.

399 In many cases, shared polymorphic inversions had promoted parallel adaptations in
400 different geographical populations (Mérot et al. 2018; Lucek et al. 2019; Morales et al. 2019;
401 Matschiner et al. 2022). This situation cannot explain the recurrent specialization on toxic noni
402 fruits in distinct islands of the Indian Ocean between *D. yakuba* and *D. sechellia*. Despite their
403 striking phenotypic convergence (Yassin et al. 2016; Ferreira et al. 2021), the divergence
404 between the two species is deep and their genomic architectures are quite different. Nearly
405 35.3% of the genome in *D. yakuba* has low number of segregating sites extending from the
406 telomeres and centromeres (Yassin et al. 2016; Pettie et al. 2022). In *D. simulans* and *D.*
407 *mauritiana*, only 10.5% of the genome has low number of segregating sites (Brand et al. 2018).
408 The three species of the *simulans* clade are homosequential and chromosomal inversions were
409 very rarely reported in *D. simulans* (Chakraborty et al. 2021). As functional genetics analyses
410 will continue to be undergone in both species, the degree of genic parallelism underlying the
411 recurrent specialization on a toxic resource despite distinct genomic architectures will be
412 clarified.
413

References

Alyokhin, A., and Y. H. Chen. 2017. Adaptation to toxic hosts as a factor in the evolution of insecticide resistance. Current Opinion in Insect Science 21:33–38.

Amlou, M., B. Moreteau, and J. R. David. 1998. Genetic Analysis of Drosophila sechellia Specialization: Oviposition Behavior Toward the Major Aliphatic Acids of Its Host Plant. Behav Genet 28:455–464.

Amlou, M., E. Pla, B. Moreteau, and J. R. David. 1997. Genetic analysis by interspecific crosses of the tolerance of Drosophila sechellia to major aliphatic acids of its host plant. Genet Sel Evol 29:511.

Andrade López, J. M., S. M. Lanno, J. M. Auerbach, E. C. Moskowitz, L. A. Sligar, P. J. Wittkopp, and J. D. Coolon. 2017. Genetic basis of octanoic acid resistance in Drosophila sechellia: functional analysis of a fine-mapped region. Mol Ecol 26:1148–1160.

Arguello, J. R., L. Abuin, J. Armida, K. Mika, P. C. Chai, and R. Benton. 2021. Targeted molecular profiling of rare olfactory sensory neurons identifies fate, wiring, and functional determinants. eLife 10:e63036. eLife Sciences Publications, Ltd.

Auer, T. O., M. A. Khallaf, A. F. Silbering, G. Zappia, K. Ellis, R. Álvarez-Ocaña, J. R. Arguello, B. S. Hansson, G. S. X. E. Jefferis, S. J. C. Caron, M. Knaden, and R. Benton. 2020. Olfactory receptor and circuit evolution promote host specialization. Nature 579:402–408.

Auer, T. O., M. P. Shahandeh, and R. Benton. 2021. Drosophila sechellia: A Genetic Model for Behavioral Evolution and Neuroecology. Annual Review of Genetics 55:527–554.

Berdan, E. L., A. Blanckaert, R. K. Butlin, and C. Bank. 2021. Deleterious mutation accumulation and the long-term fate of chromosomal inversions. PLOS Genetics 17:e1009411. Public Library of Science.

Brand, C. L., M. V. Cattani, S. B. Kingan, E. L. Landeen, and D. C. Presgraves. 2018. Molecular Evolution at a Meiosis Gene Mediates Species Differences in the Rate and Patterning of Recombination. Current Biology 28:1289-1295.e4.

Chakraborty, M., C.-H. Chang, D. E. Khost, J. Vedanayagam, J. R. Adrion, Y. Liao, K. L. Montooth, C. D. Meiklejohn, A. M. Larracuente, and J. J. Emerson. 2021. Evolution of genome structure in the Drosophila simulans species complex. Genome Res. 31:380–396.

Cruickshank, T. E., and M. W. Hahn. 2014. Reanalysis suggests that genomic islands of speciation are due to reduced diversity, not reduced gene flow. Molecular Ecology 23:3133–3157.

David, J. R., E. A. Ferreira, L. Jabaud, D. Ogereau, H. Bastide, and A. Yassin. 2022. Evolution

447 of assortative mating following selective introgression of pigmentation genes between two
448 Drosophila species. Ecology and Evolution 12:e8821.
449 David, J. R., A. Yassin, N. Gidaszewski, and V. Debat. 2014. Drosophilids (Diptera) from
450 Mayotte island: an annotated list of species collected in 2013 and comments on the colonisation
451 of Indian Ocean Islands. Annales de la Société entomologique de France (N.S.) 50:336–342.
452 Drum, Z., S. Lanno, S. M. Gregory, S. Shimshak, W. Barr, A. Gatesman, M. Schadt, J. Sanford,
453 A. Arkin, B. Assignon, S. Colorado, C. Dalgarno, T. Devanny, T. Ghandour, R. Griffin, M.
454 Hogan, E. Horowitz, E. McGhie, J. Multer, H. O'Halloran, K. Ofori-Darko, D. Pokushalov, N.
455 Richards, K. Sagarin, N. Taylor, A. Thielking, P. Towle, and J. Coolon. 2022. Genomics
456 analysis of Drosophila sechellia response to Morinda citrifolia fruit diet. G3
457 Genes|Genomes|Genetics 12:jkac153.
458 Earley, E. J., and C. D. Jones. 2011. Next-Generation Mapping of Complex Traits with
459 Phenotype-Based Selection and Introgression. Genetics 189:1203–1209.
460 Ellegren, H., L. Smeds, R. Burri, P. I. Olason, N. Backström, T. Kawakami, A. Künstner, H.
461 Mäkinen, K. Nadachowska-Brzyska, A. Qvarnström, S. Uebbing, and J. B. W. Wolf. 2012. The
462 genomic landscape of species divergence in Ficedula flycatchers. Nature 491:756–760. Nature
463 Publishing Group.
464 Faria, R., K. Johannesson, R. K. Butlin, and A. M. Westram. 2019. Evolving Inversions. Trends
465 in Ecology & Evolution 34:239–248.
466 Feder, J. L., and P. Nosil. 2009. Chromosomal inversions and species differences: when are
467 gene affecting adaptive divergence and reproductive isolation expected to reside within
468 inversions? Evolution 63:3061–3075.
469 Feder, J. L., and P. Nosil. 2010. The efficacy of divergence hitchhiking during ecological
470 speciation. Evolution 64:1729–1747.
471 Ferreira, E. A., S. Lambert, T. Verrier, F. Marion-Poll, and A. Yassin. 2021. Soft Selective
472 Sweep on Chemosensory Genes Correlates with Ancestral Preference for Toxic Noni in a
473 Specialist Drosophila Population. Genes 12:32. Multidisciplinary Digital Publishing Institute.
474 Forbes, A. A., S. N. Devine, A. C. Hippee, E. S. Tvedte, A. K. G. Ward, H. A. Widmayer, and
475 C. J. Wilson. 2017. Revisiting the particular role of host shifts in initiating insect speciation.
476 Evolution 71:1126–1137.
477 Futuyma, D. J., and G. Moreno. 1988. The Evolution of Ecological Specialization. Annual
478 Review of Ecology and Systematics 19:207–233.
479 Garrigan, D., S. B. Kingan, A. J. Geneva, P. Andolfatto, A. G. Clark, K. R. Thornton, and D.
480 C. Presgraves. 2012. Genome sequencing reveals complex speciation in the Drosophila

481 simulans clade. Genome Res. 22:1499–1511.
482 Glaser-Schmitt, A., and J. Parsch. 2018. Functional characterization of adaptive variation
483 within a cis-regulatory element influencing Drosophila melanogaster growth. PLOS Biology
484 16:e2004538. Public Library of Science.
485 Hager, E. R., O. S. Harringmeyer, T. B. Wooldridge, S. Theingi, J. T. Gable, S. McFadden, B.
486 Neugeboren, K. M. Turner, J. D. Jensen, and H. E. Hoekstra. 2022. A chromosomal inversion
487 contributes to divergence in multiple traits between deer mouse ecotypes. Science 377:399–
488 405. American Association for the Advancement of Science.
489 Harringmeyer, O. S., and H. E. Hoekstra. 2022. Chromosomal inversion polymorphisms shape
490 the genomic landscape of deer mice. Nat Ecol Evol 6:1965–1979. Nature Publishing Group.
491 Hivert, V., R. Leblois, E. J. Petit, M. Gautier, and R. Vitalis. 2018. Measuring Genetic
492 Differentiation from Pool-seq Data. Genetics 210:315–330.
493 Hoang, K., L. M. Matzkin, and J. M. Bono. 2015. Transcriptional variation associated with
494 cactus host plant adaptation in Drosophila mettleri populations. Molecular Ecology 24:5186–
495 5199.
496 Huang, K., R. L. Andrew, G. L. Owens, K. L. Ostevik, and L. H. Rieseberg. 2020. Multiple
497 chromosomal inversions contribute to adaptive divergence of a dune sunflower ecotype.
498 Molecular Ecology 29:2535–2549.
499 Huang, Y., and D. Erezyilmaz. 2015. The Genetics of Resistance to Morinda Fruit Toxin During
500 the Postembryonic Stages in Drosophila sechellia. G3: Genes, Genomes, Genetics 5:1973–
501 1981.
502 Hungate, E. A., E. J. Earley, I. A. Boussy, D. A. Turissini, C.-T. Ting, J. R. Moran, M.-L. Wu,
503 C.-I. Wu, and C. D. Jones. 2013. A Locus in Drosophila sechellia Affecting Tolerance of a Host
504 Plant Toxin. Genetics 195:1063–1075.
505 Iida, K., D. L. Cox-Foster, X. Yang, W.-Y. Ko, and D. R. Cavener. 2007. Expansion and
506 evolution of insect GMC oxidoreductases. BMC Evolutionary Biology 7:75.
507 Inomata, N., K. R. Takahasi, and N. Koga. 2019. Association between duplicated maltase genes
508 and the transcriptional regulation for the carbohydrate changes in Drosophila melanogaster.
509 Gene 686:141–145.
510 Jay, P., M. Chouteau, A. Whibley, H. Bastide, H. Parrinello, V. Llaurens, and M. Joron. 2021.
511 Mutation load at a mimicry supergene sheds new light on the evolution of inversion
512 polymorphisms. Nat Genet 53:288–293. Nature Publishing Group.
513 Jay, P., E. Tezenas, A. Véber, and T. Giraud. 2022. Sheltering of deleterious mutations explains
514 the stepwise extension of recombination suppression on sex chromosomes and other

515 supergenes. PLOS Biology 20:e3001698. Public Library of Science.
516 Kapun, M., and T. Flatt. 2019. The adaptive significance of chromosomal inversion
517 polymorphisms in Drosophila melanogaster. Molecular Ecology 28:1263–1282.
518 Kofler, R., R. V. Pandey, and C. Schlötterer. 2011. PoPoolation2: identifying differentiation
519 between populations using sequencing of pooled DNA samples (Pool-Seq). Bioinformatics
520 27:3435–3436.
521 Lanno, S. M., and J. D. Coolon. 2019. Derived esterase activity in Drosophila sechellia
522 contributes to evolved octanoic acid resistance. Insect Molecular Biology 28:798–806.
523 Lanno, S. M., S. M. Gregory, S. J. Shimshak, M. K. Alverson, K. Chiu, A. L. Feil, M. G.
524 Findley, T. E. Forman, J. T. Gordon, J. Ho, J. L. Krupp, I. Lam, J. Lane, S. C. Linde, A. E.
525 Morse, S. Rusk, R. Ryan, A. Saniee, R. B. Sheth, J. J. Siranosian, L. Sirichantaropart, S. R.
526 Sternlieb, C. M. Zaccardi, and J. D. Coolon. 2017. Transcriptomic Analysis of Octanoic Acid
527 Response in Drosophila sechellia Using RNA-Sequencing. G3: Genes, Genomes, Genetics
528 7:3867–3873.
529 Lanno, S. M., S. J. Shimshak, R. D. Peyser, S. C. Linde, and J. D. Coolon. 2019. Investigating
530 the role of Osiris genes in Drosophila sechellia larval resistance to a host plant toxin. Ecology
531 and Evolution 9:1922–1933.
532 Lavista-Llanos, S., A. Svatoš, M. Kai, T. Riemensperger, S. Birman, M. C. Stensmyr, and B.
533 S. Hansson. 2014. Dopamine drives Drosophila sechellia adaptation to its toxic host. eLife
534 3:e03785.
535 Legrand, D., M. I. Tenaillon, P. Matyot, J. Gerlach, D. Lachaise, and M.-L. Cariou. 2009.
536 Species-Wide Genetic Variation and Demographic History of Drosophila sechellia, a Species
537 Lacking Population Structure. Genetics 182:1197–1206.
538 Lemeunier, F., and M. Ashburner. 1976. Relationships within the melanogaster species
539 subgroup of the genus Drosophila (Sophophora) - II. Phylogenetic relationships between six
540 species based upon polytene chromosome banding sequences. Proceedings of the Royal Society
541 of London. Series B. Biological Sciences 193:275–294. Royal Society.
542 Li, H. 2018. Minimap2: pairwise alignment for nucleotide sequences. Bioinformatics 34:3094–
543 3100.
544 Li, H., B. Handsaker, A. Wysoker, T. Fennell, J. Ruan, N. Homer, G. Marth, G. Abecasis, and
545 R. Durbin. 2009. The Sequence Alignment/Map format and SAMtools. Bioinformatics
546 25:2078–2079.
547 Llopart, A., D. Lachaise, and J. A. Coyne. 2005. An Anomalous Hybrid Zone in Drosophila.
548 Evolution 59:2602–2607.

549 Louis, J., and J. R. David. 1986. Ecological specialization in the Drosophila melanogaster
550 species subgroup: a case study of D. sechellia. Acta oecologica. Oecologia generalis 7:215–
551 229.
552 Lu, H., F. Giordano, and Z. Ning. 2016. Oxford Nanopore MinION Sequencing and Genome
553 Assembly. Genomics, Proteomics & Bioinformatics 14:265–279.
554 Lucek, K., Z. Gompert, and P. Nosil. 2019. The role of structural genomic variants in population
555 differentiation and ecotype formation in Timema cristinae walking sticks. Molecular Ecology
556 28:1224–1237.
557 Matschiner, M., J. M. I. Barth, O. K. Tørresen, B. Star, H. T. Baalsrud, M. S. O. Brieuc, C.
558 Pampoulie, I. Bradbury, K. S. Jakobsen, and S. Jentoft. 2022. Supergene origin and
559 maintenance in Atlantic cod. Nat Ecol Evol 6:469–481. Nature Publishing Group.
560 Matsuo, T., S. Sugaya, J. Yasukawa, T. Aigaki, and Y. Fuyama. 2007. Odorant-Binding
561 Proteins OBP57d and OBP57e Affect Taste Perception and Host-Plant Preference in
562 Drosophila sechellia. PLOS Biology 5:e118.
563 Mérot, C., E. L. Berdan, C. Babin, E. Normandeau, M. Wellenreuther, and L. Bernatchez. 2018.
564 Intercontinental karyotype–environment parallelism supports a role for a chromosomal
565 inversion in local adaptation in a seaweed fly. Proceedings of the Royal Society B: Biological
566 Sciences 285:20180519. Royal Society.
567 Mérot, C., E. L. Berdan, H. Cayuela, H. Djambazian, A.-L. Ferchaud, M. Laporte, E.
568 Normandeau, J. Ragoussis, M. Wellenreuther, and L. Bernatchez. 2021. Locally Adaptive
569 Inversions Modulate Genetic Variation at Different Geographic Scales in a Seaweed Fly.
570 Molecular Biology and Evolution 38:3953–3971.
571 Mérot, C., R. A. Oomen, A. Tigano, and M. Wellenreuther. 2020. A Roadmap for
572 Understanding the Evolutionary Significance of Structural Genomic Variation. Trends in
573 Ecology & Evolution 35:561–572.
574 Michel, A. P., S. Sim, T. H. Q. Powell, M. S. Taylor, P. Nosil, and J. L. Feder. 2010. Widespread
575 genomic divergence during sympatric speciation. Proceedings of the National Academy of
576 Sciences 107:9724–9729. Proceedings of the National Academy of Sciences.
577 Miller, D. E., C. Staber, J. Zeitlinger, and R. S. Hawley. 2018. Highly Contiguous Genome
578 Assemblies of 15 Drosophila Species Generated Using Nanopore Sequencing. G3
579 Genes|Genomes|Genetics 8:3131–3141.
580 Morales, H. E., R. Faria, K. Johannesson, T. Larsson, M. Panova, A. M. Westram, and R. K.
581 Butlin. 2019. Genomic architecture of parallel ecological divergence: Beyond a single
582 environmental contrast. Science Advances 5:eaav9963. American Association for the

583 Advancement of Science.
584 Pettie, N., A. Llopart, and J. M. Comeron. 2022. Meiotic, genomic and evolutionary properties
585 of crossover distribution in Drosophila yakuba. PLOS Genetics 18:e1010087. Public Library
586 of Science.
587 Peyser, R. D., S. M. Lanno, S. J. Shimshak, and J. D. Coolon. 2017. Analysis of cytochrome
588 P450 contribution to evolved plant toxin resistance in Drosophila sechellia. Insect Molecular
589 Biology 26:715–720.
590 Prieto-Godino, L. L., R. Rytz, S. Cruchet, B. Bargeton, L. Abuin, A. F. Silbering, V. Ruta, M.
591 D. Peraro, and R. Benton. 2017. Evolution of Acid-Sensing Olfactory Circuits in Drosophilids.
592 Neuron 93:661-676.e6.
593 Prigent, S. R., P. Le Gall, S. W. Mbunda, and M. Veuille. 2013. Seasonal and altitudinal
594 structure of drosophilid communities on Mt Oku (Cameroon volcanic line). Comptes Rendus
595 Geoscience 345:316–326.
596 Quilodrán, C. S., K. Ruegg, A. T. Sendell-Price, E. C. Anderson, T. Coulson, and S. M. Clegg.
597 2020. The multiple population genetic and demographic routes to islands of genomic
598 divergence. Methods in Ecology and Evolution 11:6–21.
599 Rio, B., G. Couturier, F. Lemeunier, and D. Lachaise. 1983. Evolution d'une spécialisation
600 saisonnière chez Drosophila erecta (Dipt., Drosophilidae). Ann. Soc. entomol. Fr 19:235–248.
601 R'Kha, S., P. Capy, and J. R. David. 1991. Host-plant specialization in the Drosophila
602 melanogaster species complex: a physiological, behavioral, and genetical analysis. PNAS
603 88:1835–1839.
604 Rogers, R. L., J. M. Cridland, L. Shao, T. T. Hu, P. Andolfatto, and K. R. Thornton. 2014.
605 Landscape of Standing Variation for Tandem Duplications in Drosophila yakuba and
606 Drosophila simulans. Mol Biol Evol 31:1750–1766.
607 Sanchez-Donoso, I., S. Ravagni, J. D. Rodríguez-Teijeiro, M. J. Christmas, Y. Huang, A.
608 Maldonado-Linares, M. Puigcerver, I. Jiménez-Blasco, P. Andrade, D. Gonçalves, G. Friis, I.
609 Roig, M. T. Webster, J. A. Leonard, and C. Vilà. 2022. Massive genome inversion drives
610 coexistence of divergent morphs in common quails. Current Biology 32:462-469.e6.
611 Schaal, S. M., B. C. Haller, and K. E. Lotterhos. 2022. Inversion invasions: when the genetic
612 basis of local adaptation is concentrated within inversions in the face of gene flow.
613 Philosophical Transactions of the Royal Society B: Biological Sciences 377:20210200. Royal
614 Society.
615 Schrider, D. R., J. Ayroles, D. R. Matute, and A. D. Kern. 2018. Supervised machine learning
616 reveals introgressed loci in the genomes of Drosophila simulans and D. sechellia. PLOS

617 Genetics 14:e1007341. Public Library of Science.
618 Sherman, B. T., M. Hao, J. Qiu, X. Jiao, M. W. Baseler, H. C. Lane, T. Imamichi, and W.
619 Chang. 2022. DAVID: a web server for functional enrichment analysis and functional
620 annotation of gene lists (2021 update). Nucleic Acids Research 50:W216–W221.
621 Siddiq, M. A., and J. W. Thornton. 2019. Fitness effects but no temperature-mediated balancing
622 selection at the polymorphic Adh gene of Drosophila melanogaster. Proceedings of the National
623 Academy of Sciences 116:21634–21640. Proceedings of the National Academy of Sciences.
624 Sodeland, M., P. E. Jorde, S. Lien, S. Jentoft, P. R. Berg, H. Grove, M. P. Kent, M. Arnyasi, E.
625 M. Olsen, and H. Knutsen. 2016. "Islands of Divergence" in the Atlantic Cod Genome
626 Represent Polymorphic Chromosomal Rearrangements. Genome Biology and Evolution
627 8:1012–1022.
628 Tepolt, C. K., E. D. Grosholz, C. E. de Rivera, and G. M. Ruiz. 2022. Balanced polymorphism
629 fuels rapid selection in an invasive crab despite high gene flow and low genetic diversity.
630 Molecular Ecology 31:55–69.
631 Thurmond, J., J. L. Goodman, V. B. Strelets, H. Attrill, L. S. Gramates, S. J. Marygold, B. B.
632 Matthews, G. Millburn, G. Antonazzo, V. Trovisco, T. C. Kaufman, B. R. Calvi, N. Perrimon,
633 S. R. Gelbart, J. Agapite, K. Broll, L. Crosby, G. dos Santos, D. Emmert, L. S. Gramates, K.
634 Falls, V. Jenkins, B. Matthews, C. Sutherland, C. Tabone, P. Zhou, M. Zytkovicz, N. Brown,
635 G. Antonazzo, H. Attrill, P. Garapati, A. Holmes, A. Larkin, S. Marygold, G. Millburn, C.
636 Pilgrim, V. Trovisco, P. Urbano, T. Kaufman, B. Calvi, B. Czoch, J. Goodman, V. Strelets, J.
637 Thurmond, R. Cripps, and P. Baker. 2019. FlyBase 2.0: the next generation. Nucleic Acids Res
638 47:D759–D765.
639 Turissini, D. A., and D. R. Matute. 2017. Fine scale mapping of genomic introgressions within
640 the Drosophila yakuba clade. PLOS Genetics 13:e1006971.
641 Turner, T. L., M. W. Hahn, and S. V. Nuzhdin. 2005. Genomic Islands of Speciation in
642 Anopheles gambiae. PLOS Biology 3:e285.
643 Vertacnik, K. L., and C. R. Linnen. 2017. Evolutionary genetics of host shifts in herbivorous
644 insects: insights from the age of genomics. Ann. N.Y. Acad. Sci. 1389:186–212.
645 Vouidibio, J. 1985. Ecologie des populations et biologie évolutive des drosophiles en Afrique
646 équatoriale (Congo). Université Paris XII Val-de-Marne, France.
647 Wiens, J. J., R. T. Lapoint, and N. K. Whiteman. 2015. Herbivory increases diversification
648 across insect clades. Nature Communications 6:8370.
649 Wolf, J. B. W., and H. Ellegren. 2017. Making sense of genomic islands of differentiation in
650 light of speciation. Nat Rev Genet 18:87–100.

651 Yassin, A. 2017. Drosophila yakuba mayottensis, a new model for the study of incipient
652 ecological speciation. Fly 11:37–45.
653 Yassin, A., V. Debat, H. Bastide, N. Gidaszewski, J. R. David, and J. E. Pool. 2016. Recurrent
654 specialization on a toxic fruit in an island Drosophila population. PNAS 113:4771–4776.
655 Yi, X., Y. Liang, E. Huerta-Sanchez, X. Jin, Z. X. P. Cuo, J. E. Pool, X. Xu, H. Jiang, N.
656 Vinckenbosch, T. S. Korneliussen, H. Zheng, T. Liu, W. He, K. Li, R. Luo, X. Nie, H. Wu, M.
657 Zhao, H. Cao, J. Zou, Y. Shan, S. Li, Q. Yang, Asan, P. Ni, G. Tian, J. Xu, X. Liu, T. Jiang, R.
658 Wu, G. Zhou, M. Tang, J. Qin, T. Wang, S. Feng, G. Li, Huasang, J. Luosang, W. Wang, F.
659 Chen, Y. Wang, X. Zheng, Z. Li, Z. Bianba, G. Yang, X. Wang, S. Tang, G. Gao, Y. Chen, Z.
660 Luo, L. Gusang, Z. Cao, Q. Zhang, W. Ouyang, X. Ren, H. Liang, H. Zheng, Y. Huang, J. Li,
661 L. Bolund, K. Kristiansen, Y. Li, Y. Zhang, X. Zhang, R. Li, S. Li, H. Yang, R. Nielsen, J.
662 Wang, and J. Wang. 2010. Sequencing of 50 Human Exomes Reveals Adaptation to High
663 Altitude. Science 329:75–78.
664 Zimin, A. V., G. Marçais, D. Puiu, M. Roberts, S. L. Salzberg, and J. A. Yorke. 2013. The
665 MaSuRCA genome assembler. Bioinformatics 29:2669–2677.
666
667
668
669

Figure legends

Figure 1 – Association of *Drosophila yakuba mayottensis* with noni (*Morinda citrifolia*) on the island of Mayotte. A) Frequency of *D. yakuba* (blue) among drosophilids (red) collected on the island of Mayotte in January 2013 (after David et al. (2014)) and April 2017. Locations with fruiting noni trees are indicated with a star. The size of the pie charts reflects the total number of collected flies per site. B) Abundance of *D. yakuba* flies among drosophilids collected using standard fermenting banana traps in Congo (after Vouidibio (1985) and in Mayotte. C) A brown lemur (*Eulemur fulvus*) eating a noni fruit in Mayotte (© Amir Yassin).

Figure 2 – Low temporal and spatial genetic differentiation among *D. y. mayottensis* populations on the island of Mayotte. A) Pairwise F_{ST} values for 10-kb windows between Pool-Seq data from Soulou collected in January 2013 and April 2017. B) Average pairwise F_{ST} values for 10-kb windows between Pool-Seq data from the three populations collected in Mayotte in April 2017. Gray areas reflect subtelomeric and subcentromeric regions with low recombination rate (Yassin et al. 2016; Pettie et al. 2022). Each chromosomal arm according to the v.1.05 reference genome is represented by different color.

Figure 3 – High genetic differentiation among the two *D. yakuba* subspecies. A) Boxplot of pairwise F_{ST} values for 10-kb windows between the two mainland *D. y. yakuba* populations (pale red), between each and a pool-of-pooled flies from *D. y. mayottensis* (pale blue and orange), and between the three Pool-Seq populations of *D. y. mayottensis* (yellow). Significance estimated from Student's *t* pairwise-tests. B) Population Branch Excess (*PBE*) identification of

693 the genomic islands of divergence (GIDs), enumerated in red, distinguishing *D. y. mayottensis*
694 from the two *D. y. yakuba* populations. Color code as in Figure 2.

695

696 **Figure 4 –** Genomic islands of divergence (GIDs) predominantly overlap with major
697 chromosomal inversions. A) Dot plot comparing *D. y. mayottensis* hybrid assembly with the *D.*
698 *yakuba* reference genome v1.05. Each *D. y. mayottensis* scaffold is given in different color with
699 the position of the GIDs indicated by their number. Shaded areas indicate subtelomeric and
700 subcentromeric regions with low recombination rate. Red arrows indicate inversion
701 breakpoints. B) Schematic presentation of major inversions on chromosomal arms *2L* (left) and
702 *2R* (right). Inversion names, breakpoints and geographical distribution into West Africa (W),
703 East Africa (E) and Mayotte (M) are given following Lemeunier and Ashburner (1976) and this
704 study.

705

706

707

708

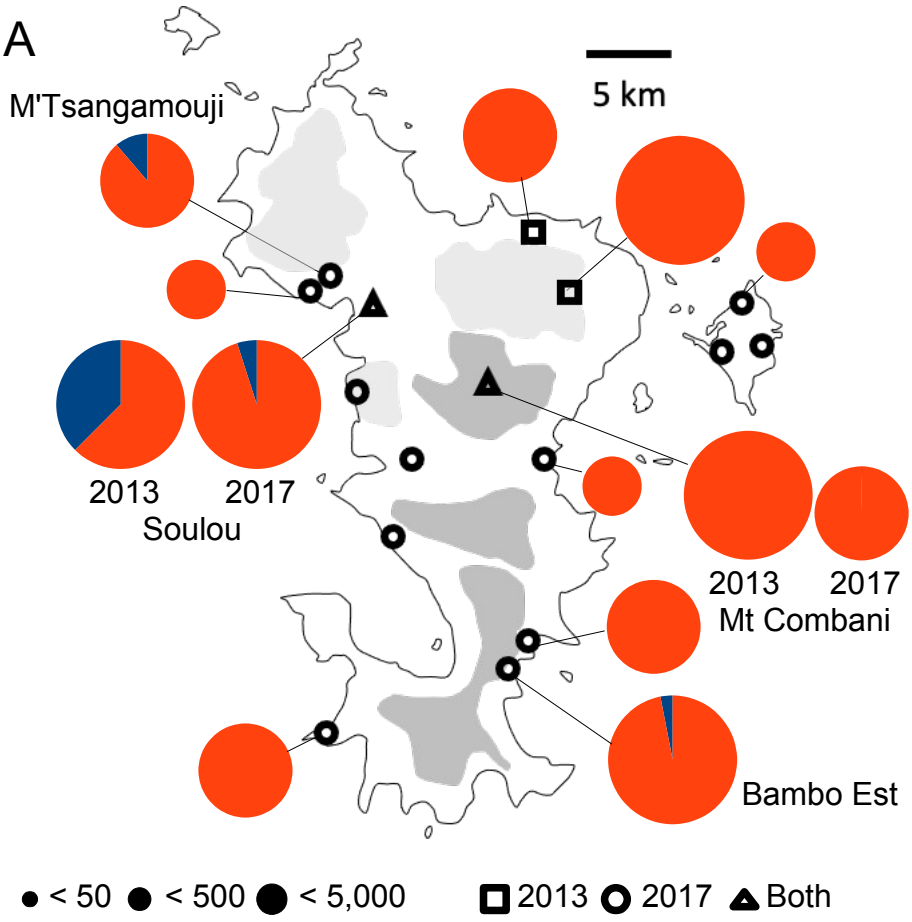
709 Table 1 – Position, structural features and candidate genes or families of the 20 genomic islands
710 of divergence (GIDs) distinguishing noni-specialist *Drosophila yakuba mayottensis* from
711 generalist mainland species. *P* levels from DAVID gene ontology enrichment analyses for each
712 GID.

GID	Coordinates (Mb)	Architecture	Candidate genes
1	X:3.63-3.98	Subtelomeric region	
2	X:8.68-9.23	Subspecies-specific *Xtb* inversion	*Flo2*; Glucose-methanol-choline oxidoreductase (ecdysteroid metabolic process, $P < 1 \times 10^{-20}$)
3	X:21.22-21.77	Subcentromeric region	
4	2L:2.21-2.27		Serine-type endopeptidase activity ($P < 3 \times 10^{-5}$)
5	2L:14.51-14.83	Subspecies-specific *2Lo* inversion	*Adh*
6	2L:16.59-17.15	Subspecies-specific *2Lo* inversion	*pdm3*; maltose alpha-glucosidase activity (sugar metabolism and transport, $P < 5 \times 10^{-13}$), cytochrome P450 ($P < 0.10$)
7	2L:18.24-18.83	Subspecies-specific *2Lo* inversion; subcentromeric region	
8	2L:20.37-20.23	Subspecies-specific *2Lo* inversion; subcentromeric region	*EcR*
9	2L:21.08-21.57	Close to subspecies-specific *2Lo* inversion; subcentromeric region	
10	2R:5.01-5.57	Subcentromeric region	*Catsup*; UDP-glycosyltransferases (metabolism of xenobiotics by cytochrome P450, $P < 0.01$)
11	2R:6.85-6.93	Close to species-common *2Rj* inversion	
12	2R:10.42-10.77	Species-common *2Rj* inversion; subspecies-specific *2Rk* inversion	Cytochrome P450 ($P < 1 \times 10^{-5}$)
13	2R:14.79-17.61	Species-common *2Rj* inversion; subspecies-specific *2Rk* inversion	LITAF ($P < 3 \times 10^{-14}$); glutathione S-transferase ($P < 0.05$); destabilase ($P < 0.05$); malic enzyme ($P < 0.10$)
14	2R:19.54-19.63	Close to species-common *2Rj* inversion	
15	3L:23.56-24.20	Subspecies-specific *3Lk* inversion; subcentromeric region	

16	3R:0.00-0.33	Subspecies-specific *3Ri* inversion; subcentromeric region	
17	3R:4.39-4.57	Subcentromeric region	
18	3R:10.82-11.92	Subcentromeric region	UDP-glucuronosyl/UDP-glucosyltransferase ($P < 5 \times 10^{-8}$)
19	3R:13.38-13.42		*Ace*, *Osi22*
20	3R:17.96-17.98		*Osi2*, *Osi3*, *Osi4*, *Osi5* (Proteins of unknown function DUF1676, $P < 1 \times 10^{-8}$)

713

A

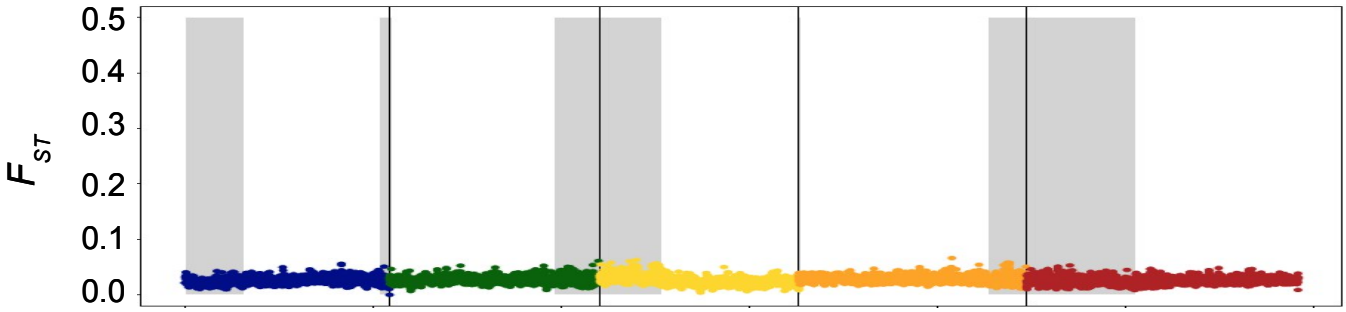


B

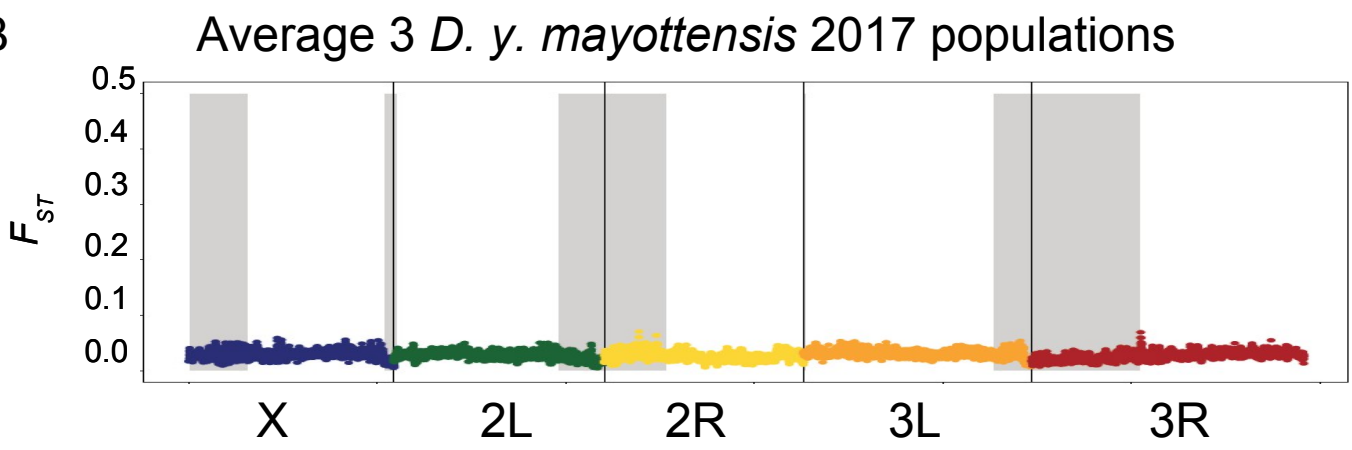
Banana bait	Not *yak*	*yak*
Mayotte	210	0
Congo	6531	383

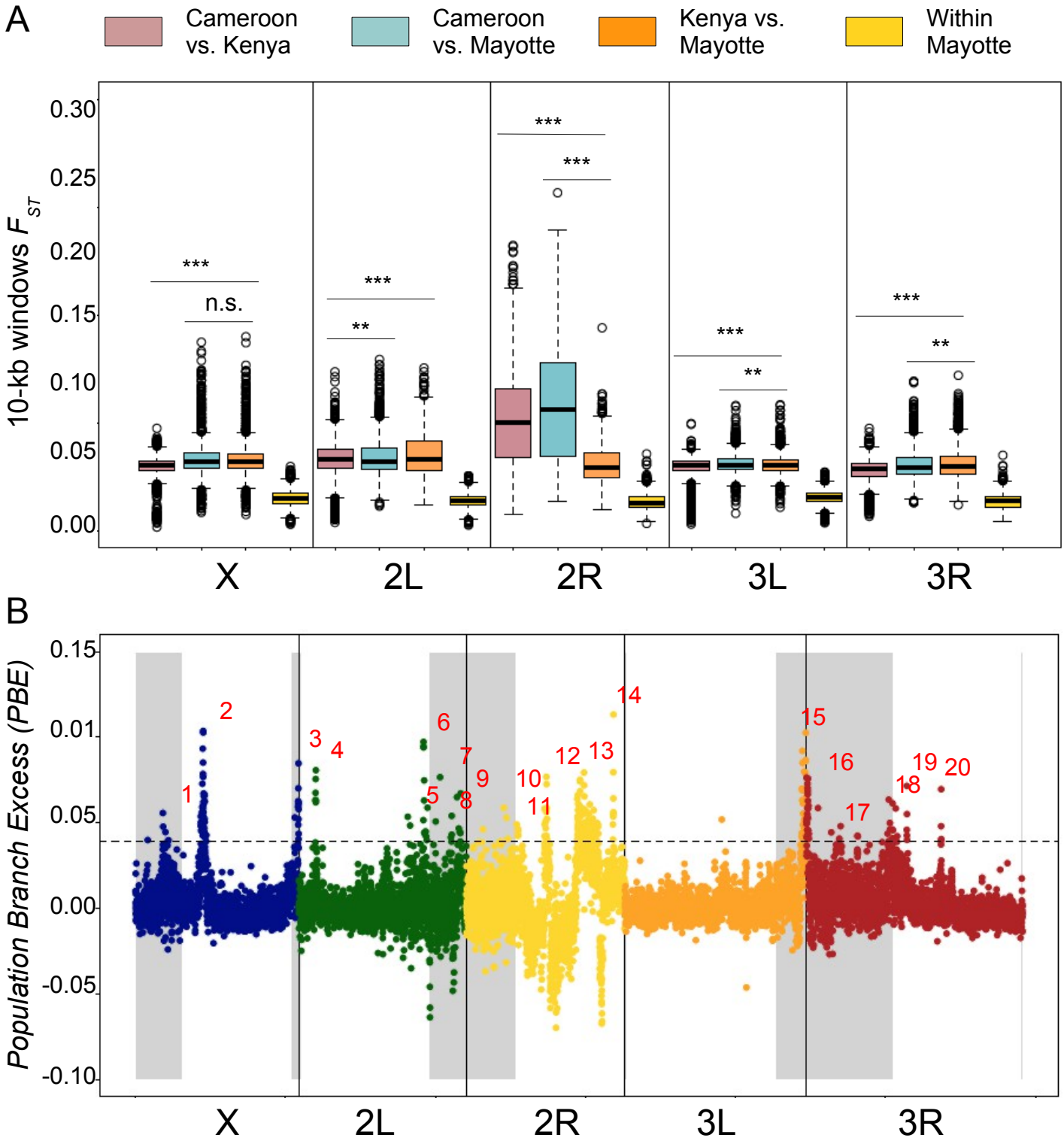
C

A

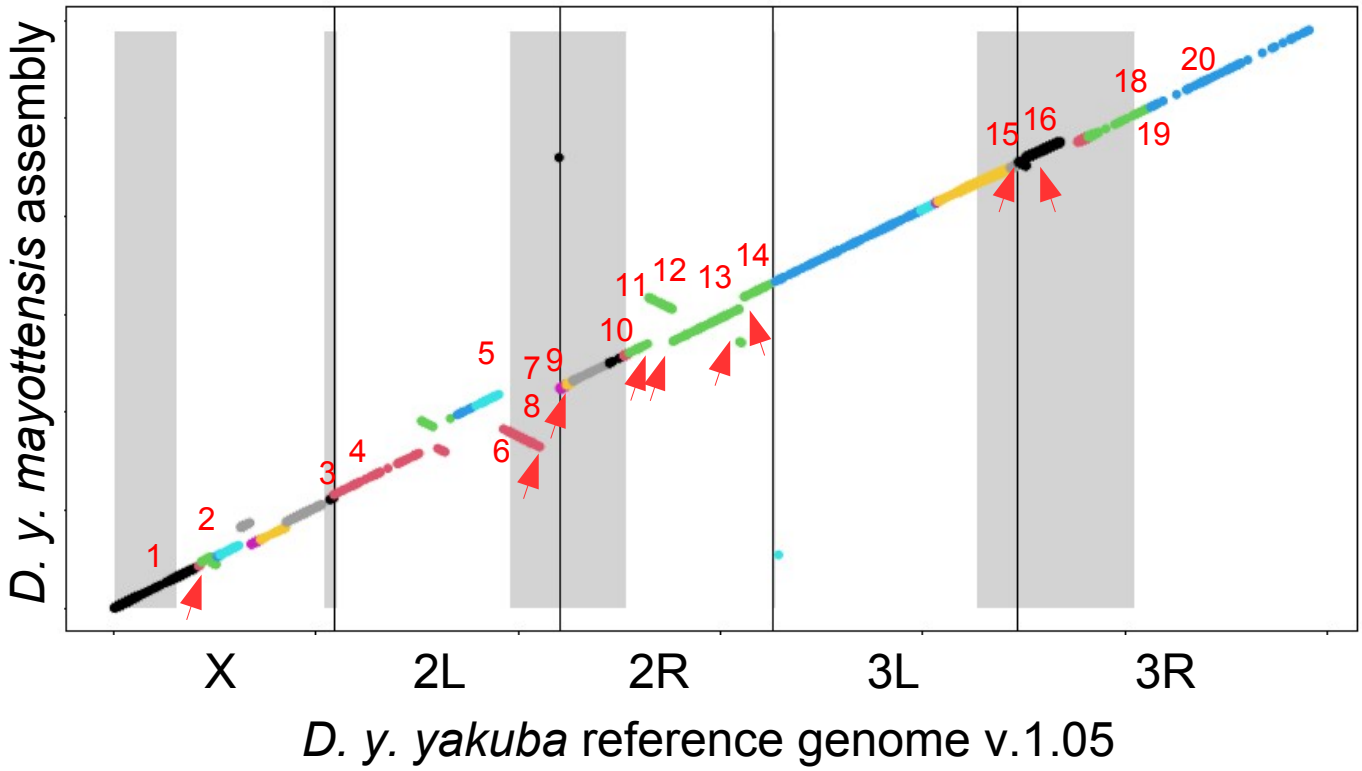


B





A



B