Published in final edited form as:
Mol Ecol. 2019 January ; 28(1): 66–85. doi:10.1111/mec.14957.

The population genomics of multiple tsetse fly (*Glossina fuscipes fuscipes*) admixture zones in Uganda

Norah P. Saarman[1], **Robert Opiro**[2], **Chaz Hyseni**[3], **Richard Echodu**[2], **Elizabeth A. Opiyo**[2], **Kirstin Dion**[1], **Thomas Johnson**[1], **Serap Aksoy**[4], **Adalgisa Caccone**[1]

[1]Department of Ecology and Evolutionary Biology, Yale University, New Haven, Connecticut

[2]Department of Biology, Faculty of Science, Gulu University, Uganda

[3]Department of Biology, University of Mississippi, Oxford, Mississippi

[4]Department of Epidemiology of Microbial Diseases, Yale School of Public Health, New Haven, Connecticut

Abstract

Understanding the mechanisms that enforce, maintain or reverse the process of speciation is an important challenge in evolutionary biology. This study investigates the patterns of divergence and discusses the processes that form and maintain divergent lineages of the tsetse fly *Glossina fuscipes fuscipes* in Uganda. We sampled 251 flies from 18 sites spanning known genetic lineages and the four admixture zones between them. We apply population genomics, hybrid zone and approximate Bayesian computation to the analysis of three types of genetic markers: 55,267 double-digest restriction site-associated DNA (ddRAD) SNPs to assess genome-wide admixture, 16 microsatellites to provide continuity with published data and accurate biogeographic modelling, and a 491-bp fragment of mitochondrial cytochrome oxidase I and II to infer maternal inheritance patterns. Admixture zones correspond with regions impacted by the reorganization of Uganda's river networks that occurred during the formation of the West African Rift system over the last several hundred thousand years. Because tsetse fly population distributions are defined by rivers, admixture zones likely represent both old and new regions of secondary contact. Our results indicate that older hybrid zones contain mostly parental types, while younger zones contain variable hybrid types resulting from multiple generations of interbreeding. These findings suggest that reproductive barriers are nearly complete in the older admixture zones, while nearly absent in the younger admixture zones. Findings are consistent with predictions of hybrid zone theory: Populations in zones of secondary contact transition rapidly from early to late stages of speciation or collapse all together.

Correspondence Norah P. Saarman, Department of Ecology and Evolutionary Biology, Yale University, New Haven, CT. norah.saarman@yale.edu.

AUTHOR CONTRIBUTIONS
A.C., S.A. and N.P.S. contributed to conceptualization.N.P.S. and C.H. performed formal analysis. A.C., S.A., R.E. and E.A.O. contributed to project administration and funding acquisition. N.P.S., R. O., K.D. and T.J. carried out investigation. A.C., C.H., E.A.O., S. A., R.O., R.E. and N.P.S. contributed to methodology. A.C. and S.A. contributed to resources. A.C., E.A.O., R.E. and S.A. supervised the study. N.P.S. contributed to visualization. N.P.S. wrote the manuscript. N.P.S., A.C., R.O., E.O. and S.A. reviewed and edited the manuscript. All authors read and approved the final version of the manuscript.

SUPPORTING INFORMATION
Additional supporting information may be found online in the Supporting Information section at the end of the article.

Keywords

ddRAD; hybridization; population genomics; speciation; trypanosomiasis; vector

INTRODUCTION

Genetic divergence occurs when the forces of mutation, genetic drift and local adaptation cause differences in genotype frequencies between isolated populations. Although divergence occurs most often during periods of geographic isolation (Coyne & Orr, 2004), periods of potential gene flow (secondary contact) are also common during speciation (Davison, Chiba, Barton, & Clarke, 2005; Martin, Dasmahapatra, Nadeau, Salazar, & Walters, 2013). Secondary contact can alter the evolutionary trajectory of the divergent lineages by enforcing, maintaining or even reversing divergence (Barton, 2001; Grant & Grant, 2008) by a balance between gene flow and the strength of selection against hybrids (Barton & Bengtsson, 1986; Barton & Hewitt, 1985; Collins & Rawlins, 2013; Mullen, Dopman, & Harrison, 2008; Turelli, Barton, & Coyne, 2001). The ecological and evolutionary processes underlying this balance appear to be contingent on the environmental, demographic and genomic context of secondary contact (reviewed in Gompert, Mandeville, & Buerkle, 2017) and remain an important question in speciation research and population genetics in general. Patterns of hybridization and introgression (gene flow among divergent lineages) across such regions of secondary contact provide an opportunity to elucidate the ecological and evolutionary processes important in divergence.

Tsetse flies (genus *Glossina*) are a promising system to investigate secondary contact because of their extensive population structure and evidence of admixture. The species *Glossina fuscipes fuscipes* is particularly well-suited for this line of questioning because there is evidence of admixture across multiple genetic breaks representing different stages of divergence (Beadell et al., 2010; Echodu et al., 2013; Opiro et al., 2017). Furthermore, *G. f. fuscipes* offers the advantage of being the object of intense applied research in population structure, ecology, genetics and physiology because of their epidemiological relevance as the obligate vectors of animal and human African trypanosomiasis (Aksoy, Caccone, Galvani, & Okedi, 2013; Kleine, 1909; Riley & Johannsen, 1932).

Previous research described four genetic clusters of *G. f. fuscipes* centred in the northwest, northeast, west and south of the country, with evidence of admixture between these clusters (Figure 1; Beadell et al., 2010; Opiro et al., 2017). We use the term “admixture zones” instead of “hybrid zones” in the context of this system because it is not clear whether they represent regions of secondary contact between reproductively isolated lineages. Instead, admixture zones could be regions undergoing primary divergence with gene flow or they could represent the remnants of old collapsed hybrid zones (e.g., in North American ravens, Webb, Marzluff, & Omland, 2011; in Galapagos giant tortoises, Garrick et al., 2014; in Galapagos finches, Grant & Grant, 2006).

The multiple admixture zones in this system provide an opportunity to compare the genetic structure, biogeographic context, and patterns of hybridization and introgression among them. One expectation established in the hybrid zone literature is that the context-dependent nature of reproductive isolation can create an array of introgression patterns in multiple hybrid zones (Coyne & Orr, 2004; Gompert et al., 2017), even within the same system (Bierne, Bonhomme, & David, 2003). Understanding introgression patterns can shed light on the evolutionary and ecological processes important in the process of speciation. For example, a long-standing question in speciation research is the relative importance of alternative forms of postzygotic reproductive barriers (i.e., intrinsic Dobzhansky–Muller incompatibilities vs. extrinsic ecological barriers: Barton, 2001, Schluter, 2000; Coyne & Orr, 2004; Funk, Nosil, & Etges, 2006; Gompert et al., 2017; Gross & Rieseberg, 2005; Rundle & Nosil, 2005). Recent theoretical work suggests that extrinsic and intrinsic barriers may not work in isolation, but instead can work synergistically in zones of secondary contact (i.e., the “coupling hypothesis”; Bierne, Welch, Loire, Bonhomme, & David, 2011). Under a model of divergence with gene flow, the coupling of multifarious barriers can rapidly advance the process of speciation from early to late stages (Gompert et al., 2017). One prediction of this hypothesis is the rare occurrence of stable long-lasting hybrid zones (Gompert et al., 2017), which implies rarity of first-generation hybrids as well as intermediate levels of introgression in nature. Here, we investigate this prediction with a comparison of introgression patterns found in *G. f. fuscipes*, a unique system with four admixture zones that exist within a ~40,000 km^2 region in Uganda (Beadell et al., 2010; Echodu et al., 2013; Opiro et al., 2017; Saarman et al., 2018). Our goal was to investigate the biogeographic context and patterns of introgression in this unique system (Figure 1) to provide insight on the evolutionary and ecological processes important in diversification.

In this study, we genotyped 251 individuals from 18 sites (Figure 1; Supporting Information Table S1) at 55,267 SNPs scored from double-digest restriction site-associated DNA (ddRAD) sequencing, 16 microsatellite markers and a 491-bp fragment of mitochondrial DNA (mtDNA) sequence from the cytochrome oxidase I and II (COI and COII) genes. We chose to use these three data types as each provides unique insight into hybrid zone dynamics. ddRAD data provided a genome-wide analysis of hybridization and introgression. Microsatellites offered continuity with previous studies and reliable models of recent evolutionary history because of their well-documented mutation rates, high molecular diversity and high minor allele frequencies (Beadell et al., 2010; Echodu et al., 2013; Hyseni et al., 2012; Opiro et al., 2016, 2017; Saarman et al., 2018). mtDNA sequences supplied inference of maternal inheritance, which can help distinguish asymmetrical reproductive barriers (Coyne & Orr, 2004) and past colonization patterns in hybrid zones (Currat, Ruedi, Petit, & Excoffier, 2008). Findings suggest that the coupling of multiple isolating barriers may play an important role in speciation in tsetse flies. This work lays the groundwork for more detailed investigation of the ecological and evolutionary processes critical to speciation and of the genetic basis for these processes.

2 | METHODS

2.1 | Samples and study area

Figure 1 shows the 18 sampling sites included in this study (details in Supporting Information Table S1). Tsetse flies were collected using bi-conical traps (Challier & Laveissiere, 1973) set out in groups of 10–15 traps within a radius of 2 km, a field protocol that reliably traps unrelated individuals (Echodu et al., 2013; Opiro et al., 2017). Sampling sites are a subset of sites from previous studies (Beadell et al., 2010; Echodu et al., 2013; Hyseni et al., 2012; Opiro et al., 2017) and were selected to represented the genetic diversity of the four geographic regions of genetically distinct populations found in Uganda (Figure 1: northwest, northeast, west and south; Beadell et al., 2010; Opiro et al., 2017) and the four admixture zones between them (Figure 1: "a" between the northwest and west, "b" between the northwest and northeast, "c" between the west and south, and "d" between the northeast and south).

Previous work has shown that watersheds facilitate dispersal in *G. f. fuscipes* (Bouyer et al., 2009) and thus play an important role in shaping genetic structure observed (Beadell et al., 2010; Hyseni et al., 2012; Opiro et al., 2017). To capture this geographic aspect, sampling was distributed across the seven major watersheds, and all descriptive analysis (i.e., genetic diversity, heterozygosity and structure) used the natural boundaries created by the watersheds (Figure 1, Table 1). We chose to define groups for descriptive analysis by watershed rather than by genetic cluster because either some sampling sites did not fall neatly within a single genetic cluster (Figure 1; Table 1; Beadell et al., 2010; Opiro et al., 2017).

From the northwest cluster (Table 1; Beadell et al., 2010; Opiro et al., 2017), we sampled from the Albert Nile watershed, which is characterized by lowland woodlands and gallery forest. This habitat extends north into South Sudan and west into the Democratic Republic of the Congo (DRC).

From the northeast cluster (Table 1; Beadell et al., 2010; Opiro et al., 2017), we sampled from the Lake Kyoga watershed, which is characterized by marsh and swampland. This region is bordered to the north and east by the edge of the *G. f. fuscipes* distribution where it is too dry to support the vector.

From the west cluster (Table 1; Beadell et al., 2010), we sampled the Lake Albert and the Kafu River watersheds, which contain patchy gallery forest along the major rivers. This region is bordered to the west by protected forests (Masindi National Park), where ecological competitors *G. pallidipes*, *G. palpalis* and *G. brevipalpis* are dominant (Abila et al., 2008), and to the west and south by elevation that is too high and temperatures that are too cool for *G. f. fuscipes*.

From the south cluster (Table 1; Beadell et al., 2010), we sampled from the Lake Victoria watershed, where urbanization has greatly reduced the available habitat down to a small strip of habitat along the lake and river shorelines. This region is bordered to the east and south by the edge of the *G. f. fuscipes* distribution where savannah habitat favours the ecological competitor *G. pallidipes* (Abila et al., 2008; Ciosi, Masiga, & Turner, 2014).

We also sampled from the Achwa and the Okole River watersheds, which are geographically intermediate to the centres of the northwest and northeast genetic clusters (Figure 1; Table 1; Beadell et al., 2010; Opiro et al., 2017). These watersheds were problematic to assign to a single genetic cluster because flies from these watersheds are a mix of the northwest, northeast and west genetic backgrounds (Figure 1: admixture zones "a" and "b"; Opiro et al., 2017; Saarman et al., 2018). The habitat in these watersheds is a mix of lowland woodlands characteristic of the northwest and swampland characteristic of the northeast.

2.2 | DNA extraction and ddRAD sequencing library preparation

DNA for the ddRAD protocol was extracted from the heads, thorax, wings and legs of 254 *G. f. fuscipes* starting individuals (later reduced to 251 after removing individuals with low-quality sequence data) using Qiagen DNeasy blood and tissue extraction kits (Qiagen, Valencia, CA), with a preliminary step added for tissue pulverization using the Qiagen Bead-beater system. We then quantified DNA extractions with a Qubit 2.0 Fluorometer (Invitrogen, Carlsbad, CA, USA) and proceeded with only individuals having higher than 500 ng total yield of genomic DNA.

ddRAD sequencing libraries were prepared following Gloria-Soria et al. (2016, 2018) using a modified version of the Peterson et al. (2012) protocol with restriction enzymes NlaIII and MluCI and a size selection step that isolated 95- to 105-bp DNA fragments. We created eight ddRAD sequencing libraries of 32 pooled individuals each, which were sent to the Yale Center for Genome Analysis for 75-bp paired-read sequencing with the Illumina Hi-Seq platform under the "high-output" mode and were sequenced in lanes shared with randomly sheared libraries. The final data set includes 251 individuals (5–20 individuals per populations for 18 populations; Supporting Information Table S1). ddRAD-based paired-end short-read sequences for each of the final 251 individuals included in this study are available on NCBI *via* BioProject Accession no. PRJNA498097.

2.3 | ddRAD sequence processing and SNP calling

The ddRAD library raw sequence reads were de-multiplexed, quality filtered and filtered for unambiguous barcodes using the "process_radtags" script from the STACKS v.1.34 software (Catchen, Hohenlohe, Bassham, Amores, & Cresko, 2013). Processed reads were aligned to the 2,395 scaffolds available in the *G. fuscipes* GfusI1 reference assembly (Giraldo-Calderon, Emrich, MacCallum, Maslen, & Dialynas, 2014) with BOWTIE2 v. 2.1.0 (Langmead & Salzberg, 2012) with the default settings, sorted with SAMTOOLS v. 0.1.19 (Li et al., 2009), and analysed for SNPs with STACKS. A VCF file of all SNPs was created by using the STACKS cstacks script with a training set of 36 individuals from each of the four genetic backgrounds, and then, using pstacks, sstacks and the population script with all of the 251 final individuals in such a way that only SNPs with more than 5X coverage, a minor allele frequency greater than 0.2% and a genotyping rate greater than 80% were retained. Details of this are available in the Supporting Information Appendix S1. Read depth statistics were estimated using VCFTOOLS version 0.1.13 (Danecek, Auton, Abecasis, Albers, & Banks, 2011; flag --geno-depth). Other necessary conversions between file formats (e.g., from VCF format to GenePop format) were performed using PGDSPIDER v. 2.0.5.2 (Lischer & Excoffier, 2012). The final VCF file contained 55,267 SNPs and 251 individuals from 18 sites that passed all filters.

We used PLINK version 1.9 (Purcell et al., 2007) to exclude physically linked SNPs with variance inflation factors (VIF) greater than 2 in 100-kb sliding windows (PLINK flag --indep 100 10 2), which was a far larger window size than the average distance of LD decay estimated at ~3 kb (Gloria-Soria et al., 2016). This reduced the ddRAD data set to 33,057 unlinked SNPs for the population structure analyses, and other analyses wherever noted because these benefit from using physically unlinked loci (Purcell et al., 2007).

2.4 | Microsatellite genotyping

New genotypic data were collected from the 16 microsatellite loci for individuals present in the 251 individual ddRAD data set (Supporting Information Tables S1 and S2) that did not have existing data available from https://doi.org/10.5061/dryad.20b01 (Brown et al., 2008; Beadell et al., 2010; Opiro et al., 2017), using the same protocol as Opiro et al. (2017), which is detailed in Supplementary Methods. In several cases, because there was not enough DNA left to fill missing data gaps, we used different individuals from the same sampling site for the different data sets (Supporting Information Table S2). Thus, the total data set included new genotypic data for 16 loci from 25 new individuals from 5 sites, new data for 5 loci from 27 individuals from 3 sites already analysed at 11 loci by Beadell et al. (2010) and Echodu et al., (2013), and existing data for 16 loci from 194 individuals from 13 sites already analysed at all 16 loci by Opiro et al., (2017), which all together created a data set of 16 loci, 251 individuals, and 18 sites (Supporting Information Tables S1 and S2). 88% of the ddRAD and microsatellite data sets shared the same individuals (222 out of 251; Supporting Information Table S2). All genotypic data included in this study were deposited to Dryad (https://doi.org/10.5061/dryad.98rf2f1).

2.5 | mtDNA sequencing

A 570-bp fragment of the mtDNA spanning the COI and COII genes was PCR-amplified for any individuals present in the final data set of 251 individuals (Supporting Information Tables S1 and S2) that did not have existing data available from GenBank (Accession nos. GU296746-GU296786; Beadell et al., 2010; Opiro et al., 2017). We used the same protocol as Opiro et al. (2017), which are detailed in Supplementary Methods. We were able to include only a subset of the individuals screened for nuclear markers (Supporting Information Table S2) because of low quantity and quality of older DNA extractions. In total, we sequenced 177 additional flies with 5–20 individuals per site and combined these with data for the same fragment from Beadell et al., 2010, Echodu et al., 2013 and Opiro et al., 2017 (Supporting Information Table S2). The total mtDNA data set consisted of 228 mtDNA sequences from 18 sites (Supporting Information Table S1). 82% of all three of the data sets (i.e., the ddRAD, microsatellites and mtDNA data sets) shared the same individuals (206 out of 251; Supporting Information Table S2). All haplotype sequences included in this study were deposited to GenBank (Accession nos. MK094112-MK094173).

2.6 | Genetic diversity and population structure

2.6.1 | ddRAD SNP genetic diversity—For the 55,267 ddRAD fragments, genetic diversity statistics were estimated with the STACKS population script. The script was run once to obtain nucleotide diversity (Pi_A) estimates among all sites at both variable (158,537 SNPs at an average density of ~2.85 SNPs per ddRAD fragment) and invariable positions in the 55,267 ddRAD fragments scored. They were also scored a second time considering only a single representative SNP from each of the 55,267 polymorphic ddRAD fragments.

This second run provided per-sample estimates of the number of polymorphic sites (PS), the number of private alleles observed (PA), the mean frequency of the major (most frequent) allele at each locus (P), observed heterozygosity (H_O) and expected heterozygosity (H_E). Diversity statistics were summarized by watershed after confirming the relevance of watersheds with an analysis of molecular variance (AMOVA; Supporting Information Appendix S1).

To help interpret the level of secondary contact among genetic units observed in the data, Hardy–Weinberg disequilibrium (HWD) was estimated in all loci (SNPs) as well as linkage disequilibrium (LD) among locus pairs. Relative HWD levels were determined as the proportion of loci in HWD in each sample below the p-value threshold of 0.05 (P_{HWD}; PLINK flag --hwe 0.05). Relative levels of LD were estimated with a PLINK command that determined the proportion of locus pairs in each sample over the D' threshold of 0.5 (P_{LD}; PLINK flag --r2 inter-chr dprime with-freqs gz --ld-window-r2 0.01). The individual inbreeding coefficient relative to the sample (F_{IS}) was estimated with the Weir and Cockerham's (1984) method using R (R Core Team, 2016) and the "HIERFSTAT" R package (Goudet & Jombart, 2015) and assessed for significance with 1,000 bootstrap samplings with corrections for multiple testing with the Benjamini–Hochberg (BH; Benjamini & Hochberg, 1995) and the Bonferroni (BF) methods (Dunn, 1961) in the "STATS" R package (R Core Team, 2016).

2.6.2 | ddRAD SNP population structure—We used the LD filtered data set of 33,057 unlinked SNPs to assess ddRAD population structure. Pairwise differentiations among samples were estimated with Wright's F-statistic (F_{ST}; Wright, 1951) in the "STAMPP" R package (Perrbleton, Cogan, & Forster, 2013) and assessed for significance with 1,000 bootstrap replicates with corrections for multiple testing using BH and BF methods in the "stats" R package. Wright's F_{ST} was chosen to allow for comparisons with previous *G. f. fuscipes* studies and among our different data sources. Pairwise F_{ST} values were used to test for isolation by distance (IBD), as implemented in the "ADEGENET" R package v. 2.0.1 (Jombart, 2008; Jombart & Ahmed, 2011). Geographic distances were generated using the Java-based "GEOGRAPHIC MATRIX GENERATOR" version 1.2.3 (Ersts, downloaded November 2017). The significance of the regression was tested by a Mantel test with 10,000 randomizations (Goudet, Raymond, & Meeüs, 1996; Jombart, 2008; Jombart & Ahmed, 2011; Mantel et al., 1967). We applied the Mantel test to estimates from the full data set, from the northwest and northeast clusters, and from the west and south clusters. We split the samples into two larger clusters instead of the four units and four admixture zones (total of eight groups) for this analysis to allow for large enough sample size for statistical power.

We ran a principal component analysis (PCA) using the "ADEGENET" R package, which is a multivariate, model-free method that makes no assumptions about absence of HWD and LD. We performed clustering analyses with STRUCTURE version 2.3.4 (Pritchard & Stephens, 2000) in "ADEGENET" in R. For STRUCTURE, we used K (number of clusters) ranging from 1 to 18 (the total number of sites) and ran ten independent replicates with 250,000 iterations and 50,000 burn-in steps were run. The most informative K was assessed with the ad hoc statistic "ΔK" (Earl & Von Holdt, 2012; Evanno, Regnaut, & Goudet, 2005). Results from the ten replicates were combined using CLUMPAK version 1.1 (Kopelman, Mayzel, Jakobsson, Rosenberg, & Mayrose, 2015). We also applied discriminant analysis of principal components (DAPC) as an alternative model-free clustering method and estimated pairwise co-ancestry coefficients with the program FINESTRUCTURE v. 2.0.7 (Lawson, Hellenthal, Myers, & Falush, 2012; Scheet & Stephens, 2006). Details of the DAPC and FINESTRUCTURE methods are available in the Supporting Information Appendix S1.

2.6.3 | Microsatellite genetic diversity—In order to provide a link between the new ddRAD results and the previous estimates from the literature, we estimated observed heterozygosity (H_O), expected heterozygosity (H_E) and allelic richness (A_R). These estimates were made using the R libraries POPPR v. 2.3.0 (Kamvar, Brooks, & Grünwald, 2015; Kamvar, Tabima, & Grünwald, 2014), PopGenReport v. 1.0.2 (Adamack & Gruber, 2014; Gruber & Adamack, 2015) and "ADEGENET." The SNP and microsatellite-based estimates of diversity were compared with a linear regression, using the "STATS" R package. In this analysis, microsatellite allelic richness was regressed against both nucleotide diversity Pi_A (Supporting Information Table S1) and Pi (Table 1).

2.6.4 | Microsatellite population structure—F_{IS} and pairwise F_{ST} estimates were made in FSTAT 2.9.3 (Goudet 1995), and significance was assessed using *p*-values calculated from 1,000 randomizations. IBD was tested with the pairwise F_{ST} values, using a Mantel test with 10,000 randomizations implemented with the "ADEGENET" R package. SNPs and microsatellite F_{ST} estimates were also compared using a Mantel test with 10,000 randomizations in the "ADEGENET" R package. As with the ddRAD data set, the microsatellite principal components analysis (PCA) was implemented in the "ADEGENET" R package, and STRUCTURE and DAPC analyses were performed with *K* ranging from 1 to 18.

2.6.5 | mtDNA genetic diversity—To investigate maternal inheritance patterns in the admixture zones, we assessed the genetic diversity and population structure of mtDNA COI sequence data using the program ARLEQUIN (Excoffier & Lischer, 2010). Genetic diversity within populations was estimated by computing haplotype (H_d) and nucleotide diversity (N_d; Nei, 1987) in DnaSP version 5.0 (Librado & Rojas, 2009).

2.6.6 | mtDNA population structure—Haplotype relationships were inferred by constructing a parsimony network using TCS (Clement, Snell, Walker, Posada, & Crandall, 2002) implemented in PopART (http://popart.otago.ac.nz/). The distribution of groups of related haplotypes (haplogroups) was visualized by watershed.

2.7 | Biogeographic and demographic history of the admixture zones

2.7.1 | Phylogenetic relationships—We evaluated evolutionary relationships among the sampled populations with the software TreeMix (Pickrell & Pritchard, 2012; Sukumaran & Mark, 2010). We used the ddRAD data set for this analysis because of the three data sets available, this marker type contained the greatest number of phylogenetically informative sites, and because the software is designed to deal with a great number of loci that have variation in their allele frequencies and evolutionary histories. The benefit of this method is that it allows for both population splits and gene flow, which has been long recognized to be the more accurate representation of relationships between populations with potential gene flow (Cavalli-Sforza & Piazza, 1975; Felsenstein 1982; Pickrell & Pritchard, 2012). Details of the TreeMix methods are provided in the Supporting Information Appendix S1.

2.7.2 | Biogeographic and demographic models—We modelled the population history of the major genetic clusters found in Uganda (Figure 1) with approximate Bayesian computation (ABC) in DIYABC version 2.0.4 (Cornuet et al., 2014) and with MIGRAINE version 0.5.2 software (http://kimura.univ-montp2.fr/~rousset/Migraine.htm; Peery et al., 2012; Supporting Information Appendix S1) for comparison. The models focused on northern Uganda where alternative explanations for the patterns of divergence and diversity observed in each genetic cluster are in debate (Beadell et al., 2010; Echodu et al., 2013; Hyseni et al., 2012; Opiro et al., 2016, 2017; Saarman et al., 2018). Both ABC and MIGRAINE simulations assume panmictic populations unless explicitly specified otherwise, so we used samples from the four watersheds that fell squarely within one of the four major genetic clusters (Figure 1; Table 1; Supporting Information Table S2). The northwest cluster was represented by two sites from the Albert Nile (01-DUK and 02-ORB), the northeast cluster was represented by two sites from Lake Kyoga (10-OCU and 11-OT), the west cluster was represented by two sites from Lake Albert (13-MF and 14-MS), and the south cluster was represented by a single site from Lake Victoria (18-WAM). We also included a sample from admixture zone “b” from the Okole River (Table 1; Supporting Information Table S2) in the ABC analysis to allow us to distinguish alternative hypotheses of the formation of this admixture zone.

We used the microsatellite data set for ABC and MIGRAINE simulations because of the three data sets available, it represented the best balance between being well-characterized (e.g., in terms of mutation model) and having high enough genetic diversity to provide statistical power. We also attempted to run the same analyses with the ddRAD data set, but were unsuccessful in parameterizing the mutation models because of lack of user control of the mutation priors in the case of DIYABC, and insufficient compute power in the case of MIGRAINE. Details of the unsuccessful ddRAD-based ABC analyses are available in Supplementary Methods.

The ABC method is limited by joint estimates of timing of divergences, population size, mutation rate and the mutation model, making it impossible to accurately model simulated data sets based on these interdependent unknown parameters without informed priors. For this reason, priors for ABC models (Supporting Information Table S3) were based on independently derived estimates wherever possible. Mutation priors were based on estimates of microsatellite mutation rates in other insects (Chapuis, Plantamp, Streiff, Blondin, & Piou, 2015) and results from a published ABC analysis in a closely related species, *Glossina pallidipes* (Ciosi et al., 2014). Population size and bottleneck priors were based on previous population genetics estimates (Beadell et al., 2010; Echodu et al., 2013; Hyseni et al., 2012; Opiro et al., 2016, 2017). Time priors considered average generation time of *G. f. fuscipes* of 8 per year (45 days; Krafsur, 2009) and were based on the geologic history. We applied these time priors with caution using a three-step preliminary analysis (Supporting Information Appendix S1). In the first step of the preliminary analysis, the time priors were minimally restricted (<650 ka; Supporting Information Table S3) based on independent records of geologic events (Supporting Information Figure S1) that are well established to have driven patterns of tsetse population connectivity in Uganda at ~400, ~20–40 and ~10–20 ka (Supporting Information Figure S1; Beadell et al., 2010; Opiro et al., 2017). However, from the first step of the preliminary analysis, posterior estimates were unrealistically recent and truncated by zero. Thus, considering that all priors were interdependent, to make the models more realistic we performed two more steps of the preliminary analysis with increased specificity based on more detailed biogeographic scenarios that were built from independent records of geologic events (Supporting Information Figure S1) and published population genetics data (Beadell et al., 2010; Opiro et al., 2017; Saarman et al., 2018). There were three major river reorganizations in the region that would have impacted tsetse distribution and connectivity: (a) Lake Victoria formed ~400 ka and reversed the direction of flow of the major river systems from west to north (Supporting Information Figure S1a; Williams, Adamson, Prescott, & Williams, 2003; Danley et al., 2012), (b) the outflow from Lake Victoria altered its course to connect through Kafu River to Lake Albert ~20–40 ka (Supporting Information Figure S1b; Bishop, 1969) and (c) the outflow from Lake Victoria altered its course a third time to bypass the Kafu River and to connect directly to Albert Nile through Murchison falls (Supporting Information Figure S1c; Talbot & Williams, 2009). We used these timings of geologic events to refine the time priors during the preliminary analysis stage and then compared three alternative scenarios in the final ABC analysis:

Scenario 1 modelled the possibility that divergence among the four genetic backgrounds in Uganda was created during geologic events ~400 (Figure 2 and Supporting Information Figure S2a: t4), ~20–40 (Figure 2 and Supporting Information Figure S2b: t3) and ~10–20 ka (Figure 2 and Supporting Information Figure S2c: t2) and that the admixture zone “b” represents secondary contact between the northwest and northeast lineages, which is an idea proposed by Opiro et al. (2017).

Scenario 2 modelled the possibility that divergence among the four genetic backgrounds in Uganda was created during geologic events ~400 (Figure 2 and Supporting Information Figure S2a: t4) and ~20–40 ka (Figure 2 and Supporting Information Figure S2b: t3) and that the admixture zone “b” represents ongoing divergence with gene flow or incomplete divergence between the northwest and northeast (Figure 2 and Supporting Information Figure S2a: t1), which is an idea proposed by Saarman et al. (2018) and is based on theoretical considerations (Falush, VanDorp, & Lawson, 2016; Frantz, Cellina, Krier, Schley, & Burke, 2009; Meirmans, 2012).

Scenario 3 modelled the possibility that divergence among the four genetic backgrounds in Uganda was created during geologic events ~400 (Figure 2 and Supporting Information Figure S2a: t4), ~20–40 (Figure 2 and Supporting Information Figure S2b: t3) and ~10–20 ka (Figure 2 and Supporting Information Figure S2c: t2) and that the signal of admixture is a false one caused by a bottleneck event in one of two closely related lineages, a scenario outlined in Falush et al. (2016). We modelled the time of the possible population bottleneck at 0–5 ka, as this is the time frame that could have allowed for some divergence to accumulate between the northwest and northeast, before changing environmental conditions due to the final desiccation event in the Lake Kyoga region ~5 ka (Danley et al., 2012) and/or anthropogenic effects (Rinderpest, population displacement and recent vector control campaigns) of the past 135 years or so (Van Acker, 2004; Carmichael, 1938; Egeru et al., 2014; Gorsevski, Kasischke, Dempewolf, Loboda, & Grossmann, 2012) may have triggered dramatic shifts in population sizes.

2.7.3 | Biogeographic and demographic parameter estimates—We estimated summary statistics from the simulated data to allow for comparison among the three final alternative scenarios and the observed microsatellite data. We recorded the number of alleles, mean heterozygosity and the *M*-index (Garza & Williamson, 2001; Excoffier, Estoupe, & Cournuet, 2005) for each lineage, and the mean pairwise F_{ST} distances (Weir & Cockerham, 1984) for each pair of lineages, for a total of 25 summary statistics. PCA was then performed on these 25 summary statistics to compare posterior estimates and to choose the scenario that had the closest match with the observed microsatellite data using the weighted logistic regression method described by Fagundes, Ray, Beaumont, Neuenschwander, and Salzano (2007). We also estimated the posterior predictive error with the method described by Cornuet et al., (2014) to confirm reliability of the model. For the ABC parameter estimates (Ne, timing of divergence events, admixture rates and mutation rates), the mean, median and 95% confidence intervals were drawn from the linear regression of the 1% of the simulations of the winning scenario that were closest to the observed microsatellite data.

For comparison, MIGRAINE analysis modelled demographic history assuming an exponential change in population size from the ancestral time point continuing until the current time point, which is the basic coalescence model (Leblois et al., 2014). This provided estimates of genetic diversity of the current sample (θ) and the hypothetical ancestral sample (θ_{anc}), and estimates of Ne (Supporting Information Appendix S1).

2.8 | Characterization of genome-wide introgression in admixture zones

To understand the degree of allele frequency divergence and hybridization occurring in each of the four admixture zones, we used the LD filtered data set of 33,057 unlinked SNPs to characterize the absolute delta allele frequency, $|\Delta p|$, across each admixture zone “a,” “b,” “c” and “d,” using the “INTROGRESS” R package v. 1.2.3 (Gompert & Buerkle, 2009). The ddRAD SNPs were the most appropriate genetic marker type to use in this analysis because they are randomly distributed at high density (~1 SNP per ~210 bp) across the genome. To establish the allele frequencies of each parental type, we used the same representatives of each of the four pure genetic clusters (Table 1 and Supporting Information Table S2) as used in the ABC analysis after confirming that each individual had high assignment (*q*-values >0.8) to a single cluster in the current study. For admixture zones “a,” “c,” and “d,” we maximized accuracy by only considering SNPs with $|\Delta p|$ greater than 0.8 to estimate “interspecific” heterozygosity and “hybrid index,” a practice recommended by the authors of INTROGRESS (Gompert & Buerkle, 2009). For admixture zone “b” there were no $|\Delta p|$ values >0.8, so we only considered SNPs with $|\Delta p| > 0.5$.

For comparison with the patterns of $|\Delta p|$, we also estimated locus-by-locus F_{ST} values and tested for F_{ST} outlier loci using the R package OUTFLANK version 0.2 (Whitlock & Lotterhos, 2015) with the same ddRAD data set used for the INTROGRESS analysis. Overlap of loci with F_{ST} with a right-tailed p-value >0.02 among the four admixture zones was assessed using the free online Venn diagram bioinformatics tool http://bioinformatics.psb.ugent.be/webtools/Venn/.

3 | RESULTS

3.1 | ddRAD sequencing statistics

Sequencing and diversity statistics of the ddRAD fragments indicated adequate coverage in all samples included in the final analyses (Supporting Information Figure S2; Table S1). Read depth per individual averaged 85.9 (range, 32.7–269.3) with its standard deviation averaging 650 (range, 62–1,446; Supporting Information Table S1). The majority (92.8%) of the aligned ddRAD fragments contained polymorphisms, as only 3,938 ddRAD fragments (7.2%) were removed during the STACKS pipeline because they were invariant. The retained variant 55,267 ddRAD fragments contained 2.85 SNPs on average. Considering only the first SNP per fragment (to avoid scoring highly linked SNPs), the number of polymorphic sites (PS) per sample averaged 28,093 (range, 13,739–36,395; Supporting Information Table S1). This estimate appeared to be driven in large part by the sample number, because the lowest estimates of polymorphic sites were at sites with small sample sizes (Supporting Information Figure S2). The number of private alleles (PA) per sample averaged 348 (range, 46–1,025; Supporting Information Table S1). The mean of the major (most frequent) allele (P) at each locus averaged 0.89 (range, 0.87–0.93; Supporting Information Table S1). Among all sites, including both variant and invariant sites in the 55,267 ddRAD fragments scored, nucleotide diversity (Pi_A) averaged 0.006 (range, 0.004–0.007; Supporting Information Table S1). This equates to a genome-wide density of about 1 SNP per 150–270 bp after accounting for the 7.2% of ddRAD fragments that did not contain polymorphism.

3.2 | Genetic diversity and population structure

3.2.1 | ddRAD SNP diversity—Among the 55,267 SNPs in the ddRAD data set, sample-wide nucleotide diversity (Pi) averaged 0.15 (range, 0.10–0.19; Table 1). Sample-wide observed (H_O) and expected (H_E) heterozygosity averaged 0.15 with similar ranges (range, 0.10–0.21 and 0.09–0.18, respectively; Table 1). The proportion of loci in HWD in this sample (P_{HWD}) averaged 0.05 (range, 0.00–0.12; Table 1). The proportion of locus pairs in linkage disequilibrium in this sample (P_{LD}) averaged 0.15 (range, 0.05–0.22; Table 1). Sample-wide inbreeding F_{IS} averaged 0.04 (range, −0.17 to 0.23; Table 1). Bootstrap resampling indicated that most samples were slightly but significantly different from zero after BH and BF correction for multiple testing (Supporting Information Table S4), indicating slight HWD. The AMOVA results confirmed that there was significant partitioning of molecular variance among the seven watersheds (p-value <0.001), with 68.9% of the variation found among watersheds, 12.9% found among sites, and 20.3% found within sites (Supporting Information Table S5).

3.2.2 | ddRAD SNP population structure—Pairwise F_{ST} indicated substantial population differentiation at small spatial scales of less than 50 km (Supporting Information Table S6). All but one of the SNP-based pairwise estimates of differentiation were significantly greater than zero (Supporting Information Table S6, above the diagonal), even after correcting for multiple comparisons (Supporting Information Table S7). The Mantel test for IBD indicated a significant effect of geographic distance on level of differentiation (*p*-value=0.047), as well as among the northwest and northeast (*p*-value=0.022), but not among the west and south (*p*-value=0.711; Supporting Information Figure S3a).

The results of the PCA on the SNPs data set identified genetic structure with a clear geographic component (Figure 3a; Supporting Information Figure S4). The first PC explained 11.64% of the total variance and separated samples north to south. The northwest and northeast (Albert Nile, Achwa River, Okole River and Lake Kyoga) and the south and west (Lake Albert, Kafu River and Lake Victoria) largely grouped together. The second PC explained 3.95% of the total variance and separated samples west to east (Figure 3a), highlighting differentiation between the west and south (i.e., Lake Albert and Kafu River from Lake Victoria) but not between the northwest and northeast ones (i.e., Albert Nile from Lake Koga). The results of STRUCTURE and DAPC clustering analyses indicated an optimal K-value of two (Supporting Information Figure S5). They agree with the clustering pattern from the PCA, finding structure between the west and south, but no structure between the northwest and northeast (Supporting Information Figures S6 and S7; Table S8).

FINESTRUCTURE identified 10 and 75 major and minor clusters, respectively (Supporting Information Figure S8; horizontal lines numbered 1–10 along the top). The 10 major clusters identified the same groups as the PCA and clustering analysis. Major clusters c1–c7 contained northwest and northeast samples only, whereas the major clusters c8–c10 contained west and south samples only (Supporting Information Figure S8; Table S9).

3.2.2 | Microsatellite diversity—Genotypic diversity of the 16 microsatellite loci (Supporting Information Table S10) was strongly correlated with the results from the SNP-based estimates (*p*-value <0.001; Supporting Information Figure S9). Allelic richness averaged 3.02 (range, 2.38–3.59; Supporting Information Table S10), observed heterozygosity averaged 0.55 (range, 0.45–0.64 (Supporting Information Table S10), and expected heterozygosity was generally higher than the observed one (average = 0.60, rage, 0.47–0.70; Supporting Information Table S10). F_{IS} averaged 0.08 (range, −0.07 to 0.21; Supporting Information Table S10). Bootstrap resampling indicated that seven of the 12 samples from the north (northwest, admixture zone “b,” or northeast) were significantly different from zero after BH correction for multiple testing (Supporting Information Tables S4), indicating HWD. In contrast, none of the west or south samples had F_{IS} values significantly different from zero (Supporting Information Table S10). The AMOVA results on this data set confirmed that there was significant partitioning of molecular variance among the seven watersheds (*p*-value <0.001), with 7.1% of the variation found among watersheds, 8.3% found among sites, and 84.6% found within sites (Supporting Information Table S5).

3.2.4 | Microsatellite population structure—Pairwise F_{ST} values indicated significant population differentiation (Supporting Information Table S11) even after correcting for multiple testing (Supporting Information Table S7). The Mantel test for IBD indicated a significant effect of geographic distance on level of differentiation among all samples (*p*-value 0.047), as well as among the northwest and northeast (*p*-value 0.016), but not among the west and south (*p*-value 0.248; Supporting Information Figure S3b). These estimates were also strongly correlated with the results from the SNP-based estimates (p-value <0.001; Supporting Information Figure S3c), and all were significant after BH correction but non-significant after BF correction (Supporting Information Table S7).

Principal component analysis with the 16 microsatellites (Figure 3b) corroborates the strong geographic component in the genetic structure found with the SNPs, with strong north-to-south, and weak west-to-east differentiation. The clustering analysis assigned individuals to two genetic clusters ($K = 2$; Supporting Information Figure S5), with the same general groups found in the SNP-based analysis but with several differences (Figure 3b; Supporting Information Figures S6b and S10). Differences included grouping of the admixture zone "a" sample with the west (Lake Albert and Kafu River) rather than with the other Okole River (Supporting Information Figure S6), and substructure between the northwest and the northeast (Albert Nile from Lake Kyoga) at $K = 4$ (Supporting Information Figure S6b).

3.2.5 | mtDNA genetic diversity—Sequencing identified 25 mtDNA haplotypes (Supporting Information Table S10) representing four groups of related haplotypes (haplogroups) in the TCS network (Supporting Information Figure S11). The number of haplotypes at each sampling site ranged from one to six (Supporting Information Table S10). Sixteen haplotypes were singletons (8 were haplogroup A, seven were haplogroup B, one was Haplogroup C, and one was haplogroup D). Haplotype diversity averaged 0.52 and ranged from 0.20 to 0.80, and nucleotide diversity averaged 0.0028 and ranged from 0.0011 to 0.0071 (Supporting Information Table S10). The mtDNA-based AMOVA results confirmed that there was significant partitioning of molecular variance among the seven watersheds (*p*-value = 0.012), with 28.2% of the variation found among watersheds, 27.6% found among sites, and 20.3% found within sites (Supporting Information Table S5).

3.2.6 | mtDNA population structure—The distribution of haplogroups indicated spatial structuring (Figure 3c). Haplogroup A (purple) was most frequent in the northwest (Albert Nile, 94.4%) and west (Lake Albert and Kafu River; 79.1%, and 93.2%), occurred less commonly in admixture zone "b" (Achwa and Okole River; 17.2% and 24.0%), rarely in the northeast (Lake Kyoga; 4.8%), and not all in the south (Lake Victoria). Haplogroup B (blue) showed the opposite pattern, with high frequency in the northeast (Lake Kyoga; 95.2%) and low frequency in the northwest (Albert Nile, 5.6%; Figure 3c). Haplogroup C occurred only in the west (Lake Albert and Kafu River; 21.1%, 6.8%) and south (Lake Victoria; 95.2%). Haplogroup D occurred only once in the south (Lake Victoria; 4.8%).

3.3 | Biogeographic and demographic history of the admixture zones

3.3.1 | ddRAD SNP phylogenetic trees—Results from TreeMix confirmed the clustering found in the PCA, STRUCTURE and DAPC analysis, with strong support for a split between the north (northwest, admixture zone "b" and northeast) samples and the combined west and south samples, but no support for separation of the northeast from the northwest (Supporting Information Figure S12a). This analysis identified migration and subsequent introgression across admixture zones "a" and "d." Migration was predicted from the west (Kafu River) into the northeast (Albert Nile; 15-KAF to 01-DUK; Supporting Information

Figure S12a) and from the south (Lake Victoria) into the northeast (Lake Kyoga; 17-NB to 12-BN; Supporting Information Figure S12a). In addition, the residual fit (Supporting Information Figure S12b) indicated positive values across admixture zone “c,” suggesting admixture beyond the two migration events identified in the tree. These positive residuals occurred between the south (Lake Victoria) and the west (Lake Albert and Kafu River; between 16-JN and 14-MS/15-KAF). There was no signal of migration or admixture across admixture zone “b.”.

3.3.2 | Biogeographic scenarios modelled by ABC—The step 1 preliminary ABC analysis yielded time parameter estimates that were unrealistically low and truncated atzero (Supporting Information Figure S13). To resolve this problem, we included two more preliminary analyses (step 2 and step 3) that restricted the priors to increasingly narrow timings (Supporting Information Table S3) consistent with what is known of the geologic history of the region (Supporting Information Figure S1; Williams et al., 2003; Bishop, 1969; Danley et al., 2012). We then used the posterior distribution from step 3 of the preliminary analysis to inform the priors in the final analysis.

The final microsatellite-based ABC analysis indicated that *Scenario 2* (Figure 2b) had the highest posterior probability (0.999) regardless of the topology used (Table 2). This scenario simulated the hypothesis that the signal of admixture in northern *G. f. fuscipes* populations was caused by ongoing gene flow or incomplete divergence between the northwest and northeast. Summary statistics from simulations under *Scenario 2* (Supporting Information Table S13c) were the closest match to the observed data as compared to the other scenarios (Supporting Information Figure S14c) and were a close fit to the observed data in general (Supporting Information Figure S14). Results from the similar analysis attempted with the ddRAD SNPs did not show the same match of simulations with real data (Supporting Information Figure S14), but they did in fact indicate agreement with the microsatellite-based ABC analysis (Supporting Information Table S14). However, these results were unreliable because the only SNP mutation model available in DIYABC (Cornuet et al., 2014) did not provide a good match of the modelled data with the observed SNP data under any preliminary settings or final scenarios modelled (Supporting Information Figure S14a) and could not be improved because of the lack of user choices in the DIYABC interface (Cornuet et al., 2014).

3.3.3 | Biogeographic and demographic parameter estimates—The final microsatellite-based ABC simulations under the winning *Scenario 2* indicated a mutation rate of ~1.15E-4 (Tables 1 and Supporting Information Table S13), and a geometric distribution of change in repeat length of ~0.2 (Tables 1 and Supporting Information Table S13). With this scenario, divergence between the northwest and northeast was estimated at ~61–123 years ago, divergence between the south and the west was estimated at ~22–23 ka, and divergence between the northwest/northeast and the south/west was estimated at ~390–416 ka (Tables 1 and Supporting Information Table S13). Ne was modelled at an average of 15,082, and ranged from a low of 8,780 in the south cluster to a high of 23,800 in the northwest cluster (Tables 1 and Supporting Information Table S13).

In the MIGRAINE simulations, Ne averaged 4,933, and ranged from a low of 2,465 in the south to highs of ~6,000 in the northeast and northwest (Supporting Information Table S12). The only significant change in population size detected in the MIGRAINE simulations was in the west, with an estimated doubling from 2,530 to 5,632 since the time of coalescence of all microsatellite diversity observed in the contemporary sample (Supporting Information Table S12).

3.4 | Characterization of introgression in the admixture zones—The INTROGRESS analysis indicated greater divergence in allele frequency ($|\Delta p|$) between parental samples for admixture zones “a,” “c,” and “d” than for admixture zone “b” (Figure 4). Of the 33,057 SNPs that passed the filters for unlinked loci, 797, 54, 605 and 1,451 SNPs had $|\Delta p| > 0.5$, and only 89, 0, 20 and 108 SNPs had $|\Delta p|$ greater than 0.8 across admixture zones “a,” “b,” “c” and “d,” respectively (Figure 4a,c,d).

Estimates of $|\Delta p|$ and introgression suggested two general patterns. In the admixture zones “a” and “d,” most individuals were classified as parental types (h-index 0 or 1 with H_O < 0.5; insets of Figure 4a,d). In contrast, in the admixture zones “b” and “c,” individuals were classified as a wide range of advanced generation hybrids and backcrosses (h-index 0–1 and H_O < 0.5; insets of Figure 4a,d), suggesting many generations of interbreeding.

No matter the admixture zone, results indicate there are no first-generation (F1) hybrids in this study, as there is a gap in data points at the apex of each of the triangle plots (h-index of 0.5 and H_O > 0.5; insets of Figure 4). The single individual with an h-index close to 0.5 in admixture zone “a” had a very low estimated H_O value (much lower than 0.5, indicating it was not a true F1; Figure 4a). Instead, this outlier may be caused by incomplete characterization of the “parental” samples for this admixture zone, which is especially likely because of the geographic proximity of admixture zones “a,” “b” and “c” to the representative sampling site (09-UWA) in central Uganda (Figure 1). We also found some asymmetry in h-index scores in admixture zone “c,” where more individuals had h-index scores closer to the west (Lake Albert) genetic background (ranging from 0 to 0.5; inset of Figure 4c) than the other way around. This could reflect a biological reality, but might also be caused by sparse sampling across admixture zone “c.” More samples from this zone, especially to the east of the current samples, are needed to resolve the biological reality.

The distribution of F_{ST} values across each of the admixture zones (Supporting Information Figure S15) mirrored the distribution of $|\Delta p|$. However, after appropriate corrections for multiple testing, outlier tests failed to identify statistically significant F_{ST} outliers in any of the admixture zones (Supporting Information Figure S16; Table S15), presumably because of small sample sizes (Table 1).

4 | DISCUSSION

Our goal was to investigate the genetic structure, biogeographic context, and rates of hybridization and introgression in four *G. f. fuscipes* admixture zones in Uganda (Figure 1) to provide insight on the evolutionary and ecological processes important in diversification in this group of flies. We demonstrate that three of the four admixture zones were likely caused by hybridization during secondary contact, but that the third admixture zone “b” could instead be the outcome of ongoing divergence with gene flow, very recent divergence, or represent the remnants of an old collapsed hybrid zone. We find no F1 hybrids in any of the admixture zones and a dichotomous pattern of introgression (Figure 4) that could reflect weak reproductive barriers and early-stage speciation or nearly complete reproductive barriers and late-stage speciation. These findings provide preliminary support for the hypothesis that coupling of multiple isolating barriers may have accelerated the transition from early to late stages of speciation in *G. f. fuscipes* admixture zones.

4.1 | Genetic structure

The SNP-based PCA (Figure 3a), clustering analyses (Figure 3a, Supporting Information Figures S6a and S7) and FINESTRUCTURE analysis (Supporting Information Figure S8) generally supported $K = 2$, with clear substructure between the west (Lake Albert) and the south (Lake Victoria), and less consistent substructure between the northwest (Albert Nile) and the northeast (Lake Kyoga) samples (Figure 3). There was a stronger signal of substructure in the microsatellites than in the SNPs. Stronger substructure could be the result of higher mutation rates in the microsatellites than in the average SNP scored across the entire genome. This is supported by the AMOVA results (Supporting Information Table S5), which indicated that the majority of the microsatellite variance occurred within sites (84.3%), while the majority of the SNP variance occurred among watersheds (68.9%; Supporting Information Table S5). This high within-site variation in the microsatellites is likely a consequence of recent mutations that arose within the last hundreds of generations and have remained at low frequency in the individuals at their site of origin because of limited exposure to the effects of genetic drift and migration. In contrast, the high among-watershed variation in the SNPs is likely a consequence of some ancient mutations that have diverged in allele frequency because of prolonged exposure to the effects of genetic drift and migration within watersheds and have caused a lower genome-wide estimate of within-site variation in the SNPs relative to the microsatellites.

Genetic structure was slightly different in the mtDNA than the nuclear markers. The major genetic break between the northeast/northwest and west/south samples occurred much further south. One possible cause of discordance could be the spatial dynamics of a moving contact zone (Barton & Turelli, 2011; Currat et al., 2008). A moving hybrid zone can result in preferential introgression of organelle genes from the local into the expanding population because of the higher probability of a hybrid backcrossing with the expanding population and the smaller Ne of organelle than nuclear genomes (Currat et al., 2008). Thus, the incongruence of mtDNA and nuclear genetic structure could suggest that there have been recent range expansions northwards of the west and south clusters. This hypothesis is supported by migration estimates from previous studies (Beadell et al., 2010; Echodu et al., 2013) as well as the TREEMIX results in this study, which identified migration events from admixture zone “c” (Kafu River) into the northeast (Albert Nile; Supporting Information Figure S12a) and from the south (Lake Victoria) into the northeast (Lake Kyoga; Supporting Information Figure S12a). This hypothesis is also supported by the population expansion in the west supported by the demographic estimates from the MIGRAINE analyses (Albert Nile; Supporting Information Table S12), which could also indicate range expansion.

Another possible cause of discordance between the mitochondrial and the nuclear markers is asymmetrical isolation. One cause of asymmetrical isolation could be maternally inherited intracellular parasite *Wolbachia*, which can cause cytoplasmic incompatibility (CI) in *G. f. fuscipes* (Aksoy et al., 2013; Alam et al., 2012). Cytoplasmic incompatibility impacts the reproductive success and inheritance patterns because *Wolbachia* can prevent the formation of viable offspring in crosses of infected with uninfected individuals. *Wolbachia* infections in *G. f. fuscipes* are associated with the most common mtDNA haplotypes in Uganda (Symula et al., 2013), suggesting a potential fitness advantage of individuals with *Wolbachia.* If this is true, then *Wolbachia* would more strongly influence the inheritance patterns of the *G. f. fuscipes* mitochondrial genome than it would the nuclear genome. This implies that *Wolbachia* could have aided in the spread of the most common mitochondrial haplotypes beyond their original native range, thus accounting for discordance in the geographic placement of the genetic break in mitochondrial vs. nuclear markers in this study (Figure 3). In this case, since the boundary between northern and southern mitochondrial haplotypes (Figure 3c) falls further south than the boundary between the major nuclear genetic clusters (Figure 3a,b), *Wolbachia*-induced CI may have caused the southward spread of the mitochondrial haplotypes originating in northern Uganda. Since the resulting pattern would look much the same as the result of the alternative hypothesis (i.e., preferential introgression southwards of organelle genes during range expansion of nuclear genotypes northwards), more data are needed to test this hypothesis. Future research that examines the maternal transmission efficiency of *Wolbachia* in *G. f. fuscipes* and CI expression will provide insights into which mechanisms may have established the observed *G. f. fuscipes* mitochondrial/nuclear discordance in central Uganda.

4.2 | Genetic diversity and linkage and Hardy–Weinberg disequilibrium

Estimates from the SNP and the mitochondrial data sets generally showed the same patterns of genetic diversity and aligned with previous studies of the nuclear genome (Beadell et al., 2010; Echodu et al., 2013; Hyseni et al., 2012; Opiro et al., 2016, 2017). SNP diversity (nucleotide diversity) and microsatellite diversity (allelic richness) shared a general trend (Supporting Information Figure S9), with higher estimates in the north than in the south (Table 1; Supporting Information Table S10). This latitudinal gradient was observed and shown to be statistically significant by Beadell et al. (2010) and Opiro et al., (2017), and it has been suggested that this pattern reflects higher connectivity of northwestern populations of *G. f. fuscipes* to the rest of the *G. f. fuscipes* geographic distribution to the north in South Sudan and west in the Demographic Republic of the Congo (DRC) and beyond, where most of its range lies. This gradient of genetic diversity could also reflect larger population sizes and or more long-term demographic stability in the northwest than other parts of Uganda. The mtDNA results did not show a latitudinal gradient, but instead showed slightly elevated diversity in the admixture zones (Supporting Information Table S10). This is expected because maternal inheritance of mitochondrial genomes prevents recombination, making haplotype and nucleotide diversity additive in localities that contain descendants from multiple lineages.

Elevated LD and HWD, specifically heterozygote deficit, are a distinguishing feature of geographic regions where there are secondary contact and interbreeding between divergent lineages (Gompert et al., 2017). We found positive F_{IS} and elevated LD and HWD in admixture zones "a" (09-UWA), "d" (12-BN and 17-NB) and "c" (15-KAF and 16-JN), with the exception of the estimates made in 16-JN and 17-NB, perhaps because of insufficient sample size of just five individuals in these samples (Table 1). Positive F_{IS} and elevated LD and HWD indicate heterozygote deficit, suggesting these admixture zones contain descendants from a mix of divergent lineages and were formed by secondary contact. In contrast, admixture zone "b" (04-BOL, 05-CHU, 06-ACA, 07-APU and 08-OCA) showed a different pattern, with variable F_{IS} estimates (Table 1), suggesting admixture zone "b" contains descendants of a more homogenous ancestral background.

4.3 | Biogeographic and demographic history of the admixture zones

SNP-based results from TreeMix supported a history of divergence and subsequent gene flow in zones "a" (ancestral migration from west to northwest; Supporting Information Figure S12a), "c" (admixture between west and south; Supporting Information Figure S12b) and "d" (ancestral migration from south to northeast; Supporting Information Figure S12a), but there was no support for divergence across admixture zone "b" (between the northeast and northwest; Supporting Information Figure S12). This biogeographic pattern was further supported in microsatellite-based ABC simulations that most closely matched the observed data under *Scenario 2*, where divergence between the northwest (Albert Nile) and west (Lake Albert) samples occurred ~400 ka, and divergence between the west (Lake Albert) and south (Lake Victoria) occurred ~23 ka (Table 2). These modelled timings are similar to previous estimates based on mtDNA (Beadell et al., 2010). Simulations also suggested divergence between the northeast and northwest was very recent (during the 1900s; Table 2), making it possible that the signal of admixture could represent divergence with gene flow, the remnants of an old and now collapsed hybrid zone, or a signal of very recent population contraction and expansion caused by anthropogenic changes.

One possible cause for the weak signal of divergence between the northwest and northeast in the TreeMix results (Supporting Information Figure S12) and the recent timing of divergence in the winning ABC scenario (Table 2) could be that the genetic structuring found among populations in northern Uganda (Figure 2) was caused by recent demographic fluctuations rather than by sustained divergence. Introduction of the Rinderpest virus in the late 1800s led to large mammal die offs with up to 80% mortality (Carmichael, 1938), which likely reduced tsetse populations because they depend on blood meals for reproduction. Recent divergence between the northwest and northeast could also be due to more modern processes than the Rinderpest outbreak, such as habitat change that resulted from the insurgence of the Lord's Resistance Army in northern Uganda in the 1990s through 2009 (Van Acker, 2004), or very recent vector control efforts since 2009. The insurgence of the Lord's resistance army in northern Uganda displaced hundreds of thousands of people and their livestock, and left millions of acres of northwestern Uganda fallow for over a decade (Egeru et al., 2014; Gorsevski et al., 2012). This disruption of the agricultural system likely indirectly caused tsetse population crashes when livestock—and the blood meals they generally provide to tsetse flies—were removed. During this same period, vector control in northern Uganda ceased for almost 20 years (Egeru et al., 2014). Since 2009, vector control activities have resumed in northern Uganda, and there is evidence that these activities have left signals of population bottlenecks, especially in the northeast (Opiro et al., 2016, 2017). Thus, Uganda's recent history of sever livestock die offs during the Rinderpest outbreak and civil unrest could have caused tsetse population fluctuations that played a role in the apparent divergence between the northwest and northeast (Figure 2).

4.4 | Characterization of introgression in the admixture zones—INTROGRESS results indicated variable levels of divergence and hybridization in the four *G. f. fuscipes* admixture zones found in Uganda (Figure 1) that fell into one of two patterns: either displaying high genome-wide divergence (|Δ*p*|) and mostly parental types, or low genome-wide divergence (|Δ*p*|) and highly variable ancestry (Figure 4). In the first case, results suggest that admixture zones “a” and “d” have well-established reproductive barriers and that successful reproduction across the admixture zone was rare (insets of Figure 4a,d). In the second case, results suggest that admixture zones “b” and “c” have very few reproductive barriers and that successful reproduction across the admixture zone and among offspring from these crosses was common (insets of Figure 4b,c). The possible physiological, behavioural, mechanical or ecological mechanisms of reproductive barriers in admixture zones “a” and “d” are difficult to interpret from the current data (Coyne & Orr, 2004), especially since there is so little previous knowledge of the mating system and ecological interactions between the different lineages in secondary contact (Aksoy et al., 2013; Krafsur, Marquez, & Ouma, 2008). Nonetheless, the observed genome-wide pattern of low introgression levels (insets of Figure 4a,d) in these two admixture zones indicates strong reproductive barriers that block introgression in a large portion of the *G. f. fuscipes* nuclear genome. On the other hand, the observed genome-wide pattern of high introgression levels (insets of Figure 4b,c) in the other two admixture zones indicates incomplete reproductive barriers.

Possible drivers of the dichotomous pattern in introgression are the coupling of multiple exogenous and endogenous barriers across geographic space (i.e., the “coupling hypothesis” Bierne et al., 2011), or the coupling of prezygotic and postzygotic barriers (Butlin, 1989; Rundle & Nosil, 2005). The accumulation of associations of genotypes (coupling) in barrier loci across geographic space is a well-documented phenomenon that has been only recently recognized as an important component in speciation (Bierne et al., 2011). Associations among barrier loci are expected to create feedback between selection and gene flow (Gompert et al., 2017), and thereby either rapidly enhance the overall barriers to gene flow, or can lead to a sudden collapse of barriers (Bimova, MacHolan, Baird, Munclinger, & Dufkova, 2011; Flaxman, Wacholder, Feder, & Nosil, 2014; Gompert et al., 2017; Schilling et al., 2018; Servedio & Noor, 2003). At one extreme, assortative mating can cause build-up of associations between genotypes and can rapidly enhance reproductive barriers. At the other extreme, gene flow and non-assortative mating can cause disassociation among genotypes at barrier and non-barrier loci and can rapidly erode reproductive barriers. It remains an open question whether feedback caused by coupling of barrier loci is often a critical component of speciation. A prediction of this feedback hypothesis is that most hybrid zones would fit the description of either extreme: hybrid zones should either contain mostly parental types with little variability in introgression because of strong barriers to gene flow that have accumulated by this feedback, or they should contain hybrid types with highly variable ancestry and introgression because of their young age or history of barrier collapse.

Other possible explanations for the dichotomous pattern in introgression include reinforcement (Butlin, 1989) or ecological speciation (Rundle & Nosil, 2005). Although both alternatives include the coupling of multiple barriers, they also specifically involve both prezygotic and postzygotic factors, while the "coupling hypothesis" involves only postzygotic factors. Untangling which prezygotic and postzygotic factors are involved in reproductive isolation in this system will require more locus-by-locus analysis and breeding experiments. More locus-by-locus analysis to identify the genomic regions involved is not possible with the current results because low statistically power prevented detection of non-neutral introgression or F_{ST} outliers (Supporting Information Table S13; Gompert & Buerkle, 2010). Breeding experiments to distinguish prezygotic from postzygotic barriers —such as mating behaviour and egg–sperm incompatibilities—are also not possible until methods are developed to breed this species in an experimental setting. Future work that includes further molecular surveys with the same markers but with larger sample sizes and breeding experiments can elucidate the specific isolating mechanisms at play and allow tests of specific hypotheses on the contribution and contingency of prezygotic and postzygotic (intrinsic and extrinsic) isolating barriers.

4.5 | Conclusion and future directions

Our findings provide insight into the evolutionary processes that generated multiple *G. f. fuscipes* admixture zones in Uganda (Figure 1). This system provides a unique opportunity to draw from more than 10 years of study and multiple genetic markers in a species that displays a range of evolutionary histories and multiple stages of divergence. Results provide preliminary support for the "coupling hypothesis" and set the stage for a more detailed analysis of variable introgression across these admixture zones. Detailed analysis of introgression can inform the genes involved in reproductive isolation and can more directly test predictions of hypotheses like the “coupling hypothesis”. Specifically, a second prediction of the “coupling hypothesis” that hybrid zones should either display little or extensive *variability* in introgression among loci (Gompert et al., 2017) could not be tested with this data set because we did not have adequate samples sizes to identify significant outlier loci in the zones with evidence of secondary contact (admixture zones “a,” “c” and “d”). To fill this knowledge gap, we need denser and more even geographic sampling across the admixture zones, especially for zone “a” (Figure 1).

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

ACKNOWLEDGEMENTS

We acknowledge financial support from the Fogarty International Center (FIC) at the National Institutes of Health (NIH) Global Infectious Diseases Training Grant (award number D43TW007391), and from the Foundation for the NIH Research Project Grant Program (award numbers AI068932 and 5T32AI007404-24). We are grateful to Alfonse Okello, Calvin Owora and Constant Khizza of the Gulu University field team for help with sample collection, and to Andrea Gloria-Soria, Augustine W. Dunn, and Carol Mariani who are associated with the Caccone Lab at Yale University for the smooth transfer of technical information from previous related projects, and for help with sample processing and laboratory analysis. Finally, we thank Katharine Walter and Joshua Miller for their help with syntax and word choice in the final version of the manuscript.

DATA ACCESSIBILITY

ddRAD sequencing data, NCBI BioSample, PRJNA498097. Microsatellite genotypes, Dryad, https://doi.org/10.5061/dryad.98rf2f1. Mitochondrial DNA sequences, Genbank Accession nos. MK094112-MK094173

REFERENCES

Abila PP, Slotman MA, Parmakelis A, Dion KB, Robinson AS, Muwanika VB, … Caccone A (2008). High levels of genetic differentiation between Ugandan Glossina fuscipes fuscipes populations separated by Lake Kyoga. PLoS Neglected Tropical Diseases, 2(5), e242. 10.1371/journal.pntd.0000242 [PubMed: 18509474]

Adamack AT, & Gruber B (2014). PopGenReport: Simplifying basic population genetic analyses in R. Methods in Ecology and Evolution, 5(4), 384–387. 10.1111/2041-210X.12158

Aksoy S, Caccone A, Galvani AP, & Okedi LM (2013). *Glossina fuscipes* populations provide insights for Human African Trypanosomiasis transmission in Uganda. Trends in Parasitology, 29, 394–406. doi.org/10.1016/j.pt.2013.06.005 [PubMed: 23845311]

Alam U, Hyseni C, Symula RE, Brelsfoard C, Wu Y, Kruglov O, … Aksoy S (2012). Implications of microfauna-host interactions for trypanosome transmission dynamics in *Glossina fuscipes fuscipes* in Uganda. Applied and Environmental Microbiology, 78, 4627–4637. 10.1128/AEM.00806-12 [PubMed: 22544247]

Barton NH (2001). The role of hybridization in evolution. Molecular Ecology, 10, 551–568. 10.1046/j.1365-294x.2001.01216.x [PubMed: 11298968]

Barton NH, & Bengtsson BO (1986). The barrier to genetic exchange between hybridising populations. Heredity, 57(3), 357–376. doi.org/10.1038/hdy.1986.135 [PubMed: 3804765]

Barton NH, & Hewitt GM (1985). Analysis of Hybrid Zones. Annual Review of Ecology and Systematics, 16, 113–148.

Barton NH, & Turelli M (2011). Spatial waves of advance with bistable dynamics, cytoplasmic and genetic analogues of Allee effects. The American Naturalist, 178, E48–E75. 10.1086/661246

Beadell JS, Hyseni C, Abila PP, Azabo R, Enyaru JCK, Ouma JO, … Caccone A (2010). Phylogeography and Population Structure of Glossina fuscipes fuscipes in Uganda: Implications for Control of Tsetse. PLoS Neglected Tropical Diseases, 4(3), e636. 10.1371/journal.pntd.0000636 [PubMed: 20300518]

Benjamini Y, & Hochberg H (1995). Controlling the false discovery rate, a practical and powerful approach to multiple testing. Journal of the Royal Statistical Society. Series B (Methodological), 57(1), 289–300.

Bierne N, Bonhomme F, & David P (2003). Habitat preference and the marine-speciation paradox. Proceedings of the Royal Society of London Biology, 270, 1399–1406.

Bierne N, Welch J, Loire E, Bonhomme F, & David P (2011). The coupling hypothesis, why genome scans may fail to map local adaptation genes. Molecular Ecology, 20, 2044–2072. 10.1111/j.1365-294X.2011.05080.x [PubMed: 21476991]

Bimova BV, MacHolan M, Baird SJ, Munclinger P, & Dufkova P (2011). Reinforcement selection acting on the European house mouse hybrid zone. Molecular Ecology, 20, 2403–2424. 10.1111/j.1365-294X.2011.05106.x [PubMed: 21521395]

Bishop WW (1969). Pleistocene stratigraphy in Uganda. Entebbe, Uganda: The Geological Survey of Uganda.

Bouyer J, Balenghien T, Ravel S, Vial L, Sidibe I, Thevenon S, De Meeus T , (2009). Population sizes and dispersal pattern of tsetse flies: Rolling on the river? Molecular Ecology, 18(13), 2787–2797. 10.1111/j.1365-294X.2009.04233.x [PubMed: 19457176]

Brown JE, Komatsu KJ, Abila PP, Robinson AS, Okedi LM, Dyer N, … Caccone A (2008). Polymorphic microsatellite markers for the tsetse fly Glossina fuscipes fuscipes (Diptera: Glossinidae), a vector of human African trypanosomiasis. Molecular Ecology Resources, 8(6), 1506–1508. [PubMed: 21586090]

Butlin RK (1989). Reinforcement of premating isolation. In Otte D, & Endler JA (Eds.), Speciation and Its Consequences (pp. 158–179). Sunderland, MA: Sinauer Associates.

Carmichael J (1938). Rinderpest in African game. Journal of Comparative Pathology and Therapeutics, 51, 264–268. 10.1016/S0368-1742(38)80025-4

Catchen J, Hohenlohe PA, Bassham S, Amores A, & Cresko WA (2013). STACKS, an analysis tool set for population genomics. Molecular Ecology 22, 3124–3140. 10.1111/mec.12354 [PubMed: 23701397]

Cavalli-Sforza LL, & Piazza A (1975). Analysis of evolution, evolutionary rates, independence and treeness. Theoretical Population Biology, 8, 127–165. 10.1016/0040-5809(75)90029-5 [PubMed: 1198349]

Challier A, & Laveissiere C (1973). A new trap for capturing Glossina flies (Diptera, Muscidae), description and field trials. Cah ORSTOM Entomol Med Parasitol., 1973(11), 251–262.

Chapuis M-P, Plantamp C, Streiff R, Blondin L, & Piou C (2015). Microsatellite evolutionary rate and pattern in *Schistocerca gregaria* inferred from direct observation of germline mutations. Molecular Ecology 24, 6107–6119. [PubMed: 26562076]

Ciosi M, Masiga DK, & Turner CM (2014). Laboratory colonization and genetic bottlenecks in the tsetse fly *Glossina pallidipes*. PLoS Neglected Tropical Diseases., 8(2), e2697. [PubMed: 24551260]

Clement M, Snell Q, Walker P, Posada D, & Crandall K (2002). Estimating gene genealogies. Parallel Distrib Process Symp Int Proc., 2, 184.

Collins M, & Rawlins JE (2013). A transect for reproductive compatibility and evidence for a “hybrid sink” in a hybrid zone of hyalophora (Insecta: Lepidoptera: Saturniidae). Annals of Carnegie Museum, 82(2), 193–223.

Cornuet JM, Veyssier J, Pudlo P, Dehne-Garcia A, Gautier M, Leblois R, … Estoup A (2014). DIYABC v2.0: A software to make Approximate Bayesian Computation inferences about population history using Single Nucleotide Polymorphism, DNA sequence and microsatellite data. Bioinformatics, 30(8), 1187–1189. [PubMed: 24389659]

Coyne JA, & Orr HA (2004). Speciation. Sunderland, MA: Sinauer Associates.

Currat M, Ruedi M, Petit RJ, & Excoffier L (2008). The hidden side of invasions, Massive introgression by local genes. Evolution, 62, 1908–1920. 10.1111/j.1558-5646.2008.00413.x [PubMed: 18452573]

Danecek P, Auton A, Abecasis G, Albers CA, & Banks E (2011). The variant call format and VCFTOOLS. Bioinformatics, 27, 2156–2158. 10.1093/bioinformatics/btr330 [PubMed: 21653522]

Danley PD, Husemann M, Ding B, Dipietro LM, Beverly EJ, & Peppe DJ (2012). The impact of the geologic history and paleoclimate on the diversification of east African cichlids. International Journal of Evolutionary Biology, 2012, 574851. 10.1155/2012/574851 [PubMed: 22888465]

Davison A, Chiba S, Barton NH, & Clarke B (2005). Speciation and gene flow between snails of opposite chirality. PLoS Biology, 3, e282. 10.1371/journal.pbio.0030282 [PubMed: 16149849]

Dunn OJ (1961). Multiple comparisons among means. Journal of American Statistical Association, 56, 52–64. 10.1080/01621459.1961.10482090

Earl DA, & Von Holdt BM (2012). STRUCTURE HARVESTER: A website and program for visualizing STRUCTURE output and implementing the Evanno method. Conservation Genetics Resources, 4(2), 359–361.

Echodu R, Sistrom M, Hyseni C, Enyaru J, Okedi L, Aksoy S, & Caccone A (2013). Genetically distinct *Glossina fuscipes fuscipes* populations in the Lake Kyoga region of Uganda and its relevance for human African trypanosomiasis. BioMed Research International, 2013, 614721. [PubMed: 24199195]

Egeru A, Wasonga O, Kyagulanyi J, Majaliwa GM, MacOpiyo L, & Mburu J (2014). Spatio-temporal dynamics of forage and land cover changes in Karamoja sub-region. Uganda. Pastoralism, 4(1), 1–21. 10.1186/2041-7136-4-6

Ersts PJ (2017). Geographic distance matrix generator. American museum of natural history, center for biodiversity and conservation. Retrieved from http://biodiversityinformat-ics.amnh.org/open_source/gdmg

Evanno G, Regnaut S, & Goudet J (2005). Detecting the number of clusters of individuals using the software STRUCTURE: A simulation study. Molecular Ecology, 14, 2611–2620. 10.1111/j.1365-294X.2005.02553.x [PubMed: 15969739]

Excoffier L, & Lischer HEL (2010). ARLEQUIN suite ver 3.5: A new series of programs to perform population genetics analyses under Linux and Windows. Molecular Ecology Resources, 10, 564–567. 10.1111/j.1755-0998.2010.02847.x [PubMed: 21565059]

Excoffier L, Estoup A, & Cornuet JM (2005). Bayesian analysis of an admixture model with mutations and arbitrarily linked markers. Genetics, 169, 1727–1738. [PubMed: 15654099]

Fagundes NJR, Ray N, Beaumont M, Neuenschwander S, & Salzano SM (2007). Statistical evaluation of alternative models of human evolution. Proceedings of the National Academy of Sciences of the United States of America, 104(45), 17614–17619. 10.1073/pnas.0708280104 [PubMed: 17978179]

Falush D, VanDorp L, & Lawson D (2016). A tutorial on how (not) to over-interpret STRUCTURE/ADMIXTURE bar plots. bioRxiv, 066431.

Felsenstein J (1982). How can we infer geography and history from gene frequencies? Journal of Theoretical Biology, 96(1), 9–20. 10.1016/0022-5193(82)90152-7 [PubMed: 7109659]

Flaxman SM, Wacholder AC, Feder JL, & Nosil P (2014). Theoretical models of the influence of genomic architecture on the dynamics of speciation. Molecular Ecology, 23, 4074–4088. doi.org/10.1111/mec.12750 [PubMed: 24724861]

Frantz AC, Cellina S, Krier A, Schley L, & Burke T (2009). Using spatial Bayesian methods to determine the genetic structure of a continuously distributed population, Clusters or isolation by distance? Journal of Applied Ecology, 46, 493–505. 10.1111/j.1365-2664.2008.01606.x

Funk DJ, Nosil P, & Etges WJ (2006). Ecological divergence exhibits consistently positive associations with reproductive isolation across disparate taxa. Proceedings of the National Academy of Sciences, 103, 3209–3213.

Garrick RC, Benavides E, Russello MA, Hyseni C, Edwards DL, Gibbs JP, … Caccone A (2014). Lineage fusion in Galápagos giant tortoises. Molecular Ecology, 23, 5276–5290. 10.1111/mec.12919 [PubMed: 25223395]

Garza J, & Williamson E (2001). Detection of reduction in population size using data from microsatellite loci. Molecular Ecology, 10, 305–318. [PubMed: 11298947]

Giraldo-Calderon GI, Emrich SJ, MacCallum RM, Maslen G, & Dialynas E (2014). VectorBase, an updated bioinformatics resource for invertebrate vectors and other organisms related with human diseases. Nucleic Acids Research, 43, D707–D713. [PubMed: 25510499]

Gompert Z, & Buerkle CA (2009). A powerful regression-based method for admixture mapping of isolation across the genome of hybrids. Molecular Ecology, 18, 1207–1224. 10.1111/j.1365-294X.2009.04098.x [PubMed: 19243513]

Gompert Z, & Buerkle CA (2010). INTROGRESS, a software package for mapping components of isolation in hybrids. Molecular Ecology Resources, 10, 378–384. 10.1111/j.1755-0998.2009.02733.x [PubMed: 21565033]

Gompert Z, Mandeville EG, & Buerkle CA (2017). Analysis of population genomic data from hybrid zones. Annual Review of Ecology, Evolution, and Systematics, 48, 207–229. 10.1146/annurev-ecolsys-110316-022652

Gorsevski V, Kasischke E, Dempewolf J, Loboda T, & Grossmann F (2012). Analysis of the Impacts of armed conflict on the Eastern Afromontane forest region on the South Sudan – Uganda border using multitemporal Landsat imagery. Remote Sensing of Environment, 118, 10–20. 10.1016/j.rse.2011.10.023

Goudet J, & Jombart T (2015). Hierfstat: Estimation and tests of hierarchical F-Statistics. R package version, 04–22. Retrieved from https://CRAN.R-project.org/package=hierfstat

Goudet J, Raymond M, & de Meeüs TRF (1996). Testing differentiation in diploid populations. Genetics, 144, 1933–1940. [PubMed: 8978076]

Goudet J (1995). FSTAT (Version 1.2): A Computer Program to Calculate F-Statistics. The Journal of Heredity, 86(6), 485–486.

Grant BR, & Grant PR (2008). Fission and fusion of Darwin's finches populations. Philosophical Transactions of the Royal Society of London Biological Sciences, 363, 2821–2829. [PubMed: 18508750]

Grant PR, & Grant BR (2006). Species before speciation is complete. Annals of the Missouri Botanical Garden, 93, 94–102.

Gross BL, & Rieseberg LH (2005). The ecological genetics of homoploid hybrid speciation. Journal of Heredity, 96, 241–252. [PubMed: 15618301]

Gruber B, & Adamack AT (2015). Landgenreport, a new R function to simplify landscape genetic analysis using resistance surface layers. Molecular Ecology Resources, 15(5), 1172–1178. 10.1111/1755-0998.12381 [PubMed: 25644761]

Hyseni C, Kato AB, Okedi LM, Masembe C, Ouma JO, Aksoy S, & Caccone A (2012). The population structure of Glossina fuscipes fuscipes in the Lake Victoria basin in Uganda, implications for vector control. Parasites & Vectors, 5, 222. 10.1186/1756-3305-5-222 [PubMed: 23036153]

Jombart T (2008). ADEGENET: A R package for the multivariate analysis of genetic markers. Bioinformatics, 24, 1403–1405. 10.1093/bioinformatics/btn129 [PubMed: 18397895]

Jombart T, & Ahmed I (2011). ADEGENET 1.3-1: New tools for the analysis of genome-wide SNP data. Bioinformatics, 27(21), 3070–3071. doi:10.1093/bioinformatics/btr521 [PubMed: 21926124]

Kamvar ZN, Brooks JC, & Grünwald NJ (2015). Novel R tools for analysis of genome-wide population genetic data with emphasis on clonality. Front. Genet, 6, 208. 10.3389/fgene.2015.00208 [PubMed: 26113860]

Kamvar ZN, Tabima JF, & Grünwald NJ (2014). POPPR: An R package for genetic analysis of populations with clonal, partially clonal, and/or sexual reproduction. PeerJ, 2, e281. 10.7717/peerj.281 [PubMed: 24688859]

Kleine FK (1909). Positiv infektionversuche mit *Trypanosoma brucei* durch *Glossina palpalis*. Detsch. Med. Wochenschr, 35, 469–470.

Kopelman NM, Mayzel J, Jakobsson M, Rosenberg NA, & Mayrose I (2015). Clumpak, A program for identifying clustering modes and packaging population structure inferences across K. Molecular Ecology Resources, 15, 1179–1191. [PubMed: 25684545]

Krafsur ES (2009). Tsetse flies: Genetics, evolution, and role as vectors. Infection, Genetics and Evolution, 9, 124–141. 10.1016/j.meegid.2008.09.010

Krafsur ES, Marquez JG, & Ouma JO (2008). Structure of some East African *Glossina fuscipes fuscipes* populations. Medical and Veterinary Entomology, 22, 222–227. [PubMed: 18816270]

Langmead B, & Salzberg SL (2012). Fast gapped-read alignment with Bowtie 2. Nature Methods, 9, 357–359. 10.1038/nmeth.1923 [PubMed: 22388286]

Lawson DJ, Hellenthal G, Myers S, & Falush D (2012). Inference of population structure using dense haplotype data. PLoS Genetics, 8(1), 1–16.

Leblois R, Pudlo P, Neron J, Bertaux F, Beeravolu CR, Vitalis R, & Rousset F (2014). Maximum-likelihood inference of population size contractions from microsatellite data. Molecular Biology and Evolution, 31, 2805–2823. 10.1093/molbev/msu212 [PubMed: 25016583]

Li H, Handsaker B, Wysoker A, Fennell T, Ruan J, Homer N, … Durbin R (2009). The sequence alignment/map format and SAMTOOLS. Bioinformatics, 25, 2078–2079. 10.1093/bioinformatics/btp352 [PubMed: 19505943]

Librado J, & Rojas P (2009). DNASP v5: A software for comprehensive analysis of DNA polymorphism data. Bioinformatics, 25, 1451–1452. 10.1093/bioinformatics/btp187 [PubMed: 19346325]

Lischer HEL, & Excoffier L (2012). PGDSPIDER, an automated data conversion tool for connecting population genetics and genomics programs. Bioinformatics, 28, 298–299. [PubMed: 22110245]

Mantel N (1967). The detection of disease clustering and a generalized regression approach. Cancer Research, 27, 209–220. [PubMed: 6018555]

Martin SH, Dasmahapatra KK, Nadeau NJ, Salazar C, & Walters JR (2013). Genome-wide evidence for speciation with gene flow in *Heliconius* butterflies. Genome Research, 23, 1817–1828. [PubMed: 24045163]

Meirmans PG (2012). The trouble with isolation by distance. Molecular Ecology, 21, 2839–2846. 10.1111/j.1365-294X.2012.05578.x [PubMed: 22574758]

Mullen SP, Dopman EB, & Harrison RG (2008). Hybrid zone origins, species boundaries, and the evolution of wing-pattern diversity in a polytypic species complex of North American admiral butterflies (Nymphalidae: Limenitis). Evolution, 62, 1400–1417. 10.1111/j.1558-5646.2008.00366.x [PubMed: 18331459]

Nei M (1987). Molecular Evolutionary Genetics. New York, NY: Columbia University Press.

Opiro R, Saarman NP, Echodu R, Opiyo EA, Dion K, Halyard A, … Caccone A (2016). Evidence of temporal stability in allelic and mitochondrial haplotype diversity in populations of Glossina fuscipes fuscipes (Diptera, Glossinidae) in northern Uganda. Parasites & Vectors, 9, 1–12. 10.1186/s13071-016-1522-5 [PubMed: 26728523]

Opiro R, Saarman NP, Echodu R, Opiyo EA, Dion K, Halyard A, … Caccone A (2017). Genetic diversity and population structure of the tsetse fly Glossina fuscipes fuscipes (Diptera, Glossinidae) in Northern Uganda, I implications for vector control. PLoS Neglected Trop/ca/ Diseases, 11(4), e0005485. 10.1371/journal.pntd.0005485

Peery MZ, Kirby R, Reid BN, Stoelting R, Doucet-Beer E, Robinson S, … Palsboll PJ (2012). Reliability of genetic bottleneck tests for detecting recent population declines. Molecular Ecology, 21, 3403–3418. 10.1111/j.1365-294X.2012.05635.x [PubMed: 22646281]

Pembleton LW, Cogan NOI, & Forster JW (2013). StAMPP: An R package for calculation of genetic differentiation and structure of mixed-ploidy level populations. Molecular Ecology Resources, 13, 946–952. 10.1111/1755-0998.12129 [PubMed: 23738873]

Peterson BK, Weber JN, Kay EH, Fisher HS, & Hoekstra HE (2012). Double digest RADseq: An inexpensive method for de novo SNP discovery and genotyping in model and non-model species. PloS ONE, 7(5), e37135. 10.1371/journal.pone.0037135 [PubMed: 22675423]

Pickrell JK, & Pritchard JK (2012). Inference of population splits and mixtures from genome-wide allele frequency data. PLoS Genetics, 8(11), e1002967. 10.1371/journal.pgen.1002967 [PubMed: 23166502]

Pritchard JK, & Stephens PD (2000). Inference of population structure using multilocus genotype data. Genetics, 155, 945–959. [PubMed: 10835412]

Purcell S, Neale B, Todd-Brown K, Thomas L, Ferreira MAR, Bender D, … Sham PC (2007). PLINK: A tool set for whole-genome association and population-based linkage analyses. American Journal of Eluman Genetics, 81, 559–575. 10.1086/519795

R Core Team (2016). R: A language and environment for statistical computing. Vienna, Australia: R Foundation for Statistical Computing.

Riley WA, & Johannsen OA (1932). Medical entomology (1st ed.). New York, PA: The Maple Press Co.

Rundle HD, & Nosil P (2005). Ecological speciation. Ecology Letters, 8, 336–352.

Saarman NP, Burak M, Opiro R, Hyseni C, Echodu R, Dion K, … Caccone A (2018). A spatial genetics approach to inform vector control of tsetse flies (Glossina fuscipes fuscipes) in Northern Uganda. Ecology and Evolution, 8(11), 5336–5354. 10.1002/ece3.4050 [PubMed: 29938057]

Scheet P, & Stephens M (2006). A fast and flexible statistical model for large-scale population genotype data, applications to inferring missing genotypes and haplotypic phase. Am. J. Hum. Genetics, 78, 629–644. 10.1086/502802 [PubMed: 16532393]

Schilling MP, Mullen SP, Kronforst M, Safran RJ, Nosil P, Feder JL, … Flaxman SM (2018). Transitions from single- to multi-locus processes during speciation. Genes, 9(6), 274–300. 10.3390/genes9060274

Schluter D (2000). The ecology of adaptive radiation. Oxford, UK: Oxford University Press.

Servedio MR, & Noor MAF (2003). The role of reinforcement in speciation, theory and data. Annual Review of Ecology Evolution and Systematics, 34, 339–364.

Soria A, Dunn WA, Telleria EL, Evans BR, Okedi L, Echodu R, … Caccone A (2016). Patterns of genome-wide variation in *Glossina fuscipes fuscipes* tsetse flies from Uganda. G3:Genes|Genomes|Genet/cs, 6, 1573–1584.

Soria A, Dunn WA, Yu X, Vigneron A, Lee K-Y, Li M, … Caccone A (2018). Uncovering genomic regions associated with *Trypanosoma* infections in wild populations of the tsetse fly *Glossina fuscipes*. G3:Genes|genomes|genetics, 8(3):887–897. 10.1534/g3.117.300493 [PubMed: 29343494]

Sukumaran J, & Mark TH (2010). DendroPy, A Python library for phylogenetic computing. Bioinformatics, 26, 1569–1571. 10.1093/bioinformatics/btq228 [PubMed: 20421198]

Symula RE, Alam U, Brelsfoard C, Wu Y, Echodu R, Okedi LM, … Caccone A (2013). Wolbachia association with the tsetse fly, *Glossina fuscipes fuscipes*, reveals high levels of genetic diversity and complex evolutionary dynamics. BMC Evolutionary Biology, 13(1), 31. 10.1186/1471-2148-13-31 [PubMed: 23384159]

Talbot MR, & Williams MAJ (2009). Cenozoic evolution of the Nile Basin. In Dumont HJ (Ed.), The Nile: Origin, Environments, Limnology and Human Use (pp. 37–60). Berlin, Germany: Springer Science & Business Media.

Turelli M, Barton NH, & Coyne JA (2001). Theory and speciation. Trends in Ecology & Evolution, 16, 330–343. 10.1016/S0169-5347(01)02177-2 [PubMed: 11403865]

Van Acker F (2004). Uganda and the Lord’s Resistance Army, the new order no one ordered. African Affairs, 103(412), 335–357. 10.1093/afraf/adh044

Webb WC, Marzluff JM, & Omland KE (2011). Random interbreeding between cryptic lineages of the common raven, evidence for speciation in reverse. Molecular Ecology, 20, 2390–2402. doi.org/10.1111/j.1365-294X.2011.05095.x [PubMed: 21518060]

Weir BS, & Cockerham CC (1984). Estimating F-statistics for the analysis of population structure. Evolution, 38, 1358. 10.2307/2408641 [PubMed: 28563791]

Whitlock MC, & Lotterhos KE (2015). Reliable detection of loci responsible for local adaptation: Inference of a null model through trimming the distribution of FST. The American Naturalist, 186, S24–S36. 10.1086/682949

Williams MAJ, Adamson D, Prescott JR, & Williams FM (2003). New light on the age of the White Nile. Geology, 31(11), 1001–1004. 10.1130/G19801.1

Wright S (1951). The genetical structure of populations. Annals of Eugenics, 15, 323–354. [PubMed: 24540312]

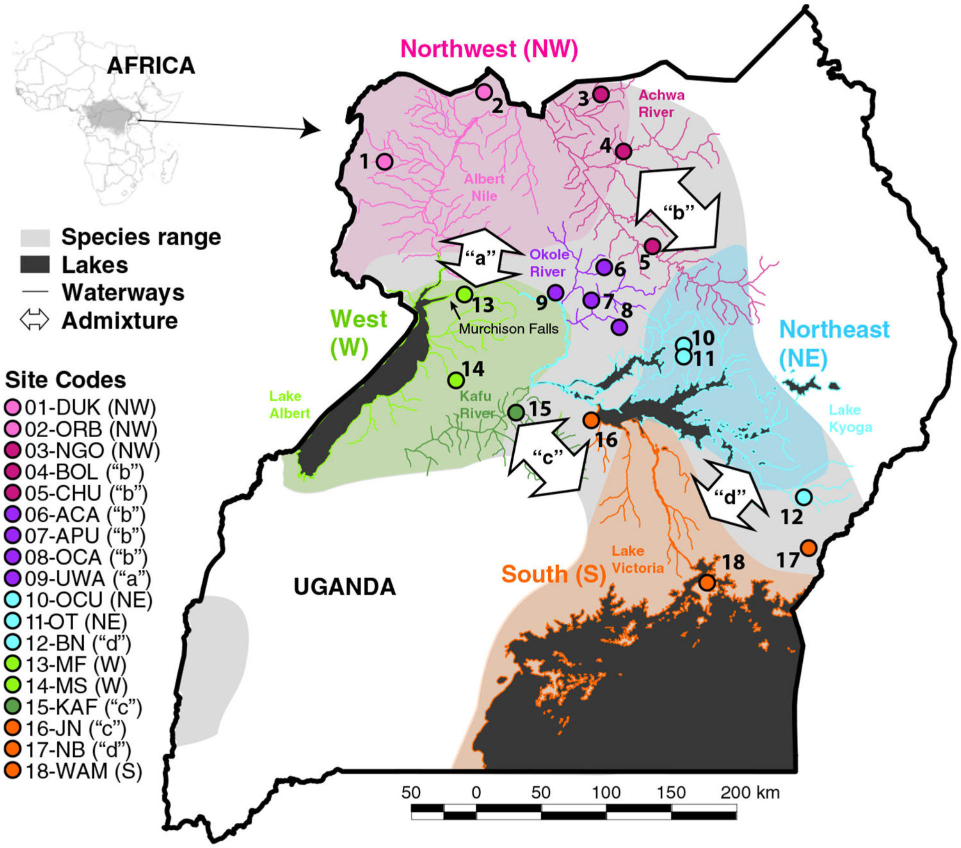

FIGURE 1.

Map of Uganda and the location of the 18 sampling sites of *Glossina fuscipes fuscipes* used in this study. Each sampling site marker indicates its placement relative to the distribution of the four pure genetic backgrounds (northwest: shaded pink, northeast: shaded blue, west: shaded green, south: shaded orange), and is color-coded by the watershed it falls within (Albert Nile: pink, Achwa River: magenta, Okole River: purple, Lake Kyoga: blue, Lake Albert: light green, Kafu River: dark green, Lake Victoria: orange). The species distribution of G. f. fuscipes beyond the distribution of the four pure genetic backgrounds is shown in light grey, and admixture zones ("a", "b", "c" and "d") within this shading are indicated with arrows pointing to the genetic backgrounds in putative secondary contact in each zone.

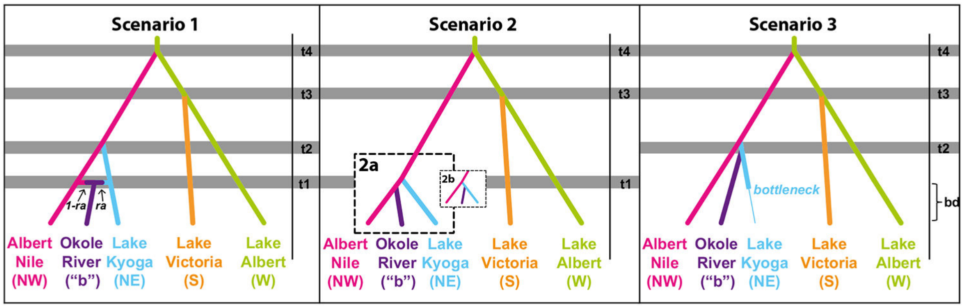

FIGURE 2.

Schematic of competing scenarios 1–3 tested in DIYABC (Cornuet et al., 2014) to evaluate the relative probability of (a) divergence followed by secondary contact (Scenario 1), (b) ongoing gene flow or very recent and incomplete divergence (Scenario 2) with the alternative topologies 2a and 2b, (c) and the role of a bottleneck (shown with the transition from a thick to thin line; Scenario 3) in shaping the patterns of genetic variation observed in G. f. fuscipes populations in Uganda. Each panel shows the topology and population fluctuations specified in the scenario, wherein Ne was free to vary 100–30,000, or 100–5,000 during a bottleneck. Time priors are shown in shading (not to scale) and are labeled on the right of each scenario (t4 = 340–360 ka, t3 = 20–40 ka, t2 = 10–20 ka, t1 = 5–135 years ago). "ra" refers to rate of admixture, and was allowed to vary from 0 to 1. "bd" refers to the bottleneck duration, and was allowed to vary from 1.25 to 5,000 years ago (Supporting Information Table S3), which corresponds to a final desiccation event that was thought to occur in the Lake Kyoga region (Danley et al., 2012).

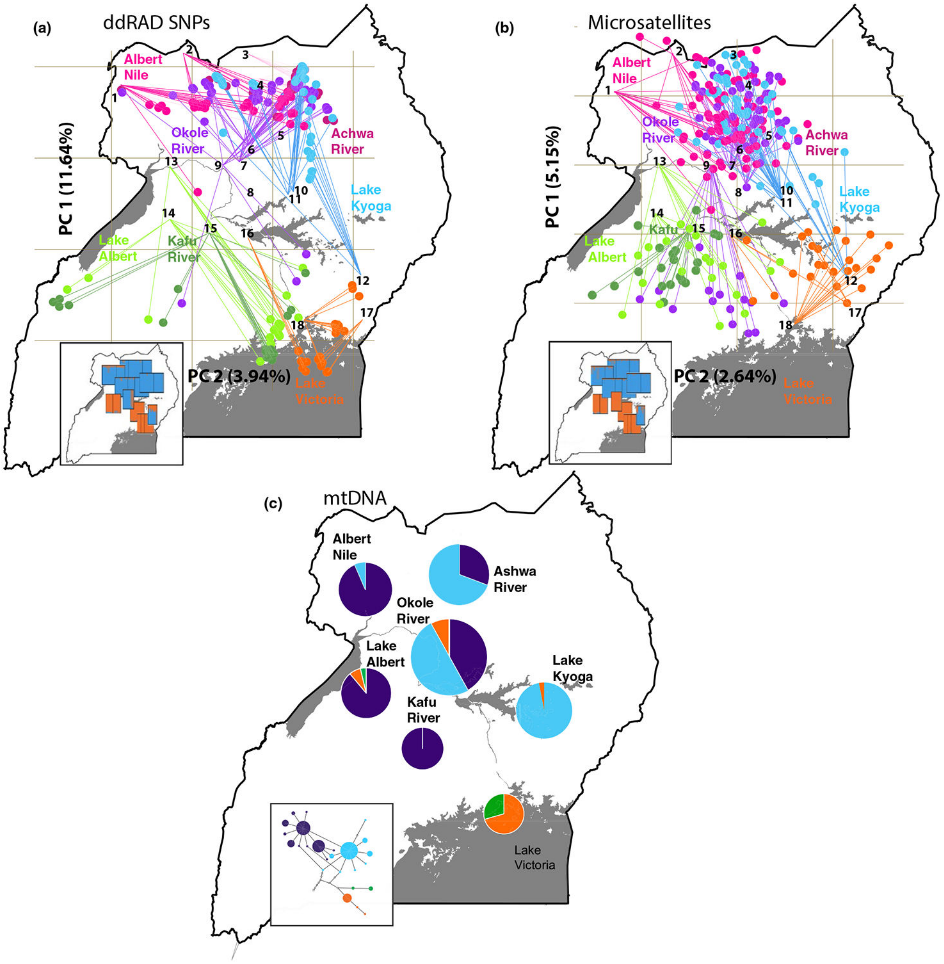

FIGURE 3.

Genetic structure of *Glossina fuscipes fuscipes* in Uganda, including (a) principle components analysis (PCA) plots for 55,267 ddRAD SNPs, (b) PCA plots for 16 microsatellites, and (c) the frequency of groups of related mtDNA COI-COII haplotypes. In (a) and (b) each individual is indicated with a point connected by a line to the site of origin, and are colored by watershed of origin (Albert Nile in pink, Ashwa River in magenta, Okole River in purple, Lake Kyoga in blue, Lake Albert in light green, Kafu River in dark green and Lake Victoria in orange; see Supporting Information Table S4) for PCA plots of PC1-PC4 without the map), and the insets display STRUCTURE version 2.3.4 (Pritchard & Stephens, 2000) results at K=2 (see Supporting Information Table S6) for full size). In (c), groups of related haplotypes are shown in the same colors (Haplogroup A in purple, Haplogroup B in blue, Haplogroup C in orange, and Haplogroup D in green), and the inset displays the TCS haplotype network (see Supporting Information Table S11) for full size).

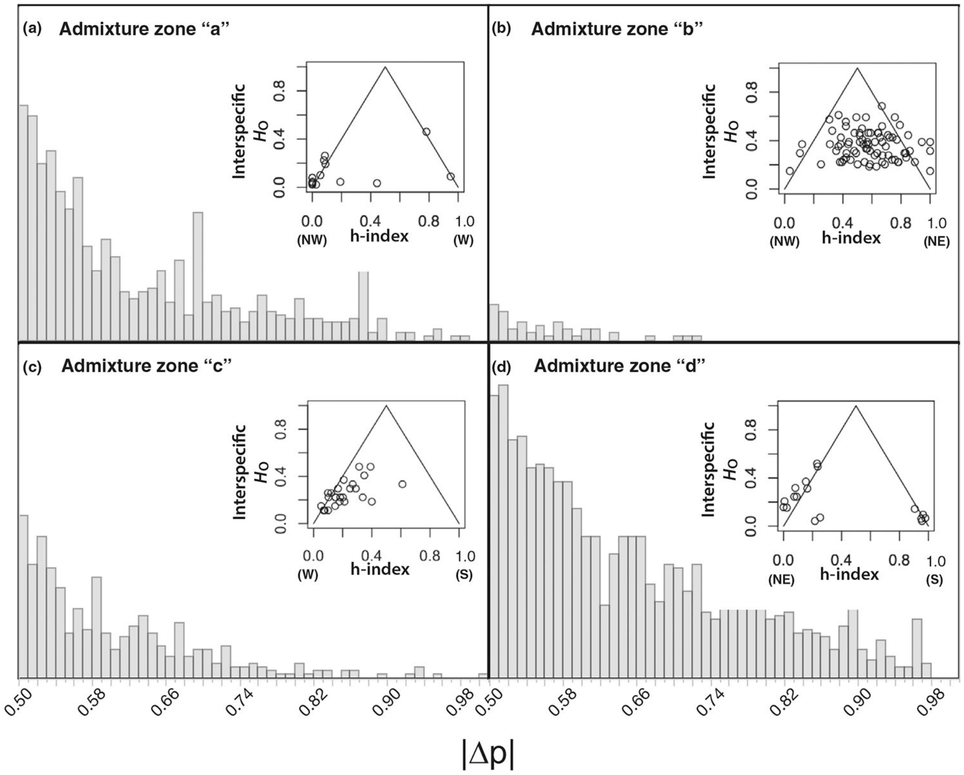

FIGURE 4.

Characterization of the *Glossina fuscipes fuscipes* admixture zones in Uganda using INTROGRESS v. 1.2.3 (Gompert & Buerkle, 2009) for (a) admixture zone “a”, (b) admixture zone “b”, (c) admixture zone “c”, and (d) admixture zone “d”. The main plots show the histograms of the absolute allele frequency difference (|Δ*p*|) estimated from 33,057 unlinked ddRAD SNPs. Insets show triangle plots of the “interspecific” heterozygosity (H_O) against “hybrid index” (h-index) estimated with a subset of SNPs with high |Δ*p*| (0.8 for admixture zones “a”, “c”, and “d”; 0.5 for admixture zone “b”), wherein first generation hybrids appear at the apex, advanced generation hybrids occur in the center, and parental types occur in the bottom left/right, with the genetic cluster that represents that parental type indicated with abbreviations: northwest (NW), northeast (NE), west (W), south (S).

TABLE 1

Sample details and summary statistics for the 55,267 ddRAD SNPs for each sample organized and summarized by watershed

	Geographic region (admixture zone)	*N*	Pi	H_O	H_E	P_{LD}	P_{HWD}	F_{IS}
Albert Nile								
01-DUK$^{\dagger}_{\text{-NW}}$	Northwest	20	0.177	0.207	0.172	0.200	0.121	−0.169[1,2,3]
02-ORB$^{\dagger}_{\text{-NW}}$	Northwest	12	0.175	0.163	0.167	0.186	0.037	0.081[1,2,3]
		32	0.179	0.189	0.176	0.192	0.117	−0.050[1,2,3]
Achwa River								
03-NGO	Northwest	12	0.169	0.139	0.162	0.176	0.093	0.206[1,2,3]
04-BOL	Northwest/Northeast (“b”)	20	0.162	0.180	0.158	0.174	0.051	−0.104[1,2,3]
05-CHU	Northwest/Northeast (“b”)	13	0.145	0.133	0.139	0.140	0.042	0.095[1,2,3]
		45	0.162	0.155	0.160	0.138	0.141	0.055[1,2,3]
Okole River								
06-ACA	Northwest/Northeast (“b”)	18	0.163	0.175	0.158	0.167	0.053	−0.071[1,2,3]
07-APU	Northwest/Northeast (“b”)	10	0.170	0.190	0.161	0.199	0.026	−0.115[1,2,3]
08-OCA	Northwest/Northeast (“b”)	20	0.163	0.172	0.158	0.159	0.047	−0.052[1,2,3]
09-UWA	Northwest/West (“a”)	15	0.185	0.153	0.178	0.222	0.057	0.190[1,2,3]
		63	0.172	0.171	0.170	0.158	0.178	0.014[1,2,3]
Lake Kyoga								
10-OCU$^{\dagger}_{\text{-NE}}$	Northeast	20	0.146	0.148	0.142	0.132	0.048	0.000
11-OT$^{\dagger}_{\text{-NE}}$	Northeast	14	0.143	0.142	0.137	0.121	0.024	0.009[1,2,3]
12-BN	Northeast/South (“d”)	12	0.161	0.154	0.154	0.144	0.040	0.059[1,2,3]
		46	0.156	0.148	0.155	0.130	0.103	0.067[1,2,3]
Lake Albert								
13-MF$^{\dagger}_{\text{-W}}$	West	10	0.155	0.147	0.147	0.185	0.021	0.063[1,2,3]

	Geographic region (admixture zone)	***N***	**Pi**	***H*** **O**	***H*** **E**	***P*** **LD**	***P*** **HWD**	***F*** **IS**
14-MS$^{\dagger\text{-w}}$	West	11	0.160	0.129	0.153	0.134	0.085	$0.230^{1,2,3}$
		21	0.165	0.137	0.160	0.164	0.120	$0.186^{1,2,3}$
Kafu River								
15-KAF	West/South (“c”)	20	0.158	0.129	0.154	0.114	0.117	$0.212^{1,2,3}$
Lake Victoria								
16-JN	West/South (“c”)	5	0.108	0.106	0.096	0.056	0.004	$0.031^{1,2,3}$
17-NB	Northeast/South (“d”)	5	0.100	0.101	0.089	0.048	0.004	0.003
18-WAM$^{\dagger\text{-s}}$	South	14	0.132	0.137	0.127	0.074	0.020	$-0.035^{1,2,3}$
		24	0.134	0.122	0.131	0.087	0.043	$0.101^{1,2,3}$

Geographic region of origin and assignment to admixture zones “a”–“d” if applicable (Beadell et al., 2010; Echodu et al., 2013; Opiro et al., 2017), sample size (*N*) and statistics from STACKS v.1.34 software (Catchen et al., 2013) including sample-wide nucleotide diversity in the 55,267 representative SNPs (Pi; one SNP per ddRAD fragment), sample-wide observed heterozygosity (H_O), and sample-wide expected heterozygosity (H_E), statistics from PLINK version 1.9 (Purcell et al., 2007) including the proportion of locus pairs in linkage disequilibrium in this sample (P_{LD}) and the proportion of loci in Hardy–Weinberg disequilibrium in this sample (P_{HWD}), and sample-wide inbreeding fixation index (F_{IS}) calculated in the “HIERFSTAT” R package (Goudet and Jombart, 2015). F_{IS} values are marked with ^{1} if significantly different from zero with $p < 0.02$ based on 1,000 bootstrap replicates, with ^{1,2} if $p < 0.02$ after Benjamini–Hochberg correction for multiple testing, and with ^{1,2,3} if $p < 0.02$ after Bonferroni correction for multiple testing (Supporting Information Table S4). Samples that were used to represent the four pure genetic backgrounds are marked † followed by the abbreviation of the genetic cluster it represents (northwest: NW, northeast: NE, west: W, south: S).

TABLE 2

Results from the ABC analysis on northern Uganda implemented in DIYABC (Cornuet et al., 2014) based on the main topology 2a (Figure 3). (a) Posterior probabilities of competing scenarios by the polychotomic weighted logistic regression method (Cornuet et al., 2008), and 95% confidence interval for the microsatellite data set. (b) Parameter estimates based on the winning *Scenario 2* including the mean, 95% confidence interval (CI), median and mode of the posterior distribution, and the description of each parameter estimated

(a)

Data set	Scenario 1	Scenario 2	Scenario 3	Error	Best scenario
Microsatellites	0.000 [0.000–0.000]	0.999 [0.999–1.000]	0.000 [0.000–0.000]	0.000	2

(b)

Parameter	Mean	CI	Median	Mode	Description
t1 (yrs)	95	40–128	102	118	Northwest/admixture zone “b” divergence
t2 (yrs)	123	101–135	126	135	Northwest/northeast divergence
t3 (yrs)	22,625	20,125–31,000	21,125	20,125	West/south divergence
t4 (yrs)	416,250	353,750–457,500	423,750	457,500	West/northwest divergence
N1	23,800	16,800–28,700	24,300	25,500	Ne of northwest lineage
N2	14,000	6,860–28,400	13,100	9,780	Ne of admixture zone “b”
N3	16,600	2,950–22,800	16,100	14,900	Ne of northeast lineage
N4	12,600	5,020–21,600	12,000	11,900	Ne of west lineage
N5	11,500	3,700–15,900	11,100	10,700	Ne of south lineage
μ	1.13E–04	1.00–1.58E–04	1.06E–04	1.00E–04	Mutation rate
p	0.1940	0.1340–0.2780	0.1930	0.1850	Geometric distribution of change in repeat length

Time estimates (t1, t2, t3 and t4) are expressed in years (yrs), effective population size estimates (N1, N2, N3, N4, N5) are expressed in number of breeding individuals. We also report estimates of the microsatellite mutation rate (μ) and geometric distribution of change in repeat length (*P*). See Figure 2 for schematics of the three scenarios and Supporting Information Table S3 for details of the priors used.