Don't throw out the sympatric species with the crater lake water: fine-scale investigation of introgression provides weak support for functional role of secondary gene flow in one of the clearest examples of sympatric speciation

Emilie J. Richards[1*], Jelmer W. Poelstra[1,2], and Christopher H. Martin[1]

[1] Biology Department, University of North Carolina at Chapel Hill, Chapel Hill, North Carolina, United States of America

[2] Biology Department, Duke University, Durham, North Carolina, United States of America

*corresponding author: ejr@live.unc.edu

Running title: No functional role for secondary gene flow in sympatric cichlid radiation

Key words: adaptive radiation, introgression, gene flow, sympatric speciation

Abstract: 232 words; Main text: 5,493 words; Figures: 7; Tables: 2

Abstract

Genomic data has revealed complex histories of colonization and repeated gene flow previously unrecognized in some of the most celebrated examples of sympatric speciation and radiation. However, much of the evidence for secondary gene flow into these radiations comes from genome-wide tests, which tells us little about how gene flow potentially influenced sympatric diversification. Here we investigated whole genomes of Barombi Mbo crater lake cichlids for fine-scale patterns of introgression between species with neighboring riverine cichlid populations. We did find evidence of secondary gene flow into the radiation scattered across < 0.24% of the genome; however, the functional and genetic diversity in these regions paint no clear picture of how that variation could have contributed to the ecological and morphological diversity found in the lake. Our results suggest that either variation in novel genetic pathways introduced during secondary gene flow contributed to the radiation, or that secondary gene flow was predominantly neutral with respect to the diversification processes. We also found evidence for differential assortment of ancestral polymorphism found in riverine populations between sympatric sister species, suggesting the presence of a hybrid swarm in the past. While the history of gene flow and colonization appears to be more complicated than once thought, the lack of compelling evidence for secondary gene flow influencing diversification suggests that we should not yet rule out one of the most celebrated examples of sympatric speciation in nature.

Introduction

Sympatric speciation, the extreme endpoint on the speciation-with-gene-flow continuum, is traditionally defined as the evolution of reproductive isolation without the aid of geographic barriers (Coyne and Orr 2004). Sympatric speciation has fascinated evolutionary biologists since Darwin for its illustration of the power of complex interactions between natural and sexual selection to create new species. Despite intense searches, very few case studies have been able to meet the rigorous criteria for demonstrating sympatric speciation in nature (Coyne and Orr 2004; Bolnick and Fitzpatrick 2007). Even in some of the more convincing examples that do meet these criteria, genomic data have revealed more complex evolutionary histories of multiple colonizations and repeated gene flow than previously thought (Papadopulos et al. 2011; The Heliconius Genome Consortium et al. 2012; Geiger et al. 2013; Alcaide et al. 2014; Igea et al. 2015; Malinsky et al. 2015; Martin et al. 2015a; Kautt et al. 2016).

However, much of the support for complicated histories involving repeated gene flow events into radiations comes from genome-wide tests for gene flow (e.g. (Lamichhaney et al. 2015; Martin et al. 2015a; Meier et al. 2017)). One prediction of models of speciation with gene flow is that divergence between incipient species should be heterogeneous across the genome (Turner et al. 2005; Harr 2006; Feder et al. 2012; Nosil and Feder 2012a,b). Indeed, high heterogeneity in genomic differentiation has been found across the genomes of many recent or incipient sister species (e.g. Jones et al. 2012; Martin et al. 2013; Poelstra et al. 2014; Soria-Carrasco et al. 2014; Malinsky et al. 2015; McGirr and Martin 2016), although other processes besides differential gene flow across the genome can produce similar heterogeneous patterns (Noor and Bennett 2009; Nachman and Payseur 2012; Cutter and Payseur 2013; Cruickshank and Hahn 2014; Guerrero and Hahn 2017; Ravinet et al. 2017). Only a handful of genes may

69 directly contribute to the speciation process whereas the rest of the genome is porous to gene
70 flow while reproductive isolation is incomplete (Wu 2001; Wu and Ting 2004). Therefore, gene
71 flow detected at the genome-wide level from populations outside the sympatric radiation does
72 not by itself constitute evidence that secondary gene flow was involved in the divergence process
73 among incipient species and shaped the radiation.

74 The Cameroon crater lake cichlid radiations are some of the most compelling cases for
75 sympatric speciation in the wild (Coyne and Orr 2004). The most speciose of these radiations is
76 found in the isolated 2.3 km-wide volcanic crater lake Barombi Mbo (Trewavas et al. 1972;
77 Schliewen et al. 1994; Schliewen and Klee 2004). Barombi Mbo hosts a radiation of 11 endemic
78 cichlid species, many of which have clear morphological and ecological separation from other
79 sympatric species (Schliewen et al. 1994). Some endemics have evolved unique specializations,
80 such as the spongivore *Pungu maclareni* and deep-water hypoxia specialist *Konia dikume*
81 (Trewavas et al. 1972). Other endemics, such as *Stomatepia mariae* and *S. pindu,* appear to be
82 incipient or stalled species complexes with only slight morphological and ecological divergence
83 at the extremes of a unimodal distribution of phenotypes (Martin 2012). However, evidence of
84 differential introgression, weak support for Barombi Mbo monophyly, and differences in levels
85 of shared ancestry with outgroup riverine populations from genome-wide RAD-seq data suggest
86 additional secondary gene flow into the radiation after the initial colonization, casting doubt on
87 one of the best examples of sympatric speciation in the wild (Martin et al. 2015a).

88 Here we dissect those signals of repeated gene flow to investigate their role in the
89 radiation using whole-genome sequences. We performed exhaustive searches for all genetic
90 patterns consistent with secondary gene flow into the ancestral Barombi Mbo population or into
91 subclades after their initial divergence using machine learning to finely dissect phylogenetic

92 signal across the genome and genomic scans to test for differential introgression. We find
93 evidence of both shared introgression between sister species and across subclades in the radiation
94 as well as differential introgression among sister species across small regions of the genome.
95 However, functional and genetic diversity in these regions do not paint a clear picture of how
96 introgressed variants may have contributed to speciation in these groups. Our results suggest that
97 either 1) rare introgression of variants in novel genetic pathways contributed to the
98 morphological and ecological diversity of the radiation (speciation with an allopatric phase), 2)
99 secondary gene flow was predominantly or completely neutral and did not contribute to
100 diversification in Barombi Mbo (sympatric speciation with gene flow), or 3) multiple
101 colonizations of the lake before diversification brought in genetic variation that was then
102 differentially sorted among incipient species (sympatric speciation from a hybrid swarm).
103

104 **Methods**

105 *Sampling and Genome Sequencing*

106 We sequenced whole genomes of 1-3 individuals from 10 out of the 11 species within the
107 sympatric radiation of Oreochromini cichlids in Cameroon crater lake Barombi Mbo (excluding
108 *Sarotherodon steinbachi* which is morphologically and ecologically similar to the other three
109 *Sarotherodon* species), an endemic *Sarotherodon* species pair from Lake Ejagham, and outgroup
110 *Sarotherodon* individuals from all three river drainages flanking the lake: Cross, Meme, and
111 Mungo rivers (e.g. see map in (Schliewen et al. 1994)). Details on the collection, extraction,
112 alignment to the *Oreochromis niloticus* reference genome, and variant calling protocols
113 following the standard GATK pipeline are provided in the supplementary methods.
114

115 *Characterization of introgression patterns across the genome*

116 First, we exhaustively searched the genomes for patterns of non-monophyletic Barombi Mbo
117 relationships using the machine learning program SAGUARO (Zamani et al. 2013) to identify
118 regions of the genome that contained relationships consistent with expectations from multiple
119 colorizations and secondary gene flow into the radiation (i.e. paraphyletic/polyphyletic Barombi
120 Mbo radiations). This method infers relationships among individuals in the form of genetic
121 distance matrices and assigns segments across the genomes to different topologies without a
122 priori hypotheses about these relationships. We partitioned the genome into a total of 75 unique
123 topologies (well past the inflection point at 30 topologies where the percent of genome explained
124 by each additional topology plateaus; Fig S1) to exhaustively search for relationships where
125 subclades or individual Barombi Mbo species were more closely related to riverine populations
126 than other species in the crater lake, suggesting sympatric speciation after a hybrid swarm (i.e.
127 differential sorting of ancestral polymorphism) or secondary gene flow into this subclade
128 (introgression). Details on the SAGUARO analysis and filtering strategies for calculating
129 proportions are provided in the supplementary methods.

130 We also looked for evidence of differential introgression within subclades of the radiation
131 on both a genome-wide and local level using f_4 statistics (Reich et al. 2009; Patterson et al. 2012;
132 Pickrell and Pritchard 2012). The f_4 statistic tests if branches on a four-taxon tree lack residual
133 genotypic covariance (as expected in the presence of incomplete lineage sorting and no
134 introgression) by comparing allele frequencies among the three possible unrooted trees.
135 We focused on tests of introgression with the two outgroup clades from our sample that came
136 from two main clusters: riverine populations of *Sarotherodon galilaeus* in the Mungo and Meme
137 rivers (MM) and riverine populations of *S. galilaeus* from the more distant Cross River (CR).

Based on the tree ((P1, P2),(*S. galilaeus* MM, *S. galilaeus* CR)), f_4 statistics were calculated for combinations of species among a) *Stomatepia*, b) the *Konia* + *Pungu* subclade, and c) *Myaka myaka* with *S. linnelli* as a representative of its sister *Sarotherodon* group. This subset of groupings was chosen to make these analyses more tractable by focusing on species with unique trophic ecologies within the radiation. Genome-wide f_4 statistics were calculated using the fourpop function in Treemix (Pickrell and Pritchard 2012). Standard error was estimated by jackknifing in windows of 1,000 adjacent SNPs to account for linkage disequilibrium.

We characterized heterogeneity in introgression across the genome among these same combinations and investigated whether differential introgression contributed variation potentially important in the divergence between species by calculating f_4 statistics in 10-kb sliding windows. We did this with a modified version of the ABBABABA.py and genomics.py scripts that use population allele frequencies of biallelic SNPs (https://github.com/simonhmartin/genomics_general;(Martin et al. 2015b); our modified version is provided in the supplementary materials). Significance of f_4 values in sliding windows across the genome were evaluated using the 1% tails of a null distribution generated from permutations of the f_4 test. For more details on the sliding window calculations of f_4, see supplementary methods.

For each of these regions, we looked for annotated genes using the well annotated NCBI *Oreochromis* Annotation Release 102 and searched their gene ontology in the phenotype database 'Phenoscape' (Mabee et al. 2012; Midford et al. 2013; Manda et al. 2015; Edmunds et al. 2016) and AmiGO2 (Balsa-Canto et al. 2016) for pertinent functions to the specializations and observed morphological differences among species, such as skeletal system or pigmentation.

Directionality of introgression

The sign of f_4 does not indicate the directionality of introgression because of the lack of an explicit outgroup. For example, in the tree (P1,P2),(P3,P4)), a positive f_4 value indicates gene flow either between P1 and P3 or P2 and P4. We narrowed down the directionality of introgression detected in these regions using the f_d statistic, a modified version of the *D*-statistic that looks at allele frequencies fitting two allelic patterns referred to as ABBA and BABA based on the tree ((P1,P2),P3,O)), where O is an outgroup species in which no gene flow is thought to occur with the other populations (Martin et al. 2015b). Using two individuals of *Coptodon kottae* from another Cameroon crater lake as our distantly related outgroup population and the same riverine and Barombi Mbo population combinations described above, f_d values were calculated in 10-kb windows across the genome using the same script and window settings as in the f_4 tests of introgression. The variation in f_d values per region was higher than f_4 of similar window sizes, perhaps due to the high variance between windows in number of sites that fit the desired ABBA/BABA patterns given our outgroup, so we used only those f_d outliers (the top 2%) that overlapped with significant f_4 outliers as potentially introgressed regions from which we could narrow down the two populations in which gene flow occurred.

We also visualized the directionality of genome-wide introgression detected with the f_4 statistics using Treemix (v 1.13) (Pickrell and Pritchard 2012). Treemix estimates a maximum likelihood phylogeny of the focal populations and then fits a user-specified number of migration edges to the tree by comparing genetic covariances between populations. We ran Treemix with *S. galilaeus* as root, and with 0 through 20 migration edges. To determine the most likely number of migration events, we performed likelihood-ratio tests comparing each graph to that with one

183 fewer migration event, starting with 1 versus 0 events, and took as the most likely value the first
184 non-significant comparison.
185
186 *Comparison of patterns of introgression to patterns of genetic divergence and diversity*
187 Reduced levels of genetic polymorphism in a population may indicate a strong selective sweep.
188 We can look at introgressed regions found in only a single Barombi Mbo species for evidence
189 that they have been adaptive, suggesting that secondary gene flow brought in variation
190 potentially important for speciation. To examine genetic diversity in candidate introgressed
191 regions, we calculated between-population nucleotide divergence (D_{xy}) and within-population
192 nucleotide diversity (π) for pairwise species comparisons among the Barombi Mbo focal species
193 and the riverine outgroups. D_{xy} and π were calculated over the same 10-kb windows as the f_4 tests
194 using the python script popGenWindows.py
195 (https://github.com/simonhmartin/genomics_general; (Martin et al. 2015b); see supplementary
196 methods for more details on these calculations).
197
198 **Results**
199 *Widespread polyphyletic relationships in Barombi Mbo are scattered across small regions of the*
200 *genome*
201 After conservative filtering of segments to remove uninformative regions (see supplementary
202 methods and Table S1), the Barombi Mbo cichlid radiation was a monophyletic group across
203 53% of the genome and only 0.6% was assigned to topologies indicating a polyphyletic Barombi
204 Mbo. These polyphyletic relationships are consistent with many patterns, including secondary
205 gene flow, incomplete lineage sorting, divergent selection, and ancestral population structure.
206 The most prevalent topology spanned 38.2% of the genome and featured the expected species

207 phylogeny for this group, in which all Barombi Mbo individuals form a single clade with distant
208 relationships to outgroup riverine *Sarotherodon* populations in Cameroon (Fig. 1A). The second
209 most prevalent topology (spanning 11.8% of the genome) featured identical evolutionary
210 relationships, except for a much shorter branch leading to *S. galilaeus* Mungo and Meme River
211 populations (Fig. 1B). Branch lengths produced by SAGUARO have no direct interpretation as
212 an evolutionary distance (analogous to a neighbor-joining tree), but may be useful for
213 comparison to similar topologies with different branch lengths, e.g. regions with higher
214 divergence rates (Zamani et al. 2013).

215 In 0.6% of the genome indicating polyphyletic Barombi Mbo relationships, we found
216 evidence consistent with multiple colonizations of the lake. Since we were looking for patterns
217 consistent with secondary gene flow or a hybrid swarm for subclades of the radiation, we
218 focused on topologies where single species or entire subclades were more closely related to
219 outgroups than other Barombi Mbo species, which represented only 0.24% of the genome. Some
220 topologies featured an entire subclade (e.g. *Stomatepia*) as monophyletic, but more closely
221 related to the riverine populations than other Barombi Mbo species, consistent with a hybrid
222 swarm scenario before the diversification of the *Stomatepia* subclade. Other topologies featured
223 individual species more closely related to outgroup riverine populations than sister species,
224 consistent with secondary gene flow into that lineage after the initiation of divergence. For
225 example, in *Stomatepia* we found topologies that group multiple species with riverine
226 populations (Fig. 2A-B), but we also found topologies where individual *Stomatepia* species (*S.*
227 *mariae* and *S. pindu*; Fig. 2C-D) were more closely related to riverine outgroups than other
228 *Stomatepia*. In the *Konia + Pungu* subclade, we saw a similar pattern with topologies for the
229 hypoxia and sponge-eating specialists (*K. dikume* and *P. maclareni,* respectively*;* Fig. 3A-B) but

230 also a topology where the entire subclade was sister to the riverine outgroup populations (Fig.
231 3C). In the zooplanktivore *M. myaka*, we found topologies in which *M. myaka* was sister to the
232 riverine populations (Fig. 4A-B), but also topologies where *M. myaka,* along with all the
233 Barombi Mbo *Sarotherodon* species, were sister to the riverine outgroup populations (Fig. 4C-
234 D).
235
236 *Genome-wide evidence for differential introgression into the radiation*
237 Consistent with evidence of differential introgression from RAD-seq data (Martin et al. 2015a),
238 genome-wide f_4 tests provided evidence of genome-wide differential gene flow between some
239 Barombi Mbo sister species and the outgroup riverine species (Table 1). There was significant
240 evidence of genome-wide introgression in tests involving both *S. pindu* in the *Stomatepia* species
241 complex and the hypoxia specialist *K. dikume* in the *Konia* + *Pungu* subclade. Some species pair
242 combinations within these subclades did not show evidence of differential gene flow, suggesting
243 that there may still be sympatric speciation occurring for some species, if not entire subclades.
244 For example, there was no significant secondary gene flow detected genome-wide in the tests
245 involving sister species *S. mariae* and *S. mongo* or *M. myaka* and *S. linnelli* (Table 1).
246 We also found evidence for widespread gene flow connecting populations across
247 Barombi Mbo and neighboring riverine populations in highly interconnected population graphs;
248 the likelihood of each graph did not plateau until reaching 10 admixture events (Fig S5). On the
249 Treemix population graph with 10 admixture events, gene flow from the Mungo/Meme River
250 populations of *S. galilaeus* occurred directly into individual species *S. mongo* and *K. eisentrauti*
251 rather than the ancestral node of their respective subclades (Fig S6). The proportion of admixture
252 inferred for these two events (0.1% into *S. mongo* and 0.4% into *K. eisentrauti*) was similar to

253 the small proportions of the genome assigned to topologies consistent with secondary gene flow
254 in the SAGUARO analyses. These admixture events pointing to the tips of the graphs suggest
255 secondary gene flow events between nearby riverine populations and individual species within
256 the radiation. In all population graphs allowing up to 21 migration events, any admixture from
257 outgroup riverine populations appears to be coming from the Mungo and Meme rivers rather
258 than the Cross River, consistent with the closer geographic proximity of the former drainages.
259
260 *Very few genomic regions contain signatures of differential introgression between sister species*
261 Very few regions of the genome introgressed into single species from outgroup riverine
262 populations (Fig 5A-C). In *Stomatepia*, only one region introgressed from Mungo/Meme Rivers
263 into *S. pindu* and only three regions into *S. mariae*, respectively, suggesting secondary gene flow
264 after initial diversification of *Stomatepia* (Table 2). Similarly, secondary introgression occurred
265 into the *Konia* + *Pungu* subclade (Table 2). However, there was also evidence of shared
266 introgression signals among sister species across all three subclades, where two subclade sister
267 species shared introgressed regions from a riverine population. Only 0.000017- 0.0000354% of
268 the genome appears introgressed into a single species of a Barombi Mbo subclade. This number
269 is smaller than suggested in the SAGUARO analysis perhaps due to the conservative
270 significance cut-offs and window size choice for f_4 statistic and that relationships observed in the
271 polyphyletic topologies are consistent with other patterns besides introgression.
272
273 *Evidence for sympatric sorting of ancestral polymorphism within a hybrid swarm*
274 A few of these 10-40 kb regions with peak signals of introgression were also present in multiple
275 subclades, indicating differential assortment of introgressed variation shared among clades. For

276 example, two significant f_4 outliers on linkage group 20 out of the 35 found across the genome
277 appear within the *Stomatepia*, *Konia*, and *Pungu,* suggesting that some of this introgression may
278 have occurred in the ancestral stages of the radiation and differentially sorted among species.
279 We also found 11 regions across the sister species pairs in which one species was more
280 similar to one outgroup riverine population while its sister species was more similar to the other
281 riverine population (Table 2). This signal is consistent with a hybrid swarm scenario due to
282 multiple colonizations by riverine populations before diversification of some of the sister species
283 and the sorting of polymorphisms brought in by these populations among incipient Barombi Mbo
284 species. For example, two regions that appear to be differentially sorted between *S. galilaeus* and
285 *S. mariae* and *S. pindu* from *S. galilaeus* CR versus *S. mongo* from *S.galilaeus* MM. Similar
286 patterns were found scattered across the genomes for *K. eisentrauti* and *K. dikume* versus *P.*
287 *maclareni* and *M. myaka* versus *S. linelli*.
288
289 *Weak support for functional importance of introgressed regions for species diversification*
290 Although we did find evidence of differential introgression among sister species scattered across
291 a small proportion of the genome, the types of genes found in these regions painted no clear
292 picture of how introgressed variation may have contributed to speciation (Table 2). For example,
293 differential introgression in *Stomatepia* occurred in regions with genes involved in a large range
294 of biological processes, including intracellular signal transduction, immune system response, and
295 motor neuron axon development (Table 2), with no obvious links to the highly divergent
296 morphological, ecological, or patterning traits observed between these species (Martin 2012) nor
297 to those traits normally associated with adaptive radiation in cichlid fishes such as body shape,
298 pharyngeal jaw morphology, retinal pigments, or male coloration (Kocher 2004; Barluenga et al.

299 2006; Wagner et al. 2012; Brawand et al. 2014; Malinsky et al. 2015; Meier et al. 2017).
300 Similarly, in both the *Konia* + *Pungu* and *Myaka* + *Sarotherodon* subclades, introgressed regions
301 were near genes involved in a large range of biological processes not clearly associated with
302 adaptive ecological traits in these species, such as *K. dikume*'s hypoxia tolerance, *P. maclareni's*
303 spongivory, and *M. myaka's* zooplanktivory. For example, while there appears to be differential
304 introgression in *Konia* in a region containing *pafah1b3*, a gene involved in platelet activation
305 activity, and *K. dikume's* deep water specialization includes higher blood volume with higher
306 concentrations of hemoglobin (Green et al. 1973), it is not obvious how introgressed variation in
307 *pafah1b3* would have played a role in the evolution of these traits from studies of its function in
308 model organisms, which includes spermatogenesis and sterility in mice (Prescott et al. 2000;
309 Koizumi et al. 2003; Yan et al. 2003).
310 Similarly, the amount of genetic diversity in introgressed regions does not suggest strong
311 divergent selection on introgressed genetic variation due to hard selective sweeps. In line with
312 the presence of peaks in f_4 values in these regions, between-population diversity (D_{xy}) was
313 typically high between one of the species and its sister species (Fig. 6). However, within-
314 population diversity across many of these regions was often greater or comparable to scaffold
315 and genome-wide averages (Table S2-4), suggesting these regions may not have experienced
316 hard selective sweeps that would support their role in adaptive divergence among species (Fig.
317 6). In summary, although we found evidence for differential secondary gene flow between sister
318 species in the radiation, we did not find strong functional support from gene ontology terms nor
319 signatures of selection that the introgressed alleles were important for sympatric species
320 diversification.
321

Discussion

Little evidence that secondary gene flow promoted the diversification of Barombi Mbo cichlids

Our fine-scale investigations of introgression across the genomes of a celebrated putative example of sympatric speciation are consistent with two possible scenarios: 1) sympatric speciation in the presence of continuous neutral secondary gene flow into the radiation, or 2) speciation initiated by secondary gene flow. We found little support for the latter allopatric scenario from both a learning machine and sliding-window approach. From the SAGUARO analyses, our most conservative estimate of introgression into single species of the radiation ranges from 0.013 -0.019% of the genome. Estimates are similarly small from the f_4 statistics, ranging from 0.000017- 0.0000354% of the genome (Fig 7). Furthermore, even these significant outliers may represent false positives. First, our method of selecting introgressed regions from the 1% tails of a null distribution can always find outliers, even in the absence of introgression. Second, it is also difficult to distinguish signatures of differential introgression from the biased assortment of ancestral polymorphism into modern lineages, e.g. a hybrid swarm scenario that would still result in sympatric divergence entirely within the crater lake. Finally, even if our statistical outliers represent differentially introgressed regions, their importance to the speciation process is equivocal. We found no evidence of selective sweeps in these regions that would suggest they aided in divergence between species and they contain mainly housekeeping genes that do not clearly suggest how introgressed variation would have contributed to the radiation.

This contrasts studies on other systems using similar approaches which found compelling cases for adaptive introgression contributing to diversification (e.g. Abi-Rached et al. 2011; The Heliconius Genome Consortium et al. 2012; Huerta-Sánchez et al. 2014; Lamichhaney et al. 2015; Stankowski and Streisfeld 2015; Arnold et al. 2016; Meier et al. 2017), including our own

345 previous work (Richards and Martin 2017). For example, several studies have found convincing
346 candidate genes/variants in introgressed regions to suggest that adaptive introgression played a
347 role in shaping ecological and morphological diversity. These include the detection of
348 introgressed alleles linked to wing-color patterning involved in mimicry and mate selection in
349 *Heliconius* butterflies (The Heliconius Genome Consortium et al. 2012), flower coloration
350 involved in pollinator preferences for *Mimulus* species (Stankowski and Streisfeld 2015), and
351 oral jaw size variation involved in scale-eating trophic specialization in *Cyprinodon* pupfishes
352 (Richards and Martin 2017).
353
354 *Evidence for a hybrid swarm further complicates the role of gene flow in the speciation process*
355 *in Barombi Mbo cichlids*
356 Beyond speciation scenarios involving secondary gene flow, our findings also suggest another
357 scenario for sympatric speciation in this system: sympatric speciation from a hybrid swarm
358 involving the differential sorting of ancestral polymorphism among incipient species. A hybrid
359 swarm is not easily detectable using the f_4 statistic because introgressed variation could be shared
360 among diverging sister species, leading to an f_4 value of zero (Reich et al. 2009; Patterson et al.
361 2012). However, many of the f_4 peaks appear to be shared across at least two of the sister species
362 in a subclade, shared between species of different subclades, or contain variation from both
363 riverine populations (Mungo/Meme and Cross Rivers) that has been differentially sorted among
364 sister species. All three of these patterns are consistent with an ancestral hybrid swarm before
365 divergence between sister species occurred. This pattern of differential sorting of variation from
366 a hybrid swarm from f_d analyses could also result from a lack of power in the statistic to
367 distinguish the directionality of the introgression detected in those regions when using biallelic

368 patterns and four populations (e.g. when two populations share similar allele patterns, the other
369 two populations can share the opposite allele pattern by default). However, we also found
370 evidence that entire subclades (e.g. *Stomatepia*) were more closely related to riverine populations
371 than other Barombi Mbo subclades from the SAGUARO analyses that are also consistent with a
372 hybrid swarm (e.g. Fig 2).

373 There are some caveats to our interpretations of secondary gene flow and its weak
374 functional role in the ecological and morphological diversity observed within the lake.
375 Recombination rate varies across genomes and determines the scale over which patterns of
376 admixture and differentiation vary (Smukowski and Noor 2011). In our fixed sliding window
377 size of 10-kb, we may have missed important patterns of introgression in regions of
378 recombination hotspots, where such patterns are expected to be very fine-scale. Shared variation
379 among species may reflect unsorted polymorphism from structured ancestral populations rather
380 than hybridization. Introgression events can also be hard to distinguish from ongoing balancing
381 selection of ancestral polymorphism that is sieved between species (Guerrero and Hahn 2017).
382 While we focused on searching for genetic signatures of hard selective sweeps in introgressed
383 regions, some of them with intermediate to high nucleotide diversity may have undergone soft
384 selective sweeps, when selection drives multiple adaptive haplotypes to fixation. Some of these
385 introgressed regions may have been adaptive and undergone soft selective sweeps, although the
386 relative contributions of hard sweeps versus soft sweeps during adaptation and speciation is still
387 the subject of much debate (Hermisson and Pennings 2005, 2017; Pritchard et al. 2010; Jensen
388 2014; Schrider et al. 2015).
389
390 *Best remaining cases for sympatric speciation within Barombi Mbo cichlid radiation*

391 While the radiation as a whole may not have entirely arisen from sympatric speciation, some
392 sister species within Barombi Mbo are better case studies of the process than others. Within the
393 three-species *Stomatepia* subclade, there is little evidence that secondary gene flow played an
394 important role in diversification. On a genome-wide level, we detected secondary gene flow in f_4
395 tests involving *S. pindu*. However, on a finer scale the one introgressed region unique to *S. pindu*
396 is unannotated and the three introgressed regions unique to *S. mariae* contain four housekeeping
397 genes involved in extracellular exosome activity and plasma membranes
398 (*jmjd8*,*prss1*,*cldn4*,*muc19*). Shared signals of introgression among the three species represent a
399 larger proportion of the genome than differentially introgressed regions, although both types of
400 introgressed material appear to be rare in the genome (< 0.045%; Fig 7). The high ecological and
401 morphological overlap among *Stomatepia* species suggests that this species complex may be
402 stalled in the earliest stages of divergence. For example, *S. pindu* and *S. mariae* appear be at the
403 extremes of a unimodal distribution of phenotypes; this is one major prediction of sympatric
404 speciation models in the presence of only weak disruptive selection on ecological traits (e.g.
405 (Matessi et al. 2002; Burger et al. 2006)), as measured in this species pair (Martin 2012).
406 Even for the two monotypic specialist species *M. myaka* and *P. maclareni*, there is
407 minimal evidence for a role of secondary gene flow in the evolution of their trophic
408 specializations. On a genome-wide level, the tests for differential introgression with one of the
409 specialists and another species from the radiation were not significant. For the two regions which
410 appear to be differentially introgressed between riverine populations and the zooplantivore *M.*
411 *myaka*, one has no annotations, suggesting it may be neutral gene flow while the other is an
412 introgressed region in other subclades, a signal of differential sorting of variation after a hybrid
413 swarm. For spongivore specialist *Pungu maclareni*, we found only a single region differentially

414 introgressed with riverine populations. This region contains *ddn1*, a gene involved in serine-type
415 peptidase activity, proteins which are found ubiquitously in prokaryotes and eukaryotes.
416 Among all the ecologically divergent species pairs focused on in this study, *K. eisentrauti*
417 and *K. dikume* are the least convincing as a putative case for sympatric speciation between sister
418 species. Similar to *Stomatepia*, there is more evidence for shared introgression in regions of the
419 *Konia* genome than differentially introgressed regions (Fig 8). However, differential
420 introgression between the two *Konia* species occurs in regions with the best potential candidates
421 in this study for contributing to diversification (although these regions are still not as strong as
422 the 'smoking guns' observed in our past study of introgressed variation, e.g. Richards and Martin
423 2017). The gene *pafah1b3*, which is involved in platelet activation activity, is differentially
424 introgressed between *Konia* species and may function in *K. dikume's* deep water specialization
425 of higher blood volume with higher concentrations of hemoglobin (Green et al. 1973). However,
426 studies of the phenotypic effects of *pafah1b3* in model organisms do not hint at a role in
427 hemoglobin concentration. Instead, mutations in this gene have been suggested to play a role in
428 male spermatogenesis and fertility in mice (Koizumi et al. 2003; Yan et al. 2003). We also see
429 differential introgression in a region containing the gene *ehmt2*, which is involved in
430 neurogenesis and retinal cell differentiation. While it is not as directly clear how introgressed
431 variation in this gene would have contributed to divergence among the *Konia* sister species,
432 variation in visual traits such as color perception are important axes of diversification in other
433 cichlid radiations, particularly along a depth gradient (Terai et al. 2002, 2006; Seehausen et al.
434 2008). These two species also exist in microallopatry; *K. eisentrauti* is an abundant detritivore
435 along the shallow littoral region of the lake while *K. dikume* is a deep-water specialist on

436 *Chaoborus* midge larvae which is only collected in deep-water gill nets (Trewavas et al. 1972;
437 Schliewen 1994). Both species are mouthbrooders and likely breed in non-overlapping habitats.
438
439 *Conclusion*
440 The complex history of colonization in the crater evidenced in this and a previous genome-wide
441 study suggests allopatric phases of the speciation process in the radiation, which violates one of
442 the strict criteria for demonstrating sympatric speciation in the wild (Coyne and Orr 2004).
443 Nonetheless, from our fine-scale dissection of where in the genome these signals are coming
444 from, we cannot point to a functional role for secondary gene flow in the speciation process
445 across any of the subclades. This suggests that either variation in genes with undiscovered
446 functional effects underlies the divergent ecologies and morphologies seen in the lake or that any
447 secondary gene flow was neutral with regard to its role in the speciation process. We also find
448 evidence to support sympatric speciation after a hybrid swarm that resulted from multiple
449 colonizations of the lake, still consistent with a scenario of sympatric speciation through
450 differential sorting of ancestral polymorphism. Disentangling the effects of a putative hybrid
451 swarm from secondary contact events on the speciation process will require a better
452 understanding of the timing of gene flow events compared to the diversification times of
453 Barombi Mbo species. We found evidence for gene flow into the radiation both before and after
454 initial diversification of subclades within the lake. Even without this information, weak support
455 for a functional role of secondary gene flow in the radiation of Barombi Mbo cichlids suggests
456 that we should not rule out this example of a sympatric radiation just yet.
457
458

Acknowledgements

This study was funded by a National Geographic Society Young Explorer's Grant, a Lewis and Clark Field Research grant from the American Philosophical Society, and the University of North Carolina at Chapel Hill to CHM. We gratefully acknowledge the Cameroonian government and the local chiefs of Barombi Mbo village and surrounding communities for permission to conduct this research. We thank Cyrille Dening Touokong and Jackson Waite-Himmelwright for field assistance and Nono Gonwuou for help obtaining permits.

Author contributions

EJR wrote the manuscript. EJR conducted all analyses except for Treemix; JP conducted Treemix analyses. CHM collected samples and provided the genomic data. All authors contributed to the development of ideas presented in the study and revised the manuscript.

Data Accessibility

All datasets used for this study will be deposited in Dryad and the NCBI Short Read Archive (SRA).

Competing Interests

We declare no competing financial interests.

References

Abi-Rached, L., M. J. Jobin, S. Kulkarni, A. McWhinnie, K. Dalva, L. Gragert, F. Babrzadeh, B. Gharizadeh, M. Luo, F. A. Plummer, J. Kimani, M. Carrington, D. Middleton, R. Rajalingam, M. Beksac, S. G. E. Marsh, M. Maiers, L. A. Guethlein, S. Tavoularis, A.-M. Little, R. E. Green, P. J. Norman, and P. Parham. 2011. The Shaping of Modern Human Immune Systems by Multiregional Admixture with Archaic Humans. Science (80-.). 334:89 LP-94.

Alcaide, M., E. S. C. Scordato, T. D. Price, and D. E. Irwin. 2014. Genomic divergence in a ring species complex. Nature 511:83–85. Nature Publishing Group.

Arnold, B. J., B. Lahner, J. M. DaCosta, C. M. Weisman, J. D. Hollister, D. E. Salt, K. Bomblies, and L. Yant. 2016. Borrowed alleles and convergence in serpentine adaptation. Proc. Natl. Acad. Sci. 113:8320–8325.

Balsa-Canto, E., D. Henriques, A. Gabor, and J. R. Banga. 2016. AMIGO2, a toolbox for dynamic modeling, optimization and control in systems biology. Bioinformatics 32:1–2.

Barluenga, M., K. N. Stölting, W. Salzburger, M. Muschick, and A. Meyer. 2006. Sympatric speciation in Nicaraguan crater lake cichlid fish. Nature 439:719–723.

Bolnick, D. I., and B. M. Fitzpatrick. 2007. Linked references are available on JSTOR for this article : Sympatric Speciation : Models and Empirical Evidence. Annu. Rev. Ecol. Evol. Syst. 38:459–487.

Brawand, D., C. E. Wagner, Y. I. Li, M. Malinsky, I. Keller, S. Fan, O. Simakov, A. Y. Ng, Z. W. Lim, E. Bezault, Turner-Maier, J. J. Johnson, R. Alcazar, H. J. Noh, P. Russell, B. Aken, J. Alföldi, C. Amemiya, N. Azzouzi, J.-F. Baroiller, F. Barloy-Hubler, A. Berlin, R. Bloomquist, K. L. Carleton, M. a. Conte, H. D'Cotta, O. Eshel, L. Gaffney, F. Galibert, H.

505 F. Gante, S. Gnerre, L. Greuter, R. Guyon, N. S. Haddad, W. Haerty, R. M. Harris, H. a.
506 Hofmann, T. Hourlier, G. Hulata, D. B. Jaffe, M. Lara, L. A.P., I. MacCallum, S. Mwaiko,
507 M. Nikaido, H. Nishihara, C. Ozouf-Costaz, D. J. Penman, D. Przybylski, M. Rakotomanga,
508 S. C. P. Renn, F. J. Ribeiro, M. Ron, W. Salzburger, L. Sanchez-Pulido, M. E. Santos, S.
509 Searle, T. Sharpe, R. Swofford, F. J. Tan, L. Williams, S. Young, S. Yin, N. Okada, T. D.
510 Kocher, E. a. Miska, E. S. Lander, B. Venkatesh, R. D. Fernald, A. Meyer, C. P. Ponting, J.
511 T. Streelman, K. Lindblad-Toh, O. Seehausen, and F. Di Palma. 2014. The genomic
512 substrate for adaptive radiation in African cichlid fish. Nature 513:375–381.

513 Burger, R., K. A. Schneider, and M. Willensdorfer. 2006. The Conditions for Speciation through
514 Intraspecific Competition. Evolution (N. Y). 60:2185–2206. [Society for the Study of
515 Evolution, Wiley].

516 Coyne, J. A., and H. A. Orr. 2004. Speciation. Sutherland, MA.

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

519 Cutter, A. D., and B. A. Payseur. 2013. Genomic signatures of selection at linked sites: unifying
520 the disparity among species. Nat. Rev. Genet. 14:262–274. Nature Publishing Group.

521 Edmunds, R. C., B. Su, J. P. Balhoff, B. F. Eames, W. M. Dahdul, H. Lapp, J. G. Lundberg, T. J.
522 Vision, R. A. Dunham, P. M. Mabee, and M. Westerfield. 2016. Phenoscape: Identifying
523 candidate genes for evolutionary phenotypes. Mol. Biol. Evol. 33:13–24.

524 Feder, J. L., S. P. Egan, and P. Nosil. 2012. The genomics of speciation-with-gene-flow. Trends
525 Genet. 28:342–350.

526 Geiger, M. F., J. K. McCrary, and U. K. Schliewen. 2013. Crater Lake Apoyo Revisited -
527 Population Genetics of an Emerging Species Flock. PLoS One 8:1–17.

528 Green, J., S. A. Corbet, and E. Betney. 1973. Ecological studies on crater lakes in West
529 Cameroon The blood of endemic cichlids in Barombi Mbo in relation to stratification and
530 their feeding habits. J. Zool. 170:299–308. Blackwell Publishing Ltd.
531 Guerrero, R. F., and M. W. Hahn. 2017. Speciation as a sieve for ancestral polymorphism. Mol.
532 Ecol. 1–7.
533 Harr, B. 2006. Genomic islands of differentiation between house mouse subspecies Genomic
534 islands of differentiation between house mouse subspecies. 730–737.
535 Hermisson, J., and P. S. Pennings. 2005. Soft Sweeps. Genetics 169:2335 LP-2352.
536 Hermisson, J., and P. S. Pennings. 2017. Soft sweeps and beyond: understanding the patterns and
537 probabilities of selection footprints under rapid adaptation. Methods Ecol. Evol. 8:700–716.
538 Huerta-Sánchez, E., X. Jin, Asan, Z. Bianba, B. M. Peter, N. Vinckenbosch, Y. Liang, X. Yi, M.
539 He, M. Somel, P. Ni, B. Wang, X. Ou, Huasang, J. Luosang, Z. X. P. Cuo, K. Li, G. Gao,
540 Y. Yin, W. Wang, X. Zhang, X. Xu, H. Yang, Y. Li, J. Wang, J. Wang, and R. Nielsen.
541 2014. Altitude adaptation in Tibetans caused by introgression of Denisovan-like DNA.
542 Nature 512:194–197.
543 Igea, J., D. Bogarín, A. S. T. Papadopulos, and V. Savolainen. 2015. A comparative analysis of
544 island floras challenges taxonomy-based biogeographical models of speciation. Evolution
545 (N. Y). 69:482–491.
546 Jensen, J. D. 2014. On the unfounded enthusiasm for soft selective sweeps. Nat. Commun.
547 5:5281.
548 Jones, F. C., M. G. Grabherr, Y. F. Chan, P. Russell, E. Mauceli, J. Johnson, R. Swofford, M.
549 Pirun, M. C. Zody, S. White, E. Birney, S. Searle, J. Schmutz, J. Grimwood, M. C. Dickson,
550 R. M. Myers, C. T. Miller, B. R. Summers, A. K. Knecht, S. D. Brady, H. Zhang, A. A.

551 Pollen, T. Howes, C. Amemiya, Broad Institute Genome Sequencing Platform & Whole
552 Genome Assembly Team, E. S. Lander, F. Di Palma, K. Lindblad-Toh, and D. M. Kingsley.
553 2012. The genomic basis of adaptive evolution in threespine sticklebacks. Nature 484:55–
554 61.
555 Kautt, A. F., G. Machado-Schiaffino, J. Torres-Dowdall, and A. Meyer. 2016. Incipient
556 sympatric speciation in Midas cichlid fish from the youngest and one of the smallest crater
557 lakes in Nicaragua due to differential use of the benthic and limnetic habitats? Ecol. Evol.
558 6:5342–5357.
559 Kocher, T. D. 2004. Adaptive evolution and explosive speciation: the cichlid fish model. Nat.
560 Rev. Genet. 5:288–298.
561 Koizumi, H., N. Yamaguchi, M. Hattori, T. O. Ishikawa, J. Aoki, M. M. Taketo, K. Inoue, and
562 H. Arai. 2003. Targeted disruption of intracellular type I platelet activating factor-
563 acetylhydrolase catalytic subunits causes severe impairment in spermatogenesis. J. Biol.
564 Chem. 278:12489–12494.
565 Lamichhaney, S., J. Berglund, M. S. Almén, K. Maqbool, M. Grabherr, A. Martinez-Barrio, M.
566 Promerová, C.-J. Rubin, C. Wang, N. Zamani, B. R. Grant, P. R. Grant, M. T. Webster, and
567 L. Andersson. 2015. Evolution of Darwin's finches and their beaks revealed by genome
568 sequencing. Nature 518:371–375.
569 Mabee, P., J. P. Balhoff, W. M. Dahdul, H. Lapp, P. E. Midford, T. J. Vision, and M.
570 Westerfield. 2012. 500,000 fish phenotypes: The new informatics landscape for
571 evolutionary and developmental biology of the vertebrate skeleton. J. Appl. Ichthyol.
572 28:300–305.
573 Malinsky, M., R. J. Challis, A. M. Tyers, S. Schiffels, Y. Terai, B. P. Ngatunga, E. A. Miska, R.

574 Durbin, M. J. Genner, and G. F. Turner. 2015. Genomic islands of speciation separate
575 cichlid ecomorphs in an East African crater lake. Science (80-.). 350:1493–1498.
576 Manda, P., J. P. Balhoff, H. Lapp, P. Mabee, and T. J. Vision. 2015. Using the phenoscape
577 knowledgebase to relate genetic perturbations to phenotypic evolution. Genesis 53:561–
578 571.
579 Martin, C. H. 2012. Weak Disruptive Selection and Incomplete Phenotypic Divergence in Two
580 Classic Examples of Sympatric Speciation: Cameroon Crater Lake Cichlids. Am. Nat.
581 180:E90–E109.
582 Martin, C. H., J. S. Cutler, J. P. Friel, C. Dening Touokong, G. Coop, and P. C. Wainwright.
583 2015a. Complex histories of repeated gene flow in Cameroon crater lake cichlids cast doubt
584 on one of the clearest examples of sympatric speciation. Evolution (N. Y). 69:1406–1422.
585 Martin, S. H., K. K. Dasmahapatra, N. J. Nadeau, C. Slazar, J. R. Walters, F. Simpson, M.
586 Blaxter, A. Manica, J. Mallet, and C. D. Jiggins. 2013. Genome-wide evidence for
587 speciation with gene flow in Heliconius butterflies. Genome Res. 23:1817–1828.
588 Martin, S. H., J. W. Davey, and C. D. Jiggins. 2015b. Evaluating the use of ABBA-BABA
589 statistics to locate introgressed loci. Mol. Biol. Evol. 32:244–257.
590 Matessi, C., A. Gimelfarb, and S. Gavrilets. 2002. Long-term Buildup of Reproductive Isolation
591 Promoted by Disruptive Selection: How Far Does it Go? Selection 2:41–64. Akadémiai
592 Kiadó.
593 McGirr, J. A., and C. H. Martin. 2016. Novel candidate genes underlying extreme trophic
594 specialization in Caribbean pupfishes. Mol. Biol. Evol. msw286.
595 Meier, J. I., D. A. Marques, S. Mwaiko, C. E. Wagner, L. Excoffier, and O. Seehausen. 2017.
596 Ancient hybridization fuels rapid cichlid fish adaptive radiations. Nat. Commun. 8:14363.

Nature Publishing Group.

Midford, P. E., T. Dececchi, J. P. Balhoff, W. M. Dahdul, N. Ibrahim, H. Lapp, J. G. Lundberg, P. M. Mabee, P. C. Sereno, M. Westerfield, T. J. Vision, D. C. Blackburn, S. Federhen, P. Mabee, J. Balhoff, W. Dahdul, H. Lapp, P. Midford, T. Vision, M. Westerfield, W. Dahdul, J. Balhoff, J. Engeman, T. Grande, E. Hilton, C. Kothari, H. Lapp, J. Lundberg, P. Midford, T. Vision, M. Westerfield, P. Mabee, F. Fang, M. Toledo-Piza, M. Uhen, A. Barnosky, B. Bills, J. Blois, M. Carrano, M. Carrasco, G. Erickson, J. Eronen, M. Fortelius, R. Graham, E. Grimm, M. O'Leary, A. Mast, W. Piel, P. Polly, L. Saila, J. Balhoff, P. Midford, H. Lapp, M. Ghiselin, S. Schulz, H. Stenzhorn, M. Boeker, D. Thau, N. Franz, B. Smith, M. Ashburner, C. Rosse, J. Bard, W. Bug, W. Ceusters, L. Goldberg, K. Eilbeck, A. Ireland, C. Mungall, O. Consortium, N. Leontis, P. Rocca-Seraa, A. Ruttenberg, S. Sansone, R. Scheuermann, N. Shah, P. Whetzel, S. Lewis, J. Beaulieu, R. Ree, J. Cavender-Bares, G. Weiblen, and M. Donoghue. 2013. The vertebrate taxonomy ontology: a framework for reasoning across model organism and species phenotypes. J. Biomed. Semantics 4:34.

Nachman, M. W., and B. A. Payseur. 2012. Recombination rate variation and speciation: theoretical predictions and empirical results from rabbits and mice. Philos. Trans. R. Soc. B Biol. Sci. 367:409–421.

Noor, M., and S. Bennett. 2009. Islands of speciation or mirages in the desert? Examingin the role of restricted recombination in maintaining species. Heredity (Edinb). 103:439–444.

Nosil, P., and J. L. Feder. 2012a. Genomic divergence during speciation: causes and consequences. Philos. Trans. R. Soc. B Biol. Sci. 367:332–342.

Nosil, P., and J. L. Feder. 2012b. Widespread yet heterogeneous genomic divergence. Mol. Ecol. 21:2829–2832.

620 Papadopulos, A. S. T., W. J. Baker, D. Crayn, R. K. Butlin, R. G. Kynast, I. Hutton, and V.
621 Savolainen. 2011. Speciation with gene flow on Lord Howe Island. Proc. Natl. Acad. Sci.
622 108:13188–13193.
623 Patterson, N., P. Moorjani, Y. Luo, S. Mallick, N. Rohland, Y. Zhan, T. Genschoreck, T.
624 Webster, and D. Reich. 2012. Ancient Admixture in Human History. Genetics 192:1069–
625 1093.
626 Pickrell, J. K., and J. K. Pritchard. 2012. Inference of Population Splits and Mixtures from
627 Genome-Wide Allele Frequency Data. PLoS Genet. 8.
628 Poelstra, J. W., N. Vijay, C. M. Bossu, H. Lantz, B. Ryll, I. Muller, V. Baglione, P. Unneberg,
629 M. Wikelski, M. G. Grabherr, and J. B. W. Wolf. 2014. The genomic landscape underlying
630 phenotypic integrity in the face of gene flow in crows. Science (80-.). 344:1410–1414.
631 Prescott, S. M., G. a Zimmerman, D. M. Stafforini, and T. M. Mcintyre. 2000. Platelet activating
632 factor. Annu. Rev. Biochem. 69:419–445.
633 Pritchard, J. K., J. K. Pickrell, and G. Coop. 2010. The Genetics of Human Adaptation: Hard
634 Sweeps, Soft Sweeps, and Polygenic Adaptation. Curr. Biol. 20:R208–R215.
635 Ravinet, M., R. Faria, R. K. Butlin, J. Galindo, N. Bierne, M. Rafajlović, M. A. F. Noor, B.
636 Mehlig, and A. M. Westram. 2017. Interpreting the genomic landscape of speciation: a road
637 map for finding barriers to gene flow. J. Evol. Biol. 30:1450–1477.
638 Reich, D., K. Thangaraj, N. Patterson, A. L. Price, and L. Singh. 2009. Reconstructing Indian
639 population history. Nature 461:489–94.
640 Richards, E. J., and C. H. Martin. 2017. Adaptive introgression from distant Caribbean islands
641 contributed to the diversification of a microendemic adaptive radiation of trophic specialist
642 pupfishes.

643 Schliewen, U. K., and B. Klee. 2004. Reticulate sympatric speciation in Cameroonian crater lake
644 cichlids. Front. Zool. 1:5.
645 Schliewen, U. K., D. Tautz, and S. Paabo. 1994. Sympatric speciation suggested by monophyly
646 of crater lake cichlids. Nature 368:629–631.
647 Schrider, D. R., F. K. Mendes, M. W. Hahn, and A. D. Kern. 2015. Soft Shoulders Ahead:
648 Spurious Signatures of Soft and Partial Selective Sweeps Result from Linked Hard Sweeps.
649 Genetics 200:267 LP-284.
650 Seehausen, O., Y. Terai, I. S. Magalhaes, K. L. Carleton, H. D. J. Mrosso, R. Miyagi, I. van der
651 Sluijs, M. V. Schneider, M. E. Maan, H. Tachida, H. Imai, and N. Okada. 2008. Speciation
652 through sensory drive in cichlid fish. Nature 455:620–626.
653 Smukowski, C. S., and M. A. F. Noor. 2011. Recombination rate variation in closely related
654 species. Heredity (Edinb). 107:496–508. The Genetics Society.
655 Soria-Carrasco, V., Z. Gompert, A. A. Comeault, T. E. Farkas, T. L. Parchman, J. S. Johnston, C.
656 A. Buerkle, J. L. Feder, J. Bast, T. Schwander, S. P. Egan, B. J. Crespi, and P. Nosil. 2014.
657 Stick Insect Genomes Reveal Natural Selection's Role in Parallel Speciation. Science (80-.
658). 344:738 LP-742.
659 Stankowski, S., and M. A. Streisfeld. 2015. Introgressive hybridization facilitates adaptive
660 divergence in a recent radiation of monkeyflowers. Proc. R. Soc. London B 282:20151666.
661 Terai, Y., W. E. Mayer, J. Klein, H. Tichy, and N. Okada. 2002. The effect of selection on a long
662 wavelength-sensitive (LWS) opsin gene of Lake Victoria cichlid fishes. Proc. Natl. Acad.
663 Sci. 99:15501–15506.
664 Terai, Y., O. Seehausen, T. Sasaki, K. Takahashi, S. Mizoiri, T. Sugawara, T. Sato, M.
665 Watanabe, N. Konijnendijk, H. D. J. Mrosso, H. Tachida, H. Imai, Y. Shichida, and N.

666 Okada. 2006. Divergent selection on opsins drives incipient speciation in Lake Victoria
667 cichlids. PLoS Biol. 4:2244–2251.
668 The Heliconius Genome Consortium, K. K. Dasmahapatra, J. R. Walters, A. D. Briscoe, J. W.
669 Davey, A. Whibley, N. J. Nadeau, A. V. Zimin, D. S. T. Hughes, L. C. Ferguson, S. H.
670 Martin, C. Salazar, J. J. Lewis, S. Adler, S.-J. Ahn, D. a. Baker, S. W. Baxter, N. L.
671 Chamberlain, R. Chauhan, B. a. Counterman, T. Dalmay, L. E. Gilbert, K. Gordon, D. G.
672 Heckel, H. M. Hines, K. J. Hoff, P. W. H. Holland, E. Jacquin-Joly, F. M. Jiggins, R. T.
673 Jones, D. D. Kapan, P. Kersey, G. Lamas, D. Lawson, D. Mapleson, L. S. Maroja, A.
674 Martin, S. Moxon, W. J. Palmer, R. Papa, A. Papanicolaou, Y. Pauchet, D. a. Ray, N.
675 Rosser, S. L. Salzberg, M. a. Supple, A. Surridge, A. Tenger-Trolander, H. Vogel, P. a.
676 Wilkinson, D. Wilson, J. a. Yorke, F. Yuan, A. L. Balmuth, C. Eland, K. Gharbi, M.
677 Thomson, R. a. Gibbs, Y. Han, J. C. Jayaseelan, C. Kovar, T. Mathew, D. M. Muzny, F.
678 Ongeri, L.-L. Pu, J. Qu, R. L. Thornton, K. C. Worley, Y.-Q. Wu, M. Linares, M. L.
679 Blaxter, R. H. Ffrench-Constant, M. Joron, M. R. Kronforst, S. P. Mullen, R. D. Reed, S. E.
680 Scherer, S. Richards, J. Mallet, W. Owen McMillan, and C. D. Jiggins. 2012. Butterfly
681 genome reveals promiscuous exchange of mimicry adaptations among species. Nature
682 487:94–98.
683 Trewavas, E., J. Green, and S. A. Corbet. 1972. Ecological studies on crater lakes in West
684 Cameroon fishes of Barombi Mbo. J. Zool. 41–95.
685 Turner, T. L., M. W. Hahn, and S. V. Nuzhdin. 2005. Genomic islands of speciation in
686 Anopheles gambiae. PLoS Biol. 3:1572–1578.
687 Wagner, C. E., L. J. Harmon, and O. Seehausen. 2012. Ecological opportunity and sexual
688 selection together predict adaptive radiation. Nature 487:366–369. Nature Publishing

Group.

Wu, C.-I., and C.-T. Ting. 2004. Genes and speciation. Nat. Rev. Genet. 5:114–122.

Wu, C. I. 2001. The genic view of the process of speciation. J. Evol. Biol. 14:851–865.

Yan, W., A. H. Assadi, A. Wynshaw-Boris, G. Eichele, M. M. Matzuk, and G. D. Clark. 2003. Previously uncharacterized roles of platelet-activating factor acetylhydrolase 1b complex in mouse spermatogenesis. Proc. Natl. Acad. Sci. U. S. A. 100:7189–94.

Zamani, N., P. Russell, H. Lantz, M. P. Hoeppner, J. R. Meadows, N. Vijay, E. Mauceli, F. Di Palma, K. Lindblad-Toh, P. Jern, and M. G. Grabherr. 2013. Unsupervised genome-wide recognition of local relationship patterns. BMC Genomics 14:347.

Table 1. Genome-wide f_4 statistics supporting differential introgression within Barombi Mbo radiation. Tests with significant evidence for differential introgression are highlighted in bold. The f_4 statistic was calculated for pairwise combinations among sister species of Barombi Mbo subclades and riverine populations of *S. galilaeus* from Mungo and Meme Rivers (MM) and Cross River (CR).

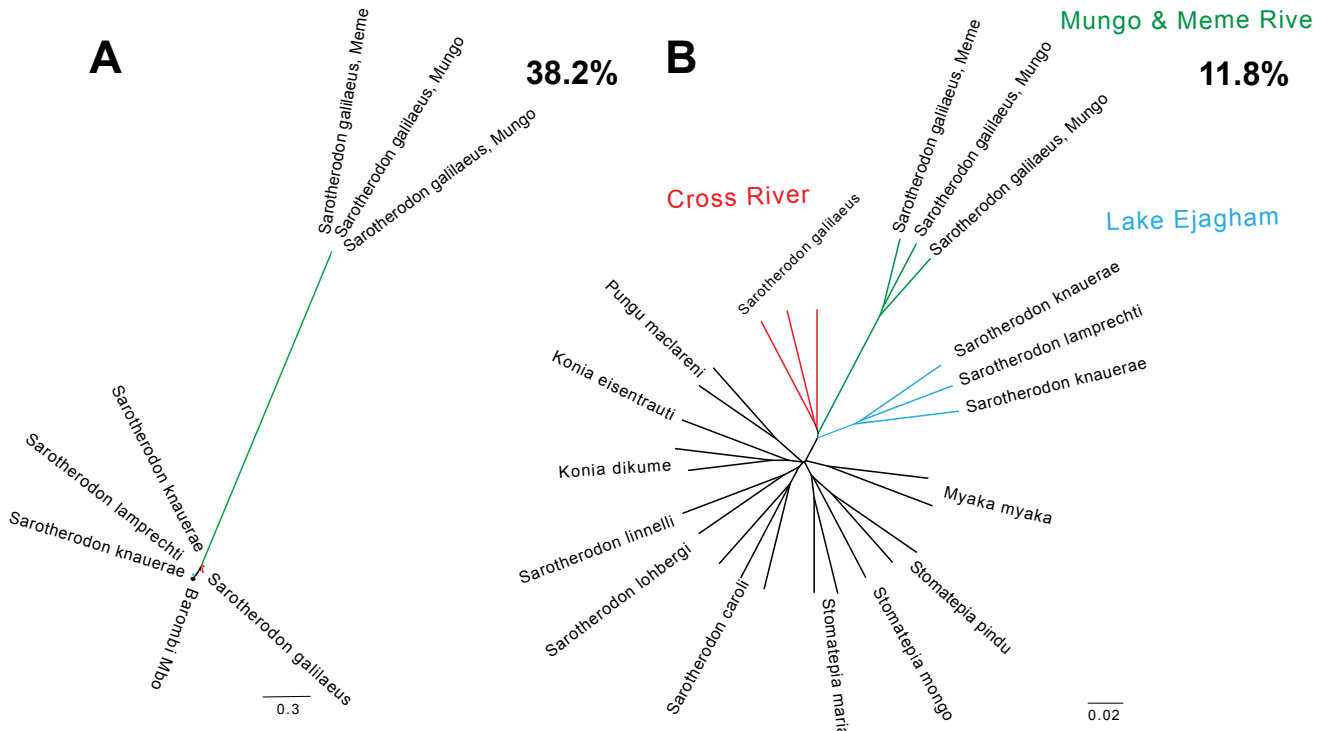
Introgression with riverine outgroups: (A,B) <-> (*MM, CR*)	***f*$_4$ statistic**	***Z*-score**	***P*-value**
S. mariae, S. mongo	$-2.04\times10^{-7} \pm -5.15\times10^{-7}$	-0.39	0.69
S. mariae, S. pindu	$\mathbf{-1.92\times10^{-6} \pm -4.48\times10^{-7}}$	**-4.29**	$\mathbf{1.8\times10^{-5}}$
S. mongo, S. pindu	$\mathbf{-1.59\times10^{-6} \pm -4.98\times10^{-7}}$	**-3.19**	**0.0014**
K. dikume, K. eisentrauti	$\mathbf{-2.4\times10^{-6} \pm -6\times10^{-7}}$	**-4.01**	$\mathbf{6.3\times10^{-5}}$
K. eisentrauti, P. maclareni	$-2.12\times10^{-7} \pm -6.15\times10^{-7}$	0.35	0.73
K. dikume, P. maclareni	$\mathbf{-2.56\times10^{-6} \pm -5.86\times10^{-7}}$	**-4.37**	$\mathbf{1.2\times10^{-5}}$
M. myaka, S. linnelli	$-4.04\times10^{-7} \pm -7.11\times10^{-7}$	0.56	0.57

730 **Table 2**. Candidate introgressed regions in Barombi Mbo cichlid radiation. These regions feature
731 significant f_4 values between riverine populations of *S. galilaeus* (MM: Mungo and Meme River;
732 CR: Cross River) and the three subclades of the radiation focused on in this study. Directionality
733 was determined from f_d outlier values (above 98th percentile) that overlapped with significant f_4
734 peaks. Unannotated regions, regions with no GO terms, and regions that were not f_d outliers are
735 marked with (-). Regions introgressed into a single species within a subclade are highlighted in
736 bold.

Linkage Group	Gene regions	GO Terms	Direction (f_d)
		Stomatepia	
LG2	*sirpb1*	intracellular signal transduction, extracellular exosome	-
LG3	**NA**	**-**	***S. galilaeus* MM <-> *S. pindu***
LG3	uncharacterized protein-coding gene	-	*S. galilaeus* MM <->S. *mongo, mariae*
LG4	***jmjd8*; *prss1***	**extracellular exosome; proteolysis,peptidase activity**	***S. galileaus* CR <-> *S. mariae***
LG7†	NA	-	*S. galileaus* CR <-> *S. pindu, mariae*; *S. galilaeus* MM <- > *S. mongo*
LG14	*cldn4*	plasma membrane; bicellular tight membrane	*S. galilaeus* MM <-> *S. mariae*
LG18	***muc19***	**hematopoietic progenitor cell differentiation; extracellular exosome**	***S. galilaeus* MM <-> *S. mariae***
LG20†	*tprxl*; *plod3*	pseudogene of *tprx1*;motor nueron axon guidance, nueral crest cell migration	*S. galileaus* CR <-> S. pindu, *mariae*; *S. galilaeus* MM<- > *S. mongo*
LG20	*samdh1*	dGTPASe, immune system process, brain hemorrhages	-
NT_168079.1†	*osbpl5*;*ptprj*	lipid transport ; artery morphogenesis, phosphatase activity	*S. galileaus* CR<-> *S. pindu,mariae*; *S. galilaeus* MM<- > *S. mongo*
		Konia + Pungu	
LG2	*sirpb1*	intracellular signal transduction	-
LG4	*anapc2*	ubiquitin dependent catabolic process	*S. galilaeus* MM <-> *P. maclareni*, *K. dikume*
LG6†	*cd209e*	endocytosis,mannose binding	*S. galileaus* CR<-> *K. eisentrauti*, *P. maclareni*; *S. galilaeus* MM<-> *K. dikume*
LG15	***adgrf4***	**g-protien coupled receptor activity, signal transduction**	***S. galilaeus* MM<-> *K. dikume***

LG20†	*klhdc8b*	ubiquitin-protein transferase activity, cytoplasm	*S. galilaeus* CR<-> *P. maclareni*; *S. galilaeus* MM<-> *K. dikume, eisentrauti*
LG20	NA	-	*S. galilaeus* MM <-> *K. dikume, eisentrauti*
LG20	*tprxl*; *plod3;clec10a*	pseudogene of *tprx1*; motor nueron axon guidance, nueral crest cell migration ; immune system response	-
LG22	*ehmt2*	neurogenesis, camera-type eye photoreceptor cell differentiation	-
LG23	***ddn1***	**serine-type endopeptidase activity,proteolysis**	***S. galilaeus* MM<-> *P. maclareni***
NT_167508.1†	*fam159a*	integral component of membrane	*S. galileaus* CR<-> *K. dikume, eisentrauti*; *S. galilaeus* MM<-> *P. maclareni*
NT_167623.1	*mepce*	methyltransferase activity	*S. galilaeus* MM<-> *K. eisentrauti*
NT_167637.1	**NA**	**-**	***S. galilaeus* MM<-> *K. dikume***
NT_167671.1	**NA**	**-**	***S. galileaus* CR<-> *K. dikume***
NT_167702.1*	***pafah1b3*; *hmcn1***	**platelet-activating factor acetyltransferase activity,brain development;fin morphogenesis, calcium ion binding**	***S. galileaus* CR<-> *K. eisentrauti***
NT_168003.1	*p2ry14*	respirotory abnormal muscle physiology, g-protien coupled purinergic nucleotide receptor activity	-
		Myaka + S. linnelli	
LG6†	*c3*	complement activation, inflammatory response	*S. galilaeus* MM<->*S. linnelli*; *S. galileaus* CR<-> *M.myaka*
LG8/LG24†	NA	-	*S. galileaus* CR<->*S. linnelli*; *S. galilaeus* MM<-> *M.myaka*
LG11†	*hfe2*;*txnip*	somite development, BMP signaling pathway; enzyme inhibitor activity	*S. galilaeus* MM<->*S. linnelli*; *S. galileaus* CR<-> *M.myaka*
LG11	*vmo1*	extracellular exosome	*S. galilaeus* MM<-> *M.myaka*
LG16/LG21	***parp4***	**ribosyltranferase activity**	***S. galileaus* CR<->*S. linnelli***
LG19†	NA	-	*S. galileaus* CR<->*S. linnelli*; *S. galilaeus* MM<-> *M.myaka*
LG20	***tprxl,plod3***	**psuedogene of tetra-peptide repeat homeobox ; motor nueron axon guidance, nueral crest cell migration**	***S. galilaeus* MM<-> *M.myaka***
LG20†	*klhdc8b*; *cd163l1*	ubiquitin-protein transferase activity, cytoplasm; receptor mediated endocytosis, scavenger receptor activity	*S. galileaus* CR<->*S. linnelli*; *S. galilaeus* MM<-> *M.myaka*
NT_167617.1†	*ssr1*	endoplasmic reticulum membrane; cotranslational protein targeting	*S. galileaus* CR<->*S. linnelli*; *S. galilaeus* MM<-> *M.myaka*
NT_168092.1	**NA**	**-**	***S. galilaeus* MM<-> *M.myaka***
NT_168092.1	*itgam*	integrin mediated signalling pathway, cell adhesion	-

*best candidate region for secondary gene flow contributing to diversification † regions with signatures of differential sorting of polymorphism from putative hybrid swarm



740

741

742

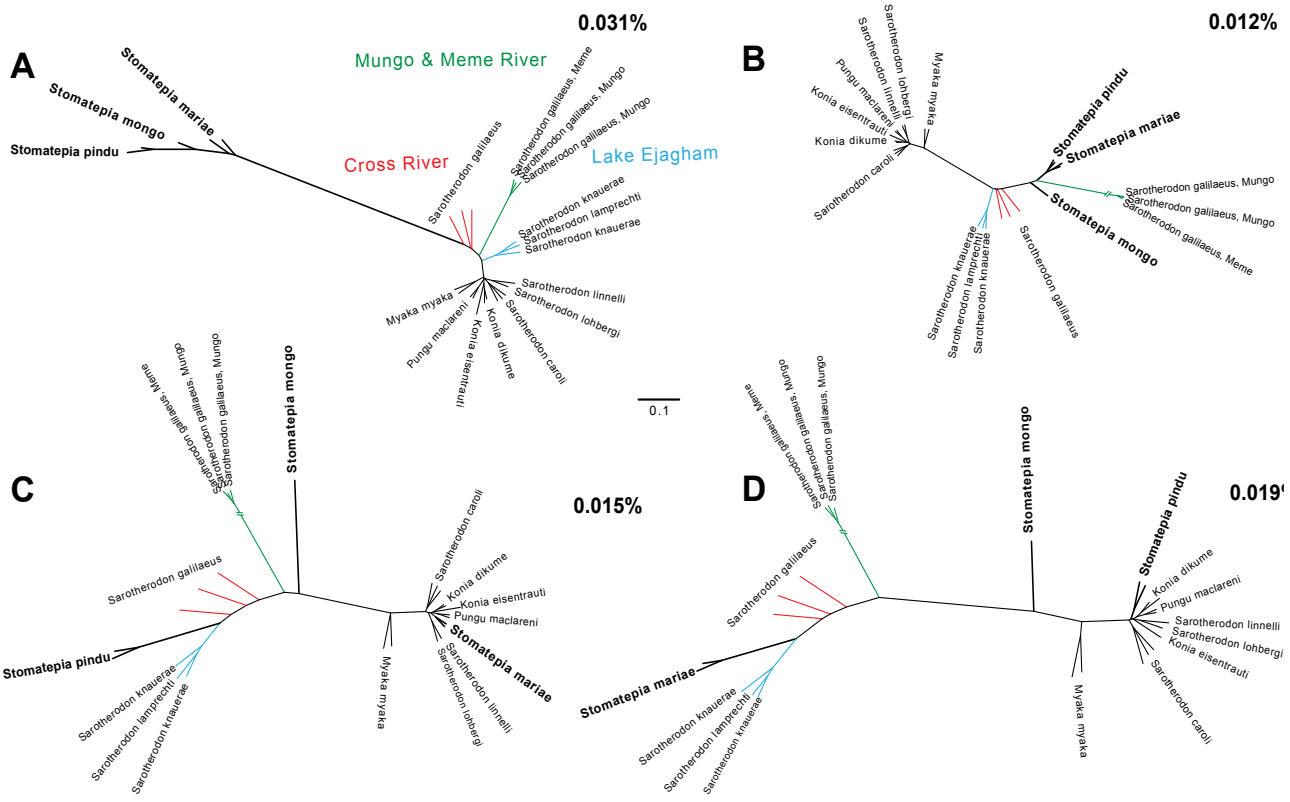
743 Fig 1. The most common topologies feature a monophyletic Barombi Mbo radiation.

744 Across 96.2% of the genome Barombi Mbo species (black) are more closely related to each other

745 than riverine outgroup populations of *S. galilaeus* Mungo and Meme River (green) and *S.*

746 *galileaus* Cross River (red), or the Lake Ejagham *Sarotherodon* radiation (blue).

747



748

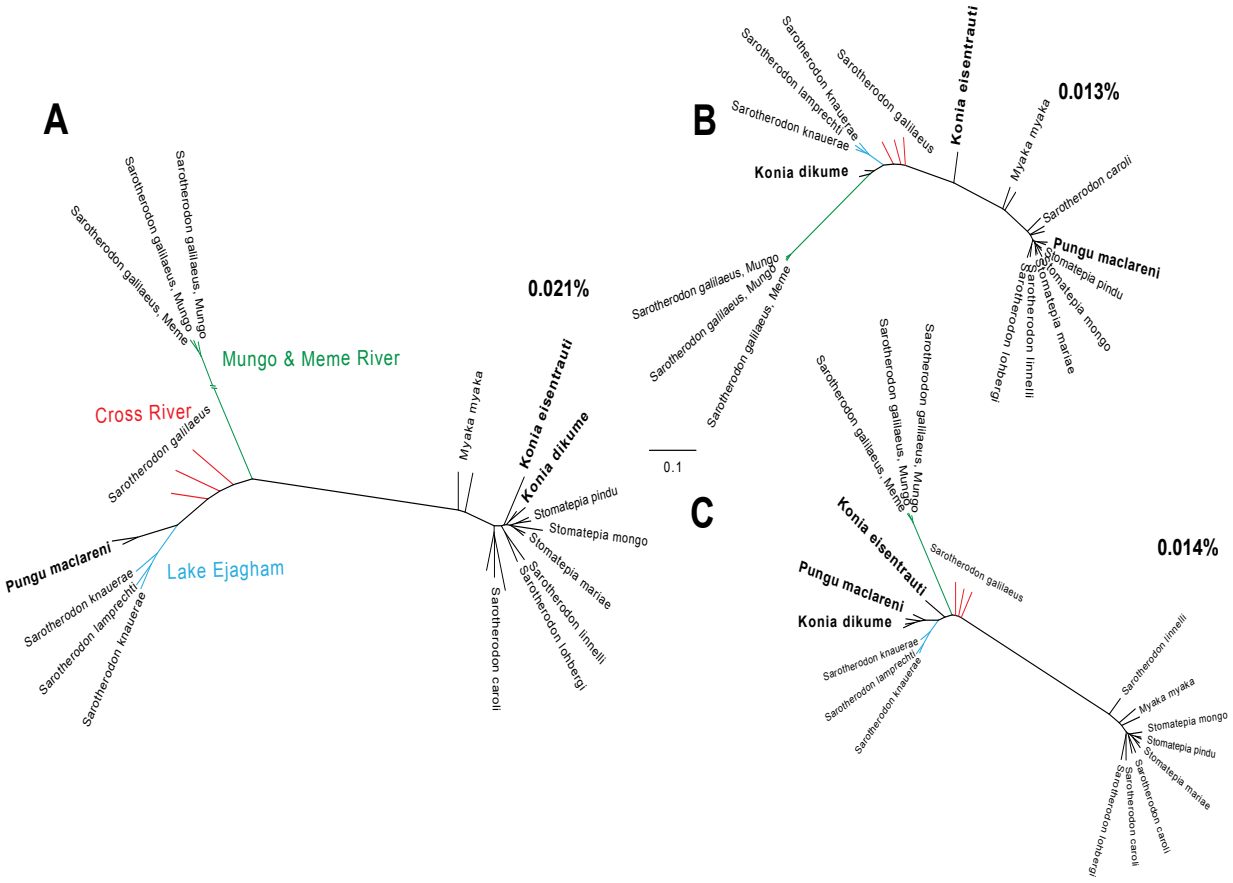
749 Fig 2. Topologies featuring Barombi Mbo polyphyly with riverine populations involving

750 the *Stomatepia* three-species complex. Across small and independent proportions of the genome

751 A-B) the entire *Stomatepia* clade, C) only *S. pindu*, and D) only *S. mariae* are more closely

752 related to outgroups than other Barombi Mbo species.

753



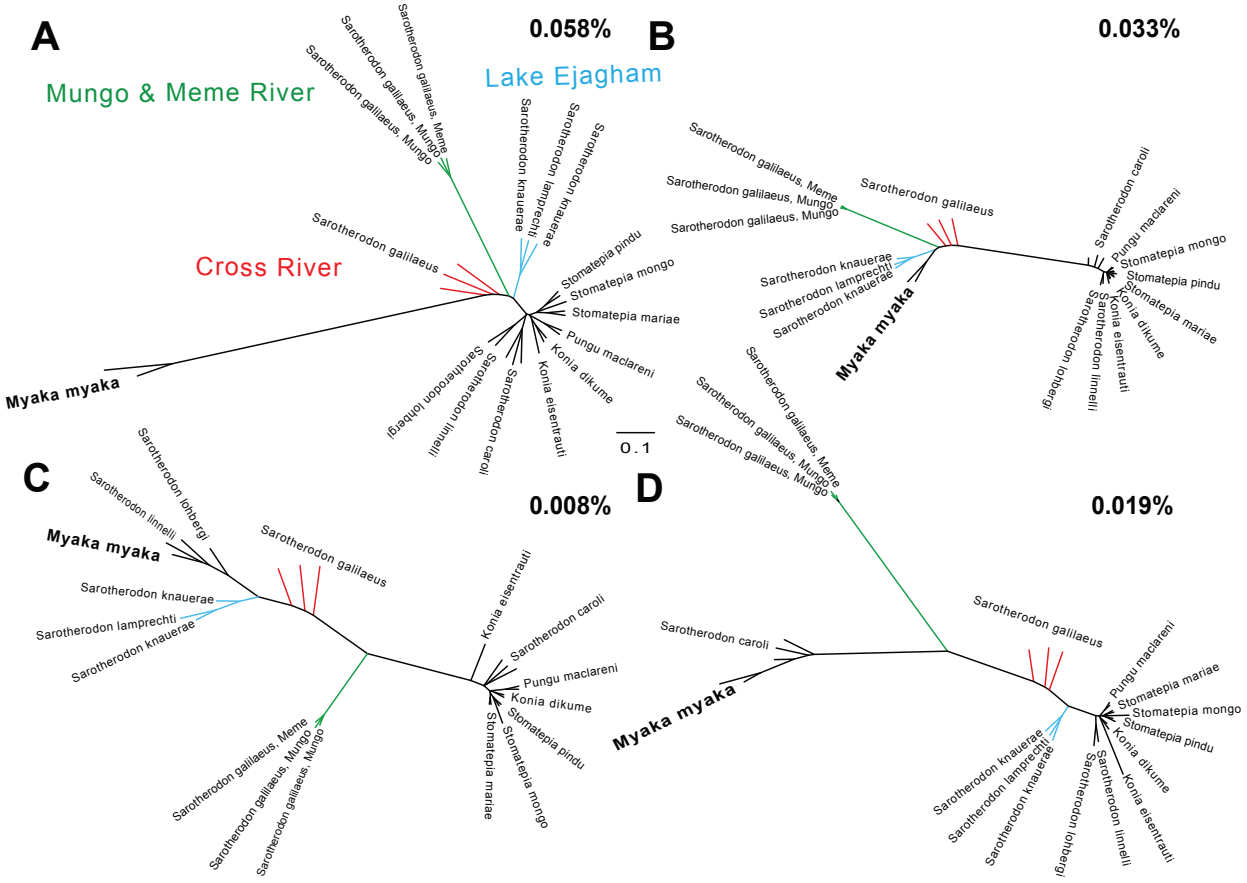
754

755 Fig 3. Topologies featuring Barombi Mbo polyphyly with riverine populations involving

756 the *Konia* + *Pungu* subclade. Across small and independent proportions of the genome A) only

757 *P. maclareni*, B) only *K. dikume*, and C) the entire *Konia* + *Pungu* subclade are more closely

758 related to outgroups than other Barombi Mbo species.



759

760

761 Fig 4. Topologies featuring Barombi Mbo polyphyly with riverine populations involving

762 the *Myaka* + *Sarotherodon* subclade. Across small and independent proportions of the genome

763 A-B) only *M. myaka*, C) *M. myaka* and two Barombi Mbo *Sarotherdon* species (*S. linelli* and *S.*

764 *lohbergi*), and D) *M. myaka* and *S. caroli* are more closely related to outgroups than other

765 Barombi Mbo species.

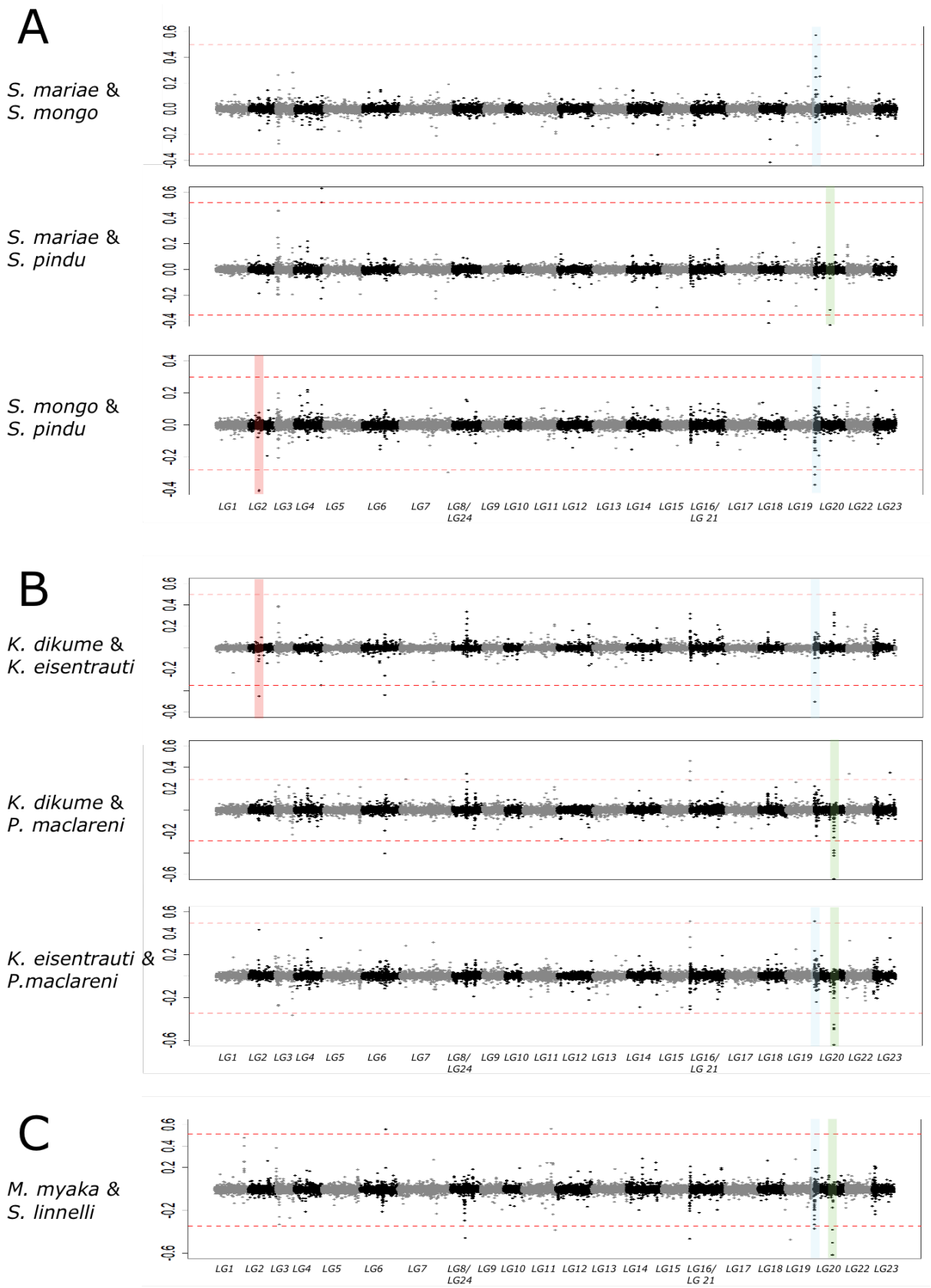
766

767

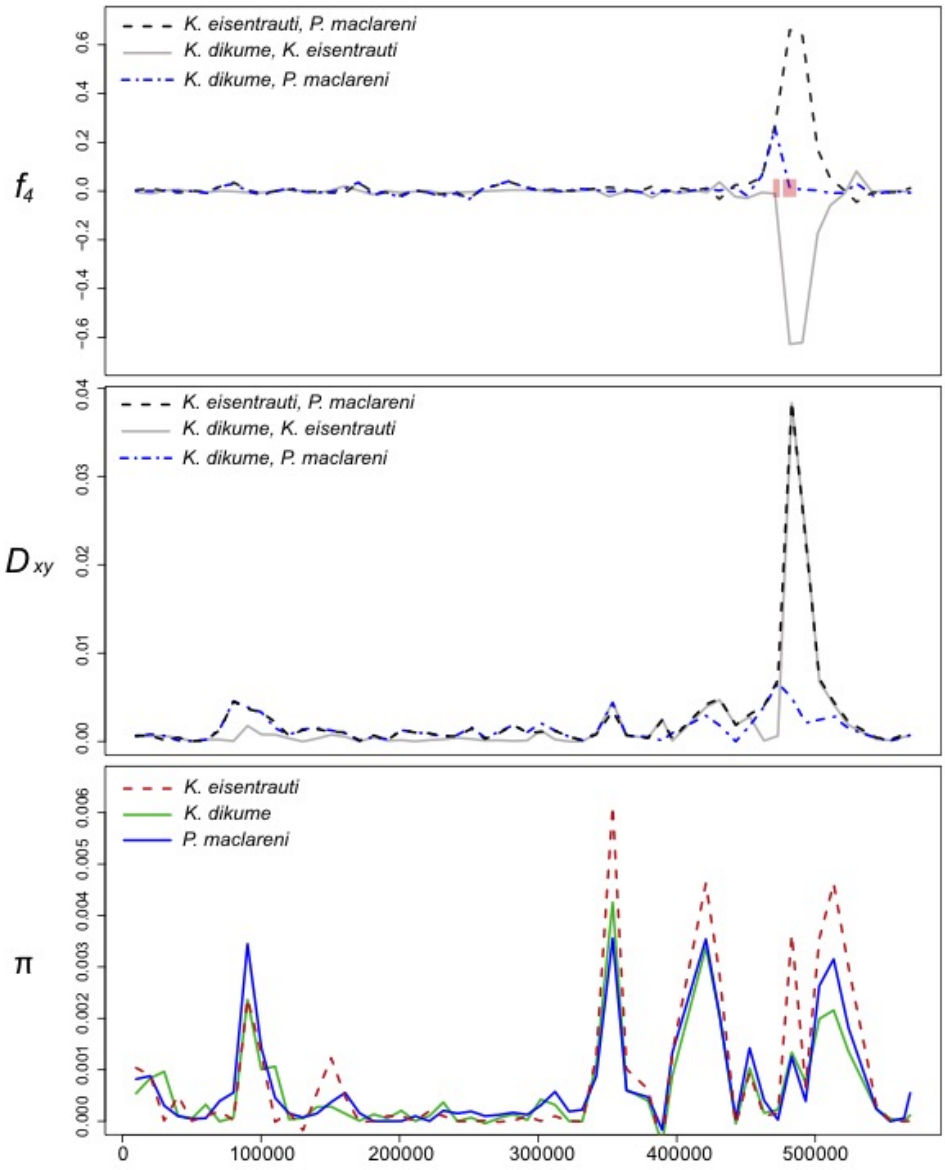
768

769

770



772 Figure 5. Manhattan plots of f_4 values between riverine populations of *S. galilaeus* from
773 Mungo and Meme river and *S. galileaus* from Cross River and A) combinations of the three
774 species of *Stomatepia*, B) combinations of the three species in the *Konia+Pungu* subclade, and
775 C) *M. myaka* and a representative from its sister *Sarotherodon* clade: *S. linnelli*. Alternating
776 gray/black colors indicate different linkage groups. Dotted red lines mark the permutation-based
777 significance thresholds for each test (P = 0.02). Peaks highlighted in colors represent those
778 signals of introgression shared across different subclades. Manhattan plots for the scaffolds not
779 assigned to the 24 linkage groups are presented in Fig S2-4.



780

781 Figure 6. Candidate introgression region in the *Konia* + *Pungu* subclade of Barombi Mbo

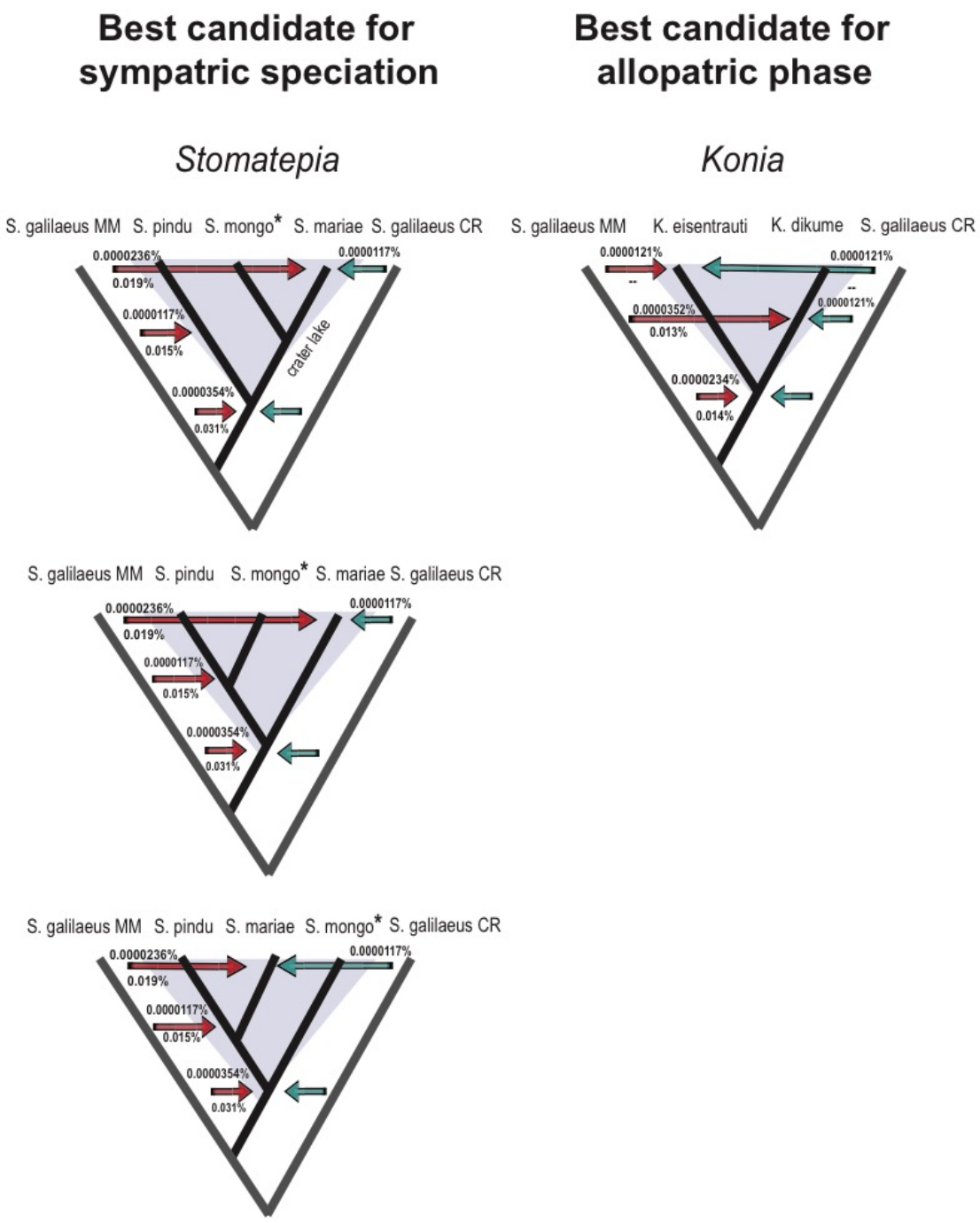
782 region containing genes *pafah1b3* and *hmcn1*. Row 1 shows the peak signal of introgression

783 across scaffold NT_167702.1 detected from the f_4 statistic across the three test combinations

784 involving the three species in the *Konia* + *Pungu* subclade and riverine populations of *S.*

785 *galilaeus* from Mungo, Meme River, and Cross River in non-overlapping 10-kb windows. The

two genes in this peak are shown in red (*pafah1b3* on the left and *hmcn1* on the right). Row 2 shows between-population divergence (D_{xy}) among the three combinations of sister species in the subclade calculated in non-overlapping 10-kb windows. Row 3 shows within-population diversity (π) in the same non-overlapping 10-kb windows.



794

795 Figure 7. The contributions of differential and shared introgressed variation to the

796 genomes of sister species in the genera *Stomatepia* and *Konia*. The left and right columns

797 represent the best candidate sister species in the radiation for sympatric speciation and speciation

798 with secondary gene flow respectively. Species with no evidence for differential introgression
799 are highlighted with an asterick (*). The true sister species in the *Stomatepia* are unknown, so all
800 three potential relationships among the complex are shown. Outgroup riverine *S. galilaeus*
801 lineages from Mungo/Meme Rivers (MM) and Cross River (CR) are shown with gray branches.
802 The percentage of introgression consistent with secondary gene flow into single species and a
803 hybrid swarm from sliding window f_4 statistic tests is show by above arrows. Proportion of
804 genome from two riverine sources (red: Mungo/Meme River; blue: Cross River) estimated by
805 SAGUARO is shown below arrows. Introgression patterns not found in the SAGUARO analysis
806 are represented with dashes (--). Blue arrows with no information that are aligned with red
807 arrows represent patterns found consistent with hybrid swarm. Note that the separation of arrows
808 along a branch is for the purpose of clarity and doesn't represent known differences in the timing
809 of introgression.
810