Supplementary Methods

Sampling and Genome Sequencing

We sequenced whole genomes of 1-3 individuals from 10 out of the 11 species within the sympatric radiation of Oreochromini cichlids in Cameroon crater lake Barombi Mbo (excluding *Sarotherodon steinbachi* which is morphologically and ecologically similar to the other three *Sarotherodon* species), an endemic *Sarotherodon* species pair from Lake Ejagham, and outgroup *Sarotherodon* individuals from all three river drainages flanking the lake: Cross, Meme, and Mungo rivers (e.g. see map in (Schliewen et al. 1994)). Individual cichlids were caught by seine, gill net, or hook-and-line from Barombi Mbo, Lake Ejagham, Cross River, Mungo River, and Meme River in January, 2010 and July, 2016. Fishes were euthanized in an overdose of buffered MS-222 (Finquel, Inc.) following approved protocols from University of California, Davis Institutional Animal Care and Use Committee (#17455) and University of North Carolina Animal Care and Use Committee (#15-179.0) and whole specimens or tissue samples were stored in 95-100% ethanol or RNAlater (Ambion, Inc.) in the field.

DNA was extracted from muscle tissue using DNeasy Blood and Tissue kits (Qiagen, Inc.) and quantified on a Qubit 3.0 fluorometer (Thermofisher Scientific, Inc.). Genomic libraries were prepared using the automated Apollo 324 system (WaterGen Biosystems, Inc.) at the Vincent J. Coates Genomic Sequencing Center (QB3). Samples were fragmented using Covaris sonication, barcoded with Illumina indices, and quality checked using a Fragment Analyzer (Advanced Analytical Technologies, Inc.). Nine to twelve samples were pooled in four different libraries for 150PE sequencing on four lanes of an Illumina Hiseq4000.

1.9 billion raw reads were mapped from 28 individuals to the *Oreochromis niloticus* reference genome v1.1 (NCBI, total sequence length = 927,679,487; number of scaffold = 5,909,

scaffold N50, = 2,766,223; contig N50 = 29,493) with the Burrows-Wheeler Alignment Tool (BWA-MEM v 0.7.15) (Li and Durbin 2009; Li 2013). Duplicate reads were identified using MarkDuplicates and BAM indices were created using BuildBamIndex in Picard Tools (v 2.10.3, http://broadinstitute.github.io/picard). We followed the best practices guide (Van der Auwera et al. 2002) recommended for the Genome Analysis Toolkit (v 3.5) (DePristo et al. 2011) to call and refine our SNP variant dataset using the program HaplotypeCaller. Because we lacked high-quality known variants for these non-model species, we filtered SNPs based on the recommended hard filter criteria (i.e. QD < 2.0; FS < 60; MQRankSum < -12.5; ReadPosRankSum < -8) (DePristo et al. 2011; Marsden et al. 2014) using vcftools "—remove-filtered-all" flag (v 0.1.14) (Danecek et al. 2011). We also removed SNPs that differed from the reference but not among focal samples using "max-non-ref-af 0.99" in vcftools, as well as SNPs with more than two alleles using "-m2 –M2" flags in bcftools (v 1.5) (Li 2011).

Characterization of introgression patterns across the genome with SAGUARO

First, we exhaustively searched the genomes for patterns of non-monophyletic Barombi Mbo relationships using the machine learning program SAGUARO (Zamani et al. 2013) to identify regions of the genome that contained relationships consistent with expectations from multiple colorizations and secondary gene flow into the radiation (i.e. paraphyletic/polyphyletic Barombi Mbo radiations). Saguaro combines a hidden Markov model with a self-organizing map to characterize variation in phylogenetic relationships among individuals across the genome without requiring *a priori* hypotheses about these relationships or the size of genomic regions. This method infers relationships among individuals in the form of genetic distance matrices and assigns segments across the genomes to different topologies. These genetic distance matrices

can then be transformed into neighborhood joining trees to visualize patterns of evolutionary relatedness across the genome. We exhaustively searched the genome for topological variation by partitioning the genome into a total of 75 unique topologies (well past the inflection point at 30 topologies where the percent of genome explained by each additional topology plateaus; Fig S1). Since smaller segments with fewer informative sites are more likely to be incorrectly assigned to a hypothesized topology by chance, we tested various minimum SNP filters (1,10,20 SNPs) for reducing the amount of short uninformative segments and their effect on the percentage of the genome assigned to topologies. We found that while the percentage of the genome assigned to topologies changes when we apply SNP filters, none of the topologies had all segments entirely removed and the relative proportions of the genome assigned to particular types of topologies were similar across filtering strategies (Table S1). This may be due to uninformative sites being assigned to topologies in a random fashion, such that removing these sites changes the percentage of the genome disproportionally across topologies. The percentages indicated in the results represent the percentage of base pairs in the genome assigned to topologies after using a 20 SNP minimum filter and represent conservative estimates. We searched these 75 topologies for evidence of relationships where subclades or individual Barombi Mbo species were more closely related to riverine populations than other species in the crater lake, suggesting sympatric speciation after a hybrid swarm (i.e. differential sorting of ancestral polymorphism) or secondary gene flow into this subclade (introgression).

Characterization of introgression patterns across the genome with sliding windows of f4 statistic

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 using a custom python script (modified from ABBABABA.py created by Simon H. Martin, available on https://github.com/simonhmartin/genomics_general; our modified version is provided in the supplementary materials of (Richards and Martin 2017)), using the population allele frequencies of biallelic SNPs, allowing for a minimum of 50 variant sites and up to 10% missing data within a population per site.

We conducted 1,000 permutations of the f_4 test to evaluate the significance of f_4 values in sliding windows across the genome. For each permutation, individuals from the four original populations were randomly reassigned without replacement to one of the four populations based on the tree ((P1,P2),(P3,P4)) to assess how likely a given f_4 value would be observed by chance within our empirical dataset. We calculated the 1% tails of this null distribution and used these thresholds for our candidate introgressed regions (i.e. significant at alpha = 0.02). We assigned each candidate introgressed region a *P*-value by counting the number of permutations with a f_4 value greater than (or lesser than for a negative f_4 value) or equal to the observed value.

Supplementary Figures

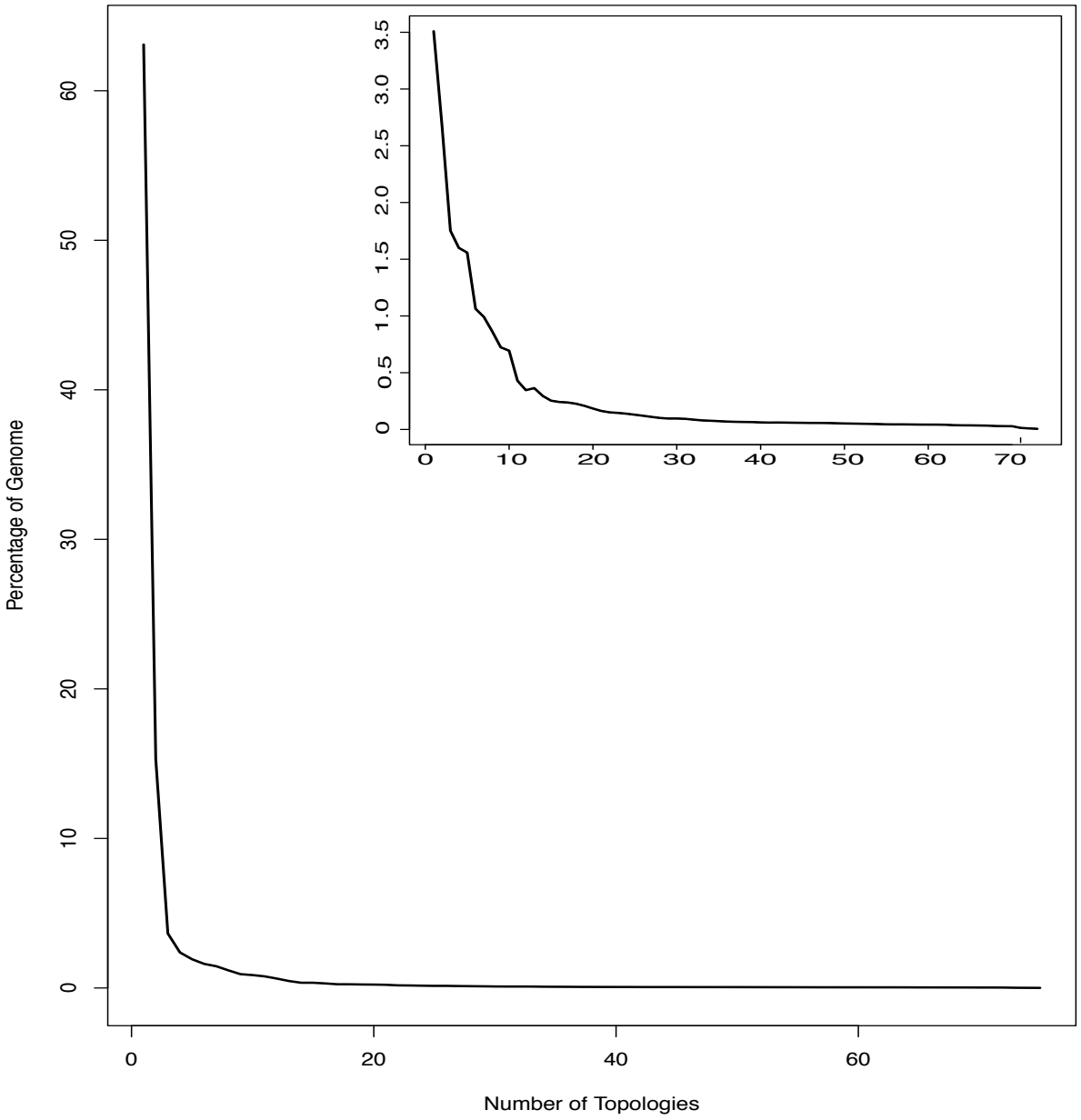

Fig S1. The proportion of the genome assigned to each topology by SAGUARO. The insert is a closer look at all but the two topologies assigned to the largest proportion of the genome (65% and 15%). This suggests saturation in the variance explained by topologies at around 20 proposed topologies (P=0.02).

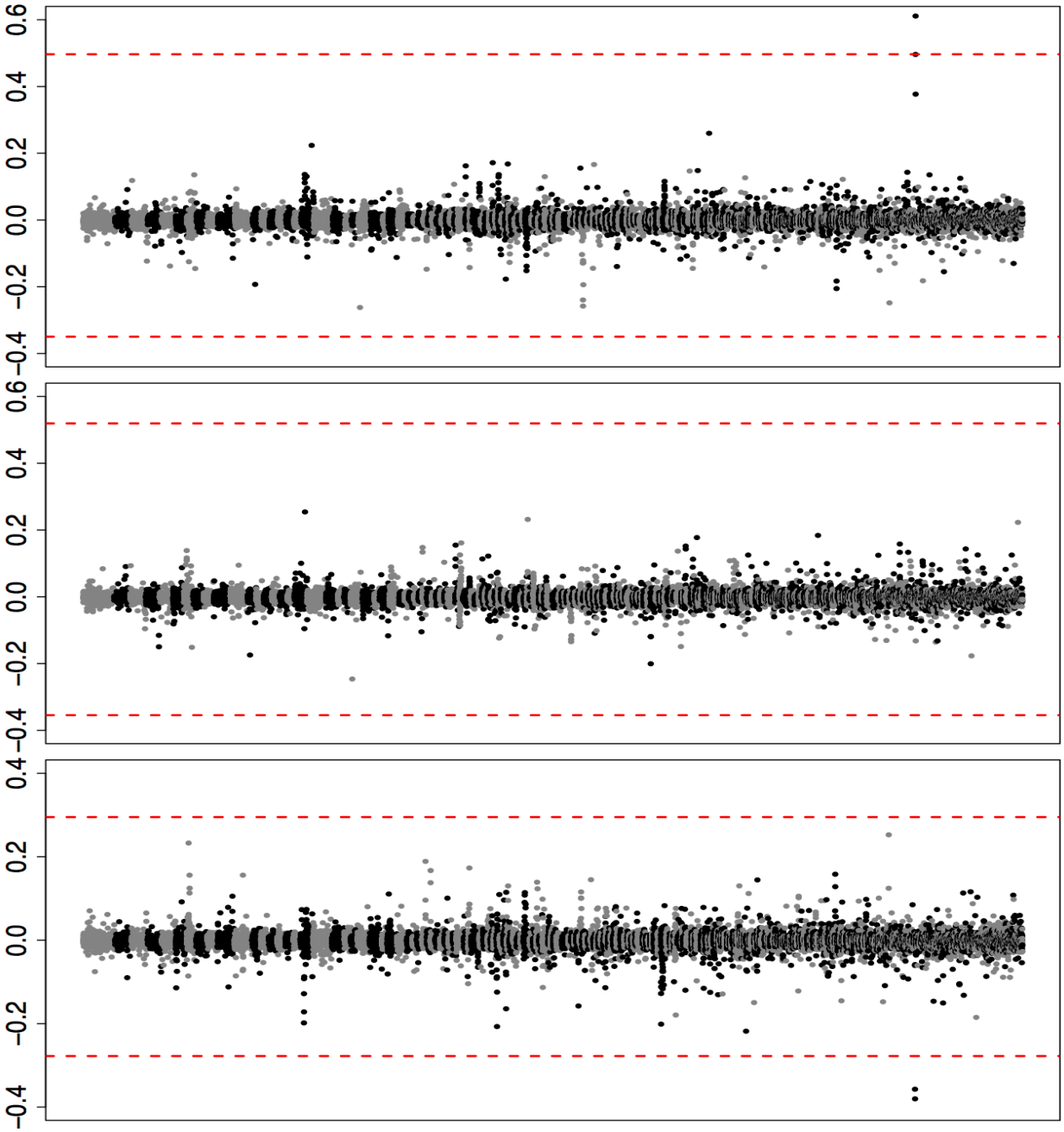

Fig S2. Visualization of introgression across unassigned scaffolds for *Stomatepia* subclade. Manhattan plots of the f_4 values between riverine populations of *S. galilaeus* from Mungo and Meme river and *S.galileaus* from Cross River with combinations of *Stomatepia* species pairs: *S. mariae* and *S. mongo* (top), *S. mongo* and *S. pindu* (middle), and *S. mariae* and *S. pindu* (bottom). Dotted red lines mark the permutation-based significance thresholds for each test (P=0.02).

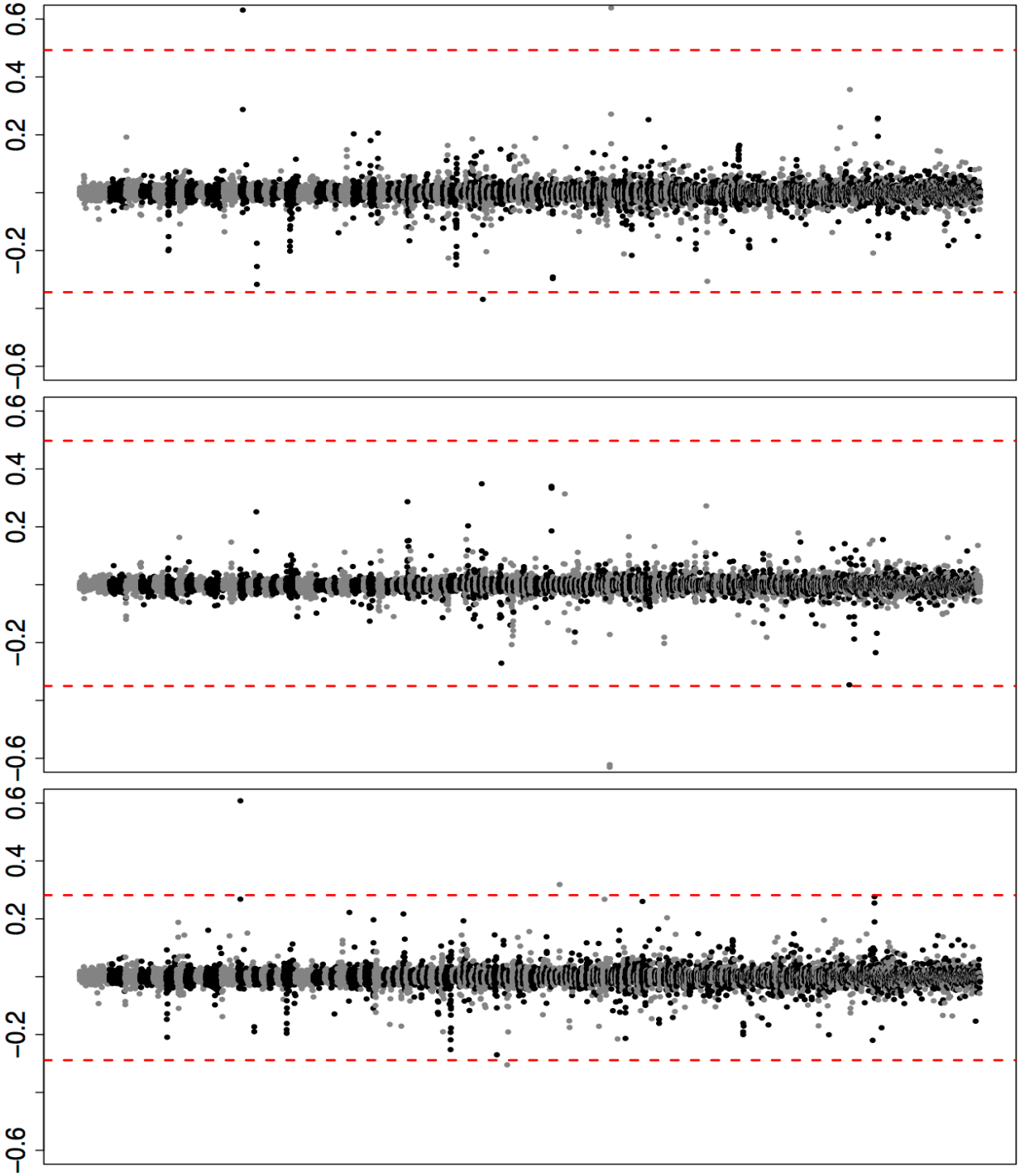

Fig S3. Visualization of introgression across unassigned scaffolds for *Konia + Pungu* subclade. Manhattan plots of the f_4 values between riverine populations of *S. galilaeus* from Mungo and Meme river and *S. galileaus* from Cross River with combinations of *Konia + Pungu* species pairs: *K. eisentrauti* and *P. maclareni* (top), *K. dikume* and *K. eisentrauti* (middle), and

K. dikume and *P. maclareni* (bottom). Dotted red lines mark the permutation-based significance thresholds for each test (P=0.02).

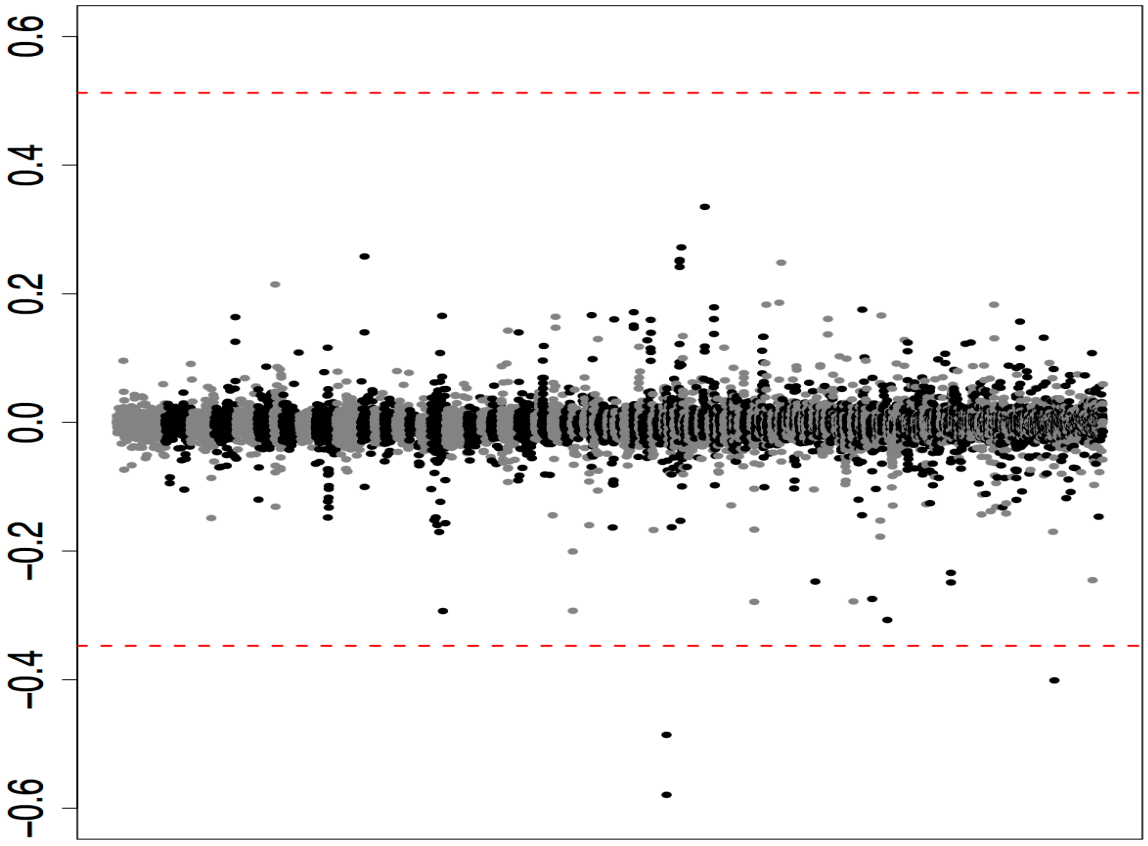

Fig S4. Visualization of introgression across unassigned scaffolds for *Myaka + Sarotherodon* subclade. Manhattan plots of the f_4 values between riverine populations of *S. galilaeus* from Mungo and Meme river and *S. galileaus* from Cross River with M. myaka + *S. linnelli* as a representative of the Barombi Mbo Sarotherodon sister clade to *M. myaka*. Dotted red lines mark the permutation-based significance thresholds for each test (P=0.02).

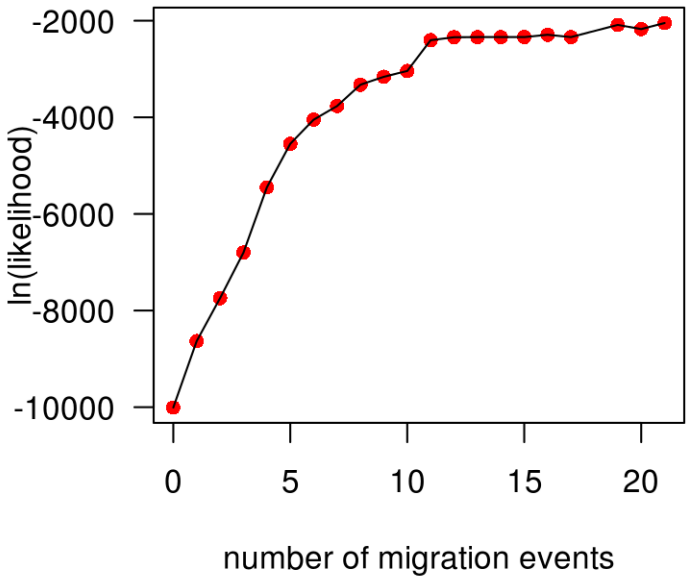

Fig S5. The log-likelihood of a Treemix population graph of Barombi Mbo cichlids as a function of the number of migration events.

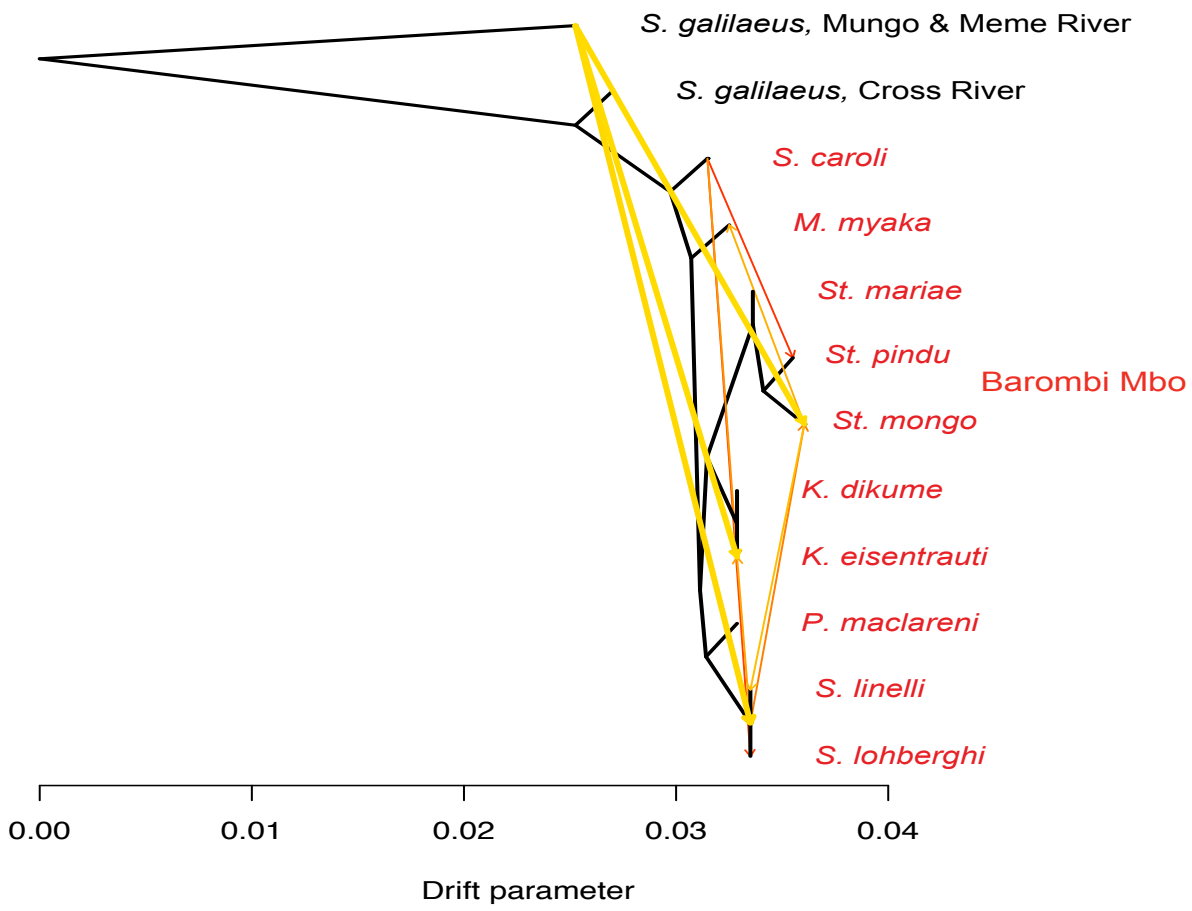

Fig S6. Visualization of genome-wide introgression from riverine *Sarotherodon* populations into Barombi Mbo radiation. Treemix graph illustrating 10 admixture events (with heat colors indicating intensity) on a population graph of the radiation. Admixture events from riverine populations into the radiation are indicated with thicker arrows.

Supplementary Tables

Table S1. Percentages of the genome assigned to topologies under various filtering criteria. The relative percentage (out of all segments assigned in SAGUARO analysis) compared to the absolute percentage (out of the entire genome length 927,679,487) for each unique topology featuring either a monophyletic or non-monophyletic Barombi Mbo. We calculated both percentages under different filtering criteria of each segment assigned to a topology having at least one SNP, greater or equal to than 10 SNPs or 20 SNPs.

Monophyletic						
Topology ID	Relative Percentage			Absolute Percentage		
	None	10 SNPs	20 SNPs	None	10 SNPs	20 SNPs
1	0.172	0.049	0.023	0.151	0.033	0.012
2	0.045	0.042	0.017	0.039	0.028	0.009
4	0.246	0.189	0.168	0.216	0.128	0.090
6	1.078	0.279	0.135	0.948	0.189	0.072
11	0.344	0.093	0.044	0.303	0.063	0.024
12	0.374	0.125	0.091	0.329	0.085	0.049
15	0.726	0.246	0.170	0.639	0.167	0.091
16	0.994	0.319	0.210	0.875	0.216	0.113
19	0.407	0.153	0.100	0.358	0.104	0.054
20	1.677	1.581	1.572	1.476	1.072	0.844
26	0.703	0.371	0.224	0.619	0.252	0.120
29	0.172	0.096	0.067	0.152	0.065	0.036
35	0.855	0.247	0.147	0.752	0.167	0.079
42	0.061	0.035	0.027	0.054	0.024	0.014
46	0.249	0.283	0.269	0.219	0.192	0.145
54	3.504	1.735	1.003	3.084	1.176	0.538
55	0.070	0.033	0.024	0.061	0.023	0.013
60	0.060	0.044	0.037	0.053	0.030	0.020
62	0.051	0.028	0.017	0.045	0.019	0.009
Fig. 1B (65)	**15.195**	**18.624**	**22.089**	**13.375**	**12.622**	**11.855**
67	1.689	0.559	0.120	1.486	0.379	0.064
68	2.700	1.016	0.287	2.377	0.069	0.154
Fig. 1B (69)	**63.184**	**71.224**	**71.340**	**55.614**	**48.271**	**38.289**
70	0.045	0.024	0.019	0.039	0.016	0.010

71	1.508	0.762	0.613	1.327	0.517	0.329
72	0.036	0.017	0.010	0.032	0.012	0.006
Non-Monophyletic						
Topology ID	Relative Percentage			Absolute Percentage		
	None	10 SNPs	20 SNPs	None	10 SNPs	20 SNPs
0	0.006	0.002	0.001	0.005	0.001	0.001
Fig. 4B (3)	**0.224**	**0.087**	**0.061**	**0.197**	**0.059**	**0.033**
5	0.058	0.053	0.027	0.051	0.036	0.014
Fig. 3B (7)	**0.147**	**0.042**	**0.024**	**0.130**	**0.029**	**0.013**
8	0.057	0.024	0.021	0.050	0.016	0.011
9	0.029	0.022	0.022	0.025	0.015	0.012
Fig. 2C (10)	**0.121**	**0.048**	**0.029**	**0.107**	**0.033**	**0.015**
Fig. 3A (13)	**0.145**	**0.065**	**0.040**	**0.127**	**0.044**	**0.021**
14	0.141	0.070	0.055	0.124	0.047	0.029
Fig. 4A (17)	**0.305**	**0.152**	**0.107**	**0.269**	**0.103**	**0.058**
18	0.095	0.042	0.031	0.084	0.028	0.016
21	0.096	0.038	0.021	0.085	0.026	0.011
22	0.044	0.030	0.024	0.039	0.020	0.013
23	0.076	0.037	0.018	0.067	0.025	0.010
24	0.043	0.017	0.010	0.038	0.011	0.005
25	0.128	0.047	0.026	0.113	0.032	0.014
27	0.059	0.031	0.022	0.052	0.021	0.012
28	0.057	0.035	0.025	0.050	0.024	0.013
Fig. 2D (30)	**0.096**	**0.048**	**0.036**	**0.084**	**0.032**	**0.019**
31	0.038	0.017	0.009	0.033	0.012	0.005
32	0.051	0.040	0.016	0.045	0.027	0.008
33	0.065	0.026	0.016	0.058	0.018	0.008
34	0.045	0.032	0.021	0.040	0.022	0.011
36	0.042	0.030	0.026	0.037	0.021	0.014
37	0.053	0.023	0.017	0.047	0.016	0.009
Fig. 4C (38)	**0.054**	**0.023**	**0.015**	**0.048**	**0.016**	**0.008**
Fig. 3C (39)	**0.101**	**0.041**	**0.026**	**0.089**	**0.028**	**0.014**
40	0.048	0.025	0.019	0.042	0.017	0.010
Fig 4D (41)	**0.110**	**0.052**	**0.035**	**0.097**	**0.035**	**0.019**
Fig. 2A (43)	**0.216**	**0.104**	**0.057**	**0.190**	**0.070**	**0.031**
44	0.081	0.041	0.020	0.072	0.027	0.011
45	0.057	0.029	0.020	0.050	0.020	0.011
47	0.073	0.048	0.031	0.064	0.033	0.017
48	0.029	0.021	0.009	0.025	0.014	0.005

49	0.035	0.019	0.015	0.031	0.013	0.008
50	0.066	0.029	0.023	0.058	0.020	0.013
Fig. 2B	**0.236**	**0.062**	**0.022**	**0.207**	**0.042**	**0.012**
52	0.031	0.017	0.014	0.027	0.011	0.007
53	0.061	0.028	0.017	0.054	0.019	0.009
56	0.067	0.036	0.025	0.059	0.025	0.013
57	0.084	0.043	0.033	0.074	0.029	0.018
58	0.036	0.021	0.012	0.032	0.014	0.006
59	0.041	0.031	0.018	0.036	0.021	0.010
61	0.061	0.023	0.015	0.054	0.015	0.008
63	0.048	0.023	0.016	0.043	0.016	0.009
64	0.034	0.018	0.011	0.030	0.012	0.006
66	0.043	0.020	0.014	0.038	0.014	0.008
73	0.014	0.006	0.003	0.012	0.004	0.002
74	0.009	0.006	0.004	0.008	0.004	0.002

Table S2. Within-population genetic diversity in introgressed regions in *Stomatepia*.

Introgressed regions with lower genetic diversity than genome-wide and linkage group averages are highlighted in bold.

Stomatepia								
Genome-wide			*S. mariae* (0.00057)		*S. pindu* (0.00055)		*S. mongo* (0.00047)	
Linkage Group	Gene	Region	f_4 region	LG avg.	f_4 region	LG avg.	f_4 region	LG avg.
LG2	sirpb1	9443045-9453042	0.00028	0.0004	0.00022	0.00039	0.00040	0.00027
LG3	uncharacterized psuedogene	3362824-3382782	0.00262	0.00074	0.00285	0.00079	0.00219	0.00059
LG3	uncharacterized protein-coding genes	3442935-3452925	0.00508	0.00074	0.00506	0.00079	0.00039	0.00059
LG4	jmjd8; prss1	26603240-26620710	0.00014	0.00048	0.00020	0.00048	0.00049	0.00041
LG7	NA	46268722-46278711	0	0.00037	0.000051	0.00031	0	0.00026
LG14	cldn4	28253272-28263253	0	0.00046	0.000078	0.00044	0.00016	0.00037
LG18	muc19	9354127-9364106	0.00032	0.00039	0.00015	0.00035	0.00051	0.00026
LG20	tprxl,plod3	174603-203401	0.00005	0.00039	0.00078	0.00037	0	0.00031

LG20	samdh1	15531162-15541130	0.00767	0.00039	0.00531	0.00037	0.01271	0.00031
NT_168079.1	osbpl5;ptprj	63823-92214	0.00038	0.0012	0.00135	0.0016	0.00031	0.0016

Table S3. Within-population genetic diversity in introgressed regions in *Konia + Pungu* subclade. Introgressed regions with lower genetic diversity than genome-wide and linkage group averages are highlighted in bold.

Konia + Pungu								
Genome-wide			*K. eisentrauti (0.00064)*		*K. dikume (0.00053)*		*P. maclareni (0.00056)*	
Linkage Group	gene	region	f_4 region	LG avg.	f_4 region	LG avg.	f_4 region	LG avg.
LG2	sirpb1	9443045-9453042	0.000219	0.00038	0.00035	0.00035	0.00054	0.00043
LG4	anapc2	25971652-25981605	0	0.00058	0.00029	0.00046	0.00016	0.0005
LG6	cd209e	21809711-21819698	0	0.00043	0	0.00036	0.00065	0.0004
LG15	adgrf4	26665600-26675599	0.000111	0.0004	0.00037	0.00033	0.00028	0.0004
LG20	klhdc8b	19333891-19393955	0.0001	0.00041	0.00442	0.00036	0.00053	0.0004
LG20	NA	19594524-19604391	0.0001	0.00041	0.00012	0.00036	0.00108	0.0004
LG20	clec10a	216267-226260	0	0.00041	0	0.00036	0	0.0004
LG22	ehmt2	2836001-2855000	0.0004	0.00051	0.00005	0.00043	0.00036	0.00047
LG23	ddn1	16052001-16071000	0	0.00057	0.00041	0.00047	0.00131	0.00047

NT_167508.1	fam159a	539219-549188	0	0.00053	0.00037	0.0004	0.00085	0.00047
NT_167623.1	mepce	633481-643444	0	0.0035	0.00264	0.0022	0.00097	0.0018
NT_167637.1	NA	285001-305000	0.006145	0.0037	0.00106	0.0016	0.00065	0.0011
NT_167671.1	NA	448001-468000	0.00005	0.0019	0.00031	0.0017	0.00029	0.0018
NT_167702.1	pafah1b3; hmcn1	472965-492982	0.002152	0.00081	0.00105	0.00056	0.00082	0.00068
NT_168003.1	p2ry14	67005-76962	0.000742	0.0022	0.00095	0.002	0.00127	0.0021

Table S4. Within-population genetic diversity in introgressed regions in *Myaka + Sarotherodon* subclade. Introgressed regions with lower genetic diversity than genome-wide and linkage group averages are highlighted in bold.

Myaka + Sarotherodon						
Genome-wide			*M. myaka (0.00071)*		*S. linnelli (0.00059)*	
Linkage Group	gene	region	f_4 region	LG avg.	f_4 region	LG avg.
LG6	complement c3	24943846-24953844	0.0013	0.0005	0.0002	0.00037
LG8/LG24	NA	14512082-14520290	0.0001	0.0005	0.00024	0.00036
LG11	hfe2;txnip	28350732-28388849	0.00036	0.00051	0.00048	0.00037
LG11	vmo1	32592649-32602646	0.00039	0.00051	0.0001	0.00037
LG16/LG21	**parp4**	**77792-87208**	**0.00026**	**0.00052**	**0.00022**	**0.00036**
LG19	NA	4530466-4537002	7.65E-05	0.00048	0	0.00033
LG20	**tprxl,plod3**	**167380-**177288	**0.00028**	**0.00048**	**0.00041**	**0.00033**
LG20	klhdc8b; CD163L1	19334324-19601586	5.97E-05	0.00048	0	0.00033
NT_167617.1	ssr1	329871-359848	0.00081	0.0016	0.00024	0.0013

NT_168092.1	itgam	26809-59646	0	0.0069	0	0.0027