

Supplementary Table S1

Table outlines species names, bower behavior type, source for description of bower behavior, source of sample, and, if wild caught, source location at Lake Malawi.

Species	Bower Behavior	Behavior source	Sample source	Location
<i>Mchenga conophorus</i>	Castle	Lab Observation	Lab bred	NA
<i>Ctenopharynx nitidus</i>	Castle	Konings(1)	Martin Genner [#17]	Metangula
<i>Copadichromis likomae</i>	Castle	Konings(1)	Martin Genner [#52]	Metangula
<i>Copadichromis sp.</i> "mloto goldcrest" sensu Konings	Castle	Konings(1)	Lab bred	NA
<i>Tramitichromis</i> "chembe"	Castle	Mckaye <i>et al.</i> (2)	Ryan York	Eastern Thumbi; Free Anchor Bay
<i>Taeniolethrinops</i> <i>furcicauda</i>	Castle	Konings(1)	Martin Genner [#25]	Metangula
<i>Nyassachromis</i> "otter"	Castle	Konings(1)	Ryan York	Otter Point
<i>Mylochromis</i> <i>anaphyrmus</i>	Castle	Konings(1)	Martin Genner [#120]	Cape Maclear
<i>Otopharynx</i> <i>argyrosoma</i>	Castle	Konings(1)	Martin Genner [#156]	L Malombe
<i>Trematocranus</i> <i>placodon</i>	Pit	Konings(1) Keenleyside(3)	Streelman finclip stock [2257]	Otter Point
<i>Tramitichromis</i> <i>intermedius</i>	Pit	Lab Observation	Lab bred	NA
<i>Copadichromis</i> <i>virginalis</i> "yellow blaze nkanda"	Pit	Lab Observation	Lab bred	NA

<i>Astatotilapia calliptera</i>	Pit	Lab Observation	Lab bred	NA
<i>Fossochromis rostratus</i>	Pit	Konings(1)	Streelman finclip stock [2549]	Otter Point
<i>Aulonocara baenschi</i>	Pit	Konings(1)	Lab Bred	NA
<i>Dimidiochromis compressiceps</i>	Pit	Konings(1)	Lab Bred	NA
<i>Mylochromis sphaerodon</i>	Pit	Konings(1)	Lab Bred	NA
<i>Nimbochromis polystigma</i>	Pit	Konings(1)	Ryan York	Eastern Thumbi; south of Free Anchor Bay
<i>Dimidiochromis kiwinge</i>	Pit	Konings(1)	Streelman finclip stock [2384]	Local Fishermen, Cape Maclear
<i>Mylochromis lateristriga</i>	Pit	Konings(1)	Ryan York	Northwestern Thumbi

Dataset S1: GeneAnalytics results for genes associated with high divergence variants between pit and castle

Dataset S2: Allele-specific expression results from MBASED

Dataset S3: Genes set enrichments resulting from the sign test

Dataset S4: Four population comparisons with significantly negative f4 statistics

Movie S1: *Mchenga conophoros* castle building behavior. The video features a male during breeding season scooping mouthfuls of sand and spitting them to construct his bower (scoops and spits are indicated by text in the video). Building behavior is alternated with aggressive encounters toward other males and courtship acts directed at females. Filmed at Cape Maclear, Malawi in July 2015.

Supplementary Figures

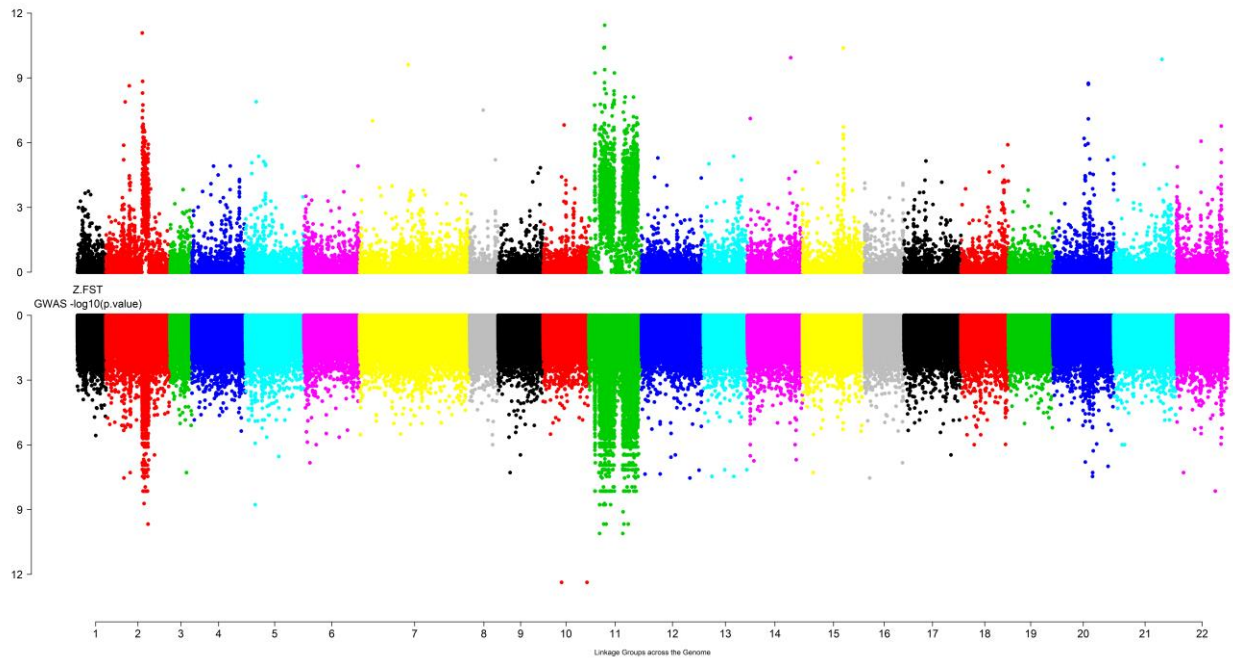


Figure S1: Comparison of genetic divergence and association patterns across the genome. Independent measures of pit versus castle divergence, F_{ST} [z-transformed F_{ST}] and ancestry-corrected GWAS [-log10 (p value)], across the genome show highly similar patterns.

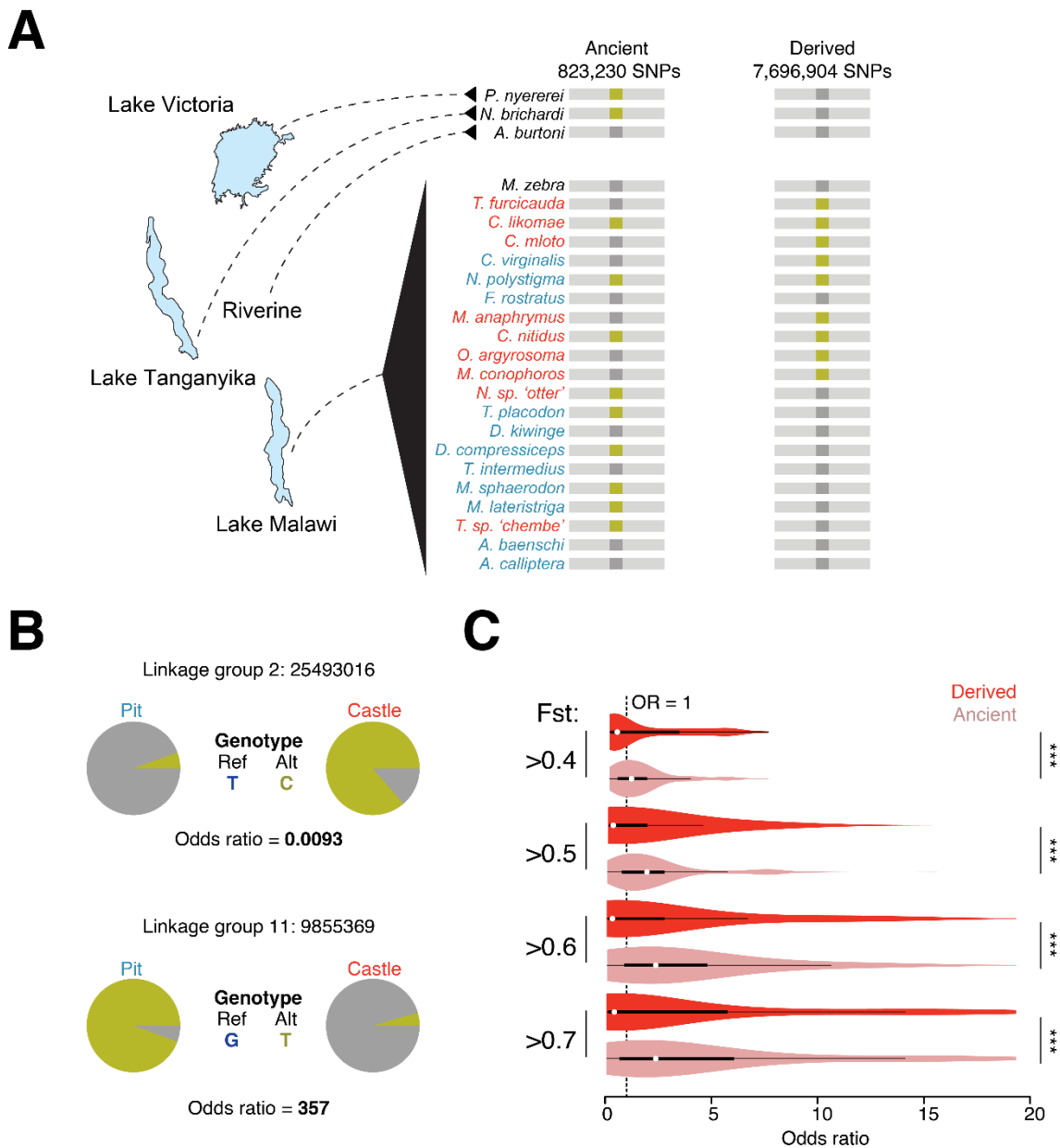


Figure S2 **Genome-wide ancestral and derived SNP enrichments** | (A) Cartoon describing the identification of derived and ancestral SNPs through whole-genome alignment with 4 non-sand dweller genomes (B) Representative high odds ratio and low odds ratio SNPs as identified by a GWAS on bower type. GWAS across the genome mirrors patterns of F_{ST} [Figure S1] (C) Violin plots of the GWAS odds ratio for SNPs at increasingly stringent p -value cutoffs, divided into derived (red) and ancestral (pink) groupings. *** = Kruskal-Wallis $p < 0.0001$.

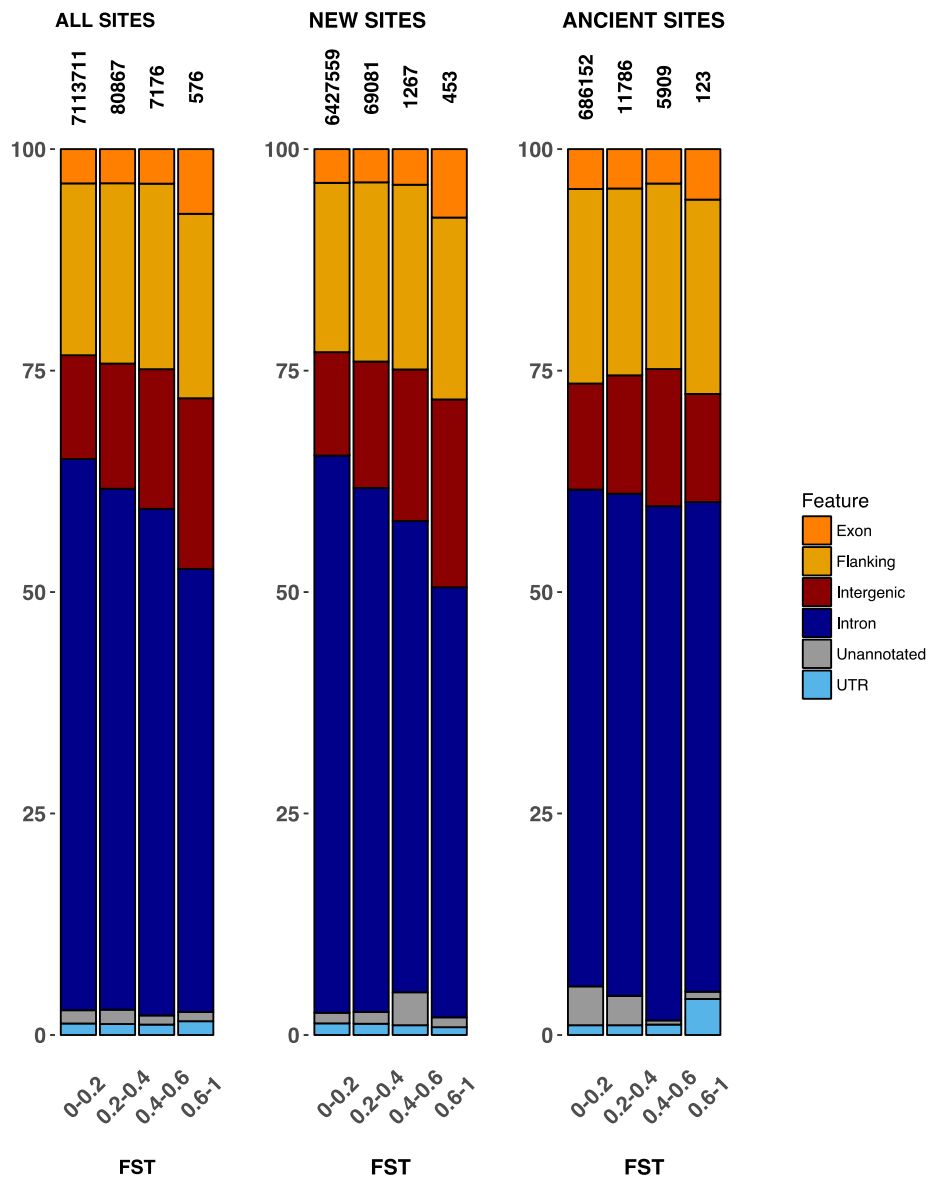


Figure S3 **Genomic distribution and FST of new and ancient SNPs** | Bars indicate proportion of genomic features represented by SNPs binned by FST values for new SNPs (polymorphic only within Lake Malawi), ancient SNPs (polymorphic within and outside of Lake Malawi), and all SNPs. 9 % of all SNPs are ancient, whereas 20% of high FST SNPs are ancient (Chi-squared test, p-value < 2.2e-16)

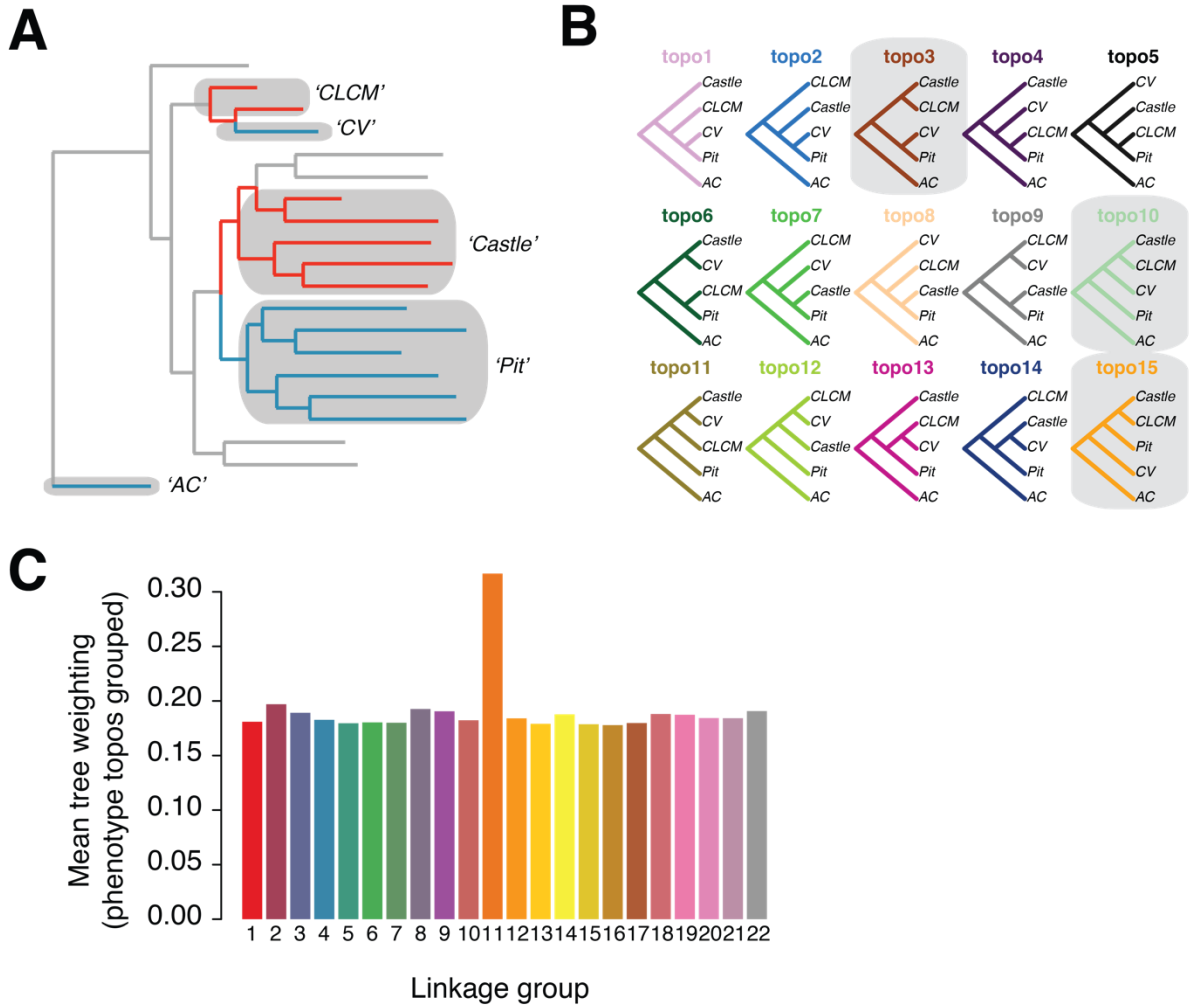


Figure S4 **Topology weighting with Twisst** | **(A)** Phylogeny of the clades used in the Twisst analyses. Branches are colored red and blue for pit and castle species, respectively. **(B)** The 15 topologies (“topos”) weighted by Twisst with the phylogenies that group by bower phenotype highlighted with grey backgrounds. **(D)** Barplot of mean combined tree weightings for the three ‘phenotype’ topos (topo3, topo10, topo15) as binned by linkage group.

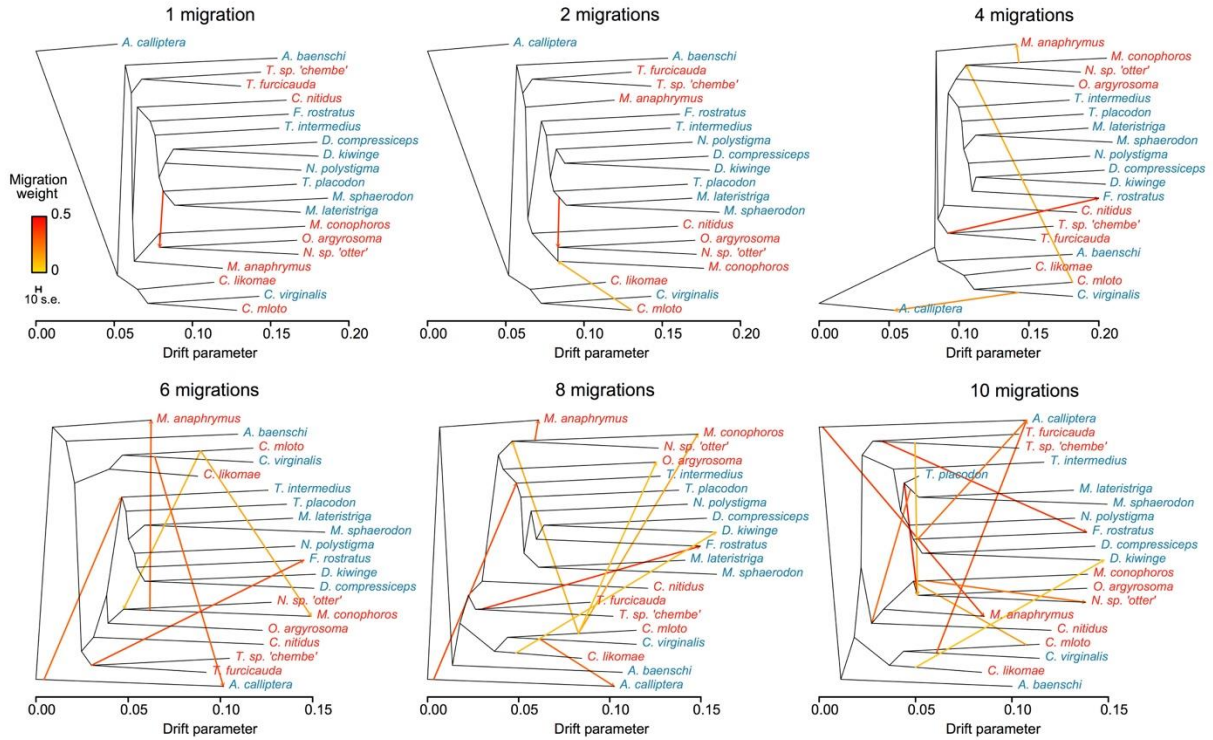


Figure S5 TREEMIX scenarios | Phylogenies are plotted with migration edges for 1, 2, 4, 6, 8, and 10 migrations. All plotted migration edges are significant. Pit and castle species names are colored blue and red, respectively.

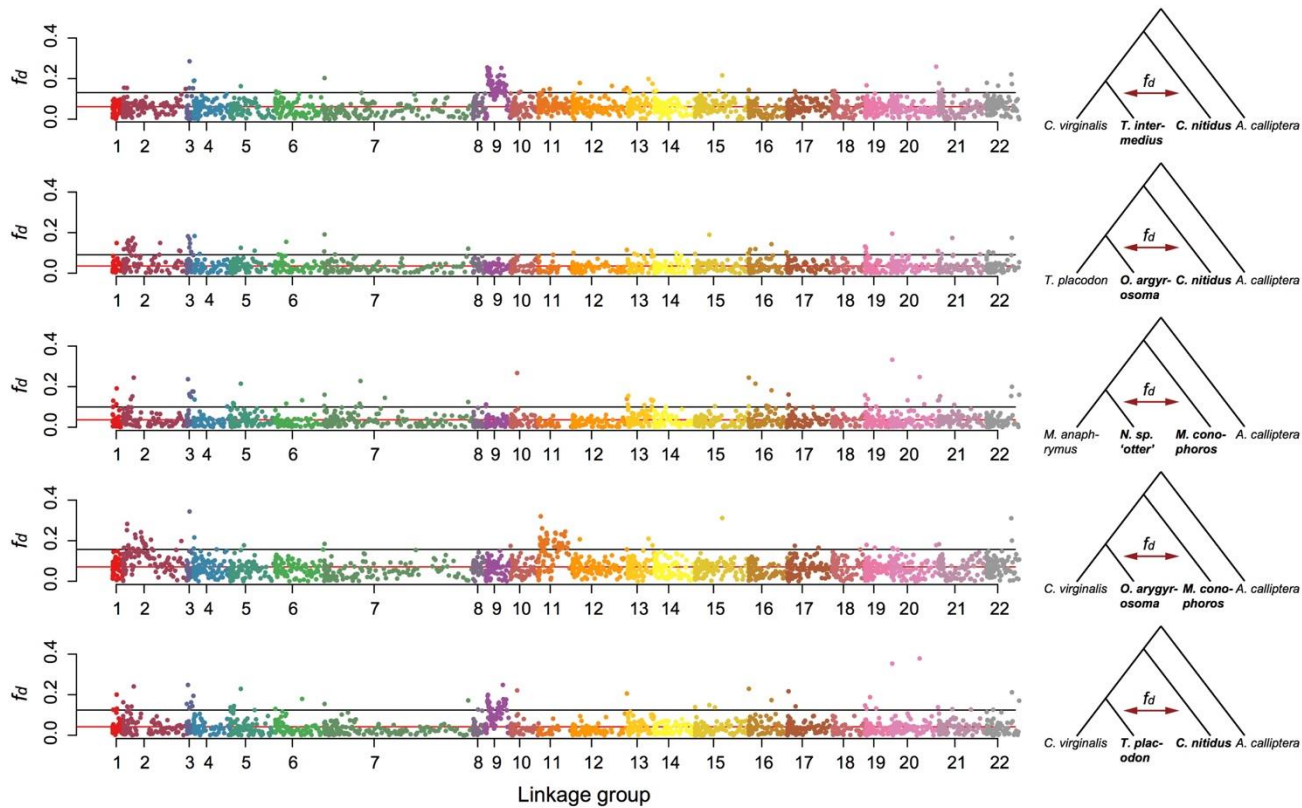


Figure S6 **Genome-wide f_d distribution** | Genome-wide scatterplots of f_d for the five most significant comparisons from the f_d analyses (cartoon comparisons presented next to the scatterplots). Higher f_d values indicate greater support for introgression at that locus.

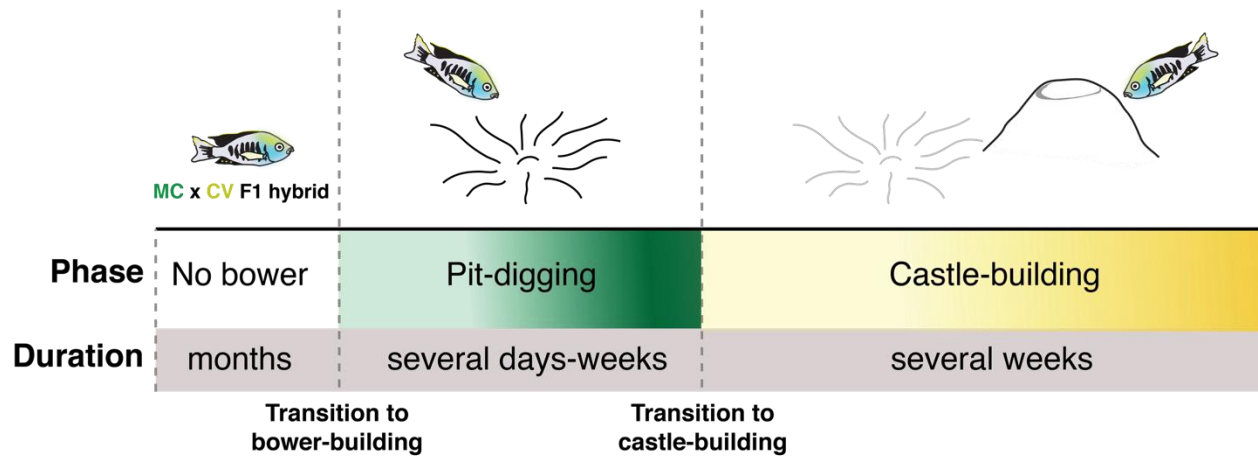


Figure S7 **Ontogeny of *Copadichromis virginalis* x *Mchenga conophoros* F₁ hybrid bower building** | The cartoon indicates the typical progression of bower building stages during a courtship “season”, proceeding from the initiation of pit-digging to the transition to castle-building.

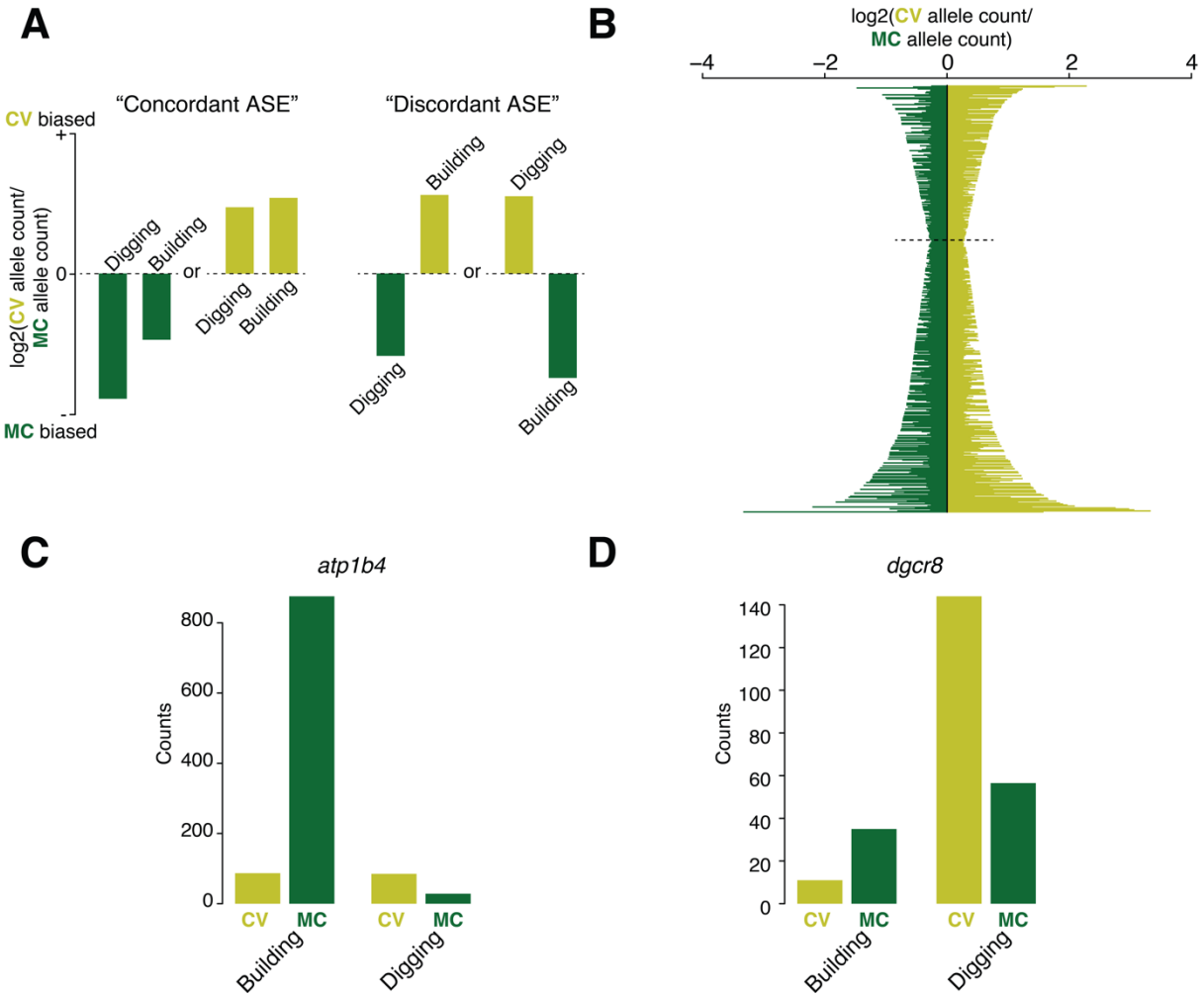


Figure S8 Genes with discordant and concordant allele-specific expression (ASE) across behavioral states | (A) Cartoon examples of the log₂ ASE ratio between *C. virginalis* (“CV”; pit) and *M. conophoros* (“MC”; castle) alleles for genes showing concordant ASE (same direction in allelic bias between behaviors) and discordant ASE (different direction in allelic bias). **(B)** Barplot of 436 genes identified as displaying discordant ASE. Each row represents an individual gene and contains two bars, one for the log₂ ASE ratio during the building phase and the other representing this ratio during the digging phase. Rows above the dashed line indicate genes that are biased toward CV during the building phase and toward MC during the digging phase. Rows below the dashed line indicate the opposite, namely an MC bias during the building phase and CV bias during the digging phase. The data presented here correspond to Dataset S3. **(C)** Barplot of gene-level RNA-seq expression counts for CV and MC alleles across digging and building behaviors for the gene *atp1b4*. **(D)** Barplot of gene-level RNA-seq expression counts for CV and MC alleles across digging and building behaviors for the gene *dgcr8*.

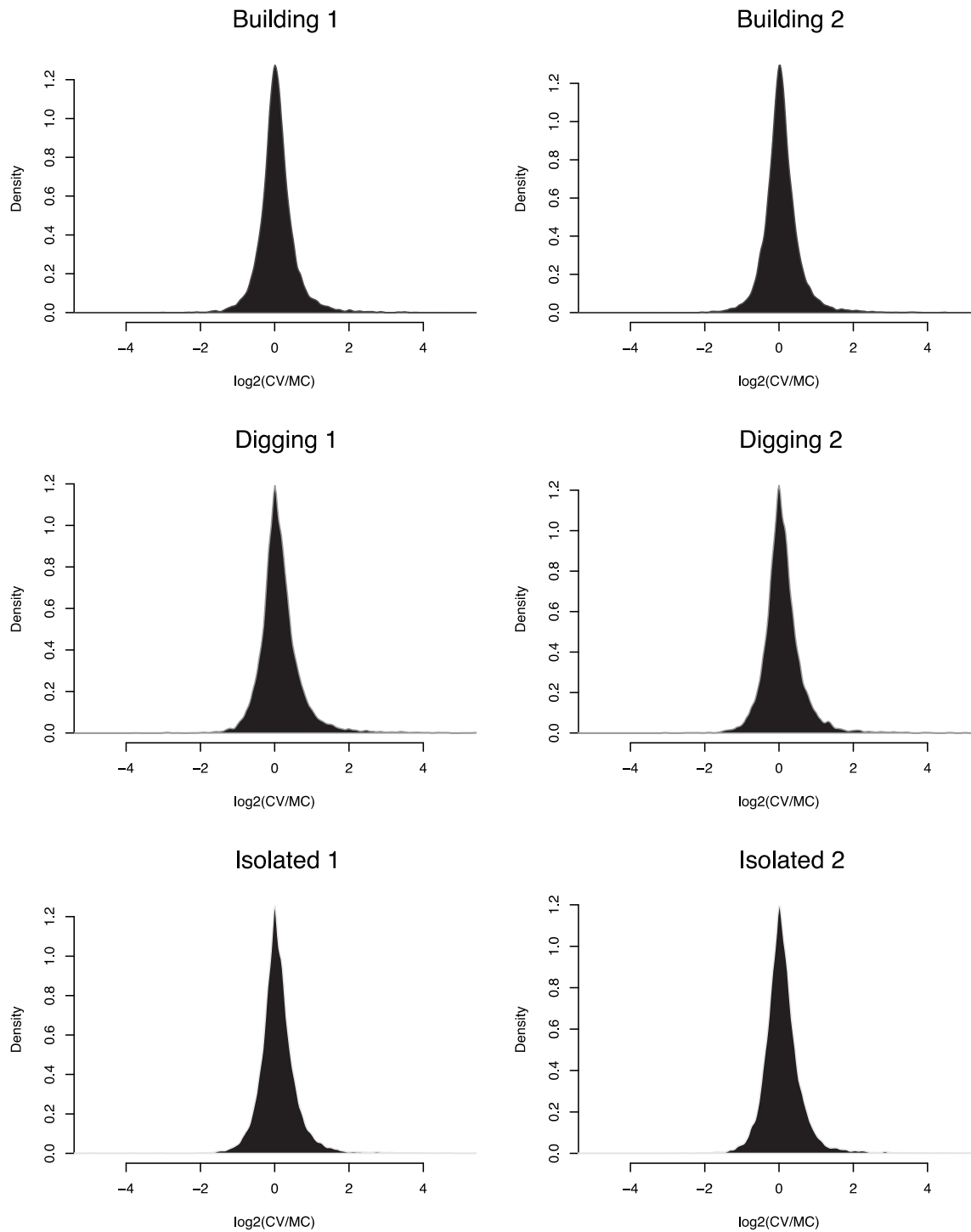


Figure S9 **The distribution of allele specific expression across F_1 hybrid samples and contexts** | Density plots representing the distribution of $\log_2(\text{CV allele counts}/\text{MC allele counts})$ for all genes measured after thresholding (see methods) in each sequencing sample.

References:

1. Konings A (2007) *Malawi Cichlids In Their Natural Habitat. 4th Edition* (Cichlid Press).
2. Mckaye KR, Howard JH, Stauffer JR, Morgan RP, & Shonhiwa F (1993) Sexual Selection and Genetic-Relationships of a Sibling Species Complex of Bower Building Cichlids in Lake Malawi, Africa. *Japanese Journal of Ichthyology* 40(1):15-21.
3. Keenleyside MHA (1991) *Cichlid Fishes: Behavior, Ecology and Evolution* (Chapman and Hall) 1st Ed.