Supplemental Information for:

Extreme copy number variation at a tRNA ligase gene affecting phenology and fitness in yellow monkeyflowers

Thomas C. Nelson, Patrick Monnahan, Mariah McIntosh, Kayli Anderson, Evan MacArthur-Waltz, Findley R. Finseth, John K. Kelly, and Lila Fishman

Supplemental Methods

PoolSeq processing -- Sequence data from each pool/line were processed with Scythe (https://github.com/vsbuffalo/scythe) to remove adaptors and with Sickle (Joshi 2011) to trim low quality sites. We then mapped reads to the unmasked v2 reference genome (retaining the 15 scaffolds corresponding to the 14 *M. guttatus* chromosomes (www.Phytozome.org) using bwa mem v 0.7.15 (Li & Durbin 2009), removed PCR duplicates, and called variant sites (SNPs and small insertions-deletions (indels) using VarScan v 2.3.6.

Example commands:

Mapping of sequence data to unmasked v2 genome build::

samtools fastq BRE13.bam | bwa mem /home/jkk/Mguttatus_256_v2.0.fa - | samtools view -bS - | samtools sort - -T xBRE13.bam.sorted -o new.BRE13.bam

Calling variant sites (snps and indels) using Varscan applied to each chromosome separately::

samtools mpileup -f ../Mguttatus_256_v2.0.fa -r scaffold_4 new.IME13.bam new.IML13.bam new.IML14.bam new.IME14.bam new.QE13.bam new.QL13.bam new.QE14.bam new.QL14.bam A.rd.bam H.rd.bam L.rd.bam NP.full.bam FP.full.bam rd.Fall.full.bam rd.Spring.full.bam BRE13.new.bam new.BRL13.bam new.BRE14.bam | java -jar /home/jkk/VarScan.v2.3.6.jar mpileup2snp - --output-vcf 1 --min-var-freq 0.001 > Bulks.4.snp.vcf

samtools mpileup -f ../Mguttatus_256_v2.0.fa -r scaffold_4 new.IME13.bam new.IML13.bam new.IML14.bam new.IME14.bam new.QE13.bam new.QL13.bam new.QE14.bam new.QL14.bam A.rd.bam H.rd.bam L.rd.bam NP.full.bam FP.full.bam rd.Fall.full.bam rd.Spring.full.bam BRE13.new.bam new.BRL13.bam new.BRE14.bam | java -jar /home/jkk/VarScan.v2.3.6.jar mpileup2indel - --output-vcf 1 --min-var-freq 0.001 > Bulks.4.indel.vcf

Coverage analyses -- To obtain an accurate median coverage for standardization (or identification as outliers), we excluded some genes from the analysis. The *M. guttatus* v2 reference does not include the sequences for either organelle, but numerous chloroplast genes have been transferred or mis-assembled into the nuclear genome (primarily in a contiguous region on Scaffold_4). Because such regions exhibit high and variable (depending on organelle number) coverage of chloroplast-derived Illumina reads with nowhere else to map to, we used BLAST to identify and exclude main scaffold genes (N = 309, with 137 on scaffold_4) with high amino-acid sequence identity (< e^{-4}) to an *Arabidopsis* chloroplast protein. In addition, we excluded Migut.C00098-C00130 and Migut.C00242-C00260, which appeared to be non-expressed (based on RNASeq data at www.Phytozome.org) repetitive DNA mistakenly annotated as coding sequence.

Genotypic analyses -- All primer pairs used for genotyping and qPCR are given in Table S1. For

genotyping, the forward primer was either 5' fluorescent- labeled directly or tailed with an M13 sequence and labeled in a 3-primer reaction including fluorescent- labeled M13. Markers were amplified using a standard 58-48°touchdown PCR protocol with the addition of bovine serum albumin (5 µg/µl final concentration) to each reaction. The PCR products, along with an in-lane size standard, were run through ABI 3130 automated sequencers (ThermoFisher, Waltham, MA) for separation and visualization of size-variable DNA fragments. Genotypes were called automatically and verified by eye using ABI Genemapper software (ThermoFisher, Waltham, MA).

Table S1. Primers for marker genotyping and qPCR analyses.

gene	name	forward primer	reverse primer
Migut.N02091	mN2091x16	TTCTAGTCCGACTCCGATGG	AAAAGATCGCCCATCAAATG
Migut.N02089	mN2089	CAGTTGCCACAGAAGCTGAA	TCGTACCATGAGGACCAATG
Migut.G00571	mG571q	CTCTCCCGATAATCGTGCTTTAG	AGCAGCCATGAAATCCTCTG
Migut.N02091	mN2091q	CCCATTTGACGCTTGCTCAT	CTCCCCATCAACCACACCAG

trip+ RLG1a expression -- mRNA was extracted using Qiagen RNEasy kits, cDNA synthesized using Maxima First Strand kits with dsDNAase step (ThermoFisher, Waltham, MA). and qPCR performed on an Agilent Stratagene qx3000using Dynamo Flash qPCR kits (ThermoFisher, WalthamMA). For quantification of RLG1a expression, we used primers within the 26th (last) exon of Migut.N02091 (mN2091q; 158bp fragment; Table S1) designed to avoid all SNP polymorphisms evident in the PoolSeq datasets, and with low identity to the other annotated tRNA ligase in the *M. guttatus* genome (Migut.D02182). As rt-qPCR controls, we used previously developed *M. guttatus* primers for EF1a and UBQ5 (Scoville *et al.* 2011) which produced similar results. However, because technical repeatability with EF1a was higher ($r^2 = 0.91$ across 2 replicates), we report the EF1a-standardized values here. (Note: the EF1a qPCR primers do not hit, at any e-value, the portion of Migut.N02091 with homology to this housekeeping gene).

high+ individual resequencing -- After extraction using our standard CTAB/chloroform protocol, genomic DNA was mechanical sheared with a Covaris E220 to a mean fragment size of 450 bp and 590 bp, respectively. We ligated sequencing adaptors and performed size exclusion of low and high molecular weight DNA with the NEB Next Ultra II Illumina library preparation kit (NEB E7645S, E7335S) following manufacturer's specifications. Libraries were quantified with the Kapa Biosystems qPCR Library Quantification Kit for Illumina (KK4824) and sequenced at the University of Pittsburgh Genomics Research Core. Raw base calls were converted to fastq using Picard tools (ExtractIlluminaBarcodes and IlluminaBasecallsToFastq) and low quality reads and adaptor sequences removed with Trimmomatic v0.35. We aligned filtered reads to the *Mimulus guttatus* v2 reference genome with bwa mem, keeping only alignments with quality scores of 29 or greater. We then used GATK IndelRealigner to improve read mapping around indels.

high+ RLG1a qPCR - For the genomic qPCR, we used Dynamo Flash SYBR green qPCR kits (ThermoFisher, Waltham, MA). For expression analyses, mRNA was extracted using Qiagen RNEasy kits

with DNAse step and rt-PCR peformed in a single step using SYBR Green Quantitative rt-qPCR kits (Sigma-Aldrich, St. Louis, MO).

Joshi NA (2011) Sickle: a sliding-window, adaptive, quality-based trimming tool for FastQ files. 1.33.

Li H, Durbin R (2009) Fast and accurate short read alignment with Burrows-Wheeler transform. *Bioinformatics*, **25**, 1754–1760.

Scoville AG, Barnett LL, Bodbyl Roels SA, Kelly JK, Hileman LC (2011) Differential regulation of a MYB transcription factor is correlated with transgenerational epigenetic inheritance of trichome density in *Mimulus guttatus*. *New Phytologist*, **191**, 251–263.