Reconciling Differences in Pool-GWAS Between Populations: A Case Study of Female Abdominal Pigmentation in *Drosophila melanogaster*

Lukas Endler, Andrea J. Betancourt, Viola Nolte, and Christian Schlötterer[1]
Institut für Populationsgenetik, Vetmeduni Vienna, 1210 Vienna, Austria

ABSTRACT The degree of concordance between populations in the genetic architecture of a given trait is an important issue in medical and evolutionary genetics. Here, we address this problem, using a replicated pooled genome-wide association study approach (Pool-GWAS) to compare the genetic basis of variation in abdominal pigmentation in female European and South African *Drosophila melanogaster*. We find that, in both the European and the South African flies, variants near the *tan* and *bric-à-brac 1* (*bab1*) genes are most strongly associated with pigmentation. However, the relative contribution of these loci differs: in the European populations, *tan* outranks *bab1*, while the converse is true for the South African flies. Using simulations, we show that this result can be explained parsimoniously, without invoking different causal variants between the populations, by a combination of frequency differences between the two populations and dominance for the causal alleles at the *bab1* locus. Our results demonstrate the power of cost-effective, replicated Pool-GWAS to shed light on differences in the genetic architecture of a given trait between populations.

KEYWORDS GWAS; genome-wide association study; pigmentation; dominance; *Drosophila melanogaster*; pooled sequencing

INSIGHT into the genetic basis of phenotypic variation within species is essential to understanding short-term evolutionary change, as phenotypic variation is an important substrate for natural selection. This endeavor has benefited from the steep decrease in the cost of rapid genotyping techniques, including DNA sequencing, which facilitates the mapping of traits using either genetic crosses (*i.e.*, quantitative trait mapping), or statistical association studies [*i.e.*, genome-wide association studies (GWAS), reviewed in Visscher *et al.* 2012]. As a result, it is now feasible to replicate association studies in multiple samples. One question we can therefore address is whether the same loci and alleles underlie phenotypic variation in different populations of species with broad geographic distributions or whether the alleles tend to be population specific. Human disease studies sometimes attempt to replicate the results of genome-wide association studies in additional populations, with varying success (*e.g.*, Li *et al.* 2008; Ng *et al.* 2008; Siontis *et al.* 2010). Results for disease traits, however, may not be representative of those for adaptive traits. Genetic diseases are probably most often due to unconditionally deleterious alleles kept at low frequency by purifying selection and thus likely to be population specific due to their rarity (Ioannidis *et al.* 2009, 2011; Gorlov *et al.* 2014). Data for adaptively varying traits are sparser than those for diseases. The best-studied examples are lactase persistence in humans, which is due to several independently derived alleles, the frequency of which differs among geographic regions (Ingram *et al.* 2009; Itan *et al.* 2010; Ranciaro *et al.* 2014), and cryptic coat coloration in mice, for which the genetic basis also differs between populations (Hoekstra and Nachman 2003; Hoekstra *et al.* 2006).

Here, we replicate a study of abdominal pigmentation in *Drosophila melanogaster* (Bastide *et al.* 2013), originally done in two European populations, in a geographically distant African population. Pigmentation in *D. melanogaster* is a useful trait for this purpose for several reasons. First, the underlying genetic pathways are well described, with most of the characterized genes playing a role in either biochemical synthesis of the pigments or sexually dimorphic and spatial regulation

doi: 10.1534/genetics.115.183376
Manuscript received October 1, 2015; accepted for publication December 21, 2015; published Early Online December 29, 2015.
Available freely online through the author-supported open access option.
Supporting information is available online at www.genetics.org/lookup/suppl/doi:10.1534/genetics.115.183376/-/DC1.
Sequence data from this article have been deposited with the European Nucleotide Archive (ENA) under accession nos. ERP001827 and ERP013300.
[1]Corresponding author: Institut für Populationsgenetik, Vetmeduni Vienna, Veterinärplatz 1, 1210 Vienna, Austria. E-mail: christian.schloetterer@vetmeduni.ac.at

of the pigment-synthesis genes (Wittkopp *et al.* 2003; Gompel *et al.* 2005). Second, this trait is both highly variable and highly heritable (Robertson *et al.* 1977; Gibert *et al.* 1998a; Pool and Aquadro 2007; Dembeck *et al.* 2015), with a reasonably simple genetic basis, qualities that render the genetics underlying this variation tractable to statistical analysis. Finally, spatial patterns of variation in this trait suggest that it is ecologically important. In particular, both thoracic and abdominal pigmentation patterns vary strongly with altitude and latitude (David *et al.* 1985; Capy *et al.* 1988; Munjal *et al.* 1997; Pool and Aquadro 2007; Telonis-Scott *et al.* 2011), with independent clines found on different continents. In the ancestral range in Africa, geographic variation is most strongly associated with levels of UV radiation, consistent with higher UV tolerance of darkly pigmented flies compared to lightly pigmented ones (Bastide *et al.* 2014).

There have been several previous studies of pigmentation variation in *Drosophila*, describing the genetic basis of variation in abdominal pigmentation between species and between sexes (Llopart *et al.* 2002; Wittkopp *et al.* 2003; Gompel *et al.* 2005; Jeong *et al.* 2006, 2008; Williams *et al.* 2008; Rogers *et al.* 2013, 2014; Salomone *et al.* 2013) and of variation in thoracic and abdominal pigmentation within species (Robertson *et al.* 1977; Pool and Aquadro 2007; Takahashi *et al.* 2007; Wittkopp *et al.* 2010; Bickel *et al.* 2011; Takahashi and Takano-Shimizu 2011; Cooley *et al.* 2012; Bastide *et al.* 2013; Rogers *et al.* 2014; Dembeck *et al.* 2015). The latter studies are the most relevant here. Thoracic pigmentation phenotypes, including differences between African strains and variation along a latitudinal cline, map to major effect loci near *ebony*, a pigment synthesis gene (Takahashi *et al.* 2007; Takahashi and Takano-Shimizu 2011; Telonis-Scott *et al.* 2011). Abdominal pigmentation variation has a slightly more complex genetic basis and has been attributed to variants at three loci: *bric-à-brac* (*bab*), *tan*, and *ebony* (Robertson *et al.* 1977; Kopp *et al.* 2003; Pool and Aquadro 2007; Bickel *et al.* 2011; Bastide *et al.* 2013; Dembeck *et al.* 2015). All three of these loci have well-described roles in pigmentation. *bab*, which comprises two protein-coding genes (*bab1* and *bab2*), is a spatial regulator of sexual dimorphic pigmentation at the tip of the abdomen (Robertson *et al.* 1977; Kopp *et al.* 2000; Williams *et al.* 2008); and *tan*, like *ebony*, is involved in pigment synthesis (True *et al.* 2005). Two independent association studies have resulted in genome-wide, fine-resolution maps of associations at these loci (Bastide *et al.* 2013; Dembeck *et al.* 2015). Both implicate variants in *cis*-regulatory elements at *tan*, *bab*, and *ebony*, although the loci with the largest effect varied between studies. In the Bastide *et al.* (2013) study, the variants with the strongest effect lay upstream of the *tan* gene, in a *cis*-regulatory element called the male-specific enhancer (MSE). The MSE has also been implicated in the pigmentation differences between *D. santomea* and *D. yakuba* (Jeong *et al.* 2008). Variants in a regulatory element of *bab1* [the dimorphic element (DME)] played a significant, but minor role, and those upstream of *ebony* were of borderline significance. Dembeck *et al.* (2015), in contrast, found the *bab* locus had a major effect on pigmentation variation, while *tan* and *ebony* and other loci had minor effects.

In this study, we replicate the Bastide *et al.* (2013) study in a South African population of *D. melanogaster*, using the pooled genome-wide association study (Pool-GWAS) approach, but with some important modifications. Although we find that the same variants appear to be causal in both populations, *bab*, rather than *tan*, shows the strongest association with pigmentation in the South African population, similar to results from the Dembeck *et al.* (2015) study and in contrast to the Bastide *et al.* (2013) study. We analyze these results in detail to explain the reasons for the contrasting genetic associations between the South African and European populations.

Materials and Methods

Sample collection, isofemale lines used, and pigmentation scoring

For the European populations ~30,000 flies were collected in Vienna in 2010 and in Bolzano, Italy, in 2011, split into replicates, used to create an F_1 generation in the laboratory at 25° and scored for pigmentation as described in Bastide *et al.* (2013). Briefly, offspring were kept at 18° for a few days posteclosion to allow adult pigmentation pattern to develop and stabilize. Pigmentation was scored by visually classifying each individual according to the extent of darkly pigmented area on the most posterior abdominal segment (A7 tergite). For classification five levels of pigmentation ranging from 0 (not pigmented) to 4 (completely pigmented) were assigned, and 100 individuals with the lightest and darkest levels (0 and 4, respectively) of each replicate were randomly chosen for pool sequencing. Around 1500 individuals were scored for each replicate. As a control three replicates of 100–160 individuals not selected for abdominal pigmentation were used.

The South African population consisted of 700 isofemale lines collected in March 2012 in Kanonkop, South Africa. These isofemale lines were kept at 25° and the offspring left for several days to allow for the adult pigmentation to mature and fully develop. For each replicate one female individual was taken from each isofemale line, leading to a replicate size of 700 individuals. The individuals were then visually scored for pigmentation of their A7 tergite, and the 135 lightest and 173 darkest flies from each replicate were pooled and sequenced (19.3% and 24.7%, respectively). The average allele frequencies in the Kanonkop population were estimated from pooled females obtained from a subset of 564 isofemale lines.

DNA extraction, library preparation, and sequencing

DNA from pooled and individual flies was extracted following a modified high-salt protocol (Miller *et al.* 1988) and fragmented using a Covaris (Woburn, MA) S2 device. Paired-end library preparation was based on the NEBNext DNA Library Prep Master Mix Set (E6040L; New England Biolabs, Ipswich,

MA) for the pooled South African sample and the individually sequenced Viennese flies, whereas a modified protocol of the NEBNext Ultra DNA Library Prep Kit (E7370L) was used for the individually sequenced South African flies. Size selection and final library purification were performed using AMPureXP beads (Beckman Coulter, Fullerton, CA) with an additional gel-based size selection for the pooled South African sample and the individually sequenced Viennese flies. The resulting insert sizes were 400 bp for these two samples and ~550 bp for the individual South African libraries. For amplification of individually barcoded libraries, the PCR master mix included in the TruSeq DNA LT Sample Prep Kit (FC-121-2001; Illumina, San Diego) or Phusion Polymerase (New England Biolabs) and 12 PCR cycles were used. All libraries were sequenced on a HiSeq2000 following a 2 × 100 bp protocol.

Read mapping, filtering of contaminants, and realignment

Sequencing reads were trimmed, mapped to a reference genome using a Hadoop-based distributed computation framework, and filtered as previously described (Kofler *et al.* 2011a; Bastide *et al.* 2013; Pandey and Schlötterer 2013). Shortly thereafter, the reads were trimmed using PoPoolation and a base quality threshold of 18 (Kofler *et al.* 2011a) and mapped to the combined genomes of *D. melanogaster* (v. 5.18), *Wolbachia pipientis* (AE017196.1), *Lactobacillus brevis* (CP000416.1), *Acetobacter pasteurianus* (AP011170), and phage *phiX174* (NC_001422.1), using BWA aln v. 0.5.9 (Li and Durbin 2009) without seeding, allowing for two gap openings (-o 2), an alignment distance threshold leading to loss of <1% of reads, assuming a 1% missing probability at a 2% error rate (-n 0.01), and up to 12-bp gap extensions (-e 12 -d 12). The aligned reads were checked for duplicates and filtered for improper pairs and a mapping quality of at least 20, using samtools v. 0.1.18 and Picard tools v. 1.79 (Li *et al.* 2009). Finally, all alignments were realigned around short insertions and deletions and in regions showing high entropy, using GATK v. 2.5.2 (McKenna *et al.* 2010), considering both the insertions and deletions occurring in the alignments and those identified in the second freeze version of the *Drosophila* Genetic Reference Panel (DGRP) lines (Mackay *et al.* 2012) (DGRP freeze 2: ftp://ftp.hgsc.bcm.edu/DGRP/freeze2_Feb_2013/ ; options RealignerTargetCreator:–maxIntervalSize 400–minReadsAtLocus 15; IndelRealigner: -entropy 0.05 -LOD 3 -maxConsensuses 50 -greedy 250 -model USE_READS).

The Austrian and Italian populations were found to contain ~1% *D. simulans* contamination when inspecting male flies for diagnostic genital differences (Sturtevant 1919). As previously described (Bastide *et al.* 2013), this can lead to false associations, as *D. melanogaster* and *D. simulans* differ from each other in the degree of abdominal pigmentation (Gibert *et al.* 1998b). To identify and remove reads potentially stemming from *D. simulans*, the mapped and filtered reads were remapped against five genomes each from *D. melanogaster* and *D. simulans*, using GMAP (ver. 2012-07-20) as previously described (Bastide *et al.* 2013), and reads mapping to a *D. simulans* genome with a higher mapping quality than that to the best-fitting *D. melanogaster* genome were removed as contamination. Previously identified fixed differences between the two fly species were used to independently estimate the amount of *D. simulans* contamination (Bastide *et al.* 2013).

The cleaned and filtered alignment files were converted into synchronized format requiring a base quality of at least 20, using PoPoolation2 (Kofler *et al.* 2011b). Repetitive regions and regions around small insertions and deletions (5 bp up- or downstream) were removed. RepeatMasker (ver. 3.2.8, www.repeatmasker.org) was used for identifying repetitive regions without considering low-complexity regions (option -nolow). Overall, after all filtering steps, an average sequencing depth of ~100 was obtained, with coverages ranging from 27 to 278 (for details see Supporting Information, Table S1).

Association mapping and false detection rate calculation

To test for associations of variants with abdominal pigmentation we used a Cochran–Mantel–Haenszel (CMH) test as previously described (Bastide *et al.* 2013). Briefly, the CMH test allows us to test for independence of values in contingency tables over multiple strata or, in our case, replicates. For each replicate a 2 × 2 contingency table containing the reference and alternative allele counts of the extremely light and dark pools was created, and the CMH test was performed over all replicates for each individual variant as implemented in PoPoolation 2 (Kofler *et al.* 2011b). For calculation of the common odds ratios and confidence intervals we used custom python scripts, using the *mantelhaen.test* function of the *stats* package of R (R Core Team 2014) and *rpy2* (rpy.sourceforge.net). All polymorphic sites were filtered for a minimum coverage of 15 for each sample and a minimum overall count of the minor allele of 8 for each sample. We also removed the 2% highest-covered sites from the calculations. To correct for multiple testing, we enforced a false detection rate of 0.05 derived from an empirical null distribution as described in Bastide *et al.* (2013).

Identification of small insertions and deletions

Small insertions and deletions were identified and quantified using the UnifiedGenotyper of GATK (ver. 2.5.2; McKenna *et al.* 2010). The alignment files of each population were analyzed together, using a sample ploidy of 25, a minimum base quality of 20, a minimum insertion/deletion (indel) fraction of 5%, and an overall occurrence of 20 counts for each indel to be considered (options: -maxAltAlleles 3 -glm GENERALPLOIDYINDEL -deletions 1.1 -mbq 20 -minIndelFrac 0.05 -minIndelCnt 20 -stand_call_conf 10.0 -stand_emit_conf 10.0 -sample_ploidy 25). A lower sample ploidy than actually occurred in the pools had to be used for the calculations as GATK was prohibitively slow for ploidies >25. While this may lead to some inaccuracies in genotyping and in scoring of low-frequency variants, it should not influence the estimated counts

of variants with minor allele frequencies >5%. The indels were filtered using GATK to perform variant quality score recalibration, taking the DGRP freeze 2 indels as a training set and choosing a sensitivity of 95%. Furthermore, all indels with a frequency <5% were removed using bcftools (v. 1.0; samtools.github.io/bcftools/). For each position, only the two most common alleles were considered and the counts of the most common indel used in a modified CMH testing against the reference allele count, a procedure similar to single-nucleotide variants. The CMH test was performed using the *mantelhaen.test* function of the *stats* package of R (R Core Team 2014), using custom python scripts and *rpy2* (rpy.sourceforge.net).

Estimation of inbreeding coefficients and genetic differentiation

Inbreeding coefficients (F_{IT}) were calculated for single females from 12 South African isofemale lines and 12 female offspring from the freshly collected Italian population. For the calculation of inbreeding coefficients (F_{IT}), 6 individual flies from each the very dark and the very light fraction of the South African samples and 12 individuals not scored for pigmentation from the Italian population were used. As we estimate F_{IT} genome-wide, we do not expect the values to be much affected by any selection on the pigmentation loci. The reads of each individual were mapped to the reference with BWA, filtered, and realigned with GATK as above. We called SNPs with the UnifiedGenotyper of GATK (McKenna *et al.* 2010), using the default parameters apart from allowing for three alleles and slightly altered call and emit thresholds (-maxAltAlleles 3 -glm SNP–output_mode EMIT_VARIANTS_ONLY -stand_call_conf 30.0 -stand_emit_conf 15.0). Subsequently we recalibrated the variant quality score based on the combined DGRP freeze 2 and Drosophila Population Genomics Project 2 (DPGP2) (Pool *et al.* 2012) variants. To calculate inbreeding coefficients, we used only high-confidence SNPs called in the 12 South African and the 12 Italian individuals, respectively (GATK variant quality score recalibration tranche sensitivity threshold 99%, allele frequency >0.15). Inbreeding coefficients were calculated for each variant from the fraction of heterozygous individuals (Het) and the frequencies of the reference (f_R) and alternative alleles (f_A) in the base populations: $F_{rr} = 1 - \mathrm{Het}/(2f_R \cdot f_A)$.

Genetic differentiation was calculated by estimating pairwise F_{ST} values from the pooled data of flies from each population that were unselected for pigmentation, using PoPoolation 2 with default settings (Kofler *et al.* 2011b) and the same filtering criteria and allele frequency cutoffs as for the CMH test.

Simulations

To better understand the combined effects of the experimental setup, allele frequencies, inbreeding coefficients, additive effects, dominance, and epistatic interactions on the results of these experiments, we performed computer simulations.

We simulated each experiment following the experimental design as closely as possible. For each simulated population, we generated individuals with genotypes for the *tan* and *bab* loci according to the measured allele frequencies, assuming Hardy–Weinberg equilibrium for the European population and using the estimated inbreeding coefficient for the South African sample ($F_{IT} = 0.3$). Phenotypic values for each individual were calculated based on its genotype and the underlying genetic architecture and adding a normally distributed random effect with an environmental variance, which was scaled by heritability. The narrow-sense heritability, h^2, for female abdominal pigmentation has been estimated to lie between 0.5 and 0.8 (Gibert *et al.* 1998a). As only two of the contributing loci were simulated, we conservatively assumed $h^2 = 0.3$ for the European population and applied the same environmental variance V_E applied to both the European and South African populations.

The genetic component of the phenotype was calculated based on the additive effect (eff) and the degree of dominance (h) for each locus. The dominance, h, ranges from 0 for a recessive light allele, to 0.5 for codominance, to 1 for dominance of the light allele. We calculated the phenotypic value P, depending on the state of loci g_1 and g_2, using $P(g_1, g_2) = a_1(g_1) + a_2(g_2) + \varepsilon$, where ε represents the environmental contribution to phenotypic variance and is a normally distributed random variable with mean 0 and variance V_E with mean zero, and $a_N(g_N)$ constitutes the genotypic effect of the locus g_N:

$$a_N(g_N) = \mathrm{eff}_N \cdot \begin{cases} -0.5 & \text{if } g_N = AA \\ h_N - 0.5 & \text{if } g_N = Aa \\ 0.5 & \text{if } g_N = aa. \end{cases}$$

In addition to dominance and additive effects, we also include potential interactions between the two loci. We defined simple pairwise epistatic interactions by an interaction constant (eps_{int}). An interaction term of zero indicates no epistasis, a negative value corresponds to negative epistasis between light alleles (*e.g.*, with the light double homozygote being darker than expected in the absence of epistasis), and a positive value corresponds to positive epistasis (*e.g.*, with the light double homozygote being lighter than expected; see Figure S17). The effect of the epistatic interaction between the loci g_1 and g_2 is calculated with $\mathrm{eps}(g_1, g_2) = \mathrm{eps}_{int} \cdot a_1(g_1) \cdot a_2(g_2) \cdot (\mathrm{eff}_1 + \mathrm{eff}_2)/2$ and the overall phenotypic value as $P(g_1, g_2) = a_1(g_1) + a_2(g_2) + \mathrm{eps}(g_1, g_2) + \varepsilon$.

For each set of dominance, epistasis, and additive effect parameters, we performed simulations corresponding to the European and South African experiments, with individuals divided into replicates as in the real experiment (1500 individuals randomly assigned to six replicates in the European experiment, or 1 fly per replicate from each of 700 isofemale lines in the South African experiment). The individuals in each replicate were then ranked according to their phenotypic values, and a CMH test was performed using the allele counts from extreme individuals (with 100 individuals selected from

each extreme for Europe and 65 individuals from each extreme from South Africa).

To infer the most likely combination of parameter values and to compare the two models, we used rejection sampling as implemented in the abc R package (Csilléry *et al.* 2012). For both models 2 million simulations were performed with parameters drawn from uniform prior distributions ([−0.125, 1.125] for dominances and [0.25, 4.5] for the additive effect of *bab1* relative to *tan* and [−1.5, 3.5] for the epistatic interaction strength). The posterior distribution of parameters was estimated using a simple rejection method, using an acceptance threshold of 0.01 and the following criteria as summary statistics: logOR(*tan*)/logOR(*bab1*): Europe = 2.51, South Africa = 0.68; logOR(tan_{EU})/logOR(tan_{SA}) = 2.28; and logOR($bab1_{EU}$)/logOR($bab1_{SA}$) = 0.62. For model selection, we used the postpr function of the abc R package with the summary statistics calculated from simulations and a tolerance of 0.01.

Data availability

All custom scripts are available for download at github under https://github.com/luenling/Pigmentation2015. Raw sequencing reads are available at the European Nucleotide Archive (ENA) (http://www.ebi.ac.uk/ena) under accession nos. ERP001827 (http://www.ebi.ac.uk/ena/data/view/ERP001827) and ERP013300 (http://www.ebi.ac.uk/ena/data/view/ERP013300).

Results and Discussion

Variation within and between samples

We used samples of *D. melanogaster* from three locations for this study. After adding an additional replicate for the Viennese population, we reanalyzed the population samples from Bastide *et al.* (2013), which consisted of two European populations, from near Bolzano, Italy, and Vienna, ∼200 miles apart. We also collected and analyzed data from a new population, from Kanonkop, near Cape Town, South Africa. We divided each of the three population samples into three replicates, which were then scored for female abdominal pigmentation. Within each replicate, we identified individuals with extreme phenotypes—those with the greatest and least extent of darkly pigmented area on segment A7, as in Bastide *et al.* (2013)—and pooled them and subjected the pools of light and dark individuals from each replicate to whole-genome sequencing (see Table S1 for coverage depth). We then analyzed these pools for allele frequency differences that suggest an association with the pigmentation phenotype. The six replicates from the European populations consisted of 1500 offspring of wild-caught females, with 100 individuals (∼7%) per replicate contributing to each extreme pool. The South African samples, for logistical reasons, were kept as 700 isofemale lines for 15 generations before they were scored for pigmentation. Each of the three replicates from this sample consisted of a single individual sampled from each isofemale line, with a similar fraction of the individuals contributing to the extreme pools as in the European populations (55–80 females or 7–11% of each replicate).

After mapping and filtering, we identified ∼3.2 million single-nucleotide polymorphisms (SNPs) segregating in the European population and ∼4 million SNPs in the South African isofemale lines with ∼1.9 million common to both. The SNPs shared by both populations had significantly higher minor allele frequencies (MAF) than the SNPs found in only one of the two populations [shared SNPs, median MAF Europe (EU) and South Africa (SA), 0.12; private SNPs, median MAF EU, 0.006 and SA, 0.015; paired Wilcoxon rank sum test, EU, $W = 1e{+}12$, $P < 2.2e{-}16$ and SA, $W = 3.5e{+}12$, $P < 2.2e{-}16$]. To estimate divergence between the populations, we calculated pairwise F_{ST} values for 200-kb windows, using PoPoolation2 (Kofler *et al.* 2011b). As differentiation between the Austrian and Italian populations was low (median $F_{ST} = 0.0107$, 95% C.I.: [0.0105, 0.011]), we combined these samples for subsequent analysis, as in Bastide *et al.* (2013). The South African sample showed higher, but still moderate, levels of divergence from the European populations {median F_{ST}(Aut/SA) = 0.044 (95% C.I.: [0.043, 0.046]), median F_{ST}(Ita/SA) = 0. 040 (95% CI: [0.039, 0.041])}, lower than those typically reported for comparisons between European and other African *D. melanogaster* samples ($F_{ST} \sim 0.20$; Pool *et al.* 2012). The low levels of population differentiation between Europe and South Africa seen here are probably at least partially due to high levels of migration from cosmopolitan flies, shown to occur in many African populations (Pool and Aquadro 2006), but which may be particularly high for our South African sample due to Cape Town's status as a shipping port.

For the South African population, we used flies from recently established isofemale lines, which have a higher degree of inbreeding than the wild-derived F_1's used in the European study. As inbreeding in the South African samples may affect the GWAS results, we first estimated the degree of inbreeding in these lines. To this end, we sequenced 12 individual females from different isofemale lines and compared these to 12 resequenced individual flies from the population collected in Bolzano, Italy. In the South African flies, the median inbreeding coefficient, F_{IT}, is 0.35 (95% C.I. = [0.337, 0.361]) see Figure S1). This level of inbreeding can be explained by an initial bottleneck of 2 progenitor flies and 15 generations at an effective population size (N_e) of ∼10–25, consistent with the conditions under which the isofemale lines were propagated. F_{IT} varied moderately but significantly among chromosomes, with *2*L and *3*L showing the lowest inbreeding (median $F_{IT} = 0.28$ both) and *X* and *2*R the highest (median $F_{IT} = 0.43$ and 0.5, respectively; Dunn's test with Benjamin–Hochberg correction; comparison, *2*L:*3*L and *2*R:*X*, $P > 0.05$; all other comparisons, $P < 10^{-6}$). In contrast, resequenced individual flies from the European populations, which have not been maintained as isofemale lines, show little evidence of inbreeding (median $F_{IT} \approx 0.00$, 95% C.I.: [−0.009,

0.001]), as expected for flies with essentially the same genotypes as wild flies.

SNPs associated with variation in abdominal pigmentation

To find SNPs associated with abdominal pigmentation, we compared the allele frequencies of the extreme light and dark pools of the European and South African populations across replicates, using the CMH test as in Bastide *et al.* (2013). In both populations, most of the 100 highest-ranked SNPs (ranked by *P*-value from the CMH test) are located in noncoding regions—73% in the European population and 95% in the South African one (see Table S2, Table S3, Table S4, Table S5, Table S6, and Table S7). Nevertheless, in the European population, the highly ranked SNPs are enriched for those in coding regions compared to random samples matched in size, chromosomal distribution, and allele frequency (100,000 samples: $P < 10^{-4}$). This enrichment is, however, mostly due to the overlap of the *tan* regulatory region with the coding sequence of various flanking genes (14 synonymous and 5 nonsynonymous SNPs).

In general, the results for the European population (Figure 1) are essentially the same as in Bastide *et al.* (2013), with minor differences likely due to the inclusion of an additional replicate for the Viennese population and a slightly altered SNP calling and filtering protocol. The strongest associations are due to SNPs found upstream of the *tan* locus, with the highest-ranking SNPs in the MSE region (Figure 1, A and C, and Table S3, Table S6, and Table S7). The remaining strongly associated SNPs are near the *bric-a-brac* locus (hereafter called *bab*), and weakly associated SNPs occur upstream of the *ebony* gene on *3*R and close to the *pdm3* gene on *2*R. While only weakly associated with pigmentation, it is plausible that both *ebony* and *pdm3* are causal, as both have been implicated in abdominal pigmentation in other studies (Pool and Aquadro 2007; Rogers *et al.* 2014).

In contrast, in the South African sample (Figure 1B and Table S2, Table S4, and Table S5), the SNPs with the strongest associations occur at *bab*, not *tan*. Specifically, these SNPs lie in the third intron of *bab1*, near a *cis*-regulatory region called the dimorphic element (Figure 1D) (Williams *et al.* 2008). The next strongest associations were with SNPs at the *tan* locus. The change in ranks for the *tan* and *bab* loci between Europe and South Africa is investigated in detail below. Finally, we again found SNPs weakly associated with pigmentation at the *pdm3* locus, as in the European sample, but none near the *ebony* locus (Figure 1, Figure S2, and Figure S3).

Using *D. simulans* as an outgroup, we inferred the ancestral states of the strongly associated SNPs via parsimony. Interestingly, the inferred ancestral phenotype differs at the *tan* and *bab* loci: for *tan*, most of the 10 highest-ranking SNPs with a consistent effect in both samples had an inferred ancestral light allele (EU, 7/10; SA, 9/10). At the *bab* locus, in contrast, the majority of SNPs had an ancestral dark allele (EU, 8/10; SA, 7/10; Table S2 and Table S3).

Non-SNP variants associated with abdominal pigmentation

Strongly associated SNPs may not directly cause abdominal pigmentation differences, but may instead be associated with pigmentation via linkage to causal alleles consisting of other kinds of genetic variation, namely, short indel variants and large chromosomal inversions. To investigate these data for differences in indel frequencies between the pools, we identified 279,016 segregating short insertions and deletions in one of the European pools and 371,146 in the South African samples, with 199,676 of these segregating in all populations (with the requirement that they segregate with minor allele frequencies of at least 15%). As with the SNPs, we compared allele counts of the indels in the extreme pools over replicates, using a CMH test. While indels and SNPs could, in principle, be combined and analyzed together, we examined the indel results separately. The reason is that the indel allele counts are likely less reliable than the SNP allele counts, as calling indels is technically difficult compared to identifying SNPs, especially in pooled sequencing samples.

Overall, we find patterns similar to those obtained with SNPs (Figure S4 and Figure S5). For the European samples, highly associated indels are found around the same regions close to *tan* and *bab1*, while for the South African flies, the only strongly associated indels were near *bab1*, with none near the *tan* locus. In general, indels associated with pigmentation variation showed higher *P*-values and smaller average allele frequency changes than the highly associated SNPs [median of the log(OR) of the 10 most significant variants: EU, SNPs = 1.8, indels = 1.3; SA, SNPs = 2.1, indels = 1.7], suggesting that the observed phenotypic variation is more likely to be due to the SNPs than to the indels.

We further investigated the pools of light and dark flies for differences in the frequencies of cosmopolitan inversions. The inversions themselves are unlikely to cause pigmentation differences, but as they suppress recombination in heterozygotes, any causal variants contained in the inverted type may be linked to many other variants also associated with the inversion. To investigate this possibility, we looked for differences in the frequencies of six common cosmopolitan inversions on chromosomes *2* and *3* between the light and dark pools. We identified inversion types indirectly, via previously identified SNP markers from Kapun *et al.* (2013).

While some inversions showed considerable frequency differences between light and dark pools in the European populations (Figure S6), these differences were not consistent between the Italian and Austrian samples and were not replicated in the South African pools. Thus, these frequency differences may lead to an overall elevation in allele frequency differences between the pools, but, assuming that the inversions are associated with similar alleles in both European samples, they are unlikely to lead to consistently associated alleles even in the European samples. In addition, linkage disequilibrium between alleles and inversions is likely strongest at the inversion breakpoints (Andolfatto *et al.* 1999).

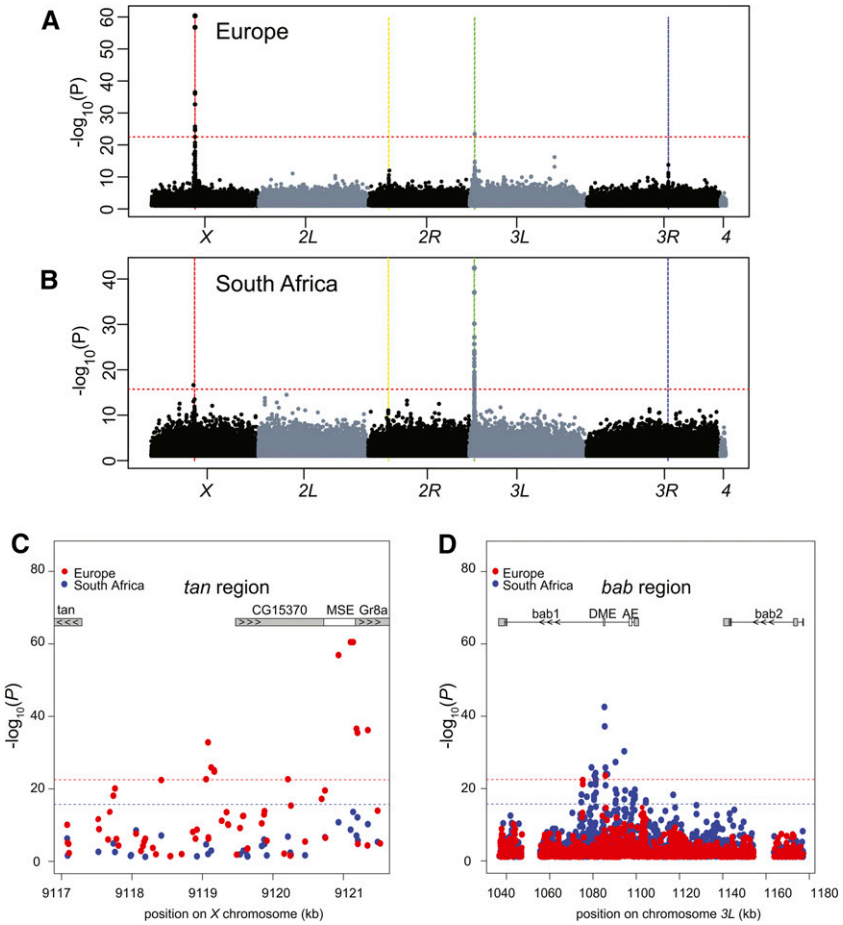


Figure 1 Manhattan plots for the effect of SNPs on female abdominal pigmentation for segment A7. The horizontal axis shows the genomic location of each tested SNP, with chromosomal location indicated alternately by black and gray and with vertical lines indicating the locations of the *tan* (red), *pdm3* (yellow), *bab1* (green), and *ebony* (blue) genes. The vertical axis shows the $-\log_{10}$ *P*-value from the CMH test, with a horizontal line indicating the 0.05 FDR threshold. (A–D) Results for the experiments with European (A) and South African (B) flies, with detailed views of regions of the *tan* (C) and *bab* (D) loci that contain SNPs showing strong associations. In the detailed views, only SNPs with *P*-values <0.1 are shown. Gray rectangles mark regions containing coding sequence of the genes, and white rectangles indicate known *cis*-regulatory elements. The arrowheads indicate the direction of transcription. Male-specific enhancer (MSE), dimorphic enhancer (DME), and anterior element (AE) indicate previously described *cis*-regulatory regions (Jeong *et al.* 2008; Williams *et al.* 2008).

The closest inversion breakpoint of any of the investigated inversion is ~2 Mb away from the *bab* locus, possibly close enough to affect allele frequencies (Stevison *et al.* 2011), but unlikely to result in strong spurious association at *bab* and not elsewhere on *3*L. The *tan* locus is *X* linked, and the *X* chromosome harbors no known common inversions (Lemeunier and Aulard 1992). In any case, the inversions are unlikely to cause artifactual signals at *tan* and *bab*.

Similarities and differences between the populations

We noted several differences between the European and South African populations. In particular, the weak effect of *ebony* is restricted to the European population, and the strongest signals occur near *tan* in the European population, but near *bab* in the South African population. Nevertheless, we identified similar sets of genes overall in the two populations, raising the question of whether pigmentation is affected by the same alleles in the two populations or by independent alleles at the same loci. We tested this by asking whether highly ranked SNPs (ranked separately by significance for each population) had consistent effects in the two samples, using two different measures of consistency. First, we asked whether the alleles at these SNPs had the same qualitative effect for Europe and South Africa, assuming that the "dark" allele at each SNP is that enriched in the dark sample of flies (and vice versa). For unassociated diallelic SNPs, we expect half of them to have the same qualitative effect in the two samples. High-ranking SNPs, however, have the same inferred dark and light alleles much more often than this (Figure 2A). Second, we asked whether high-ranking SNPs in one population had the strongest estimated phenotypic effects in the other population, as expected. We estimated and ranked phenotypic effects using the logOR for dark *vs.* light pools (Bastide *et al.* 2013) (Figure S7 and Figure S8). The results again show broad consistency between the populations; in particular, high-ranking SNPs in Europe generally had the highest-ranked phenotypic effects in South Africa (Figure 2B). For both measures of similarity, the concordance between the samples declines steeply as the effects of highly ranked SNPs are diluted by SNPs with increasingly lower ranks (Figure 2). Overall, we conclude that despite the striking differences between the populations, this analysis suggests that some SNPs contribute similarly to female abdominal pigmentation in both populations. Otherwise, we would not expect to find consistency between the high-ranking SNPs of the two populations, as each would contain two different sets of causal variants and linked SNPs.

That said, the ranks of the highest-ranking South African SNPs were poor predictors of the ranks of their effect sizes in Europe compared to the reciprocal comparison (Figure 2).

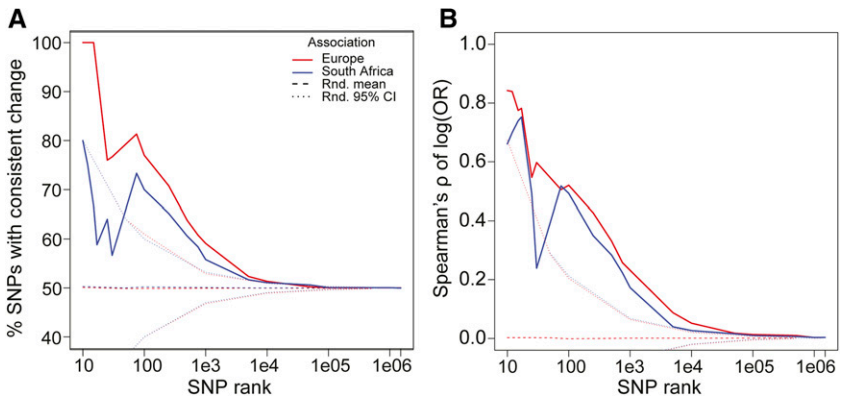


Figure 2 Consistency between European and South African samples. SNPs were ranked by CMH test *P*-values, with different cutoffs shown on the *x*-axis. SNPs were ranked according to the results in the European (red line) or South African populations (blue line). One thousand random samples of unassociated SNPs—matched in number, allele frequency, and chromosomal distribution with the analyzed SNPs—were used as a comparison (dashed lines). (A) The fraction of SNPs with consistent effects, *i.e.*, where light and dark alleles inferred between the two samples were the same. (B) Correlation coefficients (Spearman's ρ) for the logOR values of the two populations. The logOR is a measure of the effect of the SNP, with the direction of the effect indicated by its sign.

Inspection of the data reveals that this discrepancy arises mainly from the region near *bab1*, where some SNPs were high ranking in South Africa, but not in Europe. Tellingly, these SNPs segregate at intermediate frequencies in South Africa and at low frequencies in Europe (median MAFs of the 50 highest-ranked SNPs in South Africa = 0.42, median MAFs in Europe = 0.26; paired Wilcoxon rank sum test, $P = 0.0004$). The logOR estimator is akin to the average effect of an allelic substitution (Falconer and Mackay 1996), in that it is affected in a similar way by dominance and allele frequency. That is, in the absence of dominance, the logOR effect estimates are affected very little by differences in allele frequencies (Figure S7A). With dominance though, logOR estimates do change with allele frequency, reflecting the change in the proportion of individuals expressing the full homozygous effects of the recessive allele (Figure S7B). Thus, the failure to observe a strong effect for the *bab1* SNPs in Europe might be explained by a lower MAF of the recessive allele, so that its effects are rarely expressed and such that it contributes little to phenotypic variation in Europe. At the *tan* gene, in contrast, the 50 highest-ranked SNPs in the South African sample had comparable median MAFs in reference samples from both populations (median MAFs in Europe, 0.75; in SA, 0.70; paired Wilcoxon rank sum test, $P = 0.99$; Figure S9), and the effects for these SNPs were more consistent than for those at the *bab* locus.

Refining candidates using the combined analysis of multiple populations

Despite the high resolution of Pool-GWAS, for closely linked sites, several alleles from a single population are associated with a causal variant via linkage disequilibrium, resulting in false positives. The combined analysis of multiple populations may make use of more historical recombination events than have occurred in a single population. As a result, provided that the genetic architecture of the trait is identical among the populations, their joint analysis can yield fewer false positives and higher confidence candidates than the analysis of a single population. Considering only high-ranking SNPs with consistent effects in the two populations, for example, we can narrow down the SNPs at *bab* to a single best candidate SNP (3L:1085454 at *bab1*).

For SNPs in very close proximity, we can use patterns of linkage to further refine the location of SNPs that show consistent changes between populations. We achieve this by using information from SNPs on the same read or read pair and estimating short-range linkage disequilibrium between the associated SNPs as in Feder *et al.* (2012). We combined pairs of SNPs with estimated r^2 values > 0.75 to infer short haplotypes. We then determined the frequency changes and the associated *P*-values of SNPs associated with these haplotypes (Figure S9, Figure S10, and Figure S11) to determine whether candidate SNPs increase in frequency independently of nearby SNPs or as part of a haplotype. For example, the three most highly ranked *tan* SNPs in the European population appear to occur primarily on the same haplotype, consistent with their close physical distance (*e.g.*, the two most tightly linked SNPs, *X*:9121094 and *X*:9121129, show r^2(EU) ~ 0.97, r^2(SA) ~ 0.83). Unfortunately, this does not implicate a single best candidate SNP, as the response of all three SNPs is likely nonindependent. In contrast, the analysis of three groups of SNPs upstream of the *tan* gene (*X*:9119116–9119160, *X*:9120730,912683, and *X*:9123892–9123903) demonstrates the usefulness of this approach. In the analysis of the European population, the SNPs within a group show strong linkage disequilibrium, resulting in almost identical allele frequency change. In the South African population, these alleles do not show the same pattern of linkage disequilibrium, and only one SNP in each of the groups exhibits consistent allele frequency changes in the South African sample, yielding a single best candidate causal allele for each of the three haplotypes (see green boxes in Figure S10).

Contrasts between populations at the tan and bab loci

We used simulations to ask whether we can simply explain the contrasting effects of *tan* and *bab* in Europe and South Africa, assuming an identical genetic architecture of causal loci underlying the phenotype. Specifically, we investigated the role of dominance and allele frequency differences between the populations. We simulated as closely as possible the actual Pool-GWAS procedure for both populations, using the same number of sequenced individuals and replicates, and estimated

Table 1 The four best candidate SNPs at the *tan* and *bab* loci

		Europe			South Africa			
Chr	Position	AF	$-\log_{10} P$	logOR	AF	$-\log_{10} P$	logOR	logOR (EU)/logOR (SA)
X	9120922	0.22	56.78	2.39	0.29	10.64	1.26	
X	9121094	0.18	60.35	3.14	0.17	8.58	1.18	
X	9121129	0.18	60.35	2.86	0.16	13.50	1.66	
X	9121177	0.21	36.45	1.83	0.06	6.96	1.43	
tan mean		0.20		2.55	0.17		1.38	1.89
*3*L	1084990	0.90	8.38	1.20	0.52	42.41	3.03	
*3*L	1085137	0.85	4.35	0.65	0.47	37.04	2.54	
*3*L	1085454	0.85	23.41	1.64	0.58	25.65	2.00	
*3*L	1086356	0.73	4.54	0.57	0.30	23.81	2.06	
bab mean		0.83		1.02	0.47		2.41	0.44
logOR (*tan*)/logOR (*bab*)				2.51			0.57	

SNPs were selected based on their rank in both studies and had consistent allelic effects in both populations (*i.e.*, the identity of the putative light and dark alleles was consistent between Europe and South Africa). The mean allele frequencies for these SNPs and their log odds ratios were used in simulations and rejections sampling procedures, as described in the text. AF, allele frequency of the light allele; Chr, chromosome; OR, odds ratio. *P*-values and odds ratios were obtained from the CMH procedure.

allele frequencies from the European and South African reference populations (median frequency for the light allele of the four best candidates at each locus: Europe, *tan*, 0.2; *bab*, 0.83; SA, *tan*, 0.17; *bab*, 0.47; see Table 1). We assigned genotypes to individuals based on these allele frequencies, accounting for inbreeding in the South African population (using $F_{IT} = 0.35$), and phenotypes based on these genotypes, assuming a narrow-sense heritability (h^2) of 0.3. This value is conservative compared to the typical estimate of $h^2 = 0.5–0.9$ for abdominal pigmentation (Bickel *et al.* 2011; Dembeck *et al.* 2015), allowing for the effects of background loci. We assigned the same mean phenotype to individuals with the same genotype, regardless of population. As in the real experiment, simulated individuals were divided into replicates and selected for their phenotypes, and their pooled allele counts were analyzed with a CMH test.

In the simulations, we varied the degree of dominance (h) of the light allele at each locus and the relative additive effect sizes between the two loci (E) and then accepted the 1% of simulations best fitting the following four criteria: the ratio of effects of *tan* and *bab* in (i) Europe (actual results, *tan*-logOR/*bab*-logOR = 2.51) and (ii) South Africa (*tan*-logOR/*bab*-logOR = 0.57) and the ratio of effects of these SNPs across populations for (iii) *tan* (Europe-logOR/South Africa-logOR = 1.89) and (iv) *bab* (Europe-logOR/South Africa-logOR = 0.44; see Table 1). Additional details of the simulation methods are given in *Materials and Methods*. The results of rejection sampling are shown in Figure 3, Figure S12, and Figure S13. These most probable parameter combinations had dominant light alleles at both *bab* (consistent with Robertson *et al.* 1977), and *tan* and a stronger effect of *bab vs. tan* alleles. Nearly all accepted parameter combinations (92%) resulted in switching of the ranks of *tan* and *bab* for the European and South African populations.

To explore potential genetic interactions between the two loci, as suggested by the results of Robertson *et al.* (1977), we also considered a model with a simple epistatic interaction term (see *Materials and Methods*), allowing for synergistic (positive) or antagonistic (negative) effects of the light allele. Including a positive epistatic interaction between light alleles at the two loci substantially improves the model fit [Bayes factor: P(Data|Model with epistasis)/P(Data|Model without epistasis) = 644] (Csilléry *et al.* 2012). However, as other factors, such as subtle differences in experimental methods or differences between the populations at other causal loci, may yield a similar effect, we do not consider this analysis strong evidence for epistasis. In any case, the main results remain qualitatively unchanged: similar to the model without epistasis, this model also suggests dominance of the light alleles of both *tan* and *bab*, although support for *tan* dominance is reduced, and there is a stronger allelic effect of variants at the *bab* locus than at *tan* (Figure S14, Figure S15, and Figure S16).

Overall, our simulations suggest that the contrasting results between the European and the South African population can be at least partly explained by dominance and allele frequency differences between the populations: as the dark allele at *bab* is at high frequency in the South African population, the recessive homozygote is easily recovered even in the absence of inbreeding. In contrast, the recessive homozygote in Europe is rare (at an expected frequency of 2.9%), reducing its effect on phenotypic variation and in the experiment, in spite of its large allelic effect.

Comparison of the results with previous studies

We compared the results of our approach to two studies on natural genetic variation of female abdominal pigmentation. For this analysis, we chose to maximize the chance of overlap with the other two studies with a more liberal cutoff than the 5% false discovery rate (FDR) used for the main results, instead using the 1019 SNPs with *P*-values below the Bonferroni threshold ($P = 1.56e\text{-}8$ for Europe and $P = 1.25e\text{-}8$ for South Africa). We then looked for overlap between these SNPs and those identified as candidates in the other studies.

The first study, Bickel *et al.* (2011), investigated variants at the *bric-à-brac* locus (*bab1* and *bab2*) segregating in a sample

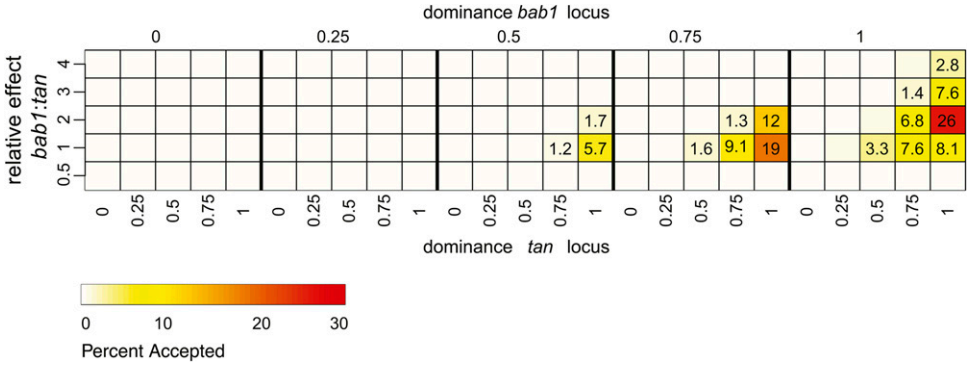


Figure 3 Rejection sampling results. Heatmap shows the percentage of accepted simulations using an ABC rejection method with 2,000,000 simulations uniformly sampling parameter space and an acceptance rate of 1%, using these criteria: logOR(tan)/logOR(bab1), Europe = 2.51, South Africa = 0.68; logOR(tan_{EU})/logOR(tan_{SA}) = 2.28; and logOR($bab1_{EU}$)/logOR($bab1_{SA}$) = 0.62. Parameters were sampled uniformly from [−0.125, 1.125] for dominance and [0.25, 4.5] for the additive effect of *bab1* relative to *tan*. For presentation in a grid, the parameter values were binned into intervals of 0.25 for dominance and 1 for relative additive effect. The axis labels indicate the center of each binning interval. The numbers in the cells give the percentage of accepted simulations in each binned parameter interval. Each cell represents a particular range of combinations of dominance of the light alleles at *bab1* (top axis) and *tan* (bottom axis) and the relative additive effect of *bab1* to *tan* loci (left axis).

of 96 inbred lines from an orchard near Winters, California, for their effects on A6 pigmentation. While this study measured pigmentation for a different tergite, genotypic values for A6 and A7 are strongly correlated (Gibert *et al.* 1998b), so an overlap between alleles affecting these traits seems plausible. In the 148-kb region covered by Bickel *et al.* (2011), there were 3066 uniquely mappable biallelic SNPs with a minor allele frequency ≥ 0.05; of these, 2367 were segregating in at least one of the populations used in this study, and 167 of these 2367 were identified as candidates by Bickel *et al.* (2011). Of these SNPs, 15 and 2 were significant after Bonferroni correction for multiple testing in the South African and European studies, respectively (Table S8), including the most strongly associated ones at the *bab* locus in the South African (*3*L:1084990) and the European (*3*L:1085454) samples. The overlapping SNPs fall into two regions, the regulatory region in the first intron of *bab1* and the noncoding intergenic region between *bab1* and *bab2*, ∼50 kb apart, much farther than the usual extent of linkage disequilibrium in *D. melanogaster* (Mackay *et al.* 2012; Franssen *et al.* 2015), suggesting that both regions may affect pigmentation variation independently.

The second study, Dembeck *et al.* (2015), looked genome-wide for variants associated with pigmentation of the fifth and sixth abdominal tergites (A5 and A6) of females in a panel of 175 inbred and sequenced lines created from a population at Raleigh, North Carolina (Mackay *et al.* 2012). We looked for overlap between our 1019 SNPs and the 154 variants with a *P*-value $<1 \times 10^{-5}$ in Dembeck *et al.* (2015) associated with effects on either or both segments. In all, we found 15 candidate SNPs in common between Dembeck *et al.* (2015) and this study; of these, 11 are in the first intron of *bab1* and 4 near the MSE upstream of *tan* (see Table S9). Encouragingly, the inferred light and dark alleles for the SNPs affecting A7 (present study) and those significantly affecting A6 (8 SNPs) or A5 (3 SNPs, with 2 of them also affecting A6) in Dembeck *et al.* (2015) were identical between the studies, consistent with similar associations and with correlated trait values for these segments (Gibert *et al.* 1998b). The remaining 6 of the 15 SNPs had a complex effect on pigmentation, affecting the contrast between the A5 and A6 segments—*e.g.*, at these SNPs, alleles that increase pigmentation in A6, reduce it in A5. As pigmentation of adjacent segments is more strongly correlated than that of more distant ones (Gibert *et al.* 1998b), we expect the effect of these alleles on A6, rather than that on A5, to predict the effect on the A7 segment measured here. Consistent with this, the inferred light and dark alleles for all 6 SNPs were identical for the A6 segment in Dembeck *et al.* (2015) and for A7 in the present study.

We next asked whether the 15 SNPs that overlap between this study and the Dembeck *et al.* (2015) study show quantitatively similar effects, again using log(OR) as a proxy for effect size in our study and focusing on the effect on A6 in the Dembeck *et al.* (2015) study. For the European population, we found no correlation, perhaps not surprisingly as the strongest associations occurred at different loci in this comparison. For the South African population, where the strongest associations were near the same locus (*i.e.*, near *bab*), the estimated effects of the alleles were strongly correlated (Spearman's $\rho = 0.86$, $P = 0.0013$).

We did not find any signs of association around the other genes identified to influence pigmentation in Dembeck *et al.* (2015), possibly due to differences in the trait measured (A6 *vs.* A7 pigmentation) or due to other differences in the methodology and power of these approaches. Finally, only one SNP—the most highly associated one in *bab1* for South Africa (*3*L:1084990)—overlapped between all three studies (Bickel *et al.* 2011; Dembeck *et al.* 2015; and this study).

Conclusions

Using the Pool-GWAS method on three different *D. melanogaster* populations, we show that variants at *tan* and *bab1* are associated with female abdominal pigmentation variation. In addition to the strong statistical support from the present study, our results are also supported by other

studies on pigmentation variation (Robertson *et al.* 1977; Kopp *et al.* 2003; Bastide *et al.* 2013; Dembeck *et al.* 2015) and by forward genetic studies showing that both *tan* and *bab* affect pigmentation when mutated. While the results for the South African and European populations are broadly consistent, the differences between them are potentially informative. The European study points to a large effect of *tan*, while the results from the South African population [and other studies on natural variants (Robertson *et al.* 1977; Dembeck *et al.* 2015)] attribute the strongest effects to *bab*. We show that this difference does not necessarily reflect alternative genetic architectures, but can be most simply explained by a combination of dominance and allele frequency: *e.g.*, in South Africa, the dark allele of *bab*, which appears to have a strong effect (present study and Robertson *et al.* 1977; Kopp *et al.* 2003; Dembeck *et al.* 2015), is at high enough frequency to occur in homozygous form reasonably often and thus can contribute substantially to phenotypic variation even if recessive. In Europe, in contrast, the dark *bab* allele is rare, and, if recessive, contributes little to phenotypic variation in wild flies. Regardless of its status in natural populations, a recessive *bab* allele can play a large role in differences between inbred lines selected for extreme phenotypes, as shown in previous studies (Robertson *et al.* 1977; Kopp *et al.* 2003).

Further, because distant populations may show different patterns of linkage disequilibrium, studying associations in different populations provides a means to help distinguish linked but noncausal variants from causal alleles contributing to the phenotype—*e.g.*, a SNP that is significant for one sample may show no association in another sample or have different inferred dark and light alleles between samples. Thus, by looking for associations that are repeatable across populations, repeated association studies allow further fine-mapping of potential causal alleles.

Finally, we also show that Pool-GWAS can be used to analyze inbred individuals from isofemale lines, a common tool in population genetics (David *et al.* 2005), offering a cost-effective way to study the genetic basis of phenotypic variation, including obtaining reasonable estimates of relative effect sizes for different loci. That this modified experimental design yields reliable results is important for logistical reasons, as it allows collection and phenotypic measurement of samples to be separated in time, such that the latter can be done under controlled laboratory conditions.

Acknowledgments

We thank A. Futschik for assistance with the statistical analysis. H. Bastide, P. Stöbe, and R. Tobler contributed to the experiment with the European populations. We are especially grateful to H. van Schalkwyk for fly collection and technical assistance. This research was supported by the Austrian Science Funds (FWF, P22725) and the European Research Council (ArchAdapt). The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

Literature Cited

Andolfatto, P., J. D. Wall, and M. Kreitman, 1999 Unusual haplotype structure at the proximal breakpoint of In(*2L*)t in a natural population of *Drosophila melanogaster*. Genetics 153: 1297–1311.

Bastide, H., A. Betancourt, V. Nolte, R. Tobler, P. Stöbe *et al.*, 2013 A genome-wide, fine-scale map of natural pigmentation variation in *Drosophila melanogaster*. PLoS Genet. 9: e1003534.

Bastide, H., A. Yassin, E. J. Johanning, and J. E. Pool, 2014 Pigmentation in *Drosophila melanogaster* reaches its maximum in Ethiopia and correlates most strongly with ultra-violet radiation in sub-Saharan Africa. BMC Evol. Biol. 14: 179.

Bickel, R. D., A. Kopp, and S. V. Nuzhdin, 2011 Composite effects of polymorphisms near multiple regulatory elements create a major-effect QTL. PLoS Genet. 7: e1001275.

Capy, P., J. David, and A. Robertson, 1988 Thoracic trident pigmentation in natural populations of *Drosophila simulans*: a comparison with *D. melanogaster*. Heredity 61: 263–268.

Cooley, A. M., L. Shefner, W. N. McLaughlin, E. E. Stewart, and P. J. Wittkopp, 2012 The ontogeny of color: developmental origins of divergent pigmentation in *Drosophila americana* and *D. novamexicana*. Evol. Dev. 14: 317–325.

Csilléry, K., O. François, and M. G. B. Blum, 2012 abc: an R package for approximate Bayesian computation (ABC). Methods Ecol. Evol. 3: 475–479.

David, J., P. Capy, V. Payant, and S. Tsakas, 1985 Thoracic trident pigmentation in *Drosophila melanogaster*: differentiation of geographical populations. Genet. Sel. Evol. 17: 211–224.

David, J. R., P. Gibert, H. Legout, G. Pétavy, P. Capy *et al.*, 2005 Isofemale lines in Drosophila: an empirical approach to quantitative trait analysis in natural populations. Heredity 94: 3–12.

Dembeck, L., W. Huang, M. Magwire, F. Lawrence, R. Lyman *et al.*, 2015 Genetic architecture of abdominal pigmentation in *Drosophila melanogaster*. PLoS Genet. 11: e1005163.

Falconer, D. S., and T. F. C. Mackay, 1996 *Introduction to Quantitative Genetics*, Ed. 4. Benjamin-Cummings, Menlo Park, CA.

Feder, A. F., D. A. Petrov, and A. O. Bergland, 2012 LDx: estimation of linkage disequilibrium from high-throughput pooled resequencing data. PLoS One 7: e48588.

Franssen, S. U., V. Nolte, R. Tobler, and C. Schlötterer, 2015 Patterns of linkage disequilibrium and long range hitchhiking in evolving experimental *Drosophila melanogaster* populations. Mol. Biol. Evol. 32: 495–509.

Gibert, P., B. Moreteau, J. Moreteau, and J. David, 1998a Genetic variability of quantitative traits in *Drosophila melanogaster* (fruit fly) natural populations: analysis of wild-living flies and of several laboratory generations. Heredity 80: 326–335.

Gibert, P., B. Moreteau, S. M. Scheiner, and J. R. David, 1998b Phenotypic plasticity of body pigmentation in Drosophila: correlated variations between segments. Genet. Sel. Evol. 30: 181.

Gompel, N., B. Prud'homme, P. J. Wittkopp, V. Kassner, and S. B. Carroll, 2005 Chance caught on the wing: cis-regulatory evolution and the origin of pigment patterns in Drosophila. Nature 433: 481–487.

Gorlov, I. P., J. H. Moore, B. Peng, J. L. Jin, O. Y. Gorlova *et al.*, 2014 SNP characteristics predict replication success in association studies. Hum. Genet. 133: 1477–1486.

Hoekstra, H. E., and M. W. Nachman, 2003 Different genes underlie adaptive melanism in different populations of rock pocket mice. Mol. Ecol. 12: 1185–1194.

Hoekstra, H. E., R. J. Hirschmann, R. A. Bundey, P. A. Insel, and J. P. Crossland, 2006 A single amino acid mutation contributes to adaptive beach mouse color pattern. Science 313: 101–104.

Ingram, C. J. E., T. O. Raga, A. Tarekegn, S. L. Browning, M. F. Elamin *et al.*, 2009 Multiple rare variants as a cause of a common phenotype: several different lactase persistence associated alleles in a single ethnic group. J. Mol. Evol. 69: 579–588.

Ioannidis, J. P., G. Thomas, and M. J. Daly, 2009 Validating, augmenting and refining genome-wide association signals. Nat. Rev. Genet. 10: 318–329.

Ioannidis, J. P., R. Tarone, and J. K. McLaughlin, 2011 The false-positive to false-negative ratio in epidemiologic studies. Epidemiology 22: 450–456.

Itan, Y., B. L. Jones, C. J. E. Ingram, D. M. Swallow, and M. G. Thomas, 2010 A worldwide correlation of lactase persistence phenotype and genotypes. BMC Evol. Biol. 10: 36.

Jeong, S., A. Rokas, and S. B. Carroll, 2006 Regulation of body pigmentation by the abdominal-b *hox* protein and its gain and loss in Drosophila evolution. Cell 125: 1387–1399.

Jeong, S., M. Rebeiz, P. Andolfatto, and T. Werner, 2008 The evolution of gene regulation underlies a morphological difference between two Drosophila sister species. Cell 132: 783–793.

Kapun, M., H. Van Schalkwyk, B. McAllister, T. Flatt, H. van Schalkwyk *et al.*, 2013 Inference of chromosomal inversion dynamics from Pool-Seq data in natural and laboratory populations of *Drosophila melanogaster*. Mol. Ecol. 23: 1813–1827.

Kofler, R., P. Orozco-terWengel, N. De Maio, R. V. Pandey, V. Nolte *et al.*, 2011a PoPoolation: a toolbox for population genetic analysis of next generation sequencing data from pooled individuals. PLoS One 6: e15925.

Kofler, R., R. V. Pandey, and C. Schlötterer, 2011b PoPoolation2: identifying differentiation between populations using sequencing of pooled DNA samples (Pool-Seq). Bioinformatics 27: 3435–3436.

Kopp, A., I. Duncan, and S. B. Carroll, 2000 Genetic control and evolution of sexually dimorphic characters in Drosophila. Nature 408: 553–559.

Kopp, A., R. M. Graze, S. Xu, S. B. Carroll, and S. V. Nuzhdin, 2003 Quantitative trait loci responsible for variation in sexually dimorphic traits in *Drosophila melanogaster*. Genetics 163: 771–787.

Lemeunier, F., and S. Aulard, 1992 Inversion polymorphism in *Drosophila Melanogaster*, pp. 339–407 in *Drosophila Inversion Polymorphism*, edited by C. B. Krimbas and J. R. Powell. CRC Press, Cleveland, OH/Boca Raton, FL.

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

Li, H., Y. Wu, R. J. F. Loos, and X. Lin, 2008 Variants in the Fat mass – and Obesity-associated (FTO) gene are not associated with obesity in a Chinese Han population. Diabetes 57: 264–268.

Li, H., B. Handsaker, A. Wysoker, T. Fennell, J. Ruan *et al.*, 2009 The Sequence Alignment/Map format and SAMtools. Bioinformatics 25: 2078–2079.

Llopart, A., S. Elwyn, D. Lachaise, and J. A. Coyne, 2002 Genetics of a difference in pigmentation between *Drosophila yakuba* and *Drosophila santomea*. Evolution 56: 2262–2277.

Mackay, T. F. C., S. Richards, E. Stone, A. Barbadilla, J. F. Ayroles *et al.*, 2012 The *Drosophila melanogaster* Genetic Reference Panel. Nature 482: 173–178.

McKenna, A., M. Hanna, E. Banks, A. Sivachenko, K. Cibulskis *et al.*, 2010 The Genome Analysis Toolkit: a MapReduce framework for analyzing next-generation DNA sequencing data. Genome Res. 20: 1297–1303.

Miller, S. A., D. D. Dykes, and H. F. Polesky, 1988 A simple salting out procedure for extracting DNA from human nucleated cells. Nucleic Acids Res. 16: 1215.

Munjal, A., D. Karan, P. Gibert, B. Moreteau, R. Parkash *et al.*, 1997 Thoracic trident pigmentation in *Drosophila melanogaster*: latitudinal and altitudinal clines in Indian populations. Genet. Sel. Evol. 29: 601–610.

Ng, M. C. Y., K. S. Park, B. Oh, C. H. T. Tam, Y. M. Cho *et al.*, 2008 Implication of genetic variants near TCF7L2, SLC30A8, HHEX, CDKAL1, CDKN2A/B, IGF2BP2, and FTO in type 2 diabetes and obesity in 6,719 Asians. Diabetes 57: 2226–2233.

Pandey, R. V., and C. Schlötterer, 2013 DistMap: a toolkit for distributed short read mapping on a Hadoop cluster. PLoS One 8: e72614.

Pool, J. E., and C. F. Aquadro, 2006 History and structure of sub-Saharan populations of *Drosophila melanogaster*. Genetics 174: 915–929.

Pool, J. E., and C. F. Aquadro, 2007 The genetic basis of adaptive pigmentation variation in *Drosophila melanogaster*. Mol. Ecol. 16: 2844–2851.

Pool, J. E., R. B. Corbett-Detig, R. P. Sugino, K. Stevens, C. M. Cardeno *et al.*, 2012 Population genomics of sub-Saharan *Drosophila melanogaster*: African diversity and non-African admixture. PLoS Genet. 8: e1003080.

Ranciaro, A., M. C. Campbell, J. B. Hirbo, W.-Y. Ko, A. Froment *et al.*, 2014 Genetic origins of lactase persistence and the spread of pastoralism in Africa. Am. J. Hum. Genet. 94: 496–510.

R Core Team, 2014 *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna.

Robertson, A., D. Briscoe, and J. Louw, 1977 Variation in abdomen pigmentation in *Drosophila melanogaster* females. Genetica 47: 73–76.

Rogers, W. A., J. R. Salomone, D. J. Tacy, E. M. Camino, K. A. Davis *et al.*, 2013 Recurrent modification of a conserved cis-regulatory element underlies fruit fly pigmentation diversity. PLoS Genet. 9: e1003740.

Rogers, W. A., S. Grover, S. J. Stringer, J. Parks, M. Rebeiz *et al.*, 2014 A survey of the trans-regulatory landscape for *Drosophila melanogaster* abdominal pigmentation. Dev. Biol. 385: 417–432.

Salomone, J. R., W. A. Rogers, M. Rebeiz, and T. M. Williams, 2013 The evolution of Bab paralog expression and abdominal pigmentation among Sophophora fruit fly species. Evol. Dev. 15: 442–457.

Siontis, K. C. M., N. Patsopoulos, and J. P. Ioannidis, 2010 Replication of past candidate loci for common diseases and phenotypes in 100 genome-wide association studies. Eur. J. Hum. Genet. 18: 832–837.

Stevison, L. S., K. B. Hoehn, and M. A. F. Noor, 2011 Effects of inversions on within- and between-species recombination and divergence. Genome Biol. Evol. 3: 830–841.

Sturtevant, A. H., 1919 A new species closely resembling *Drosophila melanogaster*. Psyche 26: 153–155.

Takahashi, A., and T. Takano-Shimizu, 2011 Divergent enhancer haplotype of ebony on inversion In(*3R*)Payne associated with pigmentation variation in a tropical population of *Drosophila melanogaster*. Mol. Ecol. 20: 4277–4287.

Takahashi, A., K. Takahashi, R. Ueda, and T. Takano-Shimizu, 2007 Natural variation of ebony gene controlling thoracic pigmentation in *Drosophila melanogaster*. Genetics 177: 1233–1237.

Telonis-Scott, M., A. Hoffmann, and C. M. Sgrò, 2011 The molecular genetics of clinal variation: a case study of ebony and thoracic trident pigmentation in *Drosophila melanogaster* from eastern Australia. Mol. Ecol. 20: 2100–2110.

True, J. R., S.-D. Yeh, B. T. Hovemann, T. Kemme, I. A. Meinertzhagen *et al.*, 2005 Drosophila *tan* encodes a novel hydrolase required in pigmentation and vision. PLoS Genet. 1: e63.

Visscher, P. M., M. Brown, M. I. McCarthy, and J. Yang, 2012 Five years of GWAS discovery. Am. J. Hum. Genet. 90: 7–24.

Williams, T. M., J. E. Selegue, T. Werner, N. Gompel, A. Kopp *et al.*, 2008 The regulation and evolution of a genetic switch controlling sexually dimorphic traits in Drosophila. Cell 134: 610–623.

Wittkopp, P. J., S. B. Carroll, and A. Kopp, 2003 Evolution in black and white: genetic control of pigment patterns in Drosophila. Trends Genet. 19: 495–504.

Wittkopp, P. J., G. Smith-Winberry, L. L. Arnold, E. M. Thompson, A. M. Cooley *et al.*, 2010 Local adaptation for body color in *Drosophila americana*. Heredity 106: 592–602.

Communicating editor: C. D. Jones

GENETICS

Supporting Information
www.genetics.org/lookup/suppl/doi:10.1534/genetics.115.183376/-/DC1

Reconciling Differences in Pool-GWAS Between Populations: A Case Study of Female Abdominal Pigmentation in *Drosophila melanogaster*

Lukas Endler, Andrea J. Betancourt, Viola Nolte, and Christian Schlötterer

DOI: 10.1534/genetics.115.183376

Supplementary Tables and Figures:

Table S1 Sequencing Depths. Sequencing depths of the Austrian, Italian, and South African replicates used for the Pool-GWAS. For the CMH test only the extremely light and dark individuals were used (very light and very dark).

	replicate	very light	unselected population	very dark
Austria 2010	1	85	106	65
	2	97	145	94
	3	100	110	79
Italy 2011	1	37	84	27
	2	126	233	143
	3	126	278	137
SA 2012	1	61	-	73
	2	63	73	93
	3	109	-	106

Table S2: 100 most highly associated SNPs in the South African (SA) GWAS. SNPs are ranked by their P-values in the CMH test. For each SNP the following are provided: chromosome (Chr), position, the light, dark and ancestral allele (Anc, against *D. simulans*), and the consistency of the direction of the allele frequency changes (con), the frequencies of the light allele in the South African (SA) and European (EU) populations, the negative decadic logarithm of the P-values in the CMH tests (-log10P) and the natural logarithm of the average odds ratio of the light

allele count in the very light and very dark pools (log(OR)), as well as their ranks (rank P, rank OR) for both populations, the classification of the effect of the SNP, and if applicable the genic location (Gene), and the amino acid change (AA) against the reference. Positions are colored according to whether they are close to *tan* (blue), *bab* (green), or *ebony* (orange). Inconsistent allele count changes are labeled in red in both the con and log(OR) columns. The red line indicates the FDR < 0.05 cutoff derived using an empirical null hypothesis.
(see attached excel file TableS2.xlsx)

Table S3: 100 most highly associated SNPs in the European (EU) GWAS. SNPs are ranked by their P-values in the CMH test. For each SNP the following are provided: chromosome (Chr), position, the light, dark and ancestral allele (Anc, against *D. simulans*), and the consistency of the direction of the allele frequency changes (con), the frequencies of the light allele in the South African (SA) and European (EU) populations, the negative decadic logarithm of the P-values in the CMH tests (-log10P) and the natural logarithm of the average odds ratio of the light allele count in the very light and very dark pools (log(OR)), as well as their ranks (rank P, rank OR) for both populations, the classification of the effect of the SNP, and if applicable the genic location (Gene), and the amino acid change (AA) against the reference. Positions are colored according to whether they are close to *tan* (blue), *bab* (green), or *ebony* (orange). Inconsistent allele count changes are labeled in red in both the con and log(OR) columns. The red line indicates the FDR < 0.05 cutoff derived using an empirical null hypothesis.
(see attached excel file TableS3.xlsx)

Table S4: 50 most highly associated SNPs around *tan* in the South African (SA) GWAS. SNPs are ranked by their P-values in the CMH test. For each SNP the following are provided: chromosome (Chr), position, the light, dark and ancestral allele (Anc, against *D. simulans*), and the consistency of the direction of the allele frequency changes (con), the frequencies of the light allele in the South African (SA) and European (EU) populations, the negative decadic logarithm of the P-values in the CMH tests (-log10P) and the natural logarithm of the average odds ratio of the light allele count in the very light and very dark pools (log(OR)), as well as their ranks (rank P, rank OR) for both populations, the classification of the effect of the SNP, and if applicable the genic location (Gene), and the amino acid change (AA) against the reference. Inconsistent allele count changes are labeled in red in both the con and log(OR) columns.
(see attached excel file TableS4.xls)x

Table S5: 50 most highly associated SNPs around *bab* in the South African (SA) GWAS. SNPs are ranked by their P-values in the CMH test. For each SNP the following are provided: chromosome (Chr), position, the light, dark and ancestral allele (Anc, against *D. simulans*), and the consistency of the direction of the allele frequency changes (con), the frequencies of the light allele in the South African (SA) and European (EU) populations, the negative decadic logarithm of the P-values in the CMH tests (-log10P) and the natural logarithm of the average odds ratio of the light allele count in the very light and very dark pools (log(OR)), as well as their ranks (rank P, rank OR) for both populations, the classification of the effect of the SNP, and if applicable the genic location (Gene), and the amino acid change (AA) against the

reference. The red line indicates the FDR < 0.05 cutoff derived using an empirical null hypothesis.

(see attached excel file TableS5.xlsx)

Table S6: 50 most highly associated SNPs around *tan* in the European (EU) GWAS. SNPS are ranked by their P-values in the CMH test. For each SNP the following are provided: chromosome (Chr), position, the light, dark and ancestral allele (Anc, against *D. simulans*), and the consistency of the direction of the allele frequency changes (con), the frequencies of the light allele in the South African (SA) and European (EU) populations, the negative decadic logarithm of the P-values in the CMH tests (-log10P) and the natural logarithm of the average odds ratio of the light allele count in the very light and very dark pools (log(OR)), as well as their ranks (rank P, rank OR) for both populations, the classification of the effect of the SNP, and if applicable the genic location (Gene), and the amino acid change (AA) against the reference. Inconsistent allele count changes are labeled in red in both the con and log(OR) columns. The red line indicates the FDR < 0.05 cutoff derived using an empirical null hypothesis.

(see attached excel file TableS6.xlsx)

Table S7: 50 most highly associated SNPs around *bab* in the European (EU) GWAS. SNPs are ranked by their P-values in the CMH test.

For each SNP the following are provided: chromosome (Chr), position, the light, dark and ancestral allele (Anc, against *D. simulans*), and the consistency of the direction of the allele frequency changes (con), the frequencies of the light allele in the South

African (SA) and European (EU) populations, the negative decadic logarithm of the P-values in the CMH tests (-log10P) and the natural logarithm of the average odds ratio of the light allele count in the very light and very dark pools (log(OR)), as well as their ranks (rank P, rank OR) for both populations, the classification of the effect of the SNP, and if applicable the genic location (Gene), and the amino acid change (AA) against the reference. Inconsistent allele count changes are labeled in red in both the con and log(OR) columns. The red line indicates the FDR < 0.05 cutoff derived using an empirical null hypothesis.
(see attached excel file TableS7.xlsx)

Table S8: Overlap with Bickel *et al.* 2011 SNPs significantly associated with female abdominal pigmentation (Tergite 6) in Bickel *et al.* 2011 and significantly associated after Bonferroni correction (significance level 0.05) for multiple testing in either the European (P<=1.56E-8) or the South African (P<=1.25E-8) GWAS. The colors of the **-logP** column indicate whether the SNP was found significantly associated in the respective population (EU: European, SA: South African) after Bonferroni correction (**blue**) or using an empirical null distribution for FDR correction (**red**). Positions were matched by aligning the sequenced fragment provided by Dr. Bickel to the reference 3L chromosome (v. 5.18) using progressive Mauve and SNPs checked for consistency of alleles. The last column gives the position of the SNPs on the DNA fragment of the 3L chromosome sequenced in Bickel *et. al* 2011, with the region defined in Bickel *et al.* indicated by colors (white: first intron of bab1, **blue**: region 1 around abdominal CRE in the first intron of bab1, **yellow**: highly linked region in the intergenic sequence between bab1 and bab2). Out of 3066 uniquely mappable SNPs in Bickel *et*

al. with a minor allele frequency >= 0.05, 2367 were segregating in at least one of the populations used in this study, with 167 of these being significantly associated with abdominal pigmentation according to Bickel *et al.* (2011).

Chr	Position	Allele	-logP			Position in Bickel *et al.*
		light	dark	EU	SA	
3L	1041545	T	A	0.4	**12.4**	12798
3L	1077795	T	G	0.8	**8.5**	41670
3L	1078420	C	T	0.8	**8.7**	42307
3L	1080569	A	G	0.3	**9.7**	44482
3L	1080600	G	A	1.0	**9.9**	44517
3L	1080908	C	G	0.4	**24.1**	44836
3L	1080915	G	A	0.4	**21.0**	44843
3L	1080934	T	C	0.0	**22.5**	44862
3L	1080935	C	A	0.1	**21.7**	44863
3L	1081068	G	C	0.7	**14.8**	44996
3L	1084990	G	C	**8.4**	**42.4**	48973
3L	1085454	C	T	**23.4**	**25.6**	49437
3L	1085642	A	T	0.0	**23.9**	49625
3L	1135568	A	T	0.5	**9.6**	100457
3L	1139615	G	A	1.0	**8.6**	104557

Table S9: Overlap with Dembeck *et al.* List of SNPs found associated with abdominal pigmentation in both the Dembeck *et al.* 2015 study (*p*-value < 1E-5, **bold**) and this study (*p*-value <= 1.56E-8 for the European population or *p*-value <= 1.25E-8 the South African lines). Columns C to U are adapted from the table S4 of Dembeck *et al.* (2015). For the alleles of the Dembeck *et al.* study, **dark brown** indicates the dark allele, **yellow** the light allele, as derived from the effect on T6. For chromosomal positions **green** stands for SNPs close to the bab locus, **blue** for SNPs close to the tan locus. The colors of the *p*-value columns of the EU and SA GWAS indicate whether the SNP was found significantly associated in the respective population after Bonferroni correction (**blue**) or using an empirical null distribution for FDR correction (**red**).

(see attached excel file tables_S2_to_S7_and_S9.xls)

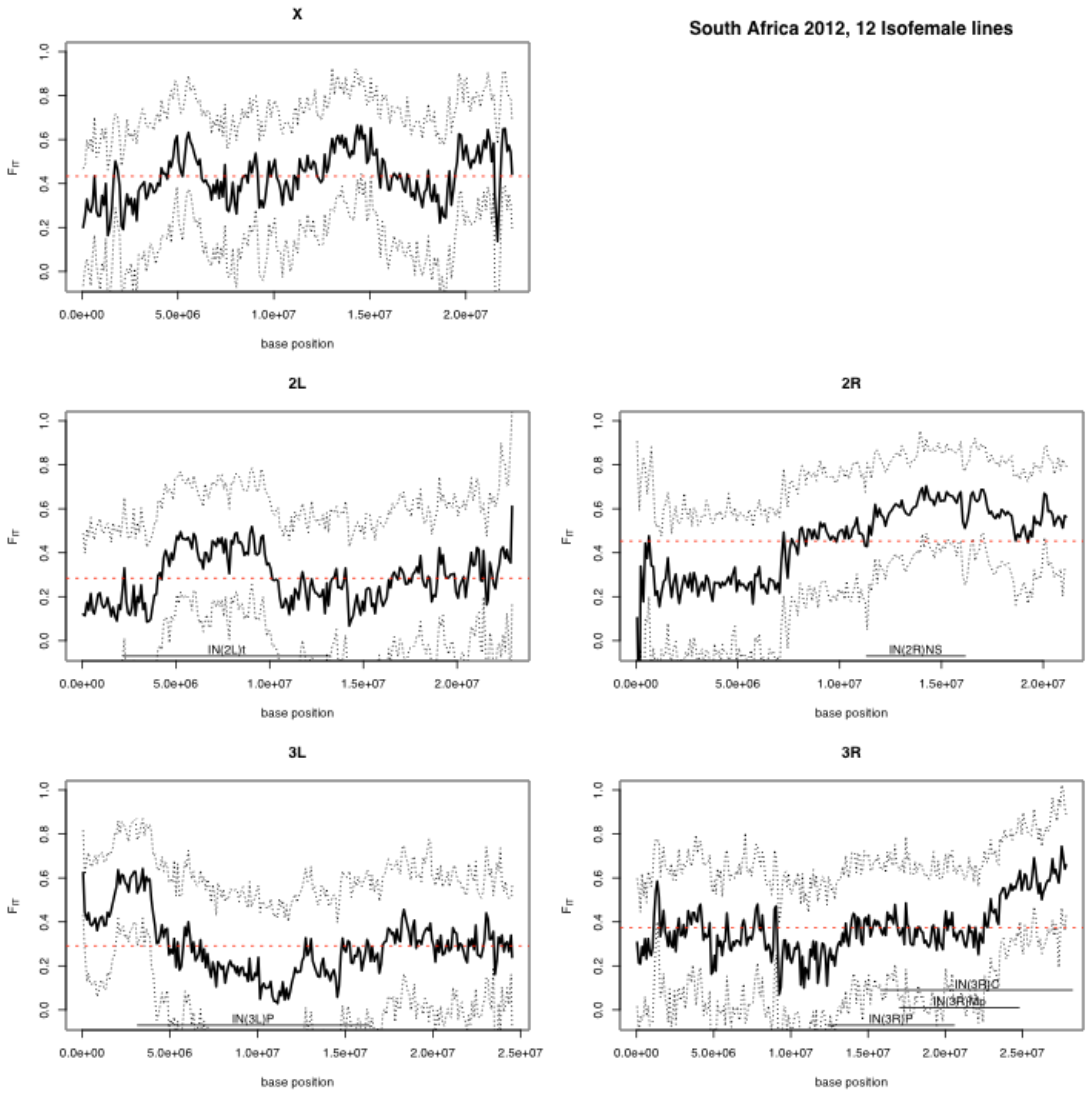


Figure S1 Inbreeding coefficients (F_{IT}). F_{IT} calculated for 12 individuals of the South African isofemale lines over non-overlapping windows of 100 kb (solid lines: window mean, dotted lines: +/- 1 standard deviation). The dashed red line shows the chromosomal median value, the horizontal black lines indicate the position of cosmopolitan inversions.

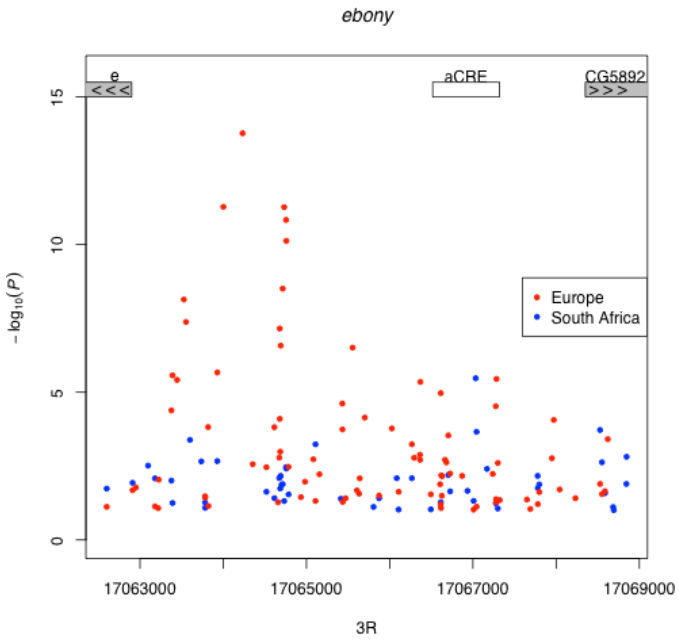


Figure S2 *ebony* region. Close-up of the region upstream the *ebony* (*e*) gene. As in Figure 1D. The white box indicates the location of a previously described abdominal cis-regulatory element (aCRE) (Rebeiz *et al.* 2009).

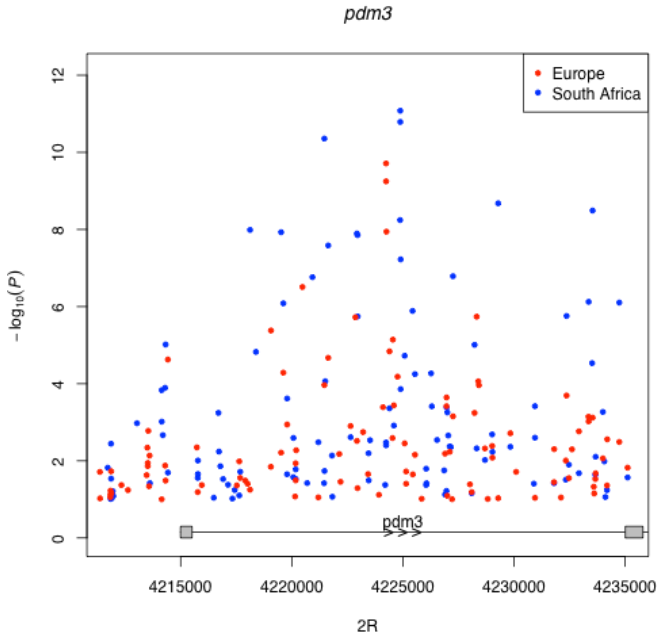


Figure S3 *pdm3* region. Close-up of the region around the first intron of the *pdm3* gene. As in Figure 1D.

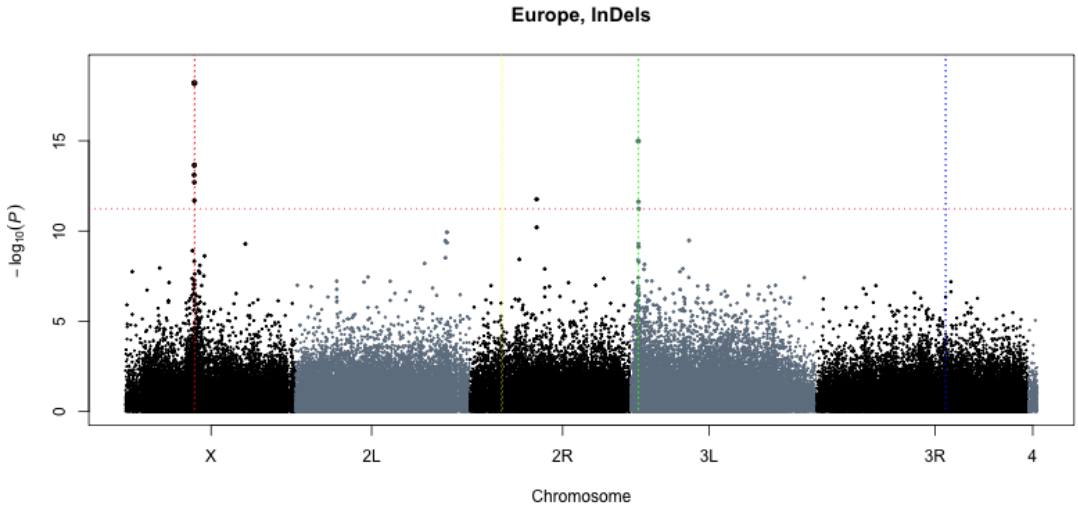


Figure S4 Indels GWAS Europe. Manhattan plot of a GWAS on short insertions and deletions in the European populations. As in Figure 1. The horizontal line indicates the FDR=0.5 threshold at $P = 10^{-11.2}$.

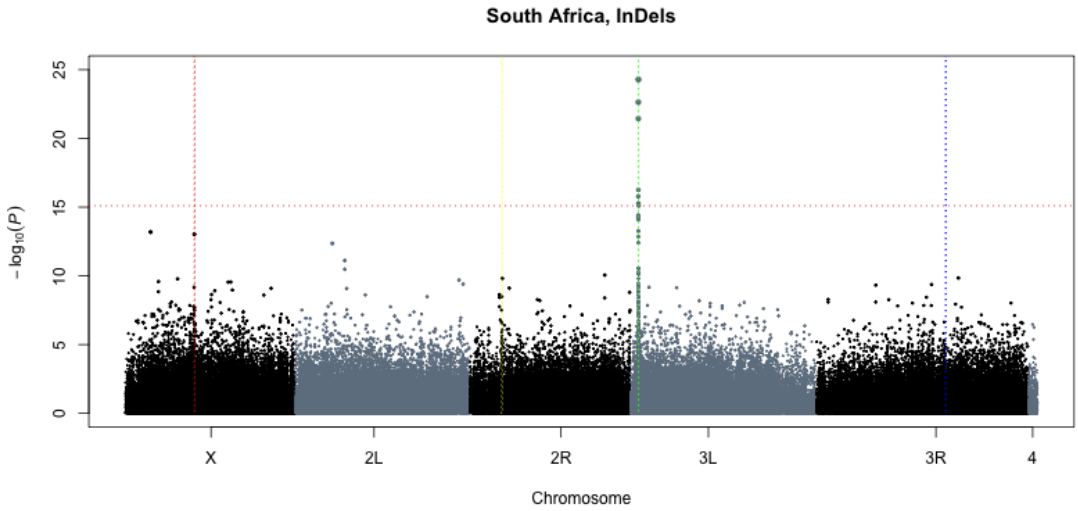


Figure S5 Indels GWAS South Africa. Manhattan plot of a GWAS on short insertions and deletions in the South African isofemale lines. As in Figure 1. The horizontal line indicates the FDR=0.5 threshold at $P = 10^{-15.1}$.

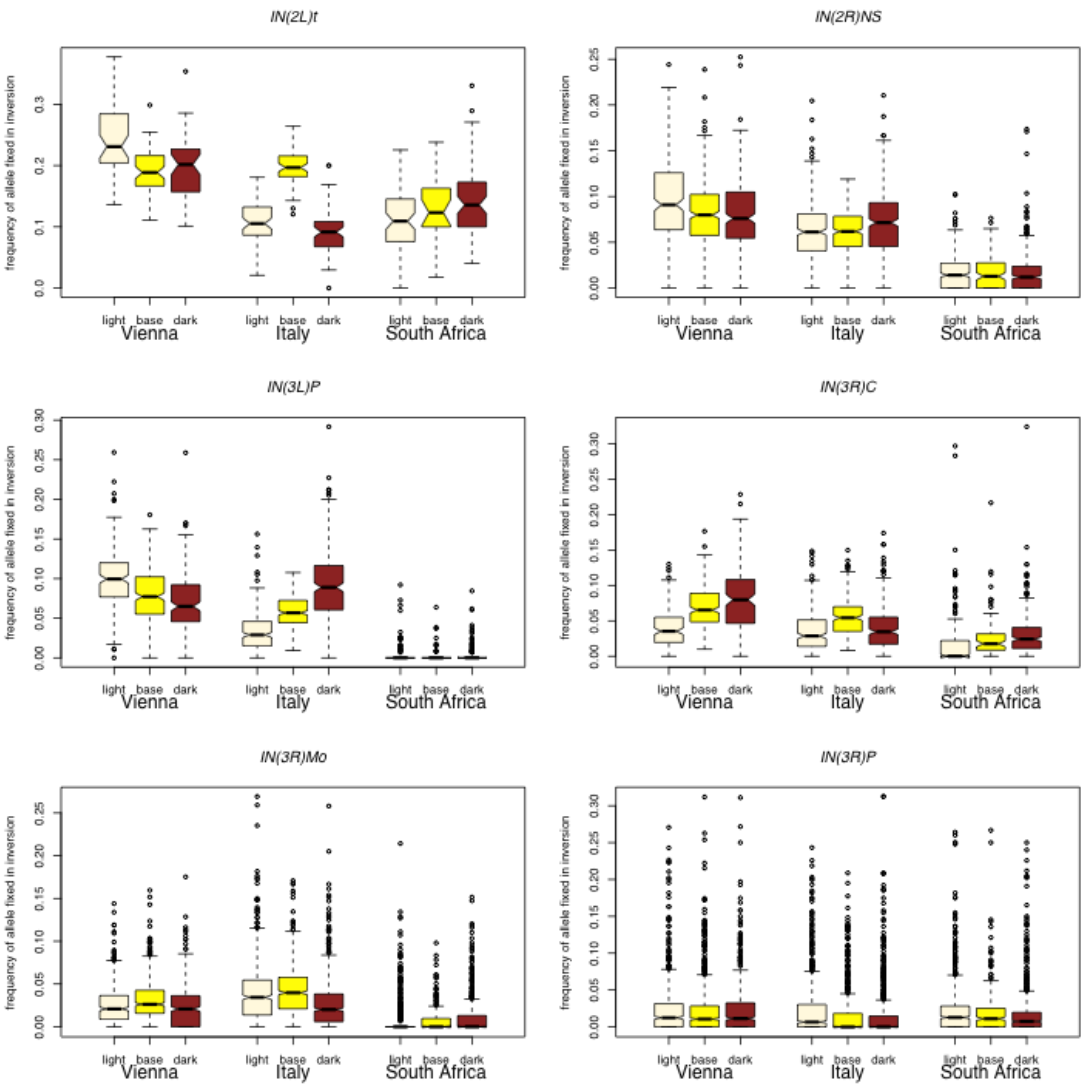
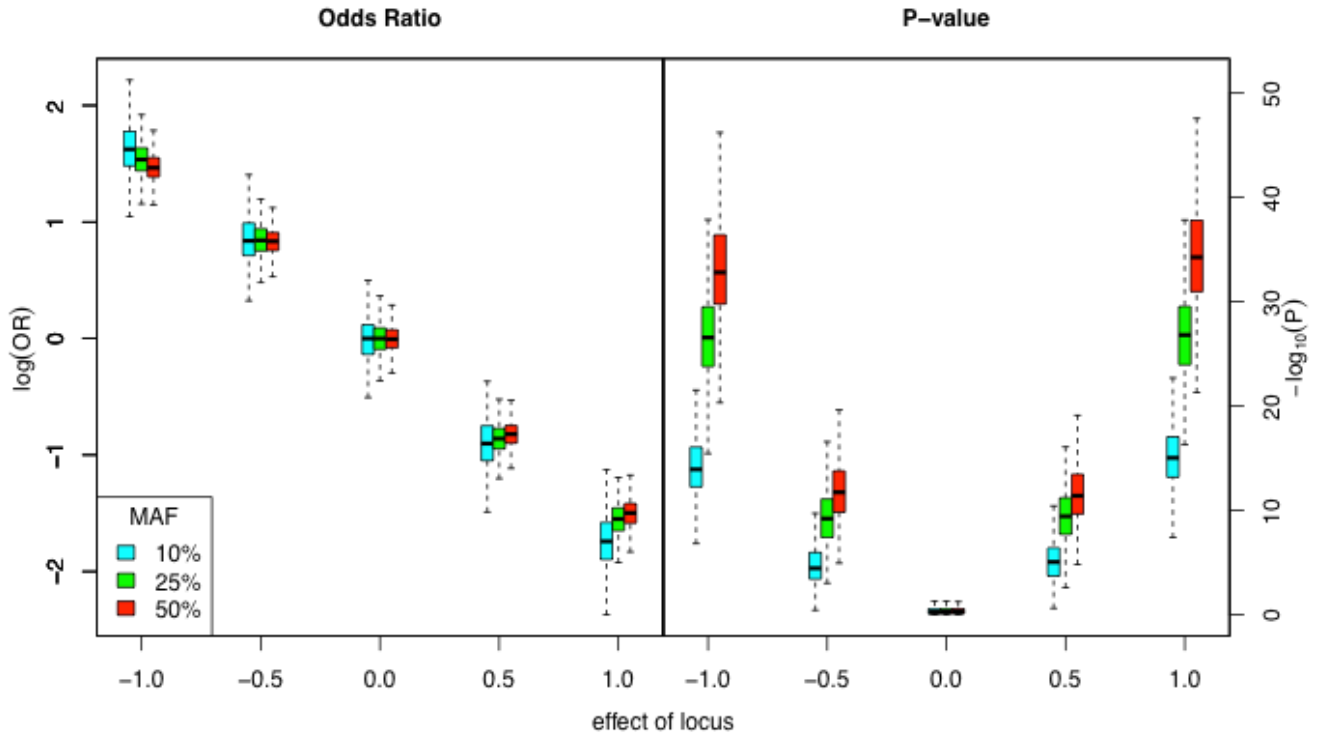


Figure S6 Inversion frequencies. Frequencies of alleles fixed in certain cosmopolitan inversions in the pools exhibiting extreme pigmentation (light:white; dark:brown) and the base populations (yellow) of the Viennese, Italian and South African (left, middle, right) flies. The markers used where taken from (Kapun *et al.* 2013) and for the graphical representation all replicates where combined. The notches approximate the 95% confidence intervals as calculated by the boxplot routine of R without accounting for the replicates.

A)


Odds Ratio
P−value
log(OR)
$-\log_{10}(P)$
MAF
10%
25%
50%
−1.0
−0.5
0.0
0.5
1.0
effect of locus

B)

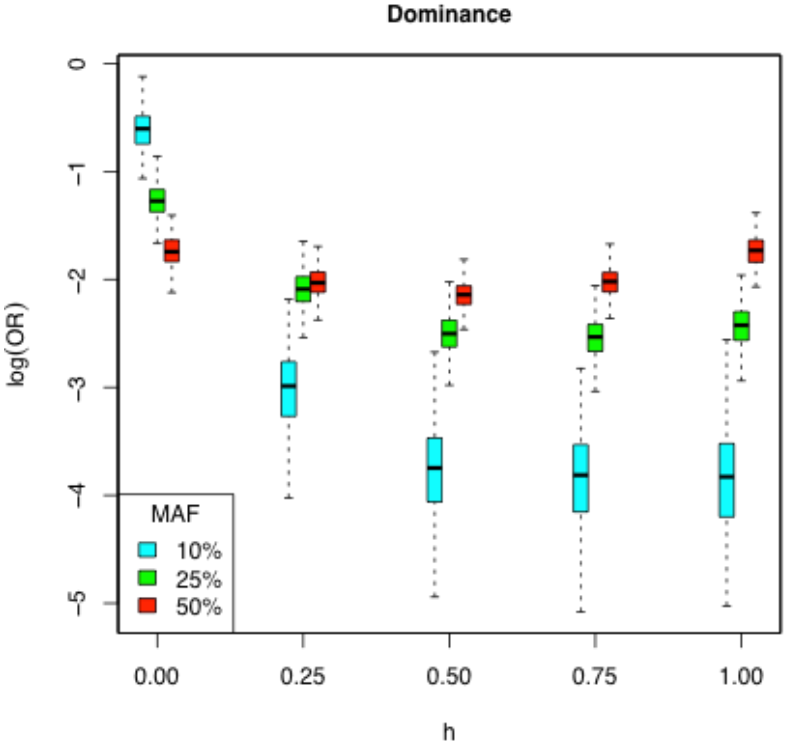

Dominance
log(OR)
MAF
10%
25%
50%
0.00
0.25
0.50
0.75
1.00
h

Figure S7 Effect of allele frequency, additive effect, and dominance. Dependence of different statistics derived from *in silico* Pool-GWAS experiments on the allele frequency and additive effect size (**A**), and dominance (**B** of loci. For **A)** 5 codominant loci with additive effects equally distributed between -1 and 1 were modeled with a minor allele frequency (MAF) of 10, 25 and 50% (blue, green and red boxplots). For each allele frequency 1000 simulations with 3 replicates of 700 individuals and a pool size of 100 individuals were performed. The trait was considered to have a heritability, h^2, of 0.25. On the left side the distributions of the logarithms of the average odds ratios (log(OR)), on the right the negative decadic logarithms of the *p*-values ($-\log_{10}(P)$) are plotted for the different loci and allele frequencies. For each combination of allele frequency and dominance in **B)** one locus with an additive effect of 0.75 was simulated 250 times with the other parameters as in **A)**.

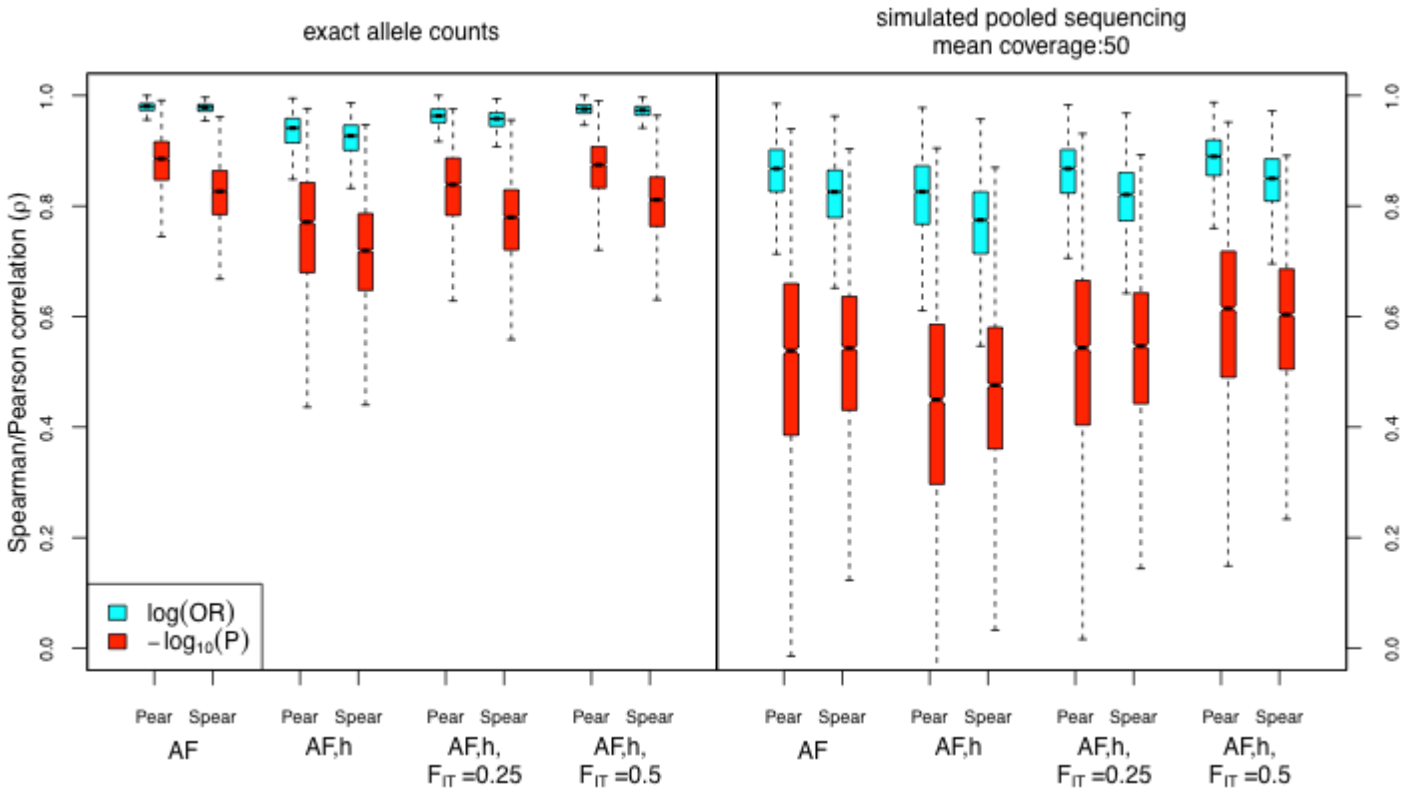


Figure S8 Correlations between logOR and P-values with additive effect. Correlations between different statistics derived from the pool-GWAS and the additive effect size for simulated pool-GWAS experiments under different settings. For each setting 5000 *in silico* pool GWAS experiments were performed as in Figure S4, but considering 17 additive loci with an additive effect between -1 and 1. The distributions of the ranked Pearson's (Pear) or Spearman's (Spear) correlation coefficients between either the log(OR) (blue) or the $-\log_{10}(P)$ (red) and the actual additive effect sizes are shown as boxplots. As the *p*-value is independent of the directionality of the additive effect, the correlation was calculated for the absolute value of the additive effect. In the settings either only the allele frequencies (AF), or the allele frequencies and the dominance (AF,h) are randomly assigned to each locus for each simulation. Additionally different degrees of inbreeding (F_{IT}=0.25 and 0.5) are simulated. For the left panel, the pool-GWAS was simulated using exact allele counts for the extreme pools, while for the right panel the additional noise introduced by pooled sequencing to an average coverage of 50 was considered, using a Poisson-distributed random variable for the coverage and binomial sampling of the alleles.

Figure S9 Allele frequencies of most significant SNPs. Most significant SNPs around *bab1* (upper figure, threshold in the European GWAS $P<10^{-12}$ and in the South African GWAS $P<10^{-22}$) and *tan* (lower figure, Europe: $P<10^{-16}$, South Africa: $P<10^{-12}$). In the upper panel the negative log(P) values of the SNPs are shown, in the lower the frequencies of the common major allele in the base populations (filled circles) and the different extreme pools (light: open triangles, dark: filled triangles) are shown. Numbers on the abscissa give the chromosomal positions of SNPs, the dotted lines indicate closely linked polymorphisms ($r^2>0.75$). The grey

rectangles show the location of genes, the white ones known regulatory regions. European: red, South African: blue

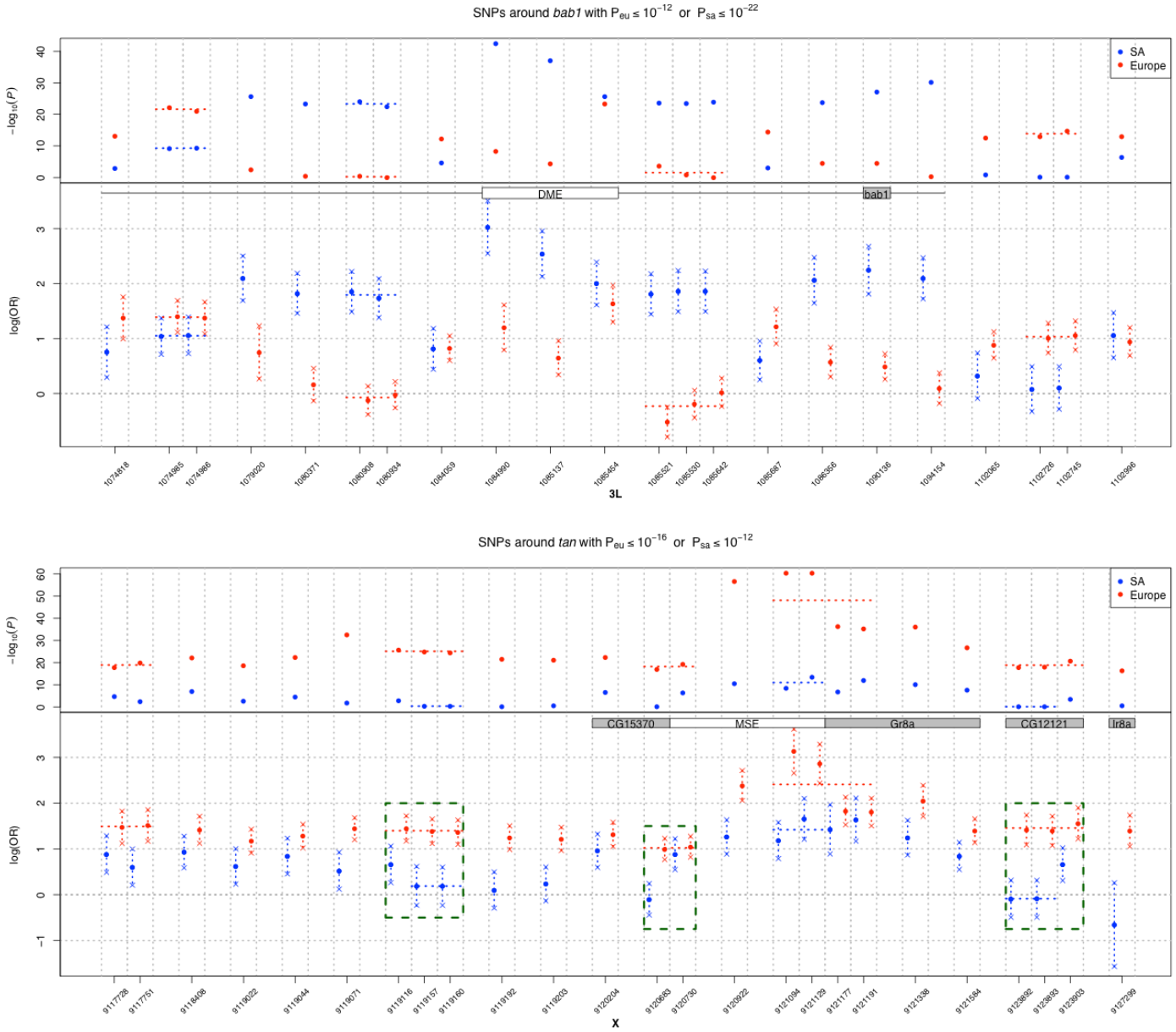


Figure S10 logOR of most significant SNPs . Highest ranking SNPs near the *bab1* (upper figure) and *tan* (lower figure) loci, as in Figure S9. In the lower panel the logarithm of the average odds ratio of the light alleles (filled circles) with their 95% confidence intervals (x) are shown. For positions with inconsistent allele changes between the two populations, the population in which the SNP exhibited a lower *p*-value was chosen for determination of the light allele. Numbers on the abscissa give the chromosomal positions of SNPs, the dotted lines indicate closely linked polymorphisms (r^2>0.75). The grey rectangles show the location of genes, the white ones known regulatory regions. The green dashed boxes indicate groups of SNPs in which cross-population comparison allows to refine the set of candidates. European: red, South African: blue

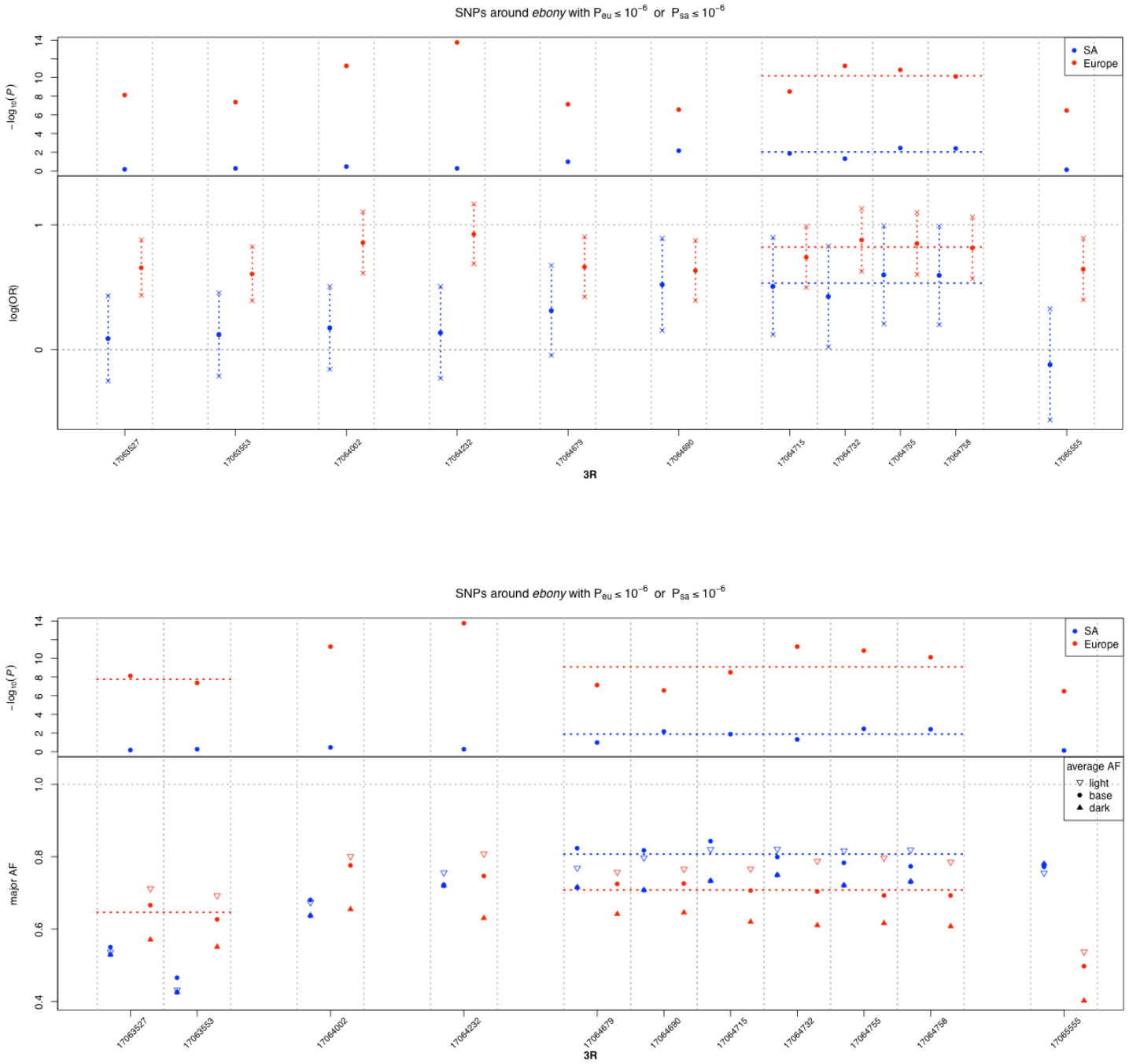


Figure S11 Allele frequency and logOR around *ebony*. Most significant SNPs upstream of *ebony* as in Figures S9 and S10. In the lower panel of each figure either the logarithm of the average odds ratio of the light alleles (filled circles) with their 95% confidence intervals (x) (upper figure) or the allele frequencies of the base populations (filled circles), or the different extreme pools (very light (empty triangles) and very dark (filled triangles)) are shown. For positions with inconsistent allele changes between the two populations, the population in which the SNP exhibited a lower *p*-value was chosen for determination of the light allele. Numbers on the

abscissa give the chromosomal positions of SNPs, the dotted lines indicate closely linked polymorphisms (r^2>0.75). European: red, South African: small extremes: blue

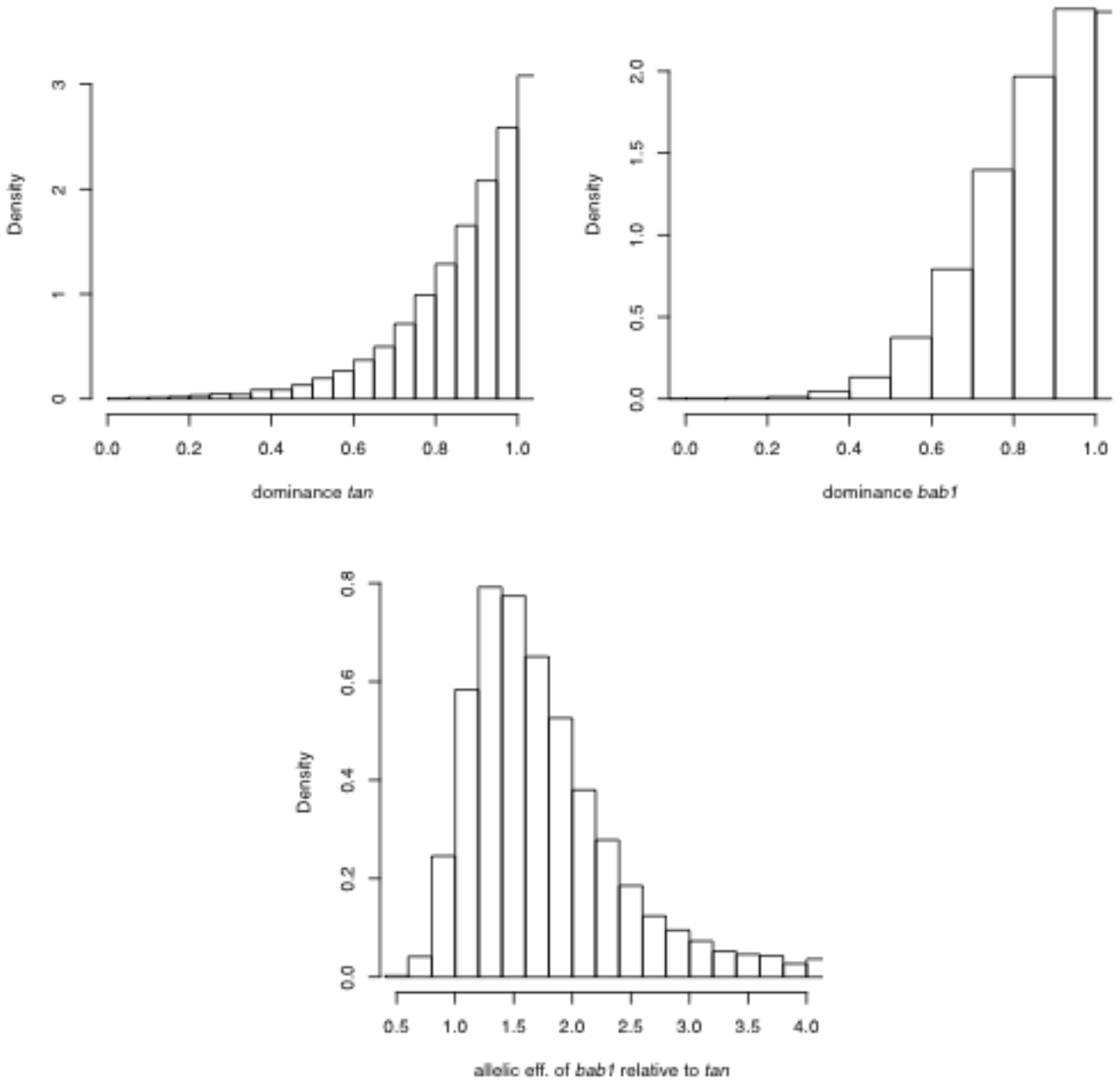


Figure S12 Distribution of accepted parameter values. Histogram of the parameter values of the simulations of the model without epistatic interactions accepted by the ABC when compared to the four criteria.

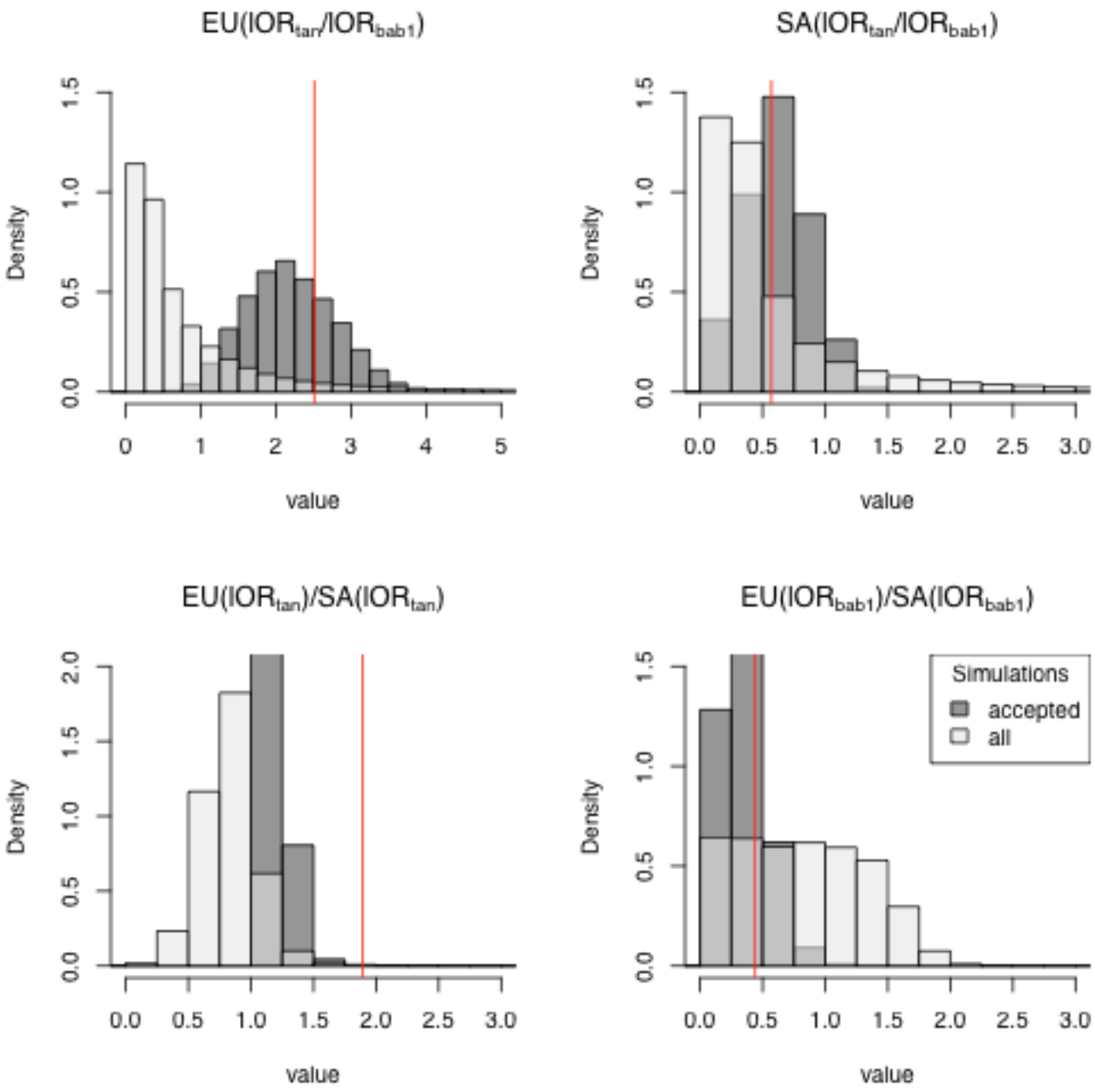


Figure S13 Distribution of accepted criteria values. Histogram of the values of the four criteria of only the accepted (dark grey) and all simulations for the model without epistatic interactions. The red vertical lines indicate the experimental values of the criteria used for selecting the simulations.

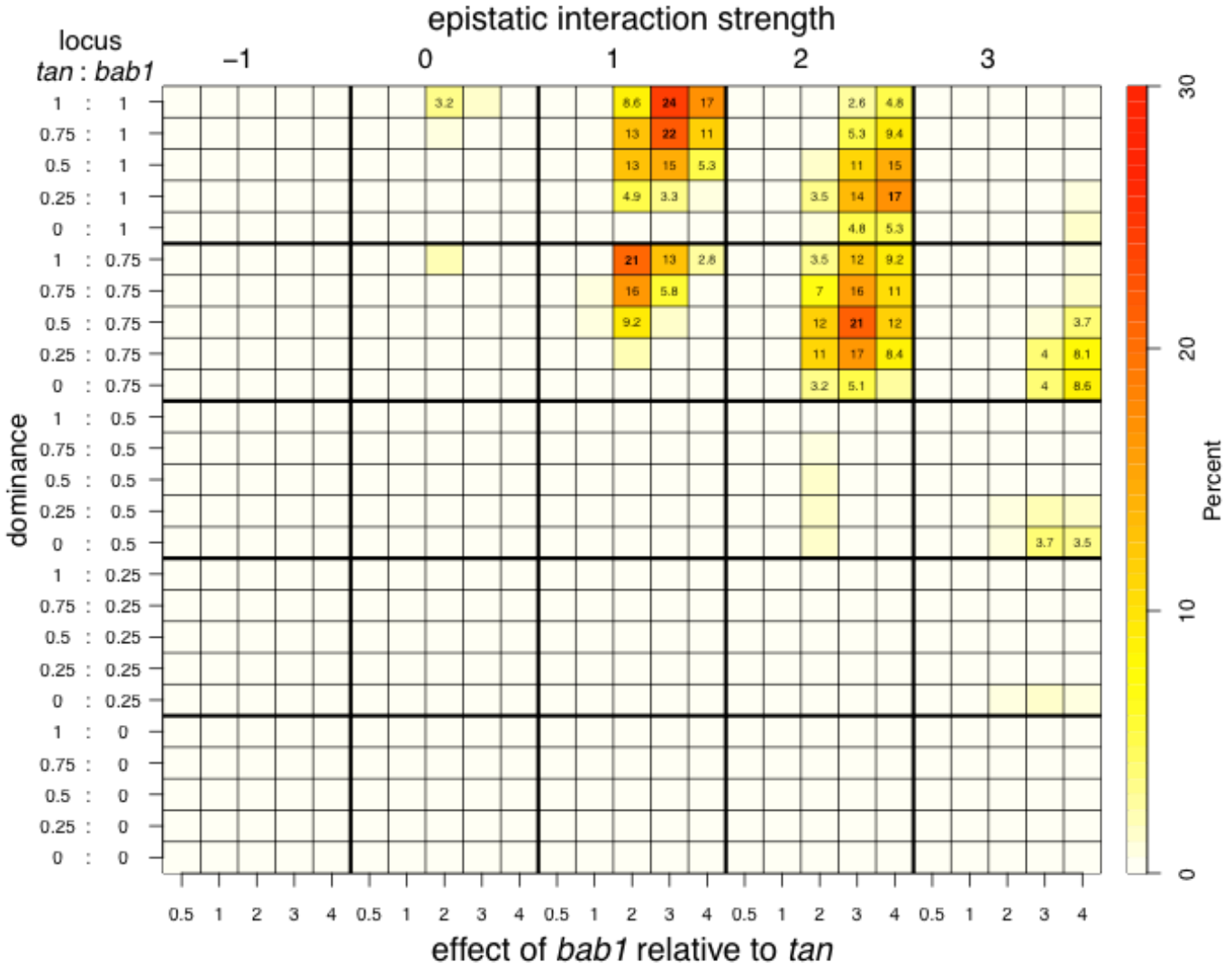


Figure S14 ABC results with epistatic interactions. Heatmap as in Figure 3 of the percentage of accepted simulations for the model with epistatic interactions using an ABC rejection method with 2,000,000 simulations uniformly sampling parameter space, an acceptance rate of 1% and the using the criteria mentioned in the text (logOR(tan)/logOR(bab1): Europe=2.51, South Africa=0.68; $\log OR(tan_{EU})/\log OR(tan_{SA})$=2.28 and $\log OR(bab1_{EU})/\log OR(bab1_{SA})$=0.62). The sampled parameter intervals were [-0.125,1.125] for the dominances of the light alleles of *tan* and *bab1* and [0.5,4] for the additive effect of *bab1* relative to *tan*, and [-1.5,3.5] for epistatic interactions. For presentation in a grid, the parameter values were binned into intervals of 0.25 for dominance and 1 for relative additive effect and epsitatic interaction strength. The axes labels indicate the center of each binning interval. The numbers in the cells give the percent of accepted simulations in each binned

parameter interval. Rows show different combinations of dominance of the light alleles at the two loci, columns different relative effect sizes of *bab1* to *tan*, and values of epistatic interaction strengths between the loci.

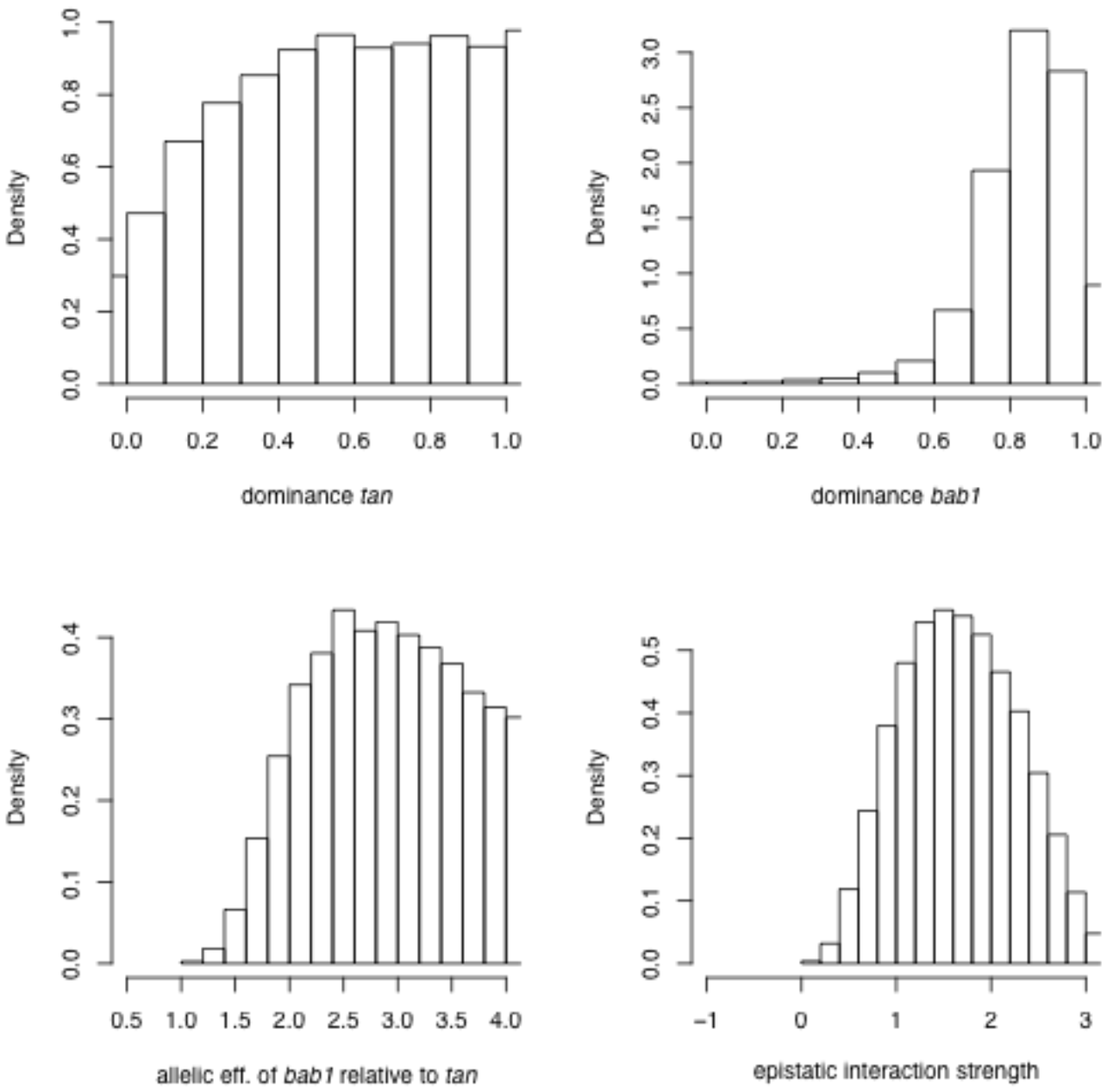


Figure S15 Distribution of accepted parameter values with epistatic interactions. Histograms of the parameter values of the accepted simulations of the model with epistatic interactions.

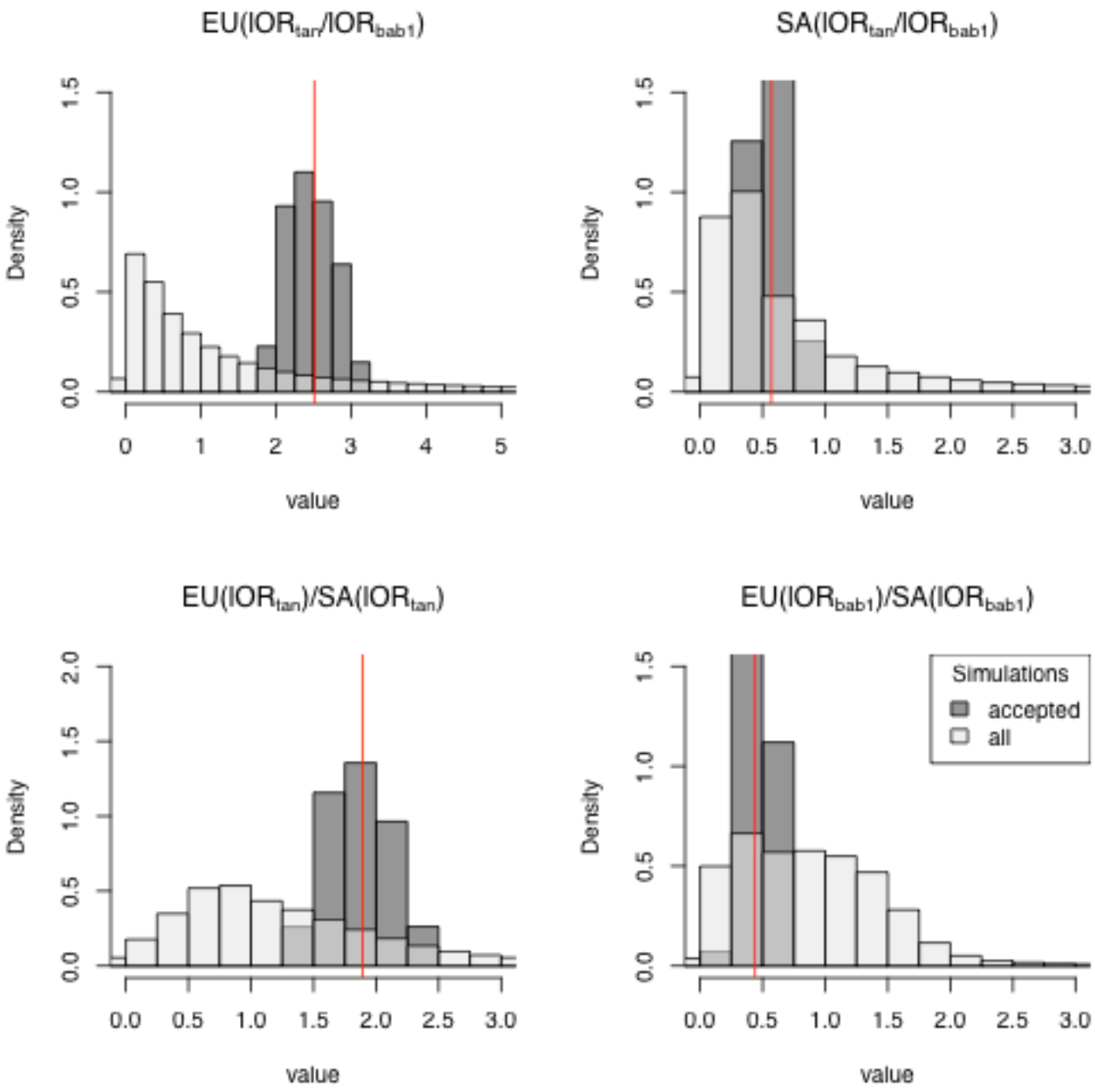


Figure S16 Accepted criteria values with epistatic interactions. Histograms of the values of the four criteria used for the accepted (dark grey) and all simulations of the model with epistatic interactions.

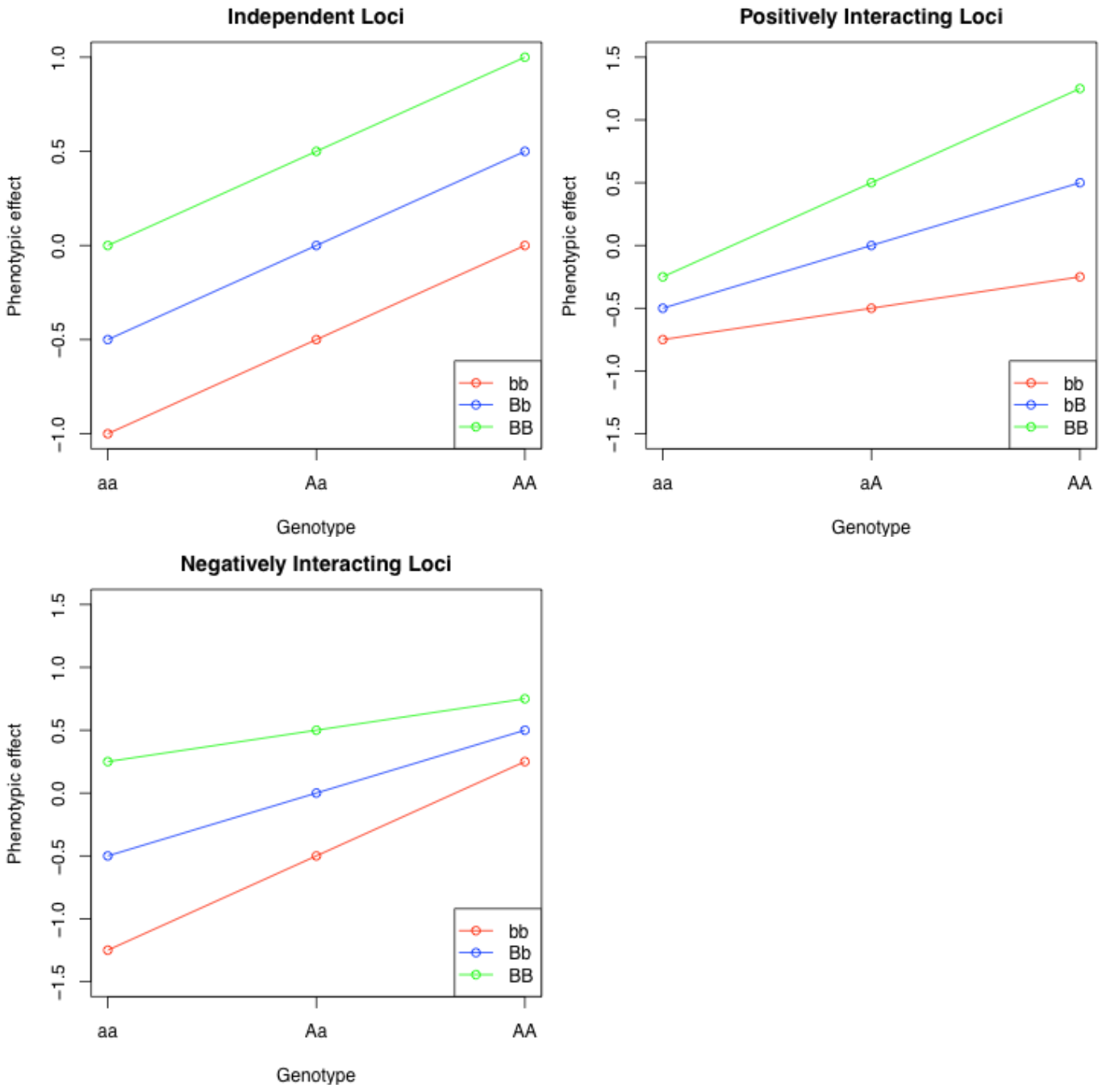


Figure S17 Phenotypic effects of epistatic interaction. Phenotypic effects for different combinations of two biallelic loci *A* and *B* using the epistatic interaction function described in the methods. The negative phenotypic values indicate dark, positive light pigmentation. The alleles are assumed to be codominant, and with the lower case allele corresponding to the dark allele and the upper case alleles the light alleles (with an additive effect of 0.5 for the light alleles). The epistatic interaction constant is taken as 0, for the independent or additive case, 1 for the positively interacting, and -1 for the negatively interacting loci.

Table S1 Sequencing Depths. Sequencing depths of the Austrian, Italian, and South African replicates used for the Pool-GWAS. For the CMH test only the extremely light and dark individuals were used (very light and very dark).

	replicate	very light	unselected population	very dark
Austria 2010	1	85	106	65
	2	97	145	94
	3	100	110	79
Italy 2011	1	37	84	27
	2	126	233	143
	3	126	278	137
SA 2012	1	61	-	73
	2	63	73	93
	3	109	-	106

Table S2: 100 most highly associated SNPs in the **South African** (**SA**) GWAS ranked by their P-values in the CMH test.
For each SNP the following are provided: chromosome (**Chr**), position, the **light**, **dark** and ancestral allele (**Anc**, against *D. simulans*), and the consistency of the direction of the allele frequency changes (**con**), the frequencies of the light allele in the South African (**SA**) and European (**EU**) populations, the negative decadic logarithm of the P-values in the CMH tests (**-log10P**) and the natural logarithm of the average odds ratio of the light allele count in the very light and very dark pools (**log(OR)**), as well as their ranks (**rank P**, **rank OR**) for both populations, the classification of the **Effect** of the SNP, and if applicable the genic location (**Gene**), and the amino acid change (**AA**) against the reference.
Positions are colored according to whether they are close to *tan* (**blue**), *bab* (**green**), or *ebony* (**orange**). Inconsistent allele count changes are labled in **red** in both the **con** and **log(OR)** columns.
The red line indicates the FDR < 0.05 cutoff derived using an empirical null hypothesis.

		Allele				Light allele freq.		-log10P		rank P		log(OR)		rank OR				
Chr	Position	light	dark	con	Anc	EU	SA	EU	SA	EU	SA	EU	SA	EU	SA	Effect	Gene	AA
3L	1084990	G	C	**y**	G	90.3%	51.5%	8.38	42.41	199	1	1.20	3.03	11	5	INTRON	bab1	.
3L	1085137	T	G	**y**	G	84.6%	47.1%	4.35	37.04	1697	2	0.65	2.54	16	7	INTRON	bab1	.
3L	1094154	T	C	**y**	T	80.5%	57.6%	0.28	30.15	4422	3	0.10	2.10	78	19	INTRON	bab1	.
3L	1090136	T	A	**y**	A	59.8%	25.7%	4.47	27.18	1657	4	0.49	2.25	26	12	INTRON	bab1	.
X	11822562	G	A	**y**	A	92.5%	63.0%	0.49	25.71	3989	5	0.24	2.51	49	8	SYNONYMOUS_CODIN	CG10362	D533
3L	1079020	G	A	**y**	A	91.5%	60.0%	2.50	25.70	2476	6	0.75	2.10	15	18	INTRON	bab1	.
3L	1085454	C	T	**y**	T	84.5%	58.1%	23.41	25.65	11	7	1.64	2.00	5	21	INTRON	bab1	.
3L	1080908	C	G	**n**	G	74.0%	41.6%	0.42	24.15	4114	8	-0.12	1.86	69	26	INTRON	bab1	.
3L	1085642	A	T	**y**	-	80.9%	49.2%	0.04	23.94	5130	9	0.02	1.86	96	25	INTRON	bab1	.
3L	1086356	T	A	**y**	A	73.1%	30.0%	4.54	23.81	1626	10	0.57	2.06	17	20	INTRON	bab1	.
3L	1085521	A	T	**n**	T	80.9%	70.5%	3.72	23.66	1941	11	-0.52	1.81	23	29	INTRON	bab1	.
3L	1085530	G	C	**n**	G	76.7%	55.6%	0.82	23.44	3535	12	-0.19	1.87	56	24	INTRON	bab1	.
3L	1080371	C	T	**y**	C	77.5%	55.7%	0.50	23.39	3969	13	0.16	1.83	62	27	INTRON	bab1	.
3L	1080934	T	C	**n**	-	69.0%	38.8%	0.04	22.49	5127	14	-0.02	1.74	97	31	INTRON	bab1	.
3L	1080935	C	A	**n**	C	70.6%	37.5%	0.06	21.68	5033	15	-0.03	1.69	95	36	INTRON	bab1	.
3L	1078685	A	C	**y**	C	87.9%	68.5%	1.87	21.01	2782	16	0.54	1.87	21	23	INTRON	bab1	.
3L	1080915	G	A	**n**	A	72.2%	36.1%	0.40	20.95	4140	17	-0.12	1.73	72	32	INTRON	bab1	.
3L	1085536	A	G	**n**	G	76.4%	53.4%	0.44	20.74	4070	18	-0.12	1.68	68	37	INTRON	bab1	.
3L	1090129	A	G	**y**	G	61.7%	37.5%	3.71	19.55	1946	19	0.45	1.67	27	39	INTRON	bab1	.
3L	1099074	A	C	**n**	C	52.5%	27.6%	0.23	19.54	4542	20	-0.07	1.58	86	47	INTRON	bab1	.
3L	1094564	A	G	**y**	G	74.0%	39.7%	2.33	19.32	2561	21	0.37	1.72	31	33	INTRON	bab1	.
3L	1081063	A	G	**y**	G	72.9%	35.6%	0.23	18.69	4521	22	0.07	1.58	85	49	INTRON	bab1	.
3L	1090646	C	T	**y**	T	52.8%	47.9%	0.02	18.44	5215	23	0.01	1.53	98	54	INTRON	bab1	.
3L	1097325	C	T	**y**	C	82.9%	43.6%	1.26	18.42	3153	24	0.26	1.44	46	66	INTRON	bab1	.
3L	1074402	A	G	**n**	A	63.9%	26.3%	0.51	18.19	3951	25	-0.12	1.58	70	48	INTRON	bab1	.
3L	1116792	C	T	**y**	C	59.1%	38.0%	0.91	18.12	3447	26	0.19	1.71	55	34	INTERGENIC	.	.
3L	1099084	T	G	**n**	G	50.9%	27.3%	0.19	17.98	4624	27	-0.07	1.49	89	59	INTRON	bab1	.
3L	1080180	C	T	**n**	T	92.0%	67.7%	0.27	17.97	4433	28	-0.14	2.17	66	13	INTRON	bab1	.
3L	1076872	G	T	**y**	G	72.8%	57.5%	2.70	17.59	2385	29	0.39	1.45	30	63	INTRON	bab1	.
3L	1080172	T	C	**n**	C	92.7%	75.9%	0.42	17.32	4109	30	-0.20	2.15	53	16	INTRON	bab1	.
3L	1099077	T	C	**n**	T	52.1%	30.3%	0.25	17.09	4474	31	-0.08	1.44	83	68	INTRON	bab1	.
3L	1094165	A	C	**y**	C	80.6%	45.8%	0.27	16.89	4429	32	0.10	1.53	77	55	INTRON	bab1	.
3L	1098471	C	A	**y**	A	58.7%	46.5%	2.33	16.79	2558	33	0.33	1.51	36	56	INTRON	bab1	.
3L	1080174	T	A	**n**	T	92.4%	75.0%	0.43	16.75	4094	34	-0.20	2.17	52	15	INTRON	bab1	.
X	8888865	G	A	**y**	A	9.0%	0.0%	17.16	16.64	24	35	2.45	3.81	1	2	INTRON	rdgA	.
3L	1104150	C	A	**y**	-	86.1%	35.0%	0.55	16.56	3885	36	0.27	2.17	43	14	INTERGENIC	.	.
3L	1090270	A	G	**y**	G	68.1%	26.5%	4.89	16.54	1476	37	0.56	1.58	18	46	INTRON	bab1	.
3L	1104190	C	T	**y**	T	58.9%	31.4%	11.32	16.42	64	38	0.94	1.92	14	22	INTERGENIC	.	.
3L	1074289	C	T	**y**	T	65.0%	52.0%	1.23	16.08	3169	39	0.22	1.38	50	76	INTRON	bab1	.
3L	1104145	G	T	**y**	-	87.8%	37.5%	0.70	15.95	3670	40	0.32	2.11	38	17	INTERGENIC	.	.
3L	1090064	C	T	**y**	T	49.9%	20.3%	5.45	15.76	1282	41	0.53	1.69	22	35	INTRON	bab1	.
3L	1098118	A	G	**y**	G	55.8%	29.8%	5.25	15.72	1356	42	0.51	1.64	25	43	INTRON	bab1	.
3L	1127547	A	C	**y**	C	46.4%	9.0%	1.70	15.25	2855	43	0.26	1.67	45	40	INTERGENIC	.	.
2L	841977	T	A	**n**	-	40.3%	42.9%	0.89	14.89	3466	44	-0.18	1.41	58	71	INTRON	drongo	.
3L	1081068	G	C	**y**	G	95.9%	85.7%	0.67	14.83	3707	45	0.33	1.76	37	30	INTRON	bab1	.
3L	1117678	A	C	**y**	C	92.5%	75.0%	6.14	14.61	1033	46	1.21	1.45	10	64	INTERGENIC	.	.
3L	1098049	G	A	**y**	G	89.3%	78.6%	0.83	14.58	3518	47	0.25	1.82	47	28	INTRON	bab1	.
2L	5922952	A	G	**y**	G	62.4%	12.3%	0.77	14.49	3595	48	0.21	2.63	51	6	UTR_3_PRIME	CG9016	.
3L	1091744	A	G	**y**	-	25.0%	7.8%	1.82	14.49	2805	49	0.40	3.06	29	4	INTRON	bab1	.
3L	1078686	A	C	**y**	A	86.0%	74.1%	1.21	14.42	3189	50	0.37	1.59	34	45	INTRON	bab1	.
3L	1085230	T	G	**y**	T	97.1%	80.4%	2.32	14.26	2570	51	1.00	2.28	12	10	INTRON	bab1	.
3L	1094738	T	C	**y**	C	57.0%	30.9%	1.76	14.15	2837	52	0.29	1.66	42	41	INTRON	bab1	.
3L	1145040	A	C	**n**	C	70.1%	34.0%	0.17	13.97	4676	53	-0.07	1.51	88	57	INTRON	bab2	.
3L	1094002	G	C	**y**	C	52.8%	23.9%	1.82	13.96	2802	54	0.29	1.48	41	60	INTRON	bab1	.
3L	1098940	G	A	**y**	A	50.9%	46.5%	0.31	13.88	4345	55	0.10	1.31	80	86	INTRON	bab1	.
2L	1394460	A	C	**y**	C	58.2%	61.4%	0.14	13.78	4753	56	0.05	1.37	91	77	INTRON	lea	.
3L	1106224	C	T	**y**	T	43.1%	34.4%	0.14	13.75	4756	57	0.04	1.45	93	65	INTERGENIC	.	.
3L	1089594	G	T	**y**	G	70.9%	60.8%	3.34	13.70	2097	58	0.42	1.43	28	69	INTRON	bab1	.
3L	1113168	A	T	**y**	T	22.9%	25.6%	0.80	13.65	3568	59	0.20	1.44	54	67	INTERGENIC	.	.
X	9121129	T	C	**y**	C	18.0%	14.3%	60.35	13.50	2	60	2.86	1.66	0	42	DOWNSTREAM	CG12121	.
2R	8144363	G	A	**n**	A	20.5%	6.9%	6.45	13.22	782	61	-1.44	4.20	7	0	DOWNSTREAM	CG13168	.
3R	12464720	A	G	**y**	G	59.7%	34.2%	0.63	13.21	3774	62	0.14	1.22	67	94	UTR_3_PRIME	Fas1	.
3L	1077967	A	G	**y**	G	87.4%	76.2%	0.62	13.21	3798	63	0.25	1.39	48	74	INTRON	bab1	.
3L	1095066	A	G	**n**	G	59.8%	30.0%	0.54	13.14	3905	64	-0.14	1.57	65	50	INTRON	bab1	.
3L	1070168	C	G	**y**	G	56.4%	19.2%	0.55	13.13	3886	65	0.15	1.40	64	72	INTRON	bab1	.
2L	1394457	G	A	**y**	G	57.9%	63.2%	0.32	13.03	4333	66	0.10	1.33	81	83	INTRON	lea	.
X	8888859	G	A	**y**	A	7.1%	1.6%	13.98	13.02	34	67	2.33	3.50	2	3	INTRON	rdgA	.
3L	1086181	G	A	**y**	A	77.9%	34.5%	11.16	13.02	67	68	0.96	1.56	13	51	INTRON	bab1	.
3L	1142850	G	A	**n**	A	90.5%	61.2%	0.22	12.92	4561	69	-0.17	1.21	59	96	SYNONYMOUS_CODIN	bab2	N773
3L	1090846	C	T	**y**	C	94.4%	68.3%	7.57	12.86	332	70	1.36	1.53	8	53	INTRON	bab1	.
3L	1127551	A	G	**y**	G	50.8%	10.3%	0.85	12.80	3508	71	0.17	1.46	61	61	INTERGENIC	.	.

FDR < 5%

2L	3889057	C	T	y	T	57.8%	53.1%	0.24	12.80	4513	72	0.07	1.21	87	95	INTRON	capu	.
3L	1073855	C	T	n	C	71.7%	30.8%	0.22	12.72	4551	73	-0.08	1.36	84	79	INTRON	bab1	.
3L	1097430	A	G	y	A	97.2%	89.5%	2.61	12.71	2430	74	1.71	2.43	4	9	INTRON	bab1	.
3L	1074001	A	G	n	G	67.8%	64.4%	1.98	12.62	2739	75	-0.34	1.24	35	92	INTRON	bab1	.
3L	1087870	G	T	y	T	41.4%	30.1%	2.37	12.61	2535	76	0.37	1.68	33	38	INTRON	bab1	.
X	5986805	C	T	y	C	14.2%	44.0%	1.40	12.55	3056	77	0.37	1.45	32	62	INTRON	Nep1	.
3L	1078558	A	G	y	G	69.9%	44.4%	1.38	12.53	3077	78	0.26	1.33	44	82	INTRON	bab1	.
3L	1136992	T	A	y	A	66.8%	43.8%	4.51	12.52	1639	79	0.51	1.25	24	90	DOWNSTREAM	bab2	.
2R	14839049	T	C	n	T	2.3%	8.4%	0.01	12.49	5296	80	-0.09	2.27	82	11	INTRON	sano	.
3L	1090847	T	C	y	T	94.4%	67.2%	7.26	12.39	405	81	1.33	1.50	9	58	INTRON	bab1	.
2R	8144365	T	C	n	C	19.7%	6.9%	7.13	12.36	438	82	-1.60	4.17	6	1	DOWNSTREAM	CG13168	.
3L	1041545	T	A	y	A	54.1%	28.1%	0.45	12.36	4064	83	0.11	1.28	74	89	INTRON	bab1	.
2L	1394453	C	G	y	C	57.1%	56.2%	0.41	12.32	4128	84	0.11	1.24	73	93	INTRON	lea	.
X	12801479	G	A	n	G	43.0%	42.1%	0.11	12.07	4845	85	-0.05	1.28	92	88	SYNONYMOUS_CODIN	mRpL49	L65
3L	1094465	G	A	y	A	96.5%	76.3%	0.23	12.07	4535	86	0.17	1.54	60	52	INTRON	bab1	.
X	9121191	T	C	y	T	20.6%	11.3%	35.32	12.00	6	87	1.81	1.64	3	44	UTR_5_PRIME	Gr8a	.
X	9084476	C	T	y	C	3.9%	23.2%	1.22	11.90	3174	88	0.56	1.34	19	81	INTRON	CG12115	.
3L	1075866	A	T	n	T	19.2%	17.7%	0.39	11.90	4172	89	-0.12	1.43	71	70	INTRON	bab1	.
X	15791922	C	T	y	C	78.8%	65.2%	0.14	11.89	4752	90	0.06	1.36	90	78	INTRON	shi	.
3L	6577042	G	T	y	T	14.0%	22.1%	0.02	11.73	5259	91	0.01	1.34	99	80	INTRON	CG18769	.
2L	4316029	T	C	y	C	31.5%	34.0%	0.39	11.65	4175	92	0.10	1.20	76	98	INTRON	tutl	.
3L	1118293	A	T	y	A	63.1%	43.2%	1.83	11.65	2800	93	0.29	1.39	40	75	INTERGENIC	.	.
3L	1073815	C	T	y	T	73.2%	45.5%	0.06	11.57	5054	94	0.03	1.31	94	85	INTRON	bab1	.
X	5883799	A	G	y	G	77.5%	58.5%	0.64	11.51	3753	95	0.19	1.28	57	87	INTRON	CG5966	.
3L	1097354	A	G	n	G	95.1%	78.7%	0.50	11.51	3971	96	-0.32	1.25	39	91	INTRON	bab1	.
X	9125067	G	A	y	G	36.3%	41.8%	4.42	11.46	1675	97	0.54	1.18	20	99	NON_SYNONYMOUS_C	CG12121	L298P
3L	1075867	T	C	n	C	19.2%	18.3%	0.30	11.46	4357	98	-0.10	1.40	79	73	INTRON	bab1	.
2L	8393767	G	C	n	G	39.5%	22.2%	0.37	11.40	4203	99	-0.10	1.33	75	84	NON_SYNONYMOUS_C	lectin-29Ca	K168N
3L	1085552	G	A	y	A	88.9%	66.1%	0.40	11.40	4142	100	0.15	1.21	63	97	INTRON	bab1	.

Table S3. 100 most highly associated SNPs in the European (EU) GWAS. (.xlsx, 87 KB)

Available for download as a .xlsx file at
www.genetics.org/lookup/suppl/doi:10.1534/genetics.115.183376/-/DC1/TableS3.xlsx

Table S4. 50 most highly associated SNPs around *tan* in the South African (SA) GWAS.
(.xlsx, 75 KB)

Available for download as a .xlsx file at
www.genetics.org/lookup/suppl/doi:10.1534/genetics.115.183376/-/DC1/TableS4.xlsx

Table S5. 50 most highly associated SNPs around *bab* in the South African (SA) GWAS.
(.xlsx, 73 KB)

Available for download as a .xlsx file at
www.genetics.org/lookup/suppl/doi:10.1534/genetics.115.183376/-/DC1/TableS5.xlsx

Table S6. 50 most highly associated SNPs around *tan* in the European (EU) GWAS. (.xlsx, 78 KB)

Available for download as a .xlsx file at
www.genetics.org/lookup/suppl/doi:10.1534/genetics.115.183376/-/DC1/TableS6.xlsx

Table S7. 50 most highly associated SNPs around *bab* in the European (EU) GWAS. (.xlsx, 188 KB)

Available for download as a .xlsx file at
www.genetics.org/lookup/suppl/doi:10.1534/genetics.115.183376/-/DC1/TableS7.xlsx

Table S8: Overlap with Bickel *et al.* 2011 SNPs significantly associated with female abdominal pigmentation (Tergite 6) in Bickel *et al.* 2011 and significantly associated after Bonferroni correction (significance level 0.05) for multiple testing in either the European (P<=1.56E-8) or the South African (P<=1.25E-8) GWAS. The colors of the **-logP** column indicate whether the SNP was found significantly associated in the respective population (EU: European, SA: South African) after Bonferroni correction (**blue**) or using an empirical null distribution for FDR correction (**red**). Positions were matched by aligning the sequenced fragment provided by Dr. Bickel to the reference 3L chromosome (v. 5.18) using progressive Mauve and SNPs checked for consistency of alleles. The last column gives the position of the SNPs on the DNA fragment of the 3L chromosome sequenced in Bickel *et. al* 2011, with the region defined in Bickel *et al.* indicated by colors (white: first intron of bab1, **blue**: region 1 around abdominal CRE in the first intron of bab1, **yellow**: highly linked region in the intergenic sequence between bab1 and bab2). Out of 3066 uniquely mappable SNPs in Bickel *et al.* with a minor allele frequency >= 0.05, 2367 were segregating in at least one of the populations used in this study, with 167 of these being significantly associated with abdominal pigmentation according to Bickel *et al.* (2011).

Chr	Position	Allele	-logP			Position in Bickel *et al.*
		light	dark	EU	SA	
3L	1041545	T	A	0.4	**12.4**	12798
3L	1077795	T	G	0.8	**8.5**	41670
3L	1078420	C	T	0.8	**8.7**	42307
3L	1080569	A	G	0.3	**9.7**	44482
3L	1080600	G	A	1.0	**9.9**	44517
3L	1080908	C	G	0.4	**24.1**	44836
3L	1080915	G	A	0.4	**21.0**	44843
3L	1080934	T	C	0.0	**22.5**	44862
3L	1080935	C	A	0.1	**21.7**	44863
3L	1081068	G	C	0.7	**14.8**	44996
3L	1084990	G	C	**8.4**	**42.4**	48973
3L	1085454	C	T	**23.4**	**25.6**	49437
3L	1085642	A	T	0.0	**23.9**	49625
3L	1135568	A	T	0.5	**9.6**	100457
3L	1139615	G	A	1.0	**8.6**	104557

Table S9. Overlap with Dembeck *et al.* (.xlsx, 62 KB)

Available for download as a .xlsx file at
www.genetics.org/lookup/suppl/doi:10.1534/genetics.115.183376/-/DC1/TableS9.xlsx

Supplementary Information

Supplementary Figure Legends

Figure S1. Ovariole number is positively correlated with early female fecundity. Number of eggs laid was counted in the first three days after eclosion (diamond: 1st day after eclosion; square: 2nd day after eclosion; circle: 3rd day after eclosion) from outbred females fed on standard food as larvae (yellow symbols) and outbred females fed on 20% sucrose food as larvae at timed intervals starting between 5 h to 25 h AL3E (symbols with different shades of blue) until the end of the feeding period. Plotted values represent means and error bars show 95% confidence intervals of means.

Figure S2. Ovary development during L3 larval stages under optimal nutritional conditions. (A) Schematic drawings representing ovary development in L3 larvae reared in standard food. Terminal filaments (TFs) are represented as dark grey symbols. Axis are presented as A-P, anterior-posterior; D-V, dorsal-ventral; M-L, medial-lateral. Pictures show developing ovaries from outbred larvae during L3 larval stages under standard food. Engrailed (grey) marks terminal filament cells (TFCs). Scale bar: 20µm. (B) Number of forming terminal filaments (TFs). (C) Ovary volume. Plotted values represent means and error bars show 95% confidence intervals of means. L3: third instar larvae; AL3E: after L3 ecdysis.

Figure S3. Similar results were obtained when larvae were fed on either 1% or 20% sucrose food. (A-D) shows terminal filaments (TFs) marked with En immunostaining. Ovaries from outbred larvae reared on 1% sucrose food between: (A) 5-29 h AL3E or (C) 15-39 h AL3E. Ovaries from larvae reared on 20% sucrose food between: (B) 5-29 h AL3E or (D) 15-39 h AL3E. Scale bar: 20µm. (E) Number of forming terminal filaments (TFs) and (F) ovary volume of ovaries from larvae fed on 1% (triangles) or 20% (points) sucrose food; larvae were transferred to 20% sucrose food either at 5 h AL3E (light blue circles) or at 15 h AL3E (dark blue points). Error bars show 95% confidence intervals of means. Wilcoxon rank test: *$p<0.05$, ***$p<0.01$, ***$p<0.001$. L3: third instar larvae; AL3E: after L3 ecdysis.

Figure S4. *Traffic jam*-GAL4 is expressed in ovarian somatic cells during L3 larval stages. (A, B, C, D and A'', B'', C'', D'') Phalloidin marks F-actin to outline cell membranes (grey, red). (A', B', C', D' and A'', B'', C'', D'') GFP reporter line under the control of *traffic jam*-GAL4 driver line (grey, green). Scale bar: 20µm. In (C, D), white arrowheads denote forming terminal filaments. In (A', B' and D'), asterisks denote germ cells. L3: third instar larvae; AL3E: after L3 ecdysis.

Figure S5. *Traffic jam*-GAL4 expression patterns in the larval brain during larval stages. (A, B, C, D and A'', B'', C'', D'') Phalloidin marks F-actin to outline cell membranes (grey, red). (A', B' and A'', B'') GFP reporter line under the control of *traffic jam*-GAL4 driver line (grey, green). (C', D' and C'', D'') GFP reporter line in the *elav*GAL80, *traffic jam*-GAL4 driver line (grey, green). Scale bar: 20µm. L2: second instar larvae. L3: third instar larvae.

Figure S6. Manipulating IIS or ecdysone signalling in the larval ovary reduces adult ovariole number and female weight. (A) Adult ovariole number, (B) female pharate weight and (C) developmental times represented in hours after third instar ecdysis (h AL3E) to pupariation of individuals with disruption of IIS or ecdysone signalling under the control of *traffic jam*-GAL4 driver line (*tj > PTEN* and *tj > EcR-DN*, respectively; blue bars). (D) Adult ovariole number and (E) pharate weight of females with disruption of IIS specifically in ovarian somatic cells under the control of *traffic jam*-GAL4 (*elav-GAL80, tj > PTEN*) or in neuroblasts and neurons of the larval brain using the *elav*-GAL4 driver (*elav > PTEN*) (blue bars). Controls are either driver (*elav*-GAL80, *tj*-GAL4; *elav*-GAL4 and *tj*-GAL4; black bars) or reporter (UAS-*PTEN* and UAS-*EcR-DN*; grey bars) lines. Error bars show 95% confidence intervals of means. ANOVAs followed by Tukey's HSD test: values not sharing the same letter are significantly different ($p < 0.05$).

Figure S7. Activating IIS and/or ecdysone signalling results in different ovary volumes at 5 h AL3E. Ovaries from larvae reared on standard food: (A) *tj*-GAL4 (control), (B) *tj > EcR-IR*, (C) *tj > InR* and (D) *tj >EcR-IR, InR*. Scale bar: 20µm. (E) Ovary volume of ovaries from *tj*-GAL4 control larvae (black bar), *tj > EcR-IR, tj > InR* and *tj >EcR-IR, InR* larvae (red bars). Larvae were dissected at 5 h AL3E. Kruskall-wallis followed by Wilcoxon rank test: values not sharing the same letter are

significantly different (Holm’s correction $p < 0.05$). L3: third instar larvae; AL3E: after L3 ecdysis.