Population genetic evidence for cold adaptation in European *Drosophila melanogaster*

Vedran Božičević¹, Stephan Hutter¹, Wolfgang Stephan¹, and Andreas Wollstein¹*

¹Section of Evolutionary Biology, Department of Biology II, University of Munich, D-82152 Planegg-Martinsried, Germany

Keywords: local adaptation, polygenic traits

*Correspondence: vedran.bozicevic@gmail.com, andreas.wollstein@gmail.com

Running title:

Cold adaptation in European *D. melanogaster*
Abstract

We studied *Drosophila melanogaster* populations from Europe (the Netherlands and France) and Africa (Rwanda and Zambia) to uncover genetic evidence of adaptation to cold. We present here four lines of evidence for genes involved in cold adaptation from four perspectives: (1) the frequency of SNPs at genes previously known to be associated with chill-coma recovery time (CCRT), startle reflex (SR), and resistance to starvation stress (RSS) vary along environmental gradients and therefore among populations; (2) SNPs of genes that correlate significantly with latitude and altitude in African and European populations overlap with SNPs that correlate with a latitudinal cline from North America; (3) at the genome-wide level, the top candidate genes are enriched in gene ontology (GO) terms that are related to cold tolerance; (4) GO enriched terms from North American clinal genes overlap significantly with those from Africa and Europe. Each SNP was tested in 10 independent runs of *Bayenv2*, using the median Bayes factors to ascertain candidate genes. None of the candidate genes were found close to the breakpoints of cosmopolitan inversions, and only four candidate genes were linked to QTLs related to CCRT. To overcome the limitation that we used only four populations to test correlations with environmental gradients, we performed simulations to estimate the power of our approach for detecting selection. Based on our results we propose a novel network of genes that is involved in cold adaptation.
Introduction

Uncovering the genetic basis of adaptation in natural populations is a major aim of evolutionary biology (Dobzhansky 1970; Lewontin 1974; Stinchcombe & Hoekstra 2008; Barrett & Hoekstra 2011; Wray et al. 2013). Adaptation results in differences in phenotypes that are often distributed gradually along clines of populations adjusted to different environments (Williams 1966; Kawecki & Ebert 2004; Le Corre & Kremer 2012; Savolainen et al. 2013). The study of local adaptation has been recently fuelled by the advancements of sequencing technologies that now allow for population genomics approaches in various organisms to identify individual adaptive single nucleotide polymorphisms (SNPs) (Akey et al. 2002; Kayser et al. 2003; Bonin et al. 2006; Joost et al. 2007; Wood et al. 2008; Hancock et al. 2011; Duforet-Frebourg et al. 2014). We can gain insight into the adaptation of phenotypes by observing frequency changes of functionally relevant alleles in populations fine-tuned to different environments. Most genome-wide scans so far have focused on the detection of selective sweeps, but sweeps apply mostly to monogenic phenotypes or phenotypes for which most loci have large effects (Jain & Stephan 2015). However, such patterns can be expected to be rare in polygenic adaptation (Orr & Betancourt 2001; Orr 2005; Pritchard & Di Rienzo 2010; Pavlidis et al. 2012; Wollstein & Stephan 2014; Jain & Stephan 2015; Stephan 2015). On the other hand, to distinguish small shifts in allele frequencies from genetic drift is quite cumbersome (Chevin & Hospital 2008; Pritchard et al. 2010; Hancock et al. 2010).

Here we investigate cold adaptation by focusing on genetic variation related to known adaptive phenotypes between African and European Drosophila melanogaster. A common approach to detect adaptive footprints has been to use a genome-wide scan of a large number of molecular markers to screen for \( F_{ST} \) outliers (Luikart et al. 2003; Storz 2005; Stinchcombe...
& Hoekstra 2008; Holsinger & Weir 2009; Lotterhos & Whitlock 2014). For example, to account for frequency differences between populations that were caused by drift (Elhaik 2012), local $F_{ST}$ values can be compared to a genomic background distribution (Holsinger & Weir 2009). Other approaches have been described in (Riebler et al. 2008; Foll & Gaggiotti 2008; Coop et al. 2010; Villemereuil et al. 2014; Foll et al. 2014). For multiple populations, assuming a phenotype that differs among them, a powerful approach is to test frequency shifts for correlation with environmental variables (Günther & Coop 2013). In addition to environmental variables, different phenotypic distributions can be used (Mendizabal et al. 2012; Villemereuil et al. 2014). We focused on chill-coma recovery time (CCRT), resistance to starvation stress (RSS), and startle response (SR), phenotypes known to be adaptive and related to cold. CCRT is a metric used to quantify the tolerance to low, but non-lethal temperatures in ectothermic organisms (David et al. 1998; Anderson et al. 2005; Rako & Hoffmann 2006; Macmillan & Sinclair 2011). Most insects, including Drosophila melanogaster, cannot tolerate the formation of ice particles in their cells (e.g. using ice nucleating agents or antifreeze proteins) and have to cope with cold stress to survive in temperate climates (Bale 1993). Chill-coma is a narcosis state induced in D. melanogaster at 0°C as a result of disruption in nerve and muscle excitability (Gibert et al. 2001b; Macmillan & Sinclair 2011). The comatose state is a consequence of the loss of ion and water balance due to low temperatures, but since it is largely reversible, it has become an important mechanism for winter survival (David et al. 1998; Macmillan & Sinclair 2011). RSS has also been used to investigate susceptibility of D. melanogaster to environmental fluctuations because the ability to survive food shortages has direct implications on fitness (Hoffmann 2010; Goenaga et al. 2010, 2012, 2013). Moreover, RSS is related to cold adaptation by conferring a fitness cost to cold tolerance (Hoffmann et al. 2005; Ayroles et al. 2009). SR is a vigorous reaction activity in D. melanogaster induced by a mechanical disturbance (Meehan
that is measured as the time of total activity within a window of time (e.g. 45 seconds) that follows immediately to a mechanical handling disturbance (Ayroles et al. 2009). Locomotory behavioral traits are generally related to finding food and mates, escape from predators, and defense of territory. Hence, locomotory behaviors are extremely important components of overall fitness (Gilchrist et al. 1997; Jordan et al. 2006).

Our aim is to accumulate evidence for annotated genes that may have been under local adaptation in the aforementioned phenotypes. In all analyses we investigate genetic variation by (1) estimating pairwise $F_{ST}$ between African and European populations of $D. melanogaster$, and by (2) quantifying correlations with environmental variables by means of Bayenv2 (Coop et al. 2010; Günther & Coop 2013). We first focus specifically on SNPs from genome-wide association (GWA) studies of CCRT, RSS, and SR by the $Drosophila melanogaster$ Genetic Reference Panel (DGRP) (Mackay et al. 2012; Huang et al. 2014).

Materials and Methods

Data analyzed

Genome-wide SNP data was generated for two African and two European populations. Data for Gikongoro, Rwanda (RG, 27 lines), Siavonga, Zambia (ZI, 27 lines) and Lyon, France (FR, 8 lines) were taken from the Drosophila Genome Nexus (http://johnpool.net/genomes.html; Lack et al. 2015). For these populations female flies were collected between 2008 and 2010 and iso-female lines were established. Whole genome sequences for individual lines were created by next-generation sequencing of haploid embryos (Langley et al. 2011) and assembling of sequencing reads into full genomes by mapping them to the $D. melanogaster$ reference sequence (Lack et al. 2015). Some of the African lines were found to harbor stretches of European admixture (Pool et al. 2012; Lack et
These genomic regions were masked in our analysis. Additionally, we analyzed a population from Leiden, the Netherlands (NL, 11 lines), for which females were collected in 1999 (Bubliy & Loeschcke 2000) and kept in the lab as iso-females lines. These fly lines were subjected to fifteen generations of full-sib mating and whole genome sequences were generated by next-generation sequencing of adult flies and genome assembly as detailed above (Voigt et al. 2015). Over all populations and chromosomes, we observed of 9,995,420 polymorphic sites. Genotypes with a PHRED score of less than 31 (=0.08% chance of calling a base incorrectly) were considered as missing data (Ewing & Green 1998). We excluded sites that were segregating for more than two alleles, and sites that contained more than 10% missing data across the populations considered. For Bayenv2 analyses, we further excluded sites that were not polymorphic across the tested populations, as well as singletons (over all populations). After quality filtering, we finally maintained 3,663,890 polymorphic sites for the autosomes and 867,049 for the X chromosome for pairwise $F_{ST}$ analyses, and 313,972 polymorphic sites for the autosomes and 39,304 for the X chromosome for Bayenv2 analyses (Table 1).

From the filtered genomic background, we defined subsets of SNPs (Table 1) that were previously associated with the following quantitative traits: (i) time to recover from chill coma (CCRT), (ii) resistance to starvation stress (RSS), and (iii) startle-induced locomotor response (SR) (Huang et al. 2014). We defined these SNPs, which were originally ascertained from a North American population, by using their corresponding positions in our European and African populations of interest. We used the annotation data as recorded in Flybase release 5 (Pierre et al. 2014).

Statistical analyses
We estimated the amount of population differentiation between pairs of populations on SNPs across the genome using the Weir and Cockerham estimator of $F_{ST}$ (Weir & Cockerham 1984), which is known to be unbiased with respect to sample size (Willing et al. 2012). It is further known that the distribution of $F_{ST}$ can be skewed, for example by sub-selection of SNPs from GWA studies (Clark et al. 2005; Elhaik 2012). Demographic events, such as population bottlenecks, affect the allele frequencies patterns that are also reflected in $F_{ST}$ estimates. Therefore, to derive unconfounded $P$-values from the observed $F_{ST}$ distribution, we decided to use a resampling approach to assess the null distribution (i.e. genomic background). For every trait-associated set of SNPs (3 traits × 4 population pairs = 12 SNP sets), we calculated the mean $F_{ST}$ over each of 12 × 10000 SNP sets of random background SNPs. The quantile of mean $F_{ST}$ of trait-associated SNPs in the empirical distribution of 10000 equally sized sets represented its empirical $P$-value that needs no further adjustment for false discovery rate (e.g. Noble 2009). Testing for different phenotypes in different pairs of populations represents independent statistical trials.

To identify adaptive loci as a response to known environmental changes, we used Bayenv (Coop et al. 2010; Günther & Coop 2013). This method takes into account covariance of allele frequencies across tested populations, which arises as a consequence of demographic history and spatial expansion. Bayenv2 performs well for accurate identification of loci under spatially varying selection (Villemereuil et al. 2014; Lotterhos & Whitlock 2014). In Bayenv2, a covariance matrix between all pairs of populations from putatively neutral sites is used as a null model to test for relationships of the population frequencies of a given site to an environmental variable and the SNP-specific allele distribution. We estimated the covariance matrix from 5000 randomly sampled SNPs that are in linkage equilibrium and used the mean from ten matrices as the null model. To test the convergence of Bayenv2, we used several independent Markov Chain Monte Carlo runs with a maximum chain length of
10000 iterations. We observed convergence after about 5000 iterations (Figure S1). However, these chains might converge to different solutions. To be most stringent, we used the median results from 10 independent runs (Blair et al. 2014). We then tested for correlations between each single SNP and six environmental variables: geographical latitude, height above mean sea level, and four temperature measures (average daily minimum of the coldest and warmest month, and average daily minimum and maximum throughout the year) (Table 2). The results for environmental variables are given as Bayes factors (BFs). A higher BF gives higher support to the model where the environmental variable has a significant effect on allele frequency distribution over an alternative model with no effect (Coop et al. 2010). Similar as above, we finally report the median BF out of ten independent runs of each SNP that has been described to improve the proportion of false positives (Blair et al. 2014; Lotterhos & Whitlock 2014). With BF values, we also used a resampling approach, analogously to the one applied on $F_{ST}$. We randomly sampled sets of SNPs of the same size from the genomic background and then assessed the null distribution for BF of CCRT-, RSS-, and SR-associated SNPs.

To get an estimate of the extent to which random genetic drift has shaped the SNP allele frequencies, we estimated the parameters of a likely demographic model (Figure S2) by means of Approximate Bayesian Computation (Beaumont 2010; Bertorelle et al. 2010; Csilléry et al. 2010; Laurent et al. 2011; Duchen et al. 2013). Coalescent simulations were performed with Hudson’s ms (Hudson 2002). The correlation of the simulated and the observed summary statistics were high for both autosomes ($R^2 > 0.96$) and the X chromosome ($R^2 > 0.97$) (Table S1). Note that our estimates might not represent the true demographic history very well. More simulations and proper estimates would be required (Beaumont 2010). However, the combination of parameters produced by our simulations sufficiently represents the effects of genetic drift on $F_{ST}$ and Bayenv2 values due to the demographic
history. The empirical $P$-values generated from coalescent simulations were then assessed for all genome-wide calculated $F_{ST}$ and Bayenv2 values by the following formula: $P$-val = (number simulated values > observed value) / (number of simulations). Note that these single SNP empirical $P$-values are distinct from empirical $P$-values obtained over a set of SNPs using the previously described resampling method. These $P$-values were additionally adjusted for false discovery rate according to Benjamini and Hochberg (1995).

Using our demographic estimations, we further assessed the power in disentangling adaptive from neutrally evolving SNPs in our framework of four populations. For this analysis, we focused on the example of the CCRT trait, comprising 90 associated SNPs with information about phenotypic effects available as average difference of major and minor alleles (Huang et al. 2014). Effect sizes were standardized in units of standard deviations of the CCRT distribution over male and female lines. From the set of associated SNPs, we further excluded sites with the lower effects from pairs in high LD ($R^2 > 0.8$) resulting in 74 SNPs. We used the forward equations of allele frequencies at independent sites that are given in de Vladar and Barton (2014) (Equation 6). The evolution of a number of independent polymorphic sites that contribute to a polygenic trait is defined forward in time dependent on the following parameters: (i) the selection coefficient ($S = 0.1$), which was chosen to be reasonably high as expected for a quantitative trait (de Vladar & Barton 2014), (ii) the population optimum ($z_0$), as expected from the mean latitude, and (iii) the initial frequency of the polymorphic sites as provided from our neutral simulations. Additionally, we used the same mutation rate ($\mu=1.13e-09$) and estimated population sizes from our neutral simulations ($N_{FR}= 1.8263e+05$, $N_{NL}=9.1975e+05$, $N_{RG}=1.2728e+06$, $N_{ZI}=1.3258e+06$). We then simulated the effect of genetic drift after each generation in the forward equation via multinomial resampling for all independent sites. We let each simulation run independently for 200 generations with and without selection in all four populations. Most trajectories
converged to equilibrium much earlier than 200 generations. We generated 100 replicates. From the resulting set, we applied Bayenv2 as described above to the neutrally as well as to the adaptively evolving sites. We assessed the power by describing the proportion of correctly identified adaptive sites with a BF above the threshold obtained from the distribution of neutrally evolving sites (see Figure S3). From the simulations of the putative adaptive history of CCRT, we obtained a power of about 43% at a (neutral) acceptance threshold of 0.05. Thus, we believe that we are able to achieve sufficient power to detect selection on variants with intermediate to large effects that change rapidly in frequency due to environmental fluctuations (Pritchard et al. 2010; de Vladar & Barton 2014).

For gene and SNP annotation, we used the perl script from Ensembl’s Variant Effect Predictor (VEP) tool (McLaren et al. 2010). For synonymous SNPs we determined possible codon usage bias using data from the Codon Usage Database (www.kazusa.or.jp/codon) that was originally compiled from NCBI-GenBank (www.ncbi.nlm.nih.gov/genbank). For associated SNPs, we checked for overlap of their respective genes with temperature tolerance genes from literature. Independently, we reported overlap of genes and high $F_{ST}$ SNPs for CCRT, RSS, and SR.

Ontology and pathway analyses were performed using the ClueGO (http://apps.cytoscape.org/apps/cluego) and CluePedia (Bindea et al. 2009, 2013) plugins (http://apps.cytoscape.org/apps/cluepedia) of Cytoscape version 2.1.6 (Shannon et al. 2003). We used Cohen’s Kappa score (Cohen 1968) of 0.7 as a threshold for the proportion of genes shared between enriched ontology and pathway terms to link the terms into GO networks (Bindea et al. 2009) and networks of KEGG (Kanehisa & Goto 2000) and Reactome (Croft et al. 2011) metabolic pathways. With ClueGO and CluePedia we integrated enriched terms into networks. Enrichment and depletion of single terms were calculated using a two-sided hypergeometric test. We applied the FDR correction (Benjamini & Hochberg 1995) and
retained the terms enriched with a FDR-corrected $P$-value of less than 0.01 that contained at least three candidate genes, or when the candidate genes represent at least 4% of the total number of genes related to the term.

Results

Genetic differentiation of trait-associated SNPs

We first quantified the amount of differentiation for SNPs associated with phenotypic traits that are known to differ between temperate and tropical regions (Da Lage et al. 1990; Gibert et al. 2001a; b; Hoffmann et al. 2001; Kennington et al. 2001; De et al. 2013). As a measure of differentiation, we estimated pairwise $F_{ST}$ of SNP sets that were previously associated with CCRT, RSS, and SR (Huang et al. 2012, 2014) in four pairs of populations (see Table 1 for SNP numbers, Table 3 for results). A $P$-value has been obtained from a resampling approach as described in Materials and Methods. We find a significant enrichment of elevated $F_{ST}$ values (see Table 3) comparing the Netherlands (NL) and Rwanda (RG) samples (mean target $F_{ST} = 0.133$; $P_{\text{empirical}} = 0.031$) in the CCRT dataset (Huang et al. 2014), and NL and RG (mean target $F_{ST} = 0.134$; $P_{\text{empirical}} = 0.0259$) in the RSS dataset (Huang et al. 2014). The $F_{ST}$ differences between NL and RG were not mirrored in the FR vs. ZI for RSS. However, for CCRT nearly all pairs of populations show a significant ($P_{\text{empirical}} < 0.05$), or nearly significant ($P_{\text{empirical}} < 0.08$) enrichment of $F_{ST}$ values. This may suggest that the differentiation between Africa and Europe at trait-associated SNPs cannot simply be explained by their demographic history and that adaptive forces need to be invoked. At the same time, the marginal significance displayed in some pairs of populations for CCRT, and no significant enrichment of $F_{ST}$ for SR (Table 3) requires more evidence than solely $F_{ST}$ before considering the possibility of selection. This evidence is presented in the following sections.
Environmental correlation with trait-associated SNPs

Next we used BFs instead of $F_{ST}$ and repeated the resampling approach, to test for an enrichment of high BFs in the associated SNPs. We tested fewer SNPs using Bayenv2 compared to $F_{ST}$, because fewer SNPs were polymorphic across all the populations tested. Nevertheless, the Bayenv2 results mostly follow qualitatively the results of the empirical $F_{ST}$ observations in that CCRT-associated SNPs were the most likely among the three traits to have BFs higher than the genomic background (Table 4). In the case of CCRT, we detected enrichments for altitude ($P_{\text{empirical}} = 0.015$) and for coldest month minimum ($P_{\text{empirical}} = 0.001$). For correlations with yearly minimum temperature, the empirical $P$-value of BFs associated with CCRT was still marginally significant ($P_{\text{empirical}} = 0.066$), while latitude and both yearly maximum temperature and the hottest month minimum showed no significant correlation. Additionally, the SNPs associated with RSS and SR generally do not show a significantly higher correlation with environmental variables (higher BFs) than the genomic background (Table 4).

The magnitude of BFs conveys information on the likelihood of a site being under selection (e.g. a BF > 1 means that selection is more likely than neutrality). BFs are a much more stringent measure of the likelihood of selection than pairwise $F_{ST}$. For example, when we imposed a cutoff of $\ln(\text{BF}) > 1$ (positive evidence as suggested by (Kass & Raftery 1995) or $P < 0.0063$ from our simulations), only one CCRT-associated SNP (chr2R_18586714) was still significantly correlated with environmental variables ($P < 0.0039$). For instance, chr3L_6723212 was not significant with this cutoff (BF between 0.46 for env2 and 2.46 for
env5), even though its $F_{ST}$ ranges from 0.44 (FR-RG) to 0.73 (NL-ZI). Note that a cutoff of $\ln(BF) > 1$ corresponds to $P < 0.0063$ according to our neutral simulations.

Many genes related to cold tolerance are enriched for SNPs with high BF and $F_{ST}$ values

We retrieved a list of genes that are known to be related to cold or heat tolerance from the literature (see Table S2, Table S4 for description). To mitigate possible false positives, we aimed to include in this list candidates from a range of studies that have employed a variety of different techniques, including QTL mapping, physiology, gene knockdowns, $P$-element insertions, RNA interference, mutant complementation tests, and various gene expression approaches (see Table S2 for references for each gene and the techniques used in each study).

We quantified the number of SNPs with significant $F_{ST}$ outlier (Table S4) and BF outlier (Table S4 and S5) values within these genes. We found that 17 genes ($DnaJ-1$, $AnxB9$, $Lsp1beta$, $CG16700$, $psq$, $stan$, $lola$, $Oatp30B$, $E(spl)m7$-$HLH$, $cpo$, $whd$, $CG12054$, $Dyrk2$, $shep$, $chas$, $Ire1$, and $Octbeta3R$) contained SNPs with evidence in favor of selection ($\ln(BF) > 1$, (Kass & Raftery 1995) or $P < 0.0063$) for at least one environmental variable. Even more strikingly, 13 of these genes contained SNPs with strong evidence ($\ln(BF) > 3$, (Kass & Raftery 1995) or $P < 0.0043$), which means that a model including selection is about 20 times more likely than neutrality at multiple loci within these genes.

We used the VEP tool to retrieve functional annotations of the top 1% SNPs from the genome and for each gene we reported numbers of intron variants, 3’ and 5’UTR variants, and synonymous and nonsynonymous coding variants. For nonsynonymous SNPs we also reported whether the alternate amino acid had a differently charged side chain, as this may lead to differences in the folding of the final protein product. We observed a general trend
such that most SNPs with elevated $F_{ST}$ values are located in introns, followed by SNPs in untranslated (3’ and 5’) regions. This is consistent with previously observed patterns of selective constraints in noncoding DNA (Halligan et al. 2004; Halligan & Keightley 2006).

Many genes show frequency changes suggestive of selection in multiple classes. We found 20 genes showing synonymous changes that might impact gene expression through codon usage bias. Overall, we found that from the 46 genes retrieved from literature (Table S2), 33 contained SNPs from the top 1% of the $F_{ST}$ distribution, for at least one pairwise population comparison (Table S4). We observed that in the heat-shock gene Hsp26 there were two variants (chr3L_5743995 and chr3L_5743998) that code for codons with negative codon usage bias, which were both more common in Europe (with frequency 90.9% or more) than in Africa (frequency 52.4% or less). This suggests that they might be under less selective constraint, or under positive selection in African populations (Table S4). Notably, 6 genes (CG31738, CG12943, CG30379, lola, nclb, and chas) contained nonsynonymous variants with very high $F_{ST}$, indicating selection for a different amino acid. This possibility is most apparent in CG30379 and lola because the associated amino acid changes also change the charge of their respective side chains, which could have an even greater influence on the folding of the protein. The gene lola is particularly interesting, because it contains high $F_{ST}$ variants along its entire length, including coding and noncoding, as well as regulatory regions.

**Inversion analysis**

Major cosmopolitan inversions are known to have an effect on clinal variation (Hoffmann & Weeks 2007; Hoffmann & Rieseberg 2008). In order to exclude the possibility that the differentiation at our candidate genes was due to the effect of inversions, we compared the genomic coordinates of the breakpoints of the four major cosmopolitan inversions, $In(2L)t$,
In(2R)NS, In(3L)P, and In(3R)P (Ashburner & Lemeunier 1976), with the coordinates of our candidate genes, i.e. genes with ln(BF)>5, as reported by Bayenv (P < 0.0034) for correlation with latitude and altitude. We obtained the cytogenetic-absolute coordinate mapping from http://flybase.org/static_pages/downloads/FB2014_06/map_conversion/genome-cyto-seq.txt.gz. We also checked for overlap with the candidate genes retrieved from literature. We found no genes in close proximity to any inversion breakpoint; for example, the candidate gene closest to any inversion was sty, a latitudinal candidate gene located approximately 115 kb upstream of a In(2L)P breakpoint (see inversion_analysis.xlsx on Dryad). Of the genes from the literature, Hsp83 comes closest (about 32 kb upstream of In(2L)P). It is therefore unlikely that these inversions would have a noticeable effect on our inferences.

**Cllinal genes in Europe overlap with clinal genes in North America**

We mapped the SNPs with extremely high BFs (ln(BF)>5, as reported by Bayenv2; P < 0.0448 from neutral simulations) for correlation with latitude and altitude to genes, and then reported the overlaps with genes that were detected in a North American cline (Fabian et al. 2012) (see Figure 1, Figures S5-S7, and Table 5). We found a total of eight candidate genes across both latitude and altitude overlapping with the latitudinal selection candidates of the North American cline (Fabian et al. 2012): Ets65A, Elk, sba, CG32066, dpr8, CG8177, X11Lbeta, and CG42699 (Figures 1 and 2). Four of these - Ets65A, Elk, sba, CG32066 - are also clinal genes in North America (Fabian et al. 2012), so we were able to compare the direction of change in their candidate SNPs’ allele frequencies and our own. In all four cases, the estimated frequency of the major allele in all candidate SNPs increases consistently from RG to FR to NL, while in ZI it is about the same as in RG. A literature research on the function of these genes showed a loose network of similar overlapping functions with
common biological themes (see Figure 2). Most notably, six of these genes (Ets65A, Elk, CG32066, dpr8, X11Lbeta, and CG42699) are related to various kinds of behavioral responses. Additionally, they have been found to influence temperature sensitivity (CG8177 and CG42699), oxidative stress (CG8177 and dpr8), courtship song (dpr8 and X11Lbeta), and mushroom bodies (X11Lbeta and CG32066), structures in the fly brain essential for learning and memory. Ets65A is a transcription factor expressed in a nutrition-dependent manner in the adipose tissue (Baltzer et al. 2009), which may affect lifespan (Ayroles et al. 2009; Durham et al. 2014), a life-history trait related to starvation resistance and cold resistance (Hoffmann et al. 2005; Ayroles et al. 2009). Intriguingly, 11 SNPs of Ets65A had been found to vary along a cline in North America, and those with the highest BF, chr3L_6112566 and chr3L_6106869, as well as 22 of the highest F<sub>ST</sub> variants, are intronic, suggesting an important adaptive role in gene regulation. Even more interesting in this regard is the synonymous SNP chr3L_6093812, whose predominantly European allele, G (f(G)<sub>FR</sub>=0.86, f(G)<sub>NL</sub>=1, f(G)<sub>RG</sub>=0.18, f(G)<sub>ZI</sub>=0.33), generates a more frequently used codon (AAG, used 70% of the time). Another transcription factor that showed up in our results was sba (six-banded), an important developmental gene (Zeidler & Mlodzik 1997; Blanco et al. 2010; Iyer et al. 2013) that might be involved in changing the gut morphology in response to gut microbes (Sharon et al. 2010; Ridley et al. 2012; Newell & Douglas 2014; Broderick et al. 2014). Two additional genes from this analysis are also implicated in mating behavior: dpr8 is involved in the production of male courtship song (Moran & Kyriacou 2009), while X11Lbeta is upregulated in females responding to male courtship song (Immonen & Ritchie 2012). Interestingly, the two genes have been found to vary in both a North American (Fabian et al. 2012) and an Australian cline (Kolaczkowski et al. 2011). The third American-Australian cline candidate that also showed up in our analysis was CG8177. This is interesting in the context of its possible role in hypoxia tolerance (Azad et al. 2012), which it
shares with dpr8 (Weber et al. 2012), and its temperature-sensitive effects on developmental time (Mensch et al. 2008). Sensitivity to temperature seems to impact the function of another candidate gene, CG42699, both through its interaction with Syx1A (van Swinderen & Greenspan 2005), and its expression after exposure to cold shock (Vermeulen & Bijlsma 2004; Vermeulen et al. 2013). Likewise, CG32066 might play a role in the crosstalk of the juvenile hormone and ecdysteroids (Li et al. 2007), which are known to impact reproductive diapause, an important overwintering mechanism (Mitrovski & Hoffmann 2001; Boulêtreau-Merle & Fouillet 2002) that varies clinally (Schmidt et al. 2005). It is therefore less surprising that CG32066 harbors clinal SNPs in North America (Fabian et al. 2012), and that it responds to non-optimal rearing temperatures (Chen et al. 2015). Our final candidate gene was the Ca$^{2+}$-gated K$^+$ channel Elk, whose mutations impair locomotion (Warmke & Ganetzky 1994; Littleton & Ganetzky 2000). Interestingly, it has been shown that locomotion may be correlated with temperature conditions (Crill et al. 1996; Gibert et al. 2001a), and that neurotransmitter release triggered by voltage-gated channels can depend on temperature (Chuang et al. 2004; Wu et al. 2005).

Overlap of enriched Gene Ontology terms with other studies

To gain a clearer understanding of the biological functions of our most significant clinal genes, we tested for significant enrichment of GO and KEGG/Reactome terms. Furthermore, we performed an equivalent enrichment analysis of the North American candidate genes (Fabian et al. 2012), and then assessed the overlap with our candidates. Using Cytoscape’s ClueGO and CluePedia plugins, we integrated our candidate genes into GO networks and networks with KEGG/Reactome metabolic pathways. Table 6 shows the results of the ClueGO analysis for our latitudinal selection candidate genes (N = 378). ClueGO resulted in 111 significantly enriched GO terms (for the 20 most highly enriched categories, see Table 6).
These categories are grouped into clusters using Cohen’s Kappa statistics, based on their shared genes (Figure S4). Many terms that were grouped in these clusters could be related to various aspects of the nervous system, epithelium, wing, and tube development, and cover multiple developmental stages (Figure S4). Interestingly, the majority of these terms (12 out of 20 for appendage development, 10 out of 16 for taxis and 3 out of 5 for generation of neurons) were also significantly enriched in our ClueGO analysis of all three population pairs of the North American cline (Fabian et al. 2012). All the remaining terms from the three clusters were enriched in at least one population pair of the North American cline (Table 6).

We further wanted to know if these overlaps could be generalized to the total set of enriched GO terms for all the population pairs from North America. Table 6 shows that the overlap in all three cases was not only substantial (between 43 and 51%), but also statistically significant ($P$-values between $10^{-16}$ and $10^{-5}$, hypergeometric test). Moreover, we even found some overlap between the Kappa-defined clusters, which were defined according to the most significant GO term in the cluster. Table 6 shows the numbers of overlapping GO terms and pathways, as well as Kappa-defined clusters (all at $P$-value FDR cutoff of 0.01).

**Discussion**

In this study, we used an approach incorporating multiple lines of evidence for adaptation to cold. To do this, we examined (1) SNPs related to cold tolerance from GWA studies, (2) candidate genes from literature, and (3) genes with SNPs that show strong evidence of being adaptive. We investigated the potential effects of linkage to QTLs known to affect cold tolerance, as well as cosmopolitan inversions. Furthermore, we looked for overlaps with candidate genes from other studies, and added information from GO enrichments of those
candidate genes, as well as our own. In the following, we discuss the results of these approaches, particularly in the context of selective forces and fitness trade-offs, and finally propose an adaptive network of genes in the core of the complex cold tolerance phenotype.

Evidence for adaptation to cold tolerance from GWA studies

We started our analysis by quantifying the amount of population differentiation at SNPs associated with CCRT, RSS, and SR. Each of the three traits has an important adaptive role. It is possible that the observed difference in average $F_{ST}$ between the three traits is due to different fitness trade-offs among the traits. We know from previous studies that RSS is sensitive to varying environmental conditions, especially low temperature, and that it has a strong effect on fitness (Boulétreau-Merle & Fouillet 2002; Hoffmann 2010; Goenaga et al. 2010, 2012, 2013). In natural populations, RSS depends on pre-adult resource acquisition, which varies with developmental temperature (Chippindale et al. 1998). Both artificial selection experiments (Hoffmann et al. 2005) and quantitative genetics studies of these traits (Ayroles et al. 2009) have demonstrated a trade-off between cold and starvation resistance. Lines with higher cold tolerance (i.e. shorter CCRT) mate more rapidly and have high competitive fitness, which comes at the expense of surviving starvation stress, while lines with higher RSS tend to have longer life spans at the cost of fitness (Ayroles et al. 2009). Additionally, lines able to postpone egg-laying in the autumn have a longer life expectancy and greater RSS, which is favored under low temperatures (Boulétreau-Merle & Fouillet 2002) and has implications for surviving cold conditions when food is scarce (Hoffmann et al. 2005; Goenaga et al. 2012). Another factor important for the RSS component of fitness is food availability during the cold period of the year. Interestingly, because fruit flies feed on decomposing fruit and on the bacteria and yeasts that grow on it (Markow & O’Grady 2008),
a colder environment might reduce the rate of fruit decay, shortening the period of time when food is scarce, and thus neutralizing the need for increased RSS in lower temperatures (Goenaga et al. 2013). As for SR, there might be a similar trade-off between the fitness advantages of good locomotor performance (e.g. rapid mating, defense of territory, escape from predators) and the pressure to perform better in cold conditions. Indeed, both developmental and adult temperatures have been shown to affect locomotor performance (Crill et al. 1996; Gibert et al. 2001a). In short, the $F_{ST}$ and BFs of SNPs associated with phenotypic traits are consistent with the conclusion that European climate might be at the same time reducing both RSS and SR in favor of greater cold resistance, and incurring an additional fitness cost on RSS due to food being naturally preserved by colder temperatures.

Evidence from candidate genes from literature

The second line of evidence for adaptation comes from our study of candidate genes previously described in the literature (Tables S2-S3). Both $F_{ST}$ (Tables S3) and BFs (Tables S3 and S4) showed that genes related to tolerance to cold or heat, disturbance, and starvation stress generally contain SNPs whose allele frequency patterns are better explained by selection. We found no literature candidate genes in the neighborhood of the breakpoints of the most common cosmopolitan inversions (see inversion_analysis.xlsx on Dryad). This is particularly striking in (1) heat-shock genes Hsp26 and Hsp68, whose synonymous variants suggest that they might be under fitness cost in European populations due to preferred codon usage; (2) the gene lola, which contained highly differentiated variants at both 3’ and 5’ regions, introns (including the intron variant chr2R_6394221 with strong evidence ($ln(BF)>3$; $P < 0.03$) for 4 environmental variables), as well as synonymous and nonsynonymous sites; and (3) another 6 genes with highly differentiated nonsynonymous variants (CG18140, CG31738, CG12943, CG30379, nclb, and chas). While in the case of some genes not all
population comparisons showed an elevated $F_{ST}$, the results are nevertheless indicative. Furthermore, our aim here was not to show that genes are disproportionately associated with CCRT, rather than the other two traits. Indeed, previous research has shown that there are fitness trade-offs between CCRT and starvation resistance, and likewise that locomotion is related to metabolism rate and dependent on the ambient temperature. We therefore expect that selection might be acting on all three traits in a concerted manner. Another interesting question is whether there is any overlap between genes related to different traits. Although the SNPs associated with the three traits in recent GWAS studies (Mackay et al. 2012; Huang et al. 2012; Huang et al. 2014) showed no overlap, literature suggests that several genes involved in lipid regulation can vary along latitudinal clines with significantly different temperature regimes. For instance, $CG12054$ codes for a transcription factor that is a target of Forkhead box $O$, itself a transcription factor that directly influences lifespan by regulating lipid metabolism (Alic et al. 2014). Furthermore, $Dyrk2$, $hdc$, $shep$ and $chas$ function in larval fat storage (Reis et al. 2010), $Ire1$ is involved in the regulation of energy metabolism (Pile et al. 2003), and $Octbeta3R$ mediates appetite for energy-rich foods (Zhang et al. 2013). Strikingly, all 7 of these genes have been related the North American latitudinal cline (Fabian et al. 2012), and most recently, all except for $Octbeta3R$ have been found to respond to a range of different suboptimal rearing temperatures (Chen et al. 2015). The gene $shep$ is particularly interesting because of functions that might make it important for both RSS and SR. As already mentioned, it plays a role in fat storage (Reis et al. 2010), but it was originally described in a forward genetic screen for gravitaxis (Armstrong et al. 2006), as its mutants impair the ability of the fly to perceive gravity. Many SNPs that map to $shep$ were $F_{ST}$ outliers (top 1%, Table S4), but more interestingly, seven SNPs also had BFs with positive evidence of selection ($ln(BF)>1$ or nominal $P < 0.0014$) and two with strong evidence ($ln(BF)>3$ or $P < 0.03$) for at least one environmental variable (Table S4 and S5). Six of these
seven SNPs are intronic, but *chr3L_5154623* might be particularly adaptively important, as it
maps to the 3’UTR region of the gene. The preponderance of candidate SNPs in the region of
*shep* suggests its importance for adaptation related to both RSS and SR, and the fact that its
SNPs also vary clinally in North America (*Fabian et al.* 2012) and that it changes expression
depending on rearing temperatures (*Chen et al.* 2015) also suggests its importance for CCRT.

Several previous QTL studies have attempted to localize regions of the genome
responsible for heat and cold resistance (*Morgan & Mackay* 2006; *Norry et al.* 2007, 2008;
*Svetec et al.* 2011; *Wilches et al.* 2014). To find out if genes from these QTLs show
signatures of selection, we compared the candidate genes from these studies with our
candidates significantly correlated with latitude and altitude (*ln*(BF)>5, as reported by
*Bayenv2, P < 0.0273). We found that four genes, *dlg1* (*Norry et al.* 2008), *CG1677* (*Wilches
*et al.* 2014), *rdgA* (*Svetec et al.* 2011), and *DnaJ-1* (*Morgan & Mackay* 2006), were also
highly correlated with latitude in our study. None of them were also correlated with altitude.
However, *rdgA* was also found to be a candidate gene in the North American cline (*Fabian et
al.* 2012). Taken together, the evidence from both high *F*<sub>ST</sub> and from significant BFs confirms
adaptive significance of several genes previously known to be important for CCRT, RSS or
SR, and in some cases even all three traits.

*Evidence from genome-wide top candidate genes*

The final line of evidence for cold adaptation comes from our genome-wide analysis (Figure
1) of genes that contained variants with particularly strong evidence of environmental
selection (*ln*(BF)>5, *(Kass & Raftery 1995) or *P < 0.0273), which means that the model
including selection is more than *e*<sup>5</sup> (or ≈148) times more likely than neutrality. Since latitude
and altitude are variables that may account for more conditions than just temperature (e.g.
length of day, amount of insolation, seasonality, amount of oxygen, pressure), and because
temperature variables were correlated to latitude, we decided to focus the analysis on genes with strong evidence of selection particularly for these two variables. An intersection of genes with overwhelming evidence for both altitudinal and latitudinal selection (Figure 1) would thus control for many potential confounding factors. As an additional level of control for false positives, we particularly closely examined genes from the overlap of latitude and altitude that have also been proposed from a North American cline (Fabian et al. 2012). We discovered that 8 genes (Ets65A, Elk, sba, CG32066, dpr8, CG8177, X11Lbeta, and CG42699) conformed to all of these strict conditions. Surprisingly, we found that these genes could be organized into a network, where each gene was functionally related to up to three of its neighbors (Figure 2). Four of these eight genes are also clinal genes in North America (Ets65A, Elk, sba, CG32066), so we were able to compare the direction of change in allele frequencies between their candidate SNPs with available data from Fabian et al. (2012). In all four cases, the estimated frequency of the major allele increases consistently from RG to FR to NL, while in ZI it is about the same as in RG. The various functions of these genes suggest that adaptation to the more temperate local conditions in Europe has been a complex process involving many factors important for fitness, from direct tolerance to cold and oxidative stress, to developmental time, nutrient storage, locomotion, mating behavior, and even learning and memory. This may suggest the presence of epistatic fitness interactions.

Finally, we examined the GO and pathway enrichment and tested for significance of overlap with equivalent enrichment analyses that we performed using the candidate genes from North America (Fabian et al. 2012). GO analyses could increase the amount of information about adaptation in certain pathways that might be revealed by the joint effect of genes that individually might contribute only slightly to the trait. Thus, GO analyses might also complement our knowledge by indicating genes that did not pass our stringent significance threshold. Our strict criteria for ascertaining genes resulted in only 27 genes that
correlated with altitude, so that we could not find enrichment of these genes. However, the
results of overlaps with latitude were quite surprising. We found significant overlaps not only
with all population pairs from North America, but even with only the genes with the steepest
frequency change with latitude, termed “significantly clinal” genes by Fabian et al. (2012).
To account for possible gene length bias, we performed GO enrichment using the software
*Gowinda* with parameters corresponding to those used by Fabian et al. (2012), but allowing
genes with multiple independent SNPs to be scored more than once for different GO terms.
We recovered 72 significantly enriched terms for latitudinal SNP candidates (at \(P_{\text{FDR}}<0.05\)).
8 of these terms overlapped with terms enriched in our ClueGO analysis (Fisher’s exact test:
\(P<0.00021\)).

Overall, the significance levels of overlaps of enriched terms were even more
pronounced than those of genes. Moreover, even the clusters of enriched terms produced by
the functional grouping in ClueGO showed overlap with the clusters we got from North
American enrichment analyses. The largest clusters from terms enriched for our latitude
candidate genes (appendage development, taxis, and generation of neurons) are in line with
selection possibly acting on the fly appendages (Bergmann’s rule) and perhaps also
influencing locomotion, behavior, memory, and learning. Taken together, the GO and
pathway analyses showed that we could replicate the significance of gene overlaps with the
candidates of Fabian et al. (2012), even when taking into account the broader functional roles
of the candidate genes. Perhaps even more importantly, the particularly enriched terms and
clusters make sense in the wider context of adaptation to colder European environments.

**Conclusions**
In this study we have detected footprints of polygenic adaptation in *Drosophila melanogaster* to temperature-related traits. Our results suggest that these traits may have responded to selection in a concerted manner, most likely as a result of complex fitness trade-offs. Additionally, we found that SNPs under strong environmental selection support genes that significantly overlap with clinal candidates from other continents. The overlaps are even more significant if assessed from common biological pathways and gene ontology terms enriched with candidate genes between studies. Lastly, we proposed a network of genes with the strongest evidence of selection, which suggests that adaptation to new environments in Europe involved a strong direct response to cold, but also changes in development, mating behavior, oxidative stress, locomotion, reproductive diapause, learning, and memory. Functional studies of these genes in the context of cold tolerance are needed to confirm these findings. Also, future studies of local adaptation to cold should take into account the intricacies of different selective pressures that may be operating on many genes across the genome simultaneously.

**Authors’ contributions**

V.B. and A.W. carried out the bioinformatics and analyzed the data. S.H. contributed bioinformatics resources and tools. V.B. and A.W. performed the statistical analysis. A.W. and W.S. initiated, designed and coordinated the study. V.B., A.W., and W.S. wrote the manuscript. All authors read and approved the manuscript.

**Data accessibility**

DNA sequences of FR and RG populations were downloaded from: [http://www.dpgp.org](http://www.dpgp.org). DNA sequences of the ZI population were downloaded from: [http://www.dpgp.org/dpgp3](http://www.dpgp.org/dpgp3).
DNA sequences of the NL population are available from:


Further data and scripts are available from Dryad (doi:10.5061/dryad.20t9v).

Acknowledgements

The authors thank Ary Hoffmann and three anonymous reviewers for constructive comments on a previous version of the manuscript, and Pablo Duchén and Stefan Laurent for bioinformatic support in the initial phases of the project. This work was supported by the Marie-Curie Initial Training Network INTERCROSSING, research project A7, “Selective sweep data and next generation sequence data analysis”. A.W. was supported by the VW Foundation (ref 80462). W.S. was supported from the German Research Foundation (Priority Program 1819, grant Ste 325/17-1).
References


experimental and in silico data. *Bioinformatics*, *29*, 661–663.


underlying local adaptation. *Genetics, 185*, 1411–1423.


among strains, local populations, and geographic regions: patterns for desiccation, starvation, cold

Hoffmann AA, Rieseberg LH (2008) Revisiting the impact of inversions in evolution: from population genetic
markers to drivers of adaptive shifts and speciation? *Annual Review of Ecology, Evolution, and
Systematics, 39*, 21–42.

Hoffmann AA, Weeks AR (2007) Climatic selection on genes and traits after a 100 year-old invasion: a critical
look at the temperate-tropical clines in *Drosophila melanogaster* from eastern Australia. *Genetica, 129*,
133–147.

Holsinger KE, Weir BS (2009) Genetics in geographically structured populations: defining, estimating and

Huang W, Massouras A, Inoue Y et al. (2014) Natural variation in genome architecture among 205 *Drosophila*

quantitative traits. *Proceedings of the National Academy of Sciences of the United States of America, 109*,
15553–15559.

*Bioinformatics, 18*, 337–338.

Immonen E, Ritchie MG (2012) The genomic response to courtship song stimulation in female *Drosophila*

ONE, 8*, e72434.

Jain K, Stephan W (2015) Response of polygenic traits under stabilizing selection and mutation when loci have
unequal effects. *G3: Genes, Genomes, Genetics, 5*, 1065–1074.

Joost S, Bonin A, Bruford MW et al. (2007) A spatial analysis method (SAM) to detect candidate loci for

Jordan KW, Morgan TJ, Mackay TFC (2006) Quantitative trait loci for locomotor behavior in *Drosophila*

27–30.


Norry FM, Scannapieco AC, Sambucetti P, Bertoli CI, Loeschcke V (2008) QTL for the thermotolerance effect of heat hardening, knockdown resistance to heat and chill-coma recovery in an intercontinental set of
recombinant inbred lines of *Drosophila melanogaster*. Molecular Ecology, **17**, 4570–4581.


**Figures and Tables**

**Figure 1** Proportions of genes supported by SNPs with strong evidence ($\ln(\text{BF}) > 5$ or $P < 0.0029$) for correlation with latitude and altitude (*Bayenv2*) that overlap with candidate genes from North America (Fabian *et al.* 2012). The most interesting genes that overlap among all three sets are shown in the bottom panels. For overlaps between North America and other environmental variables, see also Figures S5 through S7.
Figure 2 Manually drawn network of candidate genes that overlap with previous studies (Australia (Kolaczkowski et al. 2011); North America (Fabian et al. 2012); see also Figure 1 and S4 to S6). Only the genes (coloured ellipses) of SNPs related to cold tolerance (Bayenv2, ln(BF)>5 or $P < 0.0029$) were considered. Open rectangles denote functional relevance from literature with lines exemplifying relationships between genes and functions, with the relevant references.

![Gene network diagram](image-url)
Table 1 Size of the trait-associated SNP datasets before and after filtering.

<table>
<thead>
<tr>
<th>Trait</th>
<th>Association study</th>
<th>Total SNPs associated</th>
<th>Associated SNPs after filtering ($F_{ST}$)</th>
<th>Associated SNPs after filtering (Bayenv2)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Chill Coma Recovery Time (CCRT)</td>
<td>Huang et al. 2014 (DGRP Freeze 2.0)</td>
<td>119</td>
<td>59</td>
<td>14</td>
</tr>
<tr>
<td>Resistance to Starvation Stress (RSS)</td>
<td>Huang et al. 2014 (DGRP Freeze 2.0)</td>
<td>132</td>
<td>64</td>
<td>14</td>
</tr>
<tr>
<td>Startle Response (SR)</td>
<td>Huang et al. 2014 (DGRP Freeze 2.0)</td>
<td>78</td>
<td>51</td>
<td>12</td>
</tr>
<tr>
<td>Total genomic background after filtering</td>
<td></td>
<td>3 663 890 (autosomes) + 867 049 (X chromosome)</td>
<td>313 972 (autosomes) + 39 304 (X chromosome)</td>
<td></td>
</tr>
</tbody>
</table>
Table 2 The populations and environmental variables used in the analysis.

<table>
<thead>
<tr>
<th>Population</th>
<th>Latitude (degrees)</th>
<th>Altitude (m)</th>
<th>T_{min} (°C) of the coldest month*</th>
<th>T_{min} (°C) of the hottest month*</th>
<th>T_{min} (°C) yearly average*</th>
<th>T_{max} (°C) yearly average*</th>
</tr>
</thead>
<tbody>
<tr>
<td>The Netherlands, Leiden (NL)</td>
<td>52°09'29 N</td>
<td>-2</td>
<td>0.2</td>
<td>12.5</td>
<td>6.1</td>
<td>13.4</td>
</tr>
<tr>
<td>France, Lyon (FR)</td>
<td>45°45'00 N</td>
<td>198</td>
<td>0.1</td>
<td>15.6</td>
<td>7.5</td>
<td>16.3</td>
</tr>
<tr>
<td>Rwanda, Gikongoro (RG)</td>
<td>02°27’50 S</td>
<td>1796</td>
<td>13.9</td>
<td>14.9</td>
<td>14.4</td>
<td>24.8</td>
</tr>
<tr>
<td>Zambia, Siavonga (ZI)</td>
<td>16°32’17 S</td>
<td>481</td>
<td>10</td>
<td>17.9</td>
<td>14.9</td>
<td>26.4</td>
</tr>
</tbody>
</table>

* Climate data was taken from World Weather Information Service – World Meteorological Organization (worldweather.wmo.int). Climatological information is based on monthly averages for the 30-year period 1961-1990.
Table 3 Mean population differentiation ($F_{ST}$) over associated SNP sets (Huang et al. 2014). Empirical $P$-value ($P_{\text{empirical}}$) was estimated by random resampling approach (see methods).

<table>
<thead>
<tr>
<th>Trait</th>
<th>Number of associated SNPs</th>
<th>Population comparison</th>
<th>Target mean $F_{ST}$</th>
<th>Genome mean $F_{ST}$</th>
<th>$P_{\text{empirical}}$</th>
<th>significance</th>
</tr>
</thead>
<tbody>
<tr>
<td>CCRT</td>
<td>59</td>
<td>RG-NL</td>
<td>0.133</td>
<td>0.081</td>
<td>0.031</td>
<td>*</td>
</tr>
<tr>
<td></td>
<td></td>
<td>RG-FR</td>
<td>0.147</td>
<td>0.099</td>
<td>0.0654</td>
<td>°</td>
</tr>
<tr>
<td></td>
<td></td>
<td>ZI-NL</td>
<td>0.125</td>
<td>0.085</td>
<td>0.0708</td>
<td>°</td>
</tr>
<tr>
<td></td>
<td></td>
<td>ZI-FR</td>
<td>0.129</td>
<td>0.087</td>
<td>0.0765</td>
<td>°</td>
</tr>
<tr>
<td>RSS</td>
<td>64</td>
<td>RG-NL</td>
<td>0.134</td>
<td>0.081</td>
<td>0.0259</td>
<td>*</td>
</tr>
<tr>
<td></td>
<td></td>
<td>RG-FR</td>
<td>0.122</td>
<td>0.099</td>
<td>0.1963</td>
<td>ns</td>
</tr>
<tr>
<td></td>
<td></td>
<td>ZI-NL</td>
<td>0.11</td>
<td>0.085</td>
<td>0.1613</td>
<td>ns</td>
</tr>
<tr>
<td></td>
<td></td>
<td>ZI-FR</td>
<td>0.091</td>
<td>0.087</td>
<td>0.4163</td>
<td>ns</td>
</tr>
<tr>
<td>SR</td>
<td>51</td>
<td>RG-NL</td>
<td>0.083</td>
<td>0.081</td>
<td>0.4278</td>
<td>ns</td>
</tr>
<tr>
<td></td>
<td></td>
<td>RG-FR</td>
<td>0.081</td>
<td>0.099</td>
<td>0.6915</td>
<td>ns</td>
</tr>
<tr>
<td></td>
<td></td>
<td>ZI-NL</td>
<td>0.096</td>
<td>0.085</td>
<td>0.314</td>
<td>ns</td>
</tr>
<tr>
<td></td>
<td></td>
<td>ZI-FR</td>
<td>0.089</td>
<td>0.087</td>
<td>0.4336</td>
<td>ns</td>
</tr>
</tbody>
</table>
Table 4 Mean Bayes factors over associated SNPs sets calculated with Bayenv2. Empirical $P$-values were obtained by random resampling from genomic background (see methods). Asterisks indicate significance threshold ($P < 0.05 (*)$; $P < 0.01 (**)$; $P < 0.001 (***)$).

<table>
<thead>
<tr>
<th>Quantitative trait / association study</th>
<th>Environmental variable</th>
<th>Mean Bayes factor</th>
<th>$P_{\text{empirical}}$</th>
<th>significance</th>
</tr>
</thead>
<tbody>
<tr>
<td>CCRT (N = 14)</td>
<td>Latitude</td>
<td>16.8739</td>
<td>0.129</td>
<td>ns</td>
</tr>
<tr>
<td></td>
<td>Altitude</td>
<td>1532.67</td>
<td>0.015</td>
<td>*</td>
</tr>
<tr>
<td></td>
<td>$T_{\text{min}}$ of the coldest month</td>
<td>8601.49</td>
<td>0.001</td>
<td>***</td>
</tr>
<tr>
<td></td>
<td>$T_{\text{min}}$ of the hottest month</td>
<td>3373.85</td>
<td>0.755</td>
<td>ns</td>
</tr>
<tr>
<td></td>
<td>$T_{\text{min}}$ yearly average</td>
<td>17.1501</td>
<td>0.066</td>
<td>°</td>
</tr>
<tr>
<td></td>
<td>$T_{\text{max}}$ yearly average</td>
<td>15.3386</td>
<td>0.155</td>
<td>ns</td>
</tr>
<tr>
<td>RSS (N = 14)</td>
<td>Latitude</td>
<td>0.24811</td>
<td>0.300</td>
<td>ns</td>
</tr>
<tr>
<td></td>
<td>Altitude</td>
<td>0.22003</td>
<td>0.168</td>
<td>ns</td>
</tr>
<tr>
<td></td>
<td>$T_{\text{min}}$ of the coldest month</td>
<td>0.22871</td>
<td>0.430</td>
<td>ns</td>
</tr>
<tr>
<td></td>
<td>$T_{\text{min}}$ of the hottest month</td>
<td>0.21495</td>
<td>0.763</td>
<td>ns</td>
</tr>
<tr>
<td></td>
<td>$T_{\text{min}}$ yearly average</td>
<td>0.23986</td>
<td>0.475</td>
<td>ns</td>
</tr>
<tr>
<td></td>
<td>$T_{\text{max}}$ yearly average</td>
<td>0.24788</td>
<td>0.262</td>
<td>ns</td>
</tr>
<tr>
<td>SR (N = 12)</td>
<td>Latitude</td>
<td>0.21202</td>
<td>0.378</td>
<td>ns</td>
</tr>
<tr>
<td></td>
<td>Altitude</td>
<td>0.21746</td>
<td>0.170</td>
<td>ns</td>
</tr>
<tr>
<td></td>
<td>$T_{\text{min}}$ of the coldest month</td>
<td>0.20214</td>
<td>0.830</td>
<td>ns</td>
</tr>
<tr>
<td></td>
<td>$T_{\text{min}}$ of the hottest month</td>
<td>0.19755</td>
<td>0.515</td>
<td>ns</td>
</tr>
<tr>
<td></td>
<td>$T_{\text{min}}$ yearly average</td>
<td>0.21252</td>
<td>0.494</td>
<td>ns</td>
</tr>
<tr>
<td></td>
<td>$T_{\text{max}}$ yearly average</td>
<td>0.20995</td>
<td>0.379</td>
<td>ns</td>
</tr>
</tbody>
</table>
**Table 5** Number of genes with $\ln(BF)>5$ ($P < 0.0029$) for correlation with latitude that overlap with candidate genes of (Fabian *et al.* 2012). Final column gives the significance of the overlap from a hypergeometric test.

<table>
<thead>
<tr>
<th>Latitudinal candidate genes (Fabian <em>et al.</em> 2012)</th>
<th>$N$ genes (candidate, Fabian <em>et al.</em> 2012)</th>
<th>$N$ genes (total, Fabian <em>et al.</em> 2012)</th>
<th>Latitude $\ln(BF)&gt;5$ overlap (of $N_{\text{total}}=378$)</th>
<th>$P_{\text{hyper}}$</th>
</tr>
</thead>
<tbody>
<tr>
<td>Florida-Maine</td>
<td>2010</td>
<td>11314</td>
<td>78</td>
<td>0.1018</td>
</tr>
<tr>
<td>Florida-Pennsylvania</td>
<td>2051</td>
<td>11314</td>
<td>82</td>
<td>0.0541</td>
</tr>
<tr>
<td>Pennsylvania-Maine</td>
<td>720</td>
<td>11314</td>
<td>51</td>
<td>$2.36 \cdot 10^{-7}$</td>
</tr>
<tr>
<td>Clinal genes (all)</td>
<td>1973</td>
<td>11314</td>
<td>72</td>
<td>0.2191</td>
</tr>
<tr>
<td>Clinal genes (significant)</td>
<td>140</td>
<td>11314</td>
<td>11</td>
<td>0.0072</td>
</tr>
<tr>
<td>Across all populations</td>
<td>3169</td>
<td>11314</td>
<td>131</td>
<td>0.0024</td>
</tr>
</tbody>
</table>
**Table 6** Number of GO terms (upper table) and KEGG and Reactome pathways (lower table) that overlap with (Fabian *et al.* 2012) in latitudinal differentiation. With altitudinal differentiation we could not find any significant overlap (data not shown).

<table>
<thead>
<tr>
<th>Latitudinal candidate genes (Fabian <em>et al.</em> 2012)</th>
<th>N GO terms (P&lt;0.01) (Fabian <em>et al.</em> 2012)</th>
<th>N GO terms (total)</th>
<th>Latitude GO terms overlap (of N(_{\text{total}}=111))</th>
<th>(P_{\text{hyper}})</th>
</tr>
</thead>
<tbody>
<tr>
<td>Florida-Maine</td>
<td>101</td>
<td>4844</td>
<td>57</td>
<td>(4.99 \cdot 10^{-73})</td>
</tr>
<tr>
<td>Florida-Pennsylvania</td>
<td>131</td>
<td>4844</td>
<td>48</td>
<td>(8.47 \cdot 10^{-49})</td>
</tr>
<tr>
<td>Pennsylvania-Maine</td>
<td>79</td>
<td>4844</td>
<td>52</td>
<td>(2.87 \cdot 10^{-71})</td>
</tr>
<tr>
<td>Clinal genes (all)</td>
<td>152</td>
<td>4844</td>
<td>61</td>
<td>(9.98 \cdot 10^{-67})</td>
</tr>
<tr>
<td>Clinal genes (significant)</td>
<td>36</td>
<td>4844</td>
<td>6</td>
<td>(1.41 \cdot 10^{-4})</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Latitudinal candidate genes (Fabian <em>et al.</em> 2012)</th>
<th>N pathway terms (P&lt;0.01) (Fabian <em>et al.</em> 2012)</th>
<th>N pathway terms (total)</th>
<th>Latitude pathway terms overlap (of N(_{\text{total}}=75))</th>
<th>(P_{\text{hyper}})</th>
</tr>
</thead>
<tbody>
<tr>
<td>Florida-Maine</td>
<td>117</td>
<td>3251</td>
<td>14</td>
<td>(2.47 \cdot 10^{-7})</td>
</tr>
<tr>
<td>Florida-Pennsylvania</td>
<td>54</td>
<td>3251</td>
<td>17</td>
<td>(5.43 \cdot 10^{-16})</td>
</tr>
<tr>
<td>Pennsylvania-Maine</td>
<td>45</td>
<td>3251</td>
<td>3</td>
<td>(0.0840)</td>
</tr>
<tr>
<td>Clinal genes (all)</td>
<td>115</td>
<td>3251</td>
<td>13</td>
<td>(1.37 \cdot 10^{-6})</td>
</tr>
<tr>
<td>Clinal genes (significant)</td>
<td>6</td>
<td>3251</td>
<td>0</td>
<td>-</td>
</tr>
</tbody>
</table>