Universiteit
Leiden
The Netherlands

Geographic variation and thermal adaptation in Bicyclus anynana

Jong, M.A. de

Citation
Jong, M. A. de. (2010, December 16). *Geographic variation and thermal adaptation in Bicyclus anynana*. Retrieved from https://hdl.handle.net/1887/16250

Version:	Corrected Publisher's Version
License:	Licence agreement concerning inclusion of doctoral thesis in the Institutional Repository of the University of Leiden
Downloaded from:	https://hdl.handle.net/1887/16250

Note: To cite this publication please use the final published version (if applicable).

Chapter 5

Footprints of selection in wild populations of *Bicyclus anynana* along a latitudinal cline

Maaike A. de Jong, Steve Collins, Patricia Beldade, Paul M. Brakefield and Bas J. Zwaan

Manuscript

Footprints of selection in wild populations of *Bicyclus anynana* along a latitudinal cline

Maaike A. de Jong[1], Steve Collins[3], Patricia Beldade[1], Paul M. Brakefield[1] and Bas J. Zwaan[2]

[1] *Institute of Biology, Leiden University, PO Box 9505, 2300 RA Leiden, The Netherlands*

[2] *Laboratory of Genetics, Wageningen University and Research Centre, PO Box 309, 6700 AH Wageningen, The Netherlands*

[3] *African Butterfly Research Institute, PO Box 14308, 0800 Nairobi, Kenya*

ABSTRACT

One of the major questions in ecology and evolutionary biology is how variation in the genome enables species to adapt to divergent environmental parameters. The use of latitudinal clines is a powerful approach in associating genetic variation with geographically varying thermal conditions. Here, we have taken a candidate gene approach to study footprints of thermal selection in six wild populations of the afrotropical butterfly *Bicyclus anynana*, sampled along a latitudinal cline covering a distance of ~3,000 km from the equator to the subtropics. We sequenced coding regions of 19 genes that are candidates for an association with thermal adaptation, including enzymes and other proteins from the glycolytic pathway and its branches, the lipid pathway, and genes involved in pigment biosynthesis. In addition, six genes from the heat shock family and five genes involved in developmental pathways for which we did not expect structural variation associated with a thermal gradient, were included as a type of negative control. We identified non-synonymous nucleotide polymorphisms in 11 candidate genes and tested these for significant clinal variation by correlation analysis of allele frequencies with latitude. As an additional analysis to infer evidence of selection we implemented the Beaumont-Nichols F_{ST} outlier method. Two metabolic enzymes of the glycolytic pathway, *Treh* and *UGPase*, showed significant clinal variation of which *UGPase* remained significant after multiple testing correction. In addition, the outlier analysis indicated a significantly higher F_{ST} value for the same amino acid polymorphism in *UGPase* than expected under a model of neutral evolution. In contrast, we found no evidence of clines with latitude in the heat shock proteins and developmental genes. The underlying phylogeographic structure of the populations based on the mtDNA COI gene and the silent SNPs of the candidate genes did not show a clinal pattern. Our results thus indicate that the observed clinal variation in *UGPase* and *Treh* may reflect adaptation to a geographic thermal gradient.

INTRODUCTION

A central goal in our quest to unravel the mechanisms of natural selection is to understand how variation in the genome enables species or populations to cope with, and adapt to divergent environmental conditions. Studying genes that are putatively under selection can shed light on the quantity and nature of genetic changes involved in adaptive differentiation, as well as on potential constraints on evolutionary responses. Identifying genetic variation involved in the local adaptation of wild populations is especially relevant in the context of present-day human-induced environmental change, including habitat fragmentation and climate change. As past evolutionary change is recorded in the genome, we can potentially use this information to make predictions about the evolutionary potential of wild populations in response to future environmental change. Moreover, genes involved in adaptive responses could be used as genetic markers in conservation efforts and to monitor species' molecular responses to selection imposed by environmental change (Hoffmann & Willi 2008).

An increasing number of studies indicate that genetic changes in particular loci can play an important role in the performance of organisms and fitness in relation to their environment (e.g. Mitton & Duran 2004, Hoekstra et al. 2006, Campbell 2010). Genetic changes leading to variation in responses to environmental conditions include changes in regulatory regions that induce the differential expression of genes. For example, variation in the expression of heat shock protein (Hsp) genes has been linked to the presence of transposable elements in the promoter regions in *Drosophila* species, with a likely role in thermal adaptation (Chen et al. 2007). Alternatively, changes in the coding regions of the genes can lead to amino acid variation, and potential structural differences, in the transcribed proteins. A well-studied example of this is amino acid polymorphism in the metabolic enzyme phosphoglucose isomerase (Pgi), which has been associated with fitness and performance differences particularly in relation to temperature in various Arthropod taxa, including butterflies (Hanski & Saccheri 2006), beetles (Rank et al. 2007) and amphipods (Patarnello & Battaglia 1992). In the Glanville fritillary butterfly, *Melitaea cinxia*, different Pgi alleles correlate with life history traits including dispersal, metabolic rate and population growth, and are linked to thermal performance (Hanski & Saccheri 2006, Saastamoinen & Hanski 2008).

One approach to infer evidence of selection at the molecular level is to associate genetic polymorphisms in populations with geographically varying environmental parameters. Temperature is generally recognized as one of the main environmental variables influencing and limiting organismal performance and fitness, and consequently determines the distribution and range of species (e.g. Fields 2001, Angilletta 2009). Latitudinal clines are a powerful tool in demonstrating patterns of past natural selection associated with temperature (Endler 1977), and have been described at the phenotypic and molecular level in a wide range of organisms, including flies (Hoffmann & Weeks 2007), fish (Schmidt et al. 2008), and plants (Hall et al. 2007). Although molecular clinal variation has been well studied in *Drosophila* (Sezgin et al. 2004, Hoffmann & Weeks 2007), there are few comparable studies in other insect groups with distinct biological properties. Butterflies provide model species in evolutionary and ecological studies (Brakefield & Frankino 2009), and are important bio-indicators because of their

sensitivity to environmental changes (Parmesan 2003). However, studies of molecular clinal variation in butterflies are very limited, apart from those on altitudinal variation in copper butterflies (Fischer & Karl 2010).

In the present study, we took a candidate gene approach to study footprints of selection in wild populations of the butterfly *Bicyclus anynana* along a latitudinal cline. *B. anynana* is an emerging model species for developmental and life history studies, with genetic information and tools becoming increasingly available (Brakefield et al. 2009). The species occurs in East Africa, where its range extends from equatorial Kenya to subtropical South Africa. Temperature values for means, minima and maxima increase towards the equator, while daily and annual mean temperature amplitudes decrease. We sampled six populations from the equator to the southernmost part of the range covering approximately 3,000 km, thereby extending over most of the species' latitudinal range (Condamin 1973). From the EST derived gene collection for *B. anynana* (Beldade et al. 2006, 2009), we selected a set of candidate genes potentially involved in temperature adaptation, mainly based on findings from research on *Drosophila*. The majority of candidate genes for which single-locus latitudinal clines are reported in the literature are metabolic enzymes (e.g. Sezgin et al. 2004). We selected genes coding for enzymes and other proteins involved in the glycolytic pathway and its important branches (*Gapdh2*, *Gdh*, *GlyP*, *Tpi*, *Treh*, and *UGPase*) and in the lipid pathway (*Apolphorin precursor*, *desat1*, *Lipase like*, *LpR*, *TAG Lipase*, and *Vg*), as well as an antioxidant gene (*Cat*).

In addition to the metabolic genes, we included several genes involved in the pigment biosynthesis of wing pattern pigmentation (*black*, *Catsup*, *ddc*, *light*, *ovo*, and *yellow*). Latitudinal and altitudinal clines for pigmentation are widely documented for a range of species, including mice (Hoekstra 2006), flies (Wittkopp et al. 2003), and butterflies (Ellers & Boggs 2002, Karl et al. 2009). The adaptive value of these patterns of variation along thermal gradients has been ascribed to the regulation of body temperature (Ellers & Boggs 2002, Karl et al. 2009), protection from UV radiation (Gunn 1998) and desiccation resistance (Rajpurohit et al. 2007). Studies have linked polymorphisms in coding regions of genes to phenotypic variation in pigmentation genes (e.g. Sturm et al. 2001, Hoekstra et al. 2006). Previously, in a two population comparison, we have shown population differentiation for thermal reaction norms of wing pattern elements (De Jong et al. 2010). In *B. anynana*, wing pattern is a fitness-related trait that plays a role in predator avoidance (Lyytinen et al. 2004) and sexual signalling (Robertson & Monteiro 2005). Although a relationship between wing pattern and thermal adaptation has not been described for this species thus far, the wing pattern pigmentation genes are an interesting group to screen for molecular clinal variation.

Lastly, we selected several genes from the heat shock protein family (*Hsp23*, *Hsp60*, *Hsp68*, *Hsp83*, *Hsc70-3*, *Hsc70-4*) and genes involved in developmental pathways (*Apc*, *dll*, *en*, *ovo*, and *wg*). Heat shock proteins and heat shock cognates have been shown to play an important role in temperature adaptation in a variety of organisms, but this is mainly through differential gene expression (e.g. Chen et al. 2007, Rinehart et al. 2007). Hence, these genes are generally considered less likely candidates to show clinal variation in the coding regions. Similarly, although the selected developmental genes play an important role in wing development and patterning, and may thus be

linked to geographic variation in pigmentation, research is revealing that phenotypic variation is mainly associated with these genes through gene regulation (Wittkopp & Beldade 2009). Therefore, we do not expect, *a priori*, a correlation between the structural sequences and latitude. This group can thus be seen as a type of negative control.

To detect patterns of selection related to temperature adaptation, we analysed clinal variation of amino acid polymorphisms in the studied genes by testing for significant correlations between population latitude and allele frequencies. To address the possibility of confounding the phylogenetic history of the species with patterns of natural selection, we compare our findings with the geographic structure of the populations based on the COI mitochondrial gene and putatively neutral silent SNPs of the nuclear genes. Finally, we implemented the Beaumont-Nichols F_{ST} outlier method (Beaumont & Nichols 1996) to identify loci under directional or balancing selection as an additional method to infer evidence of selection for the studied candidate genes.

MATERIALS AND METHODS

Populations and samples

In 2005 and 2006, six populations were sampled along a latitudinal cline from the following locations (from North to South, followed by their abbreviations, coordinates and sample sizes): Lake Mburo in Uganda (LM, 0° 38′ S 30° 57′ E, n = 50); Watamu in Kenya (WA, 3° 21′ S 40° 1′ E, n = 40); Ngezi forest on Pemba Island, Tanzania (PE, 4° 55′ S 39° 42′ E, n = 50); Zomba in Malawi (ZO, 15° 22′ S 35° 19′ E, n = 43); Mpaphuli Cycad Reserve in Limpopo, South Africa (LP, 22° 47′ S 30° 37′ E, n = 50); and False Bay Park of the Greater St. Lucia Wetland Area, KwaZulu Natal, South Africa (FB, 27° 58′ S 32° 21′ E, n = 50). Fig. 1 gives an overview of the locations of the populations on the African continent. For each population, the gender structure of the samples was

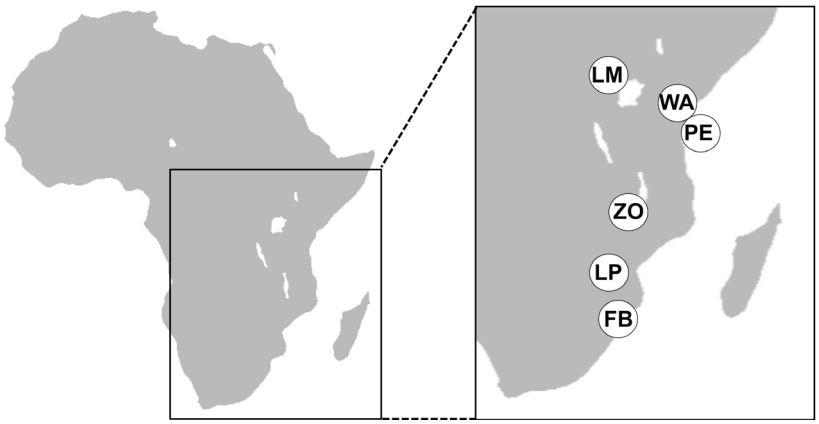

Figure 1. Overview of the locations of the analysed populations on the African continent. FB: False Bay; LM: Lake Mburo; LP: Limpopo; PE: Pemba; WA: Watamu; ZO: Zomba.

approximately evenly balanced between males and females. Samples were stored at -80°C until they were used for DNA extraction.

Candidate gene selection and primer design

We selected 19 candidate genes putatively involved in thermal adaptation from a published EST-derived gene collection of *B. anynana* (Beldade et al. 2006). This gene collection contains over 4,000 genes in the form of singletons and contigs assembled from expressed sequence tags, and was created using material from a *B. anynana* laboratory stock established from a wild population from Malawi in 1988. Although most candidate genes for this study were selected on the basis of *Drosophila* literature, we searched our *B. anynana* gene collection by tblastx analysis (e-score cut-off value of 1.0×10^{-5}) with *Bombyx mori* orthologs of the candidates to increase the chance of finding them in our gene collection. We obtained the *B. mori* orthologs from the Silkworm Genome Database (SilkDB, www.silkworm.genomics.org.cn).

The selected candidates include genes from the glycolytic pathway and its branches (*Gapdh2*, *Gdh*, *GlyP*, *Tpi*, *Treh*, and *UGPase*), the lipid pathway (*Apolphorin precursor*, *desat1*, *Lipase like*, *LpR*, *TAG Lipase*, and *Vg*), an antioxidant enzyme (*Cat*) and genes involved in pigmentation biosynthesis (*black*, *Catsup*, *ddc*, *light*, and *yellow*). In addition, we included six genes from the heat shock protein family (*Hsc70-3*, *Hsc70-4*, *Hsp23*, *Hsp60*, *Hsp68*, and *Hsp83*) and five genes involved in developmental pathways (*Apc*, *dll*, *en*, *ovo*, and *wg*) using the same method of selection from the EST-derived gene collection (Table 1 in Appendix).

For each gene, primers were designed to amplify lengths of 200 – 1,650 bp of the exonic regions, excluding the introns, resulting in one to three primer pairs per gene (Table 1 in Appendix). Based on the assumption of conserved intron positions between *Bicyclus anynana* and *Bombyx mori*, intron positions were identified by comparing coding sequence of the *B. mori* candidate orthologs with the genomic sequence from the SilkDB database. We designed the primers using the primer-BLAST tool with default settings on the website of the National Center for Biotechnology Information (NCBI, www.ncbi.nlm.nih.gov). Primers were tested on five individuals per populations and accepted for further use when the PCR products yielded bands of the same size. Table 1 (Appendix) gives an overview of the genes included in the study, including their full names and abbreviations, contig numbers from the EST-derived gene collection for *Bicyclus anynana* (Beldade et al. 2006, 2009), functional assignment, sequences of the primers used for amplification of the genes, and sequence lengths of the amplicons.

DNA extraction, PCR and sequencing

Genomic DNA was extracted from individual thoraces and legs using Qiagen's DNeasy tissue kit and following the manufacturer's instructions. DNA concentrations were measured with Nanodrop spectrophotometry and Picogreen fluorometry. Various studies have shown that accurate estimates of allele frequencies based on PCR results can be obtained by precise pooling of the DNA of several individuals combined (reviewed in Sham et al. 2002). Thus, individual samples of identical concentrations

were combined in pools of three, resulting in 17 pools per population. PCRs were conducted in a 50 µl volume, with 3 µl DNA, 0.25 µl ExTaq (Takara), 5 µl 10× ExTaq buffer, 5 µl of each dNTP (2.5 mM), and 1 µl of each primer (10 pmol). PCR conditions consisted of an initial cycle at 94°C for 3 min, 33 cycles of 94°C for 30 s, 60°C for 30 s, 72°C for 45 s, and a final extension at 72°C for 5 min. Primer combinations for which the PCR product did not yield a band on agarose gel for one or more of the pooled groups were excluded from the analysis. After cleaning, PCR product concentrations were measured using Picogreen fluorometry and subsequently standardised, after which the PCR products for each population were pooled with equal molar concentration. Each population sample was prepared according to standard Illumina protocols for genomic DNA samples (more information can be found on the website www.illumina.com), loaded into a single lane of an eight-lane flow cell and sequenced by pair-end 75 bp reads on a Illumina Solexa Genome Analyzer. Measuring of DNA concentrations, pooling of samples, PCR reactions, and Illumina sequencing were performed by Macrogen Inc. (Seoul, Korea).

Analysis

SNP calling

Maq software (Li et al. 2008) was used for alignment of the 75 bp reads to the reference sequence, which was the contig sequence available for each candidate gene. For SNP discovery, SNPs were called when the minor allele frequency was 4% or higher in one or more of the populations. The alignment and SNP calling analyses were carried out by Macrogen Inc. (Seoul, Korea). To further reduce the chance of calling false positive SNPs and to reduce the size of the data set, we included only the SNPs with a minor allele frequency of at least 20% in one population, or a minor allele frequency of at least 5% in three populations for all further analyses. Reading frame positions for protein translation of candidate genes were determined by BLAST alignment to *Bombyx mori* orthologs from the SilkDB, which allowed for assessment of silent (synonymous) and substitution (non-synonymous) SNPs.

Population genetic parameters

SNP allele frequency data was converted to genotype files for each locus and population using WhichLoci software (Banks et al. 2003), assuming Hardy-Weinberg equilibrium for the loci, and given the initial sample sizes of the collected populations. Haplotype structure and linkage between SNPs could not be analysed because the reconstructed genotype data were based on allele frequencies from pooled samples. The population genetic parameters of allelic richness (AR, corrected for sample size) and unbiased gene diversity (or expected heterozygosity, *H*) were calculated per population per gene using the program FSTAT (ver. 2.9.3; Goudet 2002). Pairwise population differences were calculated using analysis of variance followed by a Tukey's honest significant differences (HSD) test in SPSS (ver. 14). Population pairwise F_{ST} values based on the silent SNPs and associated p-values were calculated per gene with Arlequin (ver. 3.5;

Excoffier & Lischer 2010) following Weir & Cockerham (1984) using 1,000 permutations, and averaged over the total set of genes.

Evidence of selection: clinal variation and F_{ST} *outlier analysis*

For the analysis of clinal variation, we used the amino acid polymorphisms that were polymorphic in at least three populations. We chose to focus on the SNPs that showed at least 20% difference in allele frequency between the most divergent populations to reduce loss of statistical power due to multiple testing. For each SNP, the association between allele frequency and geographic origin of the populations (latitude) was tested using Pearson's correlation analysis in SPSS (ver. 14) software. Multiple comparisons were corrected using Benjamini & Hochberg's (1995) False Discovery Rate procedure.

As an additional method to detect evidence of selection, we implemented the Beaumont & Nichols (1996) outlier method, which uses the distribution of loci based on the relationship between F_{ST} and H in an island model. As F_{ST} is a measure of population divergence, it can be used to reveal patterns of local adaptation. When a locus is under directional or balancing selection, it is expected to show respectively higher and lower F_{ST} values than neutral loci (Beaumont & Nichols 1996). When taking into account clinal variation, a steeper cline will generally be associated with a higher F_{ST} value, indicating stronger population divergence. Here, we were primarily interested in finding loci with outlier F_{ST} values among the replacement SNPs that were analysed for clinal variation. Outlier loci are defined as loci that show significantly more (directional selection) or less (balancing selection) differentiation among populations than predicted by a neutral model. We also included all synonymous SNPs in the outlier analysis because these SNPs are expected to be mostly under neutral evolution and thus provide a type of neutral baseline with which outlier loci can be compared. We calculated upper and lower outlier F_{ST} values including all SNPs, under an infinite alleles mutation model with 10,000 simulations and a confidence interval of 0.95, using the program LOSITAN (Antao et al. 2008).

RESULTS

Population genetic diversity indices and unique SNPs

Table 2 shows the average allelic richness (AR) and gene diversity (H) per candidate gene including the standard deviation for each population, and the total number of unique SNPs per population (SNPs occurring in only one population). Allelic richness measures genetic diversity as the average number of alleles in a sample, and was corrected for the differences in sample size between the populations. An analysis of variance showed that AR is significantly lower ($p < 0.0001$) for the LM population, which belongs to a different subspecies (*B. anynana centralis*), compared to the other populations. The WA populations and the island population PE have lower values than the FB, LP and ZO populations but these differences are not significant (Table 2). Gene diversity (or expected heterozygosity, H) represents the probability that two randomly sampled alleles are different, and the populations show a very similar pattern for this

Table 2. Allelic Richness (AR) and gene diversity (*H*) averaged over all genes, and number of unique single nucleotide polymorphisms (SNPs) per population. FB: False Bay; LM: Lake Mburo; LP: Limpopo; PE: Pemba; WA: Watamu; ZO: Zomba; SD: standard deviation

Population	AR ± SD	*H* ± SD	unique SNPs
FB	1.70 ± 0.18	0.21 ± 0.08	0
LP	1.70 ± 0.21	0.21 ± 0.08	0
ZO	1.70 ± 0.20	0.21 ± 0.08	0
PE	1.63 ± 0.21	0.18 ± 0.07	52
WA	1.64 ± 0.23	0.20 ± 0.09	0
LM	1.42 ± 0.19	0.12 ± 0.06	69

Table 3. Pairwise population F_{ST} values, averaged over all genes (lower diagonal), and associated p-values (upper diagonal). FB: False Bay; LM: Lake Mburo; LP: Limpopo; PE: Pemba; WA: Watamu; ZO: Zomba; *** $p < 0.001$

	FB	LP	ZO	PE	WA	LM
FB	-	***	***	***	***	***
LP	0.03	-	***	***	***	***
ZO	0.03	0.03	-	***	***	***
PE	0.30	0.29	0.30	-	***	***
WA	0.08	0.09	0.09	0.30	-	***
LM	0.50	0.49	0.50	0.56	0.49	-

measure as for AR. Again, LM shows the lowest diversity and is significantly different ($p < 0.0001$) from the other populations. WA and PE have slightly lower *H* values than FB, LP and ZO but these differences are very small and non-significant (Table 2). None of the populations FB, LP, ZO and WA have unique SNPs (SNPs that do not occur in other populations). In contrast, PE has 52 unique SNPs, while LM has the highest number with 69 unique SNPs (Table 2).

Pairwise population differentiation

We calculated average pairwise population F_{ST} values, from F_{ST} values for each gene based on the silent SNPs (Table 3). All pairwise F_{ST} values were significant ($p < 0.001$). Pairwise F_{ST} values are lower than 0.1 for any pairwise comparison within the populations FB, LP, ZO and WA. Comparisons between FB, LP and ZO give the smallest F_{ST} values (0.03), while WA shows slightly more differentiation from the first three populations (0.08-0.09). Pairwise F_{ST} values between PE and the other populations are relatively high. Differentiation between PE and the populations FB, LP, ZO and WA are very similar (0.29-0.30), while the pairwise F_{ST} value between PE and LM is the highest of

all comparisons (0.56). LM shows the most population differentiation with similar F_{ST} values for all pairwise comparisons (between 0.49 and 0.56).

Candidate genes

Table 4 (Appendix) gives the total sequenced length in base pairs (bp), the number of synonymous and non-synonymous SNPs and the ratio of synonymous and non-synonymous SNPs per 100 bp for each candidate gene. Sequenced lengths of genes varied between 200 and 1,645 bp, with an average of 774 bp. On average, there was 1 SNP for every 38 bp, corresponding to 2.6 bp per 100 bp. The majority of the identified SNPs were synonymous: we found nearly 20-fold as many synonymous SNPs as non-synonymous SNPs. Based on our SNP selection criteria (see Materials and Methods), further analyses of evidence of selection was focused on 14 replacement SNPs in 11 genes: *Treh*, *UGPase*, *TAGLipase*, *Vg*, *Hsp23*, *Hsp83*, *black*, *en*, *light*, *yellow*, and *wg* (Table 5).

Clinal variation

We present the results of a Pearson's correlation analysis on non-synonymous SNP allele frequencies against latitude. A summary is given in Table 5, which shows the F_{ST} value, Pearson's correlation coefficient (r) and associated p-value with significance indication per SNP for each gene. We found significant clinal variation for two out of three replacement SNPs in *Treh* (*Treh*^{335}, Fig. 2A and *Treh*^{550}, Fig. 2B), and for the only replacement SNP in *UGPase*

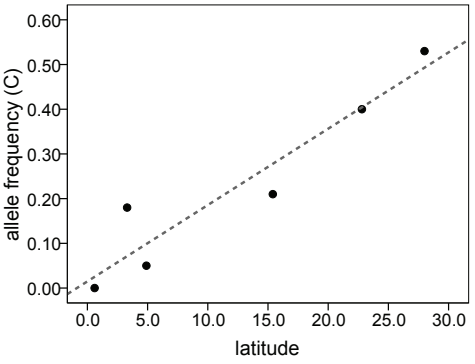

Figure 3. The latitudinal cline for SNP 408 of *UGPase*.

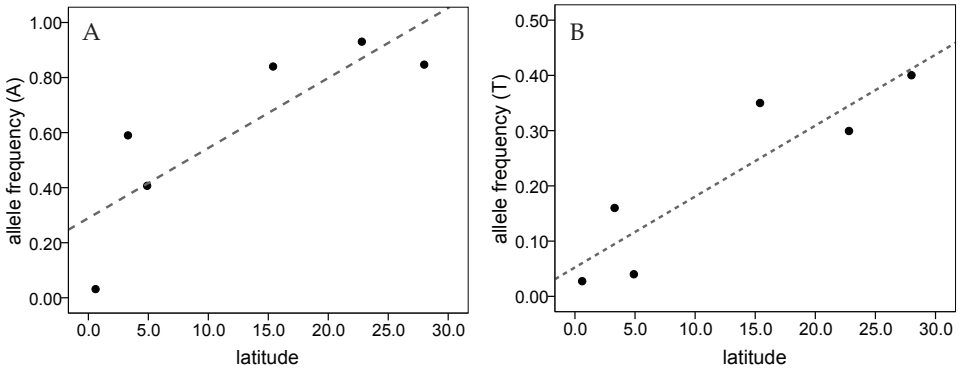

Figure 2. The latitudinal clines for (A) SNP 335 and (B) SNP 550 of *Treh*.

Table 5. Non-synonymous single nucleotide polymorphisms (SNPs) tested for significant clinal variation listed per gene, F_{ST} values, Pearson's correlation coefficient (r) and associated probabilities (p). Significant upper (+) or lower (-) F_{ST} outliers are indicated for the analysis including all populations (6) or populations FB, LP, ZO and WA (4). p-values that remained significant after False Discovery Rate (FDR) correction are indicated with *.

Gene name	SNP	F_{ST}	r	p	Outlier
Treh	335	0.451	0.840	0.036	
	549	0.536	0.388	0.116	
	550	0.137	0.905	0.013	
UGPase	248	0.306	0.966	0.002*	+ (4)
TAG Lipase I	424	0.133	0.078	0.884	
TAG Lipase II	32	0.152	0.044	0.935	
	296	0.534	0.451	0.369	
Vg	452	0.248	0.737	0.095	
	614	0.618	0.622	0.187	+ (6)
Hsp23	647	0.530	0.322	0.533	
Hsp83	551	0.006	0.098	0.854	- (6)
black	135	0.190	0.718	0.108	
light	782	0.802	0.707	0.116	+ (6)
yellow	585	0.248	0.764	0.077	
	591	0.565	0.507	0.305	+ (4)
wg	341	0.048	0.107	0.840	

(*UGP*^{248}, Fig. 3). After multiple testing correction using Benjamini & Hochberg's (1995) False Discovery Rate, only the cline for *UGP*^{248} remained significant.

Outlier analysis

To identify outlier F_{ST} values in the replacement SNPs, we first carried out the Beaumont & Nichols F_{ST} outlier analysis including all silent and replacement SNPs of the candidate genes for the six populations. This revealed one replacement SNP with a significant lower outlier F_{ST}: *Hsp83*^{551} ($p < 0.001$), and two replacement loci with significant upper outlier F_{ST} values: *Vg*^{614} ($p < 0.01$) and *Light*^{782} ($p < 0.001$; Table 5). These loci did not show significant clinal variation (Table 5), and the high F_{ST} values of the two upper outliers were caused by unique SNPs for the populations PE and LM. Overall, the upper F_{ST} values in the analysis (of both silent and replacement SNPs) were mainly determined by the large number of unique SNPs in PE and WA. Because we were more interested in outlier loci among polymorphisms shared between populations, we decided to also perform the analysis including only the four populations FB, LP, ZO and WA. These populations did not have any unique SNPs, and extend over most

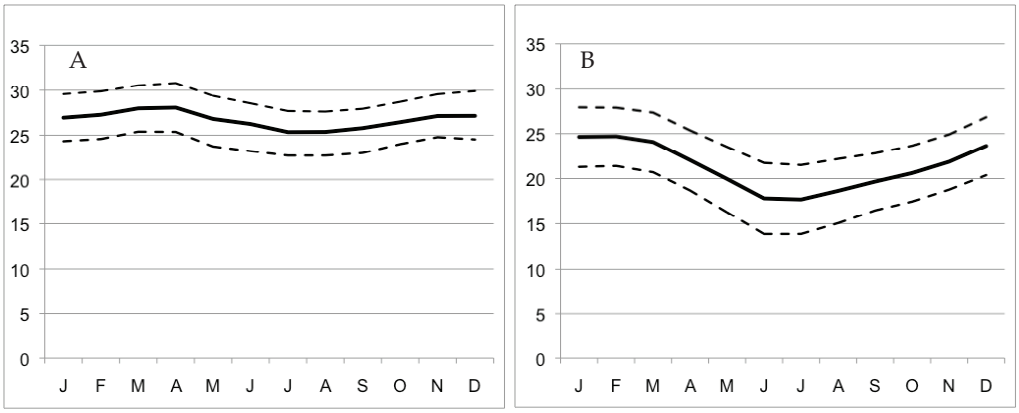

Figure 4. Average monthly temperatures for the localities of (A) the Kenyan WA population, and (B) the subtropical South African FB population. Solid line represents mean temperature in °C, and dashed lines represent average maximum and minimum temperatures. Climate data from the Global Historical Climate Network (GHCN).

of the sampled geographic area. This analysis resulted in two upper outlier F_{ST} values among the replacement SNPs: *Yellow*^{591} ($p < 0.05$) and *UGP*^{248} ($p < 0.05$), of which the latter also showed significant clinal variation (Table 5).

DISCUSSION

Climate

The distribution area of *B. anynana* spans a considerable temperature range with increasing overall temperatures and decreasing temperature amplitudes towards the equator. Fig. 4 shows climate charts for average minimum, mean and maximum temperatures for the Kenyan population near the equator (Fig. 4A) and the subtropical population from South Africa (Fig. 4B). Average monthly mean temperatures differ by as much as 8°C between these populations, and yearly temperature differences are considerably larger for South Africa than Kenya where temperatures remain fairly uniform throughout the year. Although there can be subtle deviations from a linear increase in temperature towards the equator, for example due to altitude differences, the regional scale shows a gradient in temperatures. The association of rainfall with latitude (not shown) is considerably more complex than temperature and does not show a clear gradient on a regional or local level.

Phylogeographic population structure

One potential problem with the interpretation of clinal variation is that the effects of spatially varying selection may be confounded with underlying phylogeographic patterns resulting from neutral evolution processes, such as drift and/or spatially

restricted gene flow (Gould & Johnston 1972, Endler 1977, Vasemägi 2006). To address this issue, we take into account the geographic population structure based on the mtDNA COI gene (De Jong et al. unpublished data; chapter 4 in this thesis), a marker widely used for inferring phylogenetic population structure (Avise 2000). This phylogeography indicated very little population differentiation for most populations (FB, LP, ZO and WA, further referred to as 'mainland' populations), while the island population PE and the population belonging to the different subspecies *B. anynana centralis* LM showed more differentiation. The study indicated that the low level of differentiation between the mainland populations is likely to be caused by recent population expansion from refugia during the last glacial maximum (De Jong et al. unpublished data; chapter 4 in this thesis). The most common haplotype, shared by the four mainland populations, did not show significant clinal variation in frequency with latitude.

Table 2 shows, for each population, gene diversity (expected heterozygosity) and allelic richness (average number of alleles), two statistics frequently used to measure genetic diversity. These indices (based on the nuclear genes) show a significantly lower genetic diversity for the LM population. PE also has a lower allelic richness and gene diversity than the mainland populations, but these differences are not significant. Both LM and PE have a substantial number of unique SNPs, as opposed to the mainland populations (Table 2), further indicating their relative isolation. The pairwise population differentiation based on the silent SNPs of the nuclear genes shows a similar pattern (Table 3). F_{ST} values among the four mainland populations are very small, indicating little differentiation, as opposed to the much higher values involving PE and particularly LM with these populations. The silent SNPs are likely to reflect the outcome of largely neutral evolution, although putative selection on these SNPs due to linkage/hitchhiking effects (Nielsen 2005) and other non-neutral processes (Chamary et al. 2006) cannot be excluded. An analysis with additional neutral markers could give a more conclusive perspective on the neutral population structure. To summarize, the mtDNA and silent SNPs indicate that the interpopulation structure does not show a clinal gradient. Furthermore, there is very little overall genetic divergence among the mainland populations, while the island population and especially the population of the *B. centralis* subspecies show much more differentiation.

Candidate genes

For this study, we sequenced coding sections of 20 candidate genes associated with temperature adaptation, and ten genes involved in wing pattern development. Despite conservative SNP calling, the genes were extremely polymorphic with on average 1 SNP per 38 bp. Previous studies on a single population of *B. anynana* also indicated a high level of polymorphism, with around 1 SNP per 50 bp (Beldade et al. 2006, 2009). This level of polymorphism is very high compared, for example, to humans with a reported SNP frequency in coding regions of 1 in 350bp (Cargill et al. 1999), and in the silkworm *Bombyx mori*, of 1 in 775bp (Cheng et al. 2004). On average, there were 15-fold more silent SNPs than replacement SNPs in *Bicyclus anynana*, indicating a generally strong balancing selection (Table 3). This is not surprising, considering the functional importance of these genes, which is also reflected in the relatively high level of

conservation at the protein level across species (e.g. in comparison with *Bombyx mori*). Because the silent SNPs are likely to mainly reflect the phylogeographic history of the populations, we only included the replacement SNPs for the analysis of footprints of selection.

Clinal variation in metabolic genes

The majority of evidence for adaptive clinal variation in coding polymorphisms results from studies on allozyme and candidate gene studies in metabolic genes in *Drosophila*. A well-known example is the parallel cline in the alcohol-dehydrogenase enzyme (Adh) in Australia, linked to latitudinal phenotypic variation in alcohol tolerance (Oakeshott et al. 1982). The recent advances in sequencing technology have sparked a renewed interest in studying clinal variation at the molecular level, resulting in an increase in the discovery of candidate genes displaying molecular clines. For example, Sezgin et al. (2004) reviewed and tested for clinal coding variation of metabolic enzymes in *D. melanogaster*, reporting on a total of nine genes displaying significant clines, and further reports have followed (Schmidt et al. 2008, Paaby et al. 2010).

Metabolic enzymes typically have limited thermal performance curves which are shaped by the ability to bind substrate (conformation) and the flexibility to change shape during catalysis (Hochachka & Somero 2002). The substitution of amino acids can alter the thermal properties of proteins (Fields 2001), and natural selection favours mutations that influence the conformational stability of enzymes (Marx et al. 2007). The more flexible an enzyme is, the faster it can change shape during catalysis, but this comes at the cost of a lower conformation stability. In general, higher temperatures favour greater conformational stability, while more flexible enzymes perform better at lower temperatures (Fields 2001, Hochachka & Somero 2002). Constraints on thermal performance of enzymes may be one of the most important factors determining the geographical distribution of ectotherms (Fields 2001).

In the present study, we identified significant clinal variation in replacement SNPs of two candidate genes involved in metabolic pathways: *UGPase* and *Treh*. These genes are both key enzymes in carbohydrate metabolism and widely found in plants, animals and microorganisms (Kleczkowski, 2004). Two of the three replacement SNPs in the approximately 1,000 bp sequence for *Treh* showed a significant correlation between allele frequency and latitude (Table 5; Fig. 2). The cline was very steep for $Treh^{335}$, for which the allele frequencies ranged from nearly zero to nearly one (Fig. 2A). The minor allele frequency of $Treh^{550}$ increased from nearly zero to 40% towards the south (Fig. 2B). In addition, there was also significant clinal variation for eight silent SNPs, distributed over the entire length of the fragment (data not shown). This is likely to be caused by linkage of the silent SNPs with the amino acid polymorphisms, and may reflect clinal variation in a common haplotype. *Treh* catalyzes the conversion of trehalose, an important storage carbohydrate, to glucose. In insects, *Treh* plays a crucial role in various physiological processes, including flight metabolism (Clegg & Evans 1961), and stress responses, including hypoxia (Chen & Haddad 2004), dessication (Worland et al. 1998, Timmermans et al. 2009), and thermal stress (Friedman 1978, Worland et al. 1998).

For *UGPase*, the single replacement SNP on the 500 bp sequence showed a significant cline (Fig. 3), as did two silent SNPs nearby (data not shown). *UGPase* catalyzes the reversible formation of UDP-glucose, an important step in the synthesis of glycogen, which is, like trehalose, an important storage carbohydrate (Alonso et al. 1995). The outlier analysis including the four mainland populations revealed the amino acid polymorphism in *UGPase* as an upper outlier locus. Relative to the overall low F_{ST} between these populations, *UGPase* had a much higher F_{ST} value, indicating directional selection on this locus. Crucially, for both *Treh* and *UGPase*, clines in the coding regions of the genes have also been reported for *D. melanogaster* in North America (Sezgin et al. 2004), suggesting a potential role for these enzymes in thermal adaptation across taxa. In the *Drosophila* study, an amino acid polymorphism in *Treh* and a silent SNP in *UGPase* show significant clinal variation (Sezgin 2004).

After correction for multiple testing, only the cline for the amino-acid replacement in *UGPase* remained significant. Here, we chose for an explorative design by including a large number of genes to screen for evidence of selection. This approach increases the chance of finding significant clines in individual genes, but comes at a cost of reduced statistical power due to multiple testing. Because only six populations were sampled the correlation coefficient needs to be very high in order to reach a high level of significance to withstand the stringent FDR approach. Although the clinal patterns found in *Treh* were not significant after multiple testing correction, we believe this remains an interesting candidate gene for follow-up studies, due to the clinal signal for multiple SNPs in the sequenced fragment and the reported clinal variation in *Drosophila* (Sezgin et al. 2004).

Linkage between the genes could be a potential cause of a shared pattern of clinal variation. Although we have no linkage information for the genes in this study, we do know the positions of their orthologs on the *B. mori* genome, where *UGPase* and *Treh* are each located on different chromosomes. A recently published gene-based linkage map for *Bicyclus anynana* revealed a generally strong conservation of gene assignments to chromosomes (Beldade et al. 2009).

Pigmentation genes

We included several genes involved in pigmentation biosynthesis in the analysis because of widely reported clinal variation in pigmentation across taxa (Ellers & Boggs 2002, Wittkopp et al. 2003, Hoekstra et al. 2006, Karl et al. 2009). However, we did not find significant clines for the replacement SNPs in the regions we sequenced. Interestingly, one SNP in the *yellow* pigmentation gene was an upper outlier among the mainland populations, and one SNP in the *light* pigmentation gene was an upper outlier among all six populations, indicating an increased population differentiation for these loci (Table 5). The *B. anynana centralis* subspecies has been described on the basis of differences in wing pattern from the *B. anynana anynana* subspecies (Condamin 1973). Moreover, De Jong et al. (2010) found significant population differentiation in reaction norms on rearing temperature for wing pattern between the South African FB population and a population from Malawi. Thus, it is possible that these pigmentation genes are involved in local adaptation for wing pattern in *B. anynana*.

In addition to thermal adaptation in the form of UV protection and regulation of body temperature by melanization, several adaptive explanations have been put forward to explain patterns of geographic variation in pigmentation. These include crypsis, deflection of predators, and mate choice and species recognition (Lyytinen et al. 2004, Oliver et al. 2009). In *B. anynana*, sexual selection is likely to be an important selective force shaping wing pattern, and hence may play an important role in driving population differentiation and eventually speciation (Oliver et al. 2009). Also, wing pattern is likely to be involved in crypsis and predatory deflection in this species (Lyytinen 2004). Ongoing and future research may reveal which selective forces are driving population differentiation for wing pattern in *B. anynana*.

Heat shock proteins and developmental genes

Conform our expectations, we did not find significant clinal variation in the developmental genes and the heat shock family genes. The selected developmental genes play a crucial role during embryogenesis and throughout development in *B. anynana* and other organisms (Beldade 2002), and these genes are not known to be involved in thermal adaptation. The heat shock proteins had the fewest amino-acid replacements compared to the other genes: only one in *Hsp23* and one in *Hsp83*. The one replacement SNP in *Hsp83* was a lower outlier locus (in all six populations), indicating balancing selection (Beaumont & Nichols 1996). These results are not surprising since heat shock proteins are generally highly conserved, even across taxa. In addition to their established function in (thermal) stress responses, heat shock proteins are important housekeeping genes in cellular regulation; they function as molecular chaperones and are involved in folding and transportation of other proteins. The majority of studies indicate an upregulation of the expression of heat shock proteins in response to thermal stress and other stress responses (e.g. Fangue et al. 2006, Chen et al. 2007, Rinehart et al. 2007), which has been linked to genetic variation in the regulatory regions (Chen et al. 2007). There are reports on clinal variation in the coding regions (e.g. Frydenberg et al. 2003, Hemmer-Hansen et al. 2007), although these are rare.

CONCLUSIONS

In this study, we found significant clinal variation in amino-acid polymorphisms for the metabolic enzymes *UGPase* and *Treh*. In addition, the amino-acid polymorphism in *UGPase* was an outlier loci compared to the overall F_{ST} in four populations, indicating that this locus is under selection. For these genes, our data strongly suggest adaptive population divergence along a latitudinal gradient and imply local adaptation against a background of generally low population divergence, as indicated by mtDNA and silent SNPs. Our results are paralleled by reports on clinal variation in *UGPase* and *Treh* in *D. melanogaster*. Moreover, as expected, we found no evidence of clines with latitude in the heat shock proteins and developmental genes. Taken together, our findings indicate a putative role in thermal adaptation for the genes *UGPase* and *Treh*, which are, therefore, interesting candidates for follow-up studies linking variation in phenotypic traits to molecular variation within and among populations.

Acknowledgements: We are grateful to Gavin Cohen, John Wilson, André Coetzer and Freerk Molleman for assistance in the field, to Marleen van Eijk for the molecular laboratory work, and to Jeroen Pijpe and Peter de Knijff for help with processing and analysing the data. This work was funded by the Earth and Life Sciences programme of the Netherlands Organization for Scientific Research (Grant no. 814.01.012), and additional grants for fieldwork from the Leiden University Fund, the Uyttenboogaart-Eliasen Foundation, the Treub Foundation and the Royal Netherlands Academy of Arts and Sciences (KNAW).

REFERENCES

Alonso MD, Lomako J, Lomako WM, Whelan WJ (1995) A new look at the biogenesis of glycogen. *The FASEB Journal* **9**(12): 1126-1137

Angilletta MJ (2009) *Thermal adaptation: a theoretical and empirical synthesis*. Oxford University Press, Oxford

Antao T, Lopes A, Lopes RJ, Beja-Pereira A, Luikart G (2008) LOSITAN: a workbench to detect molecular adaptation based on a F_{ST}-outlier method. *BMC Bioinformatics* **9**: 323. doi:10.1186/1471-2105-9-323

Avise JC (2000) *Phylogeography: the history and formation of species*. Harvard University Press, Cambridge, MA

Banks MA, Eichert W, Olsen JB (2003) Which genetic loci have greater population assignment power? *Bioinformatics* **19**(11): 1436-1438. doi:10.1093/bioinformatics/btg172

Beaumont MA, Nichols RA (1996) Evaluating loci for use in the genetic analysis of population structure. *Proceedings of the Royal Society of London Series B - Biological Sciences* **263**(1377): 1619-1626. doi:10.1098/rspb.1996.0237

Beldade P, Brakefield PM, Long AD (2002) Contribution of *Distal-less* to quantitative variation in butterfly eyespots. *Nature* **415**(6869): 315-318. doi:10.1038/415315a

Beldade P, Rudd S, Gruber JD, Long AD (2006) A wing expressed sequence tag resource for *Bicyclus anynana* butterflies, an evo-devo model. *BMC Genomics* **7**: 130. doi:10.1186/1471-2164-7-130.

Beldade P, Saenko SV, Pul N, Long AD (2009) A gene-based linkage map for *Bicyclus anynana* butterflies allows for a comprehensive analysis of synteny with the lepidopteran reference genome. *PLoS Genetics* **5**(2): e1000366. doi:10.1371/journal.pgen.1000366

Benjamini Y, Hochberg Y (1995) Controlling the False Discovery Rate: a practial and powerful approach to multiple testing. *Journal of the Royal Statistical Society, Series B* **57**(1): 289-300

Brakefield PM, Beldade P, Zwaan BJ (2009) The African butterfly *Bicyclus anynana*: a model for evolutionary genetics and evolutionary developmental biology. In: Behringer RR, Johnson AD and Krumlauf RE (eds) *Emerging Model Organisms: A Laboratory Manual*. Cold Spring Harbour Laboratory Press, New York, p 291-329

Brakefield PM, Frankino WA (2009) Polyphenisms in Lepidoptera: multidisciplinary approaches to studies of evolution. In: Whitman DW, Ananthakrishnan TN (eds) *Phenotypic plasticity in insects. Mechanisms and Consequences*. Cambridge University Press, Cambridge, p 112-312

Campbell KL, Roberts JE, Watson LN, Stetefeld J, Sloan AM, Signore AV, Howatt JW, Tame JR, Rohland N, Shen TJ, Austin JJ, Hofreiter M, Ho C, Weber RE, Cooper A (2010) Substitutions in woolly mammoth hemoglobin confer biochemical properties adaptive for cold tolerance. *Nature Genetics* **42**(6): 536-540. doi:10.1038/ng.574

Cargill M, Altshuler D, Ireland J, Sklar P, Ardlie K, Patil N, Shaw N, Lane CR, Lim EP, Kalyanaraman N, Nemesh J, Ziaugra L, Friedland L, Rolfe A, Warrington J, Lipshutz R,

Daley GQ, Lander ES (1999) Characterization of single-nucleotide polymorphisms in coding regions of human genes. *Nature Genetics* **22**(3): 231-238. doi:10.1038/10290

Chamary JV, Parmley JL, Hurst LD (2006) Hearing silence: non-neutral evolution at synonymous sites in mammals. *Nature Reviews Genetics* **7**(2): 98-108. doi:10.1038/nrg1770

Chen B, Walser JC, Rodgers TH, Sobota RS, Burke MK, Rose MR, Feder ME (2007) Abundant, diverse, and consequential *P* elements segregate in promoters of small heat-shock genes in *Drosophila* populations. *Journal of Evolutionary Biology* **20**(5): 2056-2066. doi:10.1111/j.1420-9101.2007.01348.x

Chen QF, Haddad GG (2004) Role of trehalose phosphate synthase and trehalose during hypoxia: from flies to mammals. *Journal of Experimental Biology* **207**(18): 3125-3129. doi:10.1242/jeb.01133

Cheng TC, Xia QY, Qian JF, Liu C, Lin Y, Zha XF, Xiang ZH (2004) Mining single nucleotide polymorphisms from EST data of silkworm, *Bombyx mori*, inbred strain *Dazao*. *Insect Biochemistry and Molecular Biology* **34**(6): 523-530. doi:10.1016/j.ibmb.2004.02.004

Clegg JS, Evans DR (1961) Blood trehalose and flight metabolism in the Blowfly. *Science* **134**(3471): 54-55. doi:10.1126/science.134.3471.54

Condamin M (1973) Monographie du genre *Bicyclus* (Lepidoptera Satyridae). *Mémoires de l'Institute Fondamental d'Afrique Noire* **88**: 1-324

De Jong MA, Kesbeke FMNH, Brakefield PM, Zwaan BJ (2010) Geographic variation in thermal plasticity of life history and wing pattern in *Bicyclus anynana*. *Climate Research* **43**(1-2): 91-102. doi:10.3354/cr00881

Ellers J, Boggs CL (2002) The evolution of wing color in *Colias* butterflies: heritability, sex linkage, and population divergence. *Evolution* **56**(4): 836-840. doi:10.1111/j.0014-3820.2002.tb01394.x

Endler JA (1977) *Geographic variation, speciation, and clines*. Princeton Unversity Press, Princeton, NJ

Excoffier L, Lischer HEL (2010) Arlequin suite ver 3.5: A new series of programs to perform population genetics analyses under Linux and Windows. *Molecular Ecology Resources* **10**(3): 564-567. doi:10.1111/j.1755-0998.2010.02847.x

Fangue NA, Hofmeister M, Schulte PM (2006) Intraspecific variation in thermal tolerance and heat shock protein gene expression in common killifish, *Fundulus heteroclitus*. *Journal of Experimental Biology* **209**(15): 2859-2872. doi:10.1242/jeb.02260

Fields PA (2001) Review: protein function at thermal extremes: balancing stability and flexibility. *Comparative Biochemistry and Physiology - Part A: Molecular & Integrative Physiology* **129**(2-3): 417-431. doi:10.1016/S1095-6433(00)00359-7

Fischer K, Karl I (2010) Exploring plastic and genetic responses to temperature variation using copper butterflies. *Climate Research* **43**(1-2): 17-30. doi:10.3354/cr00892

Friedman S (1978) Trehalose regulation, one aspect of metabolic homeostasis. *Annual Review of Entomology* **23**: 389-407. doi:10.1146/annurev.en.23.010178.002133

Frydenberg J, Hoffmann AA, Loeschcke V (2003) DNA sequence variation and latitudinal associations in *hsp23*, *hsp26* and *hsp27* from natural populations of *Drosophila melanogaster*. *Molecular Ecology* **12**(8): 2025-2032. doi:10.1046/j.1365-294X.2002.01882.x

Goudet J (2002) *FSTAT: A program to estimate and test gene diversities and fixation indices*. University de Lausanne, Switzerland

Gould SJ, Johnston RF (1972) Geographic Variation. *Annual Review of Ecology and Systematics* **3**: 457-498. doi:10.1146/annurev.es.03.110172.002325

Gunn A (1998) The determination of larval phase coloration in the African armyworm, *Spodoptera exempta* and its consequences for thermoregulation and protection from UV light. *Entomologia Experimentalis et Applicata* **86**(2): 125-133. doi:10.1023/A:1003189628726

Hall D, Luquez V, Garcia VM, St Onge KR, Jansson S, Ingvarsson PK (2007) Adaptive population differentiation in phenology across a latitudinal gradient in European aspen (*Populus tremula*, L.): a comparison of neutral markers, candidate genes and phenotypic traits. *Evolution* **61**(12): 2849-2860. doi:10.1111/j.1558-5646.2007.00230.x

Hanski I, Saccheri I (2006) Molecular-level variation affects population growth in a butterfly metapopulation. *PLoS Biology* **4**(5): e129. doi:10.1371/journal.pbio.0040129

Hemmer-Hansen J, Nielsen EE, Frydenberg J, Loeschcke V (2007) Adaptive divergence in a high gene flow environment: *Hsc70* variation in the European flounder (*Platichthys flesus* L.). *Heredity* **99**(6): 592-600. doi:10.1038/sj.hdy.6801055

Hochachka PW, Somero GN (2002) *Biochemical adaptation: mechanism and process in physiological evolution*. Oxford University Press, Oxford

Hoekstra HE (2006) Genetics, development and evolution of adaptive pigmentation in vertebrates. *Heredity* **97**(3): 222-234. doi:10.1038/sj.hdy.6800861

Hoekstra HE, Hirschmann RJ, Bundey RA, Insel PA, Crossland JP (2006) A single amino acid mutation contributes to adaptive beach mouse color pattern. *Science* **313**(5783): 101-104. doi:10.1126/science.1126121

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**(2): 133-147. doi:10.1007/s10709-006-9010-z

Hoffmann AA, Willi Y (2008) Detecting genetic responses to environmental change. *Nature Reviews Genetics* **9**(6): 421-432. doi:10.1038/nrg2339

Karl I, Schmidtt T, Fischer K (2009) Genetic differentation between alpine and lowland populations of a butterfly is related to PGI enzyme genotype. *Ecography* **32**(3): 488-496. doi:10.1111/j.1600-0587.2008.05660.x

Kleczkowski LA, Geisler M, Ciereszko I, Johansson H (2004) UDP-glucose pyrophosphorylase. An old protein with new tricks. *Plant Physiology* **134**(3): 912-918. doi:10.1104/pp.103.036053

Li H, Ruan J, Durbin R (2008) Mapping short DNA sequencing reads and calling variants using mapping quality scores. *Genome Research* **18**(11): 1851-1858. doi:10.1101/gr.078212.108

Lyytinen A, Brakefield PM, Lindström L, Mappes J (2004) Does predation maintain eyespot plasticity in *Bicyclus anynana*? *Proceedings of the Royal Society of London Series B - Biological Sciences* **271**(1536): 279-283. doi:10.1098/rspb.2003.2571

Marx JC, Collins T, D'Amico S, Feller G, Gerday C (2007) Cold-adapted enzymes from marine Antarctic microorganisms. *Journal of Marine Biotechnology* **9**(3): 293-304. doi:10.1007/s10126-006-6103-8

Mitton JB, Duran KL (2004) Genetic variation in piñon pine, *Pinus edulis*, associated with summer precipitation. *Molecular Ecology* **13**(5): 1259-1264. doi:10.1111/j.1365-294X.2004.02122.x

Nielsen R (2005) Molecular signatures of natural selection. *Annual Review of Genetics* **39**: 197-218. doi:10.1146/annurev.genet.39.073003.112420

Oakeshott JG, Gobson JB, Anderson PR, Knibb WR, Anderson DG, Chambers GK (1982) Alcohol dehydrogenase and glycerol-3-phosphate dehydrogenase clines in *Drosophila melanogaster* on different continents. *Evolution* **36**(1): 86-96

Oliver JC, Robertson KA, Monteiro A (2009) Accommodating natural and sexual selection in butterfly wing pattern evolution. *Proceedings of the Royal Society of London Series B - Biological Sciences* **276**(1666): 2369-2375. doi:10.1098/rspb.2009.0182

Paaby AB, Blacket MJ, Hoffmann AA, Schmidt PS (2010) Identification of a candidate adaptive polymorphism for *Drosophila* life history by parallel independent clines on two continents. *Molecular Ecology* **19**(4): 760-774. doi:10.1111/j.1365-294X.2009.04508.x

Parmesan C (2003) Butterflies as bioindicators for climate change. In: Boggs CL, Watts WB, Ehrlich PR (eds) *Butterflies: ecology and evolution taking flight*. Chicago University Press, Chicago, p 541-560

Patarnello T, Battaglia B (1992) Glucosephosphate isomerase and fitness: Effects of temperature on genotype dependent mortality and enzyme activity in 2 species of the genus *Gammarus* (Crustacea, Amphipoda). *Evolution* **46**(5): 1568-1573

Rajpurohit S, Parkash R, Niwas SR, Nedved O, Singh S (2007) Parallel trend in pigmentation

and dessication tolerance: altitudinal and latitudinal effects in *Drosophila melanogaster*. *Drosophila Information Service* **90**: 70-79

Rank NE, Bruce DA, McMillan DM, Barclay C, Dahlhoff EP (2007) Phosphoglucose isomerase genotype affects running speed and heat shock protein expression after exposure to extreme temperatures in a montane willow beetle. *Journal of Experimental Biology* **210**(5): 750-764. doi:10.1242/jeb.02695

Rinehart JP, Li A, Yocum GD, Robich RM, Hayward SA, Denlinger DL (2007) Up-regulation of heat shock proteins is essential for cold survival during insect diapause. *Proceedings of the National Academy of Sciences* **104**(27): 11130-11137. doi:10.1073/pnas.0703538104

Robertson KA, Monteiro A (2005) Female *Bicyclus anynana* butterflies choose males on the basis of their dorsal UV-reflective eyespot pupils. *Proceedings of the Royal Society of London Series B - Biological Sciences* **272**(1572): 1541-1546. doi:10.1098/rspb.2005.3142

Saastamoinen M, Hanski I (2008) Genotypic and environmental effects on flight activity and oviposition in the Glanville fritillary butterfly. *The American Naturalist* **171**(6): 701-712. doi:10.1086/587531

Schmidt PS, Serrão EA, Pearson GA, Riginos C, Rawson PD, Hilbish TJ, Brawley SH, Trussell GC, Carrington E, Wethey DS, Grahame JW, Bonhomme F, Rand DM (2008) Ecological genetics in the North Atlantic: environmental gradients and adaptation at specific loci. *Ecology* **89**(11 Suppl): S91-107.

Sezgin E, Duvernell DD, Matzkin LM, Duan Y, Zhu CT, Verrelli BC, Eanes WF (2004) Single-locus latitudinal clines and their relationship to temperate adaptation in metabolic genes and derived alleles in *Drosophila melanogaster*. *Genetics* **168**(2): 923-931. doi:10.1534/genetics.104.027649

Sham P, Bader JS, Craig I, O'Donovan M, Owen M (2002) DNA Pooling: a tool for large-scale association studies. *Nature Reviews Genetics* **3**(11): 862-871. doi:10.1038/nrg930

Sturm RA, Teasdale RD, Box NF (2001) Human pigmentation genes: identification, structure and consequences of polymorphic variation. *Gene* **277**(1-2): 49-62. doi:10.1016/S0378-1119(01)00694-1

Timmermans MJTN, Roelofs D, Nota B, Ylstra B, Holmstrup M (2009) Sugar sweet springtails: on the transcriptional response of *Folsomia candida* (Collembola) to desiccation stress. *Insect Molecular Biology* **18**(6): 737-746. doi:10.1111/j.1365-2583.2009.00916.x

Vasemägi A (2006) The adaptive hypothesis of clinal variation revisited: single-locus clines as a result of spatially restricted gene flow. *Genetics* **173**(4): 2411-2414. doi:10.1534/genetics.106.059881

Weir BS, Cockerham CC (1984) Estimating *F*-statistics for the analysis of population structure. *Evolution* **38**(6): 1358-1370

Wittkopp PJ, Beldade P (2009) Development and evolution of insect pigmentation: genetic mechanisms and the potential consequences of pleiotropy. *Seminars in Cell and Developmental Biology* **20**(1): 65-71. doi:10.1016/j.semcdb.2008.10.002

Wittkopp PJ, Williams BL, Selegue JE, Carroll SB (2003) *Drosophila* pigmentation evolution: Divergent genotypes underlying convergent phenotypes. *Proceedings of the National Academy of Sciences* **100**(4):1808-1813. doi:10.1073/pnas.0336368100

Worland MR, Grubor-Lajsic G, Montiel PO (1998) Partial dessication induced by sub-zero temperatures as a component of the survival strategy of the Arctic collembolan *Onychiurus arcticus* (Tullberg). *Journal of Insect Physiology* **44**(3-4): 211-219. doi:10.1016/S0022-1910(97)00166-2

Table 1. List of genes included in the study and the primers used to amplify the sequenced fragments per gene. For each gene the full name, abbreviation, contig number in published EST-Derived gene collection and functional assignment is given. Per primer pair, the forward and reverse sequences are given, and the amplicon length.

Full name	Abbreviation	Contig	Functional assignment	Forward primer	Reverse primer	Amplicon length
glyceraldehyde 3 phosphatedehydrogenase 2	*Gapdh 2*	C326	glycolytic pathway	CCCGCTAACA TGTCGAAAAT	GATACCAGCA GCAGCATCAA	903
glutamatedehydrogenase	*Gdh*	C1411	glycolytic pathway	CCAACGGCC CTACAACAC	TGGTAATTGGA TTCACGTTCA	181
glutamatedehydrogenase	*Gdh*	C1412	glycolytic pathway	GAAATCCCAG ACAGGCTCAA	ACCTCCTTTT GTTGGGGTTC	300
glutamatedehydrogenase	*Gdh*	C1412	glycolytic pathway	CGATTCTCAAC TGATGTGACG	AGCCTTTCTT CGCCAGCTC	190
glycogenphosphorylase	*Glyp*	S7458	glycolytic pathway	GAACGATACC CATCCAGCTC	TTCACTTGCA CGTCGAACAT	733
glycogenphosphorylase	*Glyp*	S6487	glycolytic pathway	TGAGCGTCCA AACAGATTCA	CGCTTGGGGT CTTTCTCATA	241
riosephosphateisomerase	*Tpi*	C7836	glycolytic pathway	GAAGCTGGGC TCAAAGTGAT	TGGCAGTCTT TCCAGTACCA	181
riosephosphateisomerase	*Tpi*	C7836	glycolytic pathway	ACATCCACG CGTCCCTTC	ACTCCGGTTT AAGGCTTGCT	171
riosephosphateisomerase	*Treh*	C6804	glycolytic pathway	ATGAAACGGT GAAGGGAGTG	ACTTCACGAG CCCTTAGCAA	931
UTP-glucose-1-phosphate uridylyltransferase	*UGPase*	C6877	glycolytic pathway	TGCAATCAGG AGAGTTGTGG	AAAGCCACTT TTCTGCCAAA	408
apolphirinprecursor	*Apolphirin precursor*	C7601	lipid pathway	CTCCGTGGTC ACCTTCTTTT	GAGAATGAGAA AGGCAAAGACA	530
apolphirinprecursor	*Apolphirin precursor*	C7015	lipid pathway	GGGACAACTCTC CTAAATCTATCC	TCAGTGACCTC AACAGTGACAA	208

Table 1 (continued).

Full name	Abbreviation	Contig	Functional assignment	Forward primer	Reverse primer	Amplicon length
desaturase 1	*desat1*	C7463	lipid pathway	AGGGGTCGTG GGTTCTGTAT	CCGTAAATATTCG TGTAGTCCACAT	198
desaturase 1	*desat1*	C7463	lipid pathway	GTACTCGATTG GGCACGAGA	AATCGCAGCA CAGGATCAG	191
lipaselike	*Lipaselike*	C5275	lipid pathway	TGGACCTTCA GTTCAATTCG	GACGTGTTTC GTGCCAGTC	890
lipaselike	*Lipaselike*	S4049	lipid pathway	AAGCCTTGGA GCTCACGTT	ATCATCCGCA TTGATTTTCC	556
lipophorinreceptor	*LpR*	S5824	lipid pathway	ATGTCGTTCGG ACGAGTTCA	CACTTCCATCA GGGCAGTCT	212
lipophorinreceptor	*LpR*	S5824	lipid pathway	CTCCGCCGAA GGTTATATCG	GACTTCATCA CTGCCGTCTG	251
lipophorinreceptor	*LpR*	C7081	lipid pathway	GACCCTCGG GACAGAACC	TCTTCTATGACA TCGCTCCAAA	194
lipophorinreceptor	*LpR*	C7081	lipid pathway	ACCAATCGAT GAAGGCAGTC	ATTCAAGGCT ATCGCACGAG	205
triacylglycerollipase	*TAG lipase*	S5789	lipid pathway	GGTTGGACCT TCGAAATGAA	GTGCCACGAG AAGTTCCAAT	618
triacylglycerollipase	*TAG lipase*	C2122	lipid pathway	CGTACGGACC AGTGCAGATA	TCCCACCCTT TGTTATCCAT	780
vitellogenin	*Vg*	C7110	lipid pathway	CATGCTGAAAAT CAATTATGACG	AAACCGCCGT TAGAGAAACA	866
catalase	*Cat*	C1371	anti oxidant	TCACCACCAA ATATGGAGCA	GTGAGAACCA TAGCCGTTCA	575
black	*black*	C4285	pigment synthesis	AAGTCTCACGG AGAGATCGAA	CTGCCATAAA CGCCAGAAG	200

Table 1 (continued).

Full name	Abbreviation	Contig	Functional assignment	Forward primer	Reverse primer	Amplicon length
black	*black*	C4285	pigment synthesis	CGCACCCTTA CTTCGTCAAC	TGGATAAAAGT AGGACCGAGCA	260
black	*black*	C4284	pigment synthesis	CAAAAAGACAAAT TTTATGATACATCC	ATATCCGATC CGAGCCTTTC	470
catecholaminesup	*Catsup*	C235	pigment synthesis	ACGTCACGTC AAGGTGAATG	AGGAGGGCGA GAATTTCTTT	1158
dopadecarboxylase	*ddc*	C3202	pigment synthesis	GACAAGTGGT GCCATCAGTG	GCTCCGCTAA GCATATCAGC	211
dopadecarboxylase	*ddc*	C3202	pigment synthesis	CTAGTCCCGC TTGCACTGA	TCCCACAAGT TTGGAAAGAA	236
dopadecarboxylase	*ddc*	C3202	pigment synthesis	GGCGACACTA GGAACGACTT	GGAGCCACA TAGCGGAAC	230
light	*light*	C3144	pigment synthesis	CTGTGAATGC GAAGGAGATG	TTGGGGTCCA CCTTCAAGTA	927
yellow	*yellow*	C4163	pigment synthesis	CGAATGTTTGT CCTTACTCCA	GGGCTTGTCG GTATTATTGG	817
heat shock protein 23	*Hsp23*	C4926	heat shock protein	TACCACACGA GAGCAACCAC	ATTCACTCGT CGGTTTCGAT	786
heat shock protein 60	*Hsp60*	C3306	heat shock protein	TAACCTACAA CCGGCGCTAA	ATGAGTTCGA GCTCGTCGTT	757
heat shock protein 60	*Hsp60*	C3307	heat shock protein	GGTAGTGAAC CGGCTGAAGA	TGGGGTTAGG CTCCTTTTCT	815
heat shock protein 68	*Hsp68*	C4558	heat shock protein	GAGCGAGGTT CGAGGAATTA	CCCTACATTCATA CTAGCGAATAAA	1088
heat shock protein 68	*Hsp68*	C4559	heat shock protein	CAAAATGCCA GCTATTGGAA	GCTTCGGTACT CGACGAAAG	837

Table 1 (continued).

Full name	Abbreviation	Contig	Functional assignment	Forward primer	Reverse primer	Amplicon length
heat shock protein 83	*Hsp83*	C1093	heat shock protein	CTTCGAGGAA CTTGCAGAGG	CCAGCGCTAG TTTCGACTTC	870
heat shock protein 83	*Hsp83*	C2544	heat shock protein	AGACGCTCTG GACAAGATCC	ATGGGATAGC CGATGAACTG	495
heat shock protein 83	*Hsp83*	S6762	heat shock protein	ATGACATCAC CCAGGAGGAG	GGCAAGTCCT CGCTGTCTAC	280
heat shock cognate 70-3	*Hsc70-3*	C3217	heat shock cognate	GGGTGTATACA AGAATGGACGA	GACCATGGCT GAGACCTCTT	319
heat shock cognate 70-3	*Hsc70-3*	C3218	heat shock cognate	ACTTTCGACG TGTCGCTTTT	TGGACTCAAC TCGCTCCTTT	1000
heat shock cognate 70-4	*Hsc70-4*	C880	heat shock cognate	AAGCACCTGC GGTAGGTATC	TGTGTTGTTG GGGTTCATTG	191
heat shock cognate 70-4	*Hsc70-4*	C880	heat shock cognate	CGTGCAAGCT GACATGAAAC	GAACTCCTGC ACGAAGTGGT	490
adenomatouspolyposis coli tumor suppressor	*Apc*	S782	developmental pathway	ATTGGAAAGT GCAGCCAGAT	CCACACTCTC GCTTCTCTCC	432
distal-less	*dll*	C3	developmental pathway	ACGAATTCAG TCCCCAAACA	TTGGTGCATAG GAGGAAAGG	375
engrailed	*en*	not published	developmental pathway	GAGCGCATCG TGTTGTTAAA	GAAGGGTATG GTTGGTGGTG	463
ovo	*ovo*	C2907	developmental pathway	CTTAGGCCCC TCGACTTAAA	GTTCTGCTGGA AGACCCAAA	842
wingless	*wg*	S2	developmental pathway	GCCTGTAAAA CGCTCCATCT	CCGAAGTTGG AACCTGTAGC	638

Table 4. List of candidate genes, length of the sequenced exon parts in base pairs (bp), number of non-synonymous (NS) single nucleotide polymorphisms (SNPs) per gene, number of synonymous (S) SNPs per gene, ratio of non-synonymous (NS) SNPs per 100 base pairs, and ratio of synonymous (S) SNPs per 100 base pairs.

Gene name	Length (bp)	# NS SNPs	# S SNPs	NS SNP per 100 bp	S SNP per 100 bp
Gapdh2	903	0	12	0	1.3
Gdh	671	0	12	0	1.79
GlyP	974	0	32	0	3.29
Tpi	352	0	5	0	1.42
Treh	931	3	30	0.32	3.22
UGPase	408	1	25	0.25	6.12
Apolphirin precursor	738	3	23	0.41	3.12
desat1	389	1	8	0.26	2.06
Lipase like	890	0	27	0	3.03
Lipase like	556	0	17	0	3.06
LpR	862	0	12	0	1.39
TAG Lipase	618	1	17	0.16	2.75
TAG Lipase	780	5	36	0.64	4.62
Vg	866	4	28	0.46	3.23
Cat	575	0	7	0	1.22
Hsc70-3	1319	0	35	0	2.65
Hsc70-4	681	0	12	0	1.76
Hsp23	786	1	12	0.13	1.53
Hsp60	1572	0	39	0	2.48
Hsp68	837	0	38	0	4.54
Hsp83	1645	1	45	0.08	3.59
Apc	432	0	11	0	2.55
black	930	1	19	0.11	2.04
Catsup	1158	0	13	0	1.12
ddc	677	1	13	0.15	1.92
dll	375	0	7	0	1.87
en	290	3	14	1.03	4.83
light	927	3	6	0.32	0.65
ovo	842	0	4	0	0.48
yellow	817	2	34	0.25	4.16
wg	200	1	3	0.50	1.50