Supporting Information

The genomic footprint of climate adaptation in *Chironomus riparius*

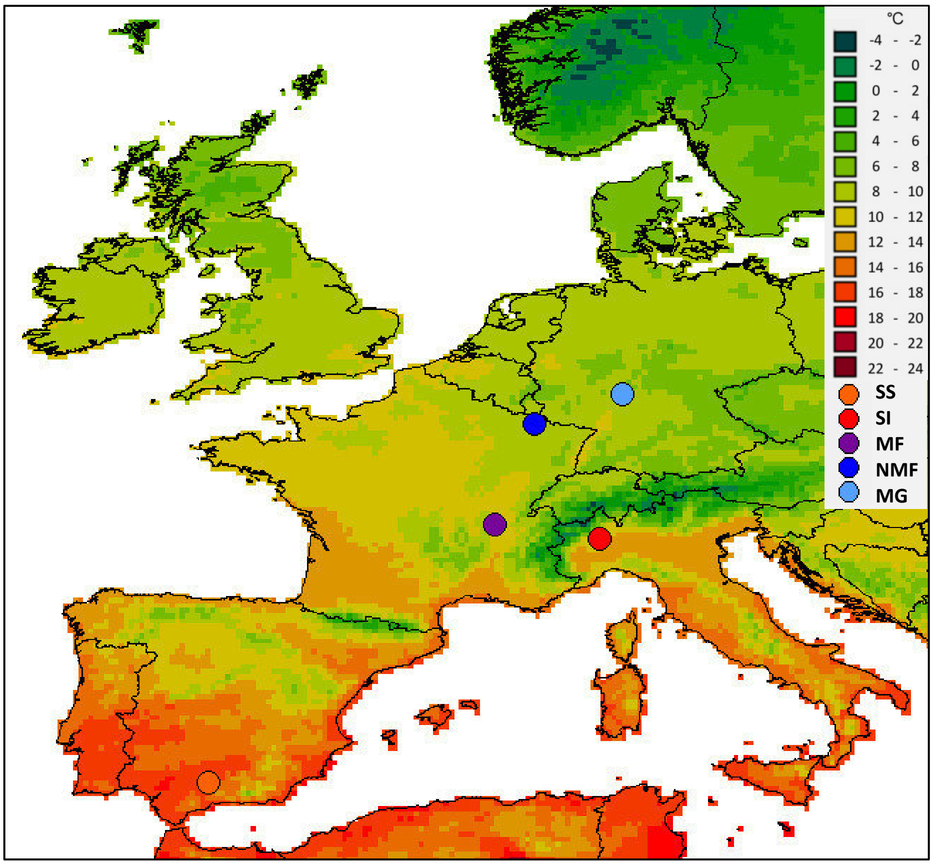
Content

1. Information about Pool-Seq data

1.4. Extended Material & Methods

Pool-Seq data was mapped with BWA using the *bwa mem* algorithm (v0.7.10-r789, Li & Durbin 2009). By increasing the minimum seed length to 30, we managed to obtain highest stringency (as recommended for Pool-Seq, Kofler *et al.* 2011a) while improving mapping success (Supporting Table S1.1) and drastically speeding up the analysis. The resulting bam-files were processed according to recommendations for the PoPoolation2 pipeline: sorting (Picard v1.119, available at http://picard.sourceforge.net), removal of duplicates (Picard), removing of low quality alignments (SAMtools utilities v1.1, Li *et al.* 2009), combining all Pool-Seq data sets to one overall *sync*-file, and subsampling the *sync*-file to a minimum coverage of 20X.

1.5. Extended Results



Supporting Figure S1.1: Geographic distribution of *C. riparius* populations sampled for this study along a climatic gradient across Europe. Climate variation is plotted as annual mean temperatures based on WorldClim climate data “bio1” (Hijmans *et al.* 2005). Population codes refer to Supporting Table S1.1 and are coloured in regard to their phenotypic temperature adaptation (warm to cold adaptation from orange to light blue, *cf.* Manuscript Figure 1).

Supporting Table S1.1:
Mapping statistics of Pool-Seq data to *C. riparius* draft genome (accession number) with *bwa mem* (-k 30).

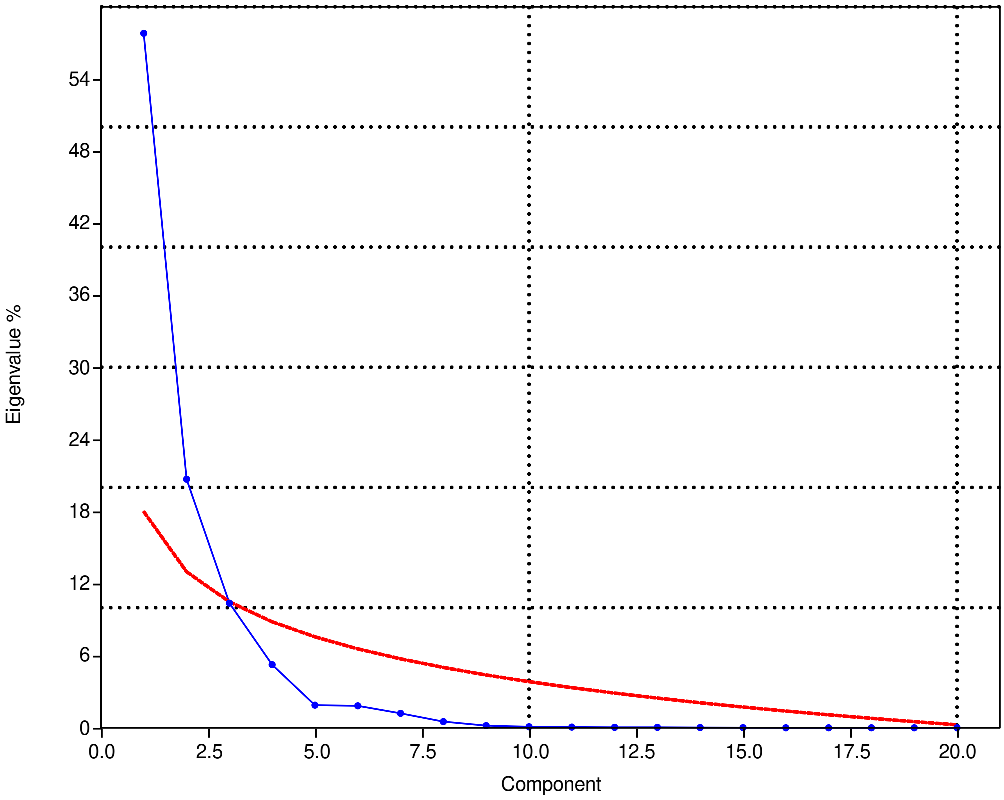
population	European region	coordinates (lat, long)	mean coverage	% mapped reads
MG	Hessen (GER)	50.1680610, 9.0819270	26.7	81.09
NMF	Lorraine (FRA)	49.1765430, 6.2156670	55.3	77.57
MF	Rhône-Alpes (FRA)	45.8616760, 4.8865000	41.0	78.03
SI	Piemont (IT)	45.4036180, 8.3473320	40.2	77.88
SS	Andalusia (SP)	37.399080, -4.5267980	36.9	80.92

2. Environmental association analysis with LFMM

2.1. Extended Material & Methods

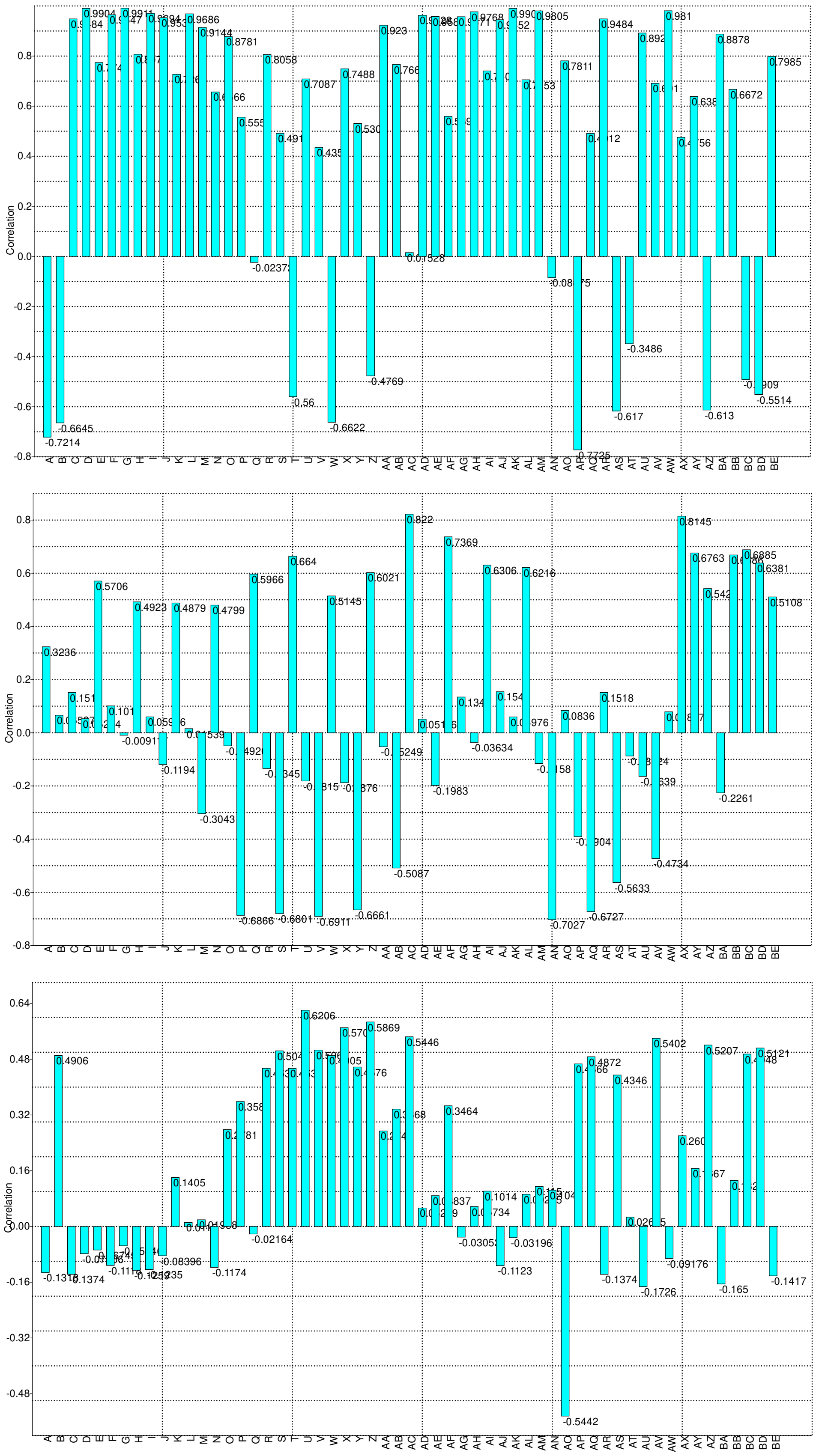
To generate the environmental input data, we extracted the complete set of current climate conditions for each sample location from WorldClim (Hijmans *et al.* 2005). To obtain meaningful, low dimensional environmental parameters, we performed a PCA (principal component analysis, software package PAST v. 3, Hammer *et al.* 2001) on all parameters (WorldClim data including BioClim data, approx. 1950-2000, Hijmans *et al.* 2005).

2.2. Extended Results



Supporting Figure S2.1:

Disribution of Eigenvalues (%) of components (blue line) after principal component analysis (PCA) with 57 climatic variables (WorldClim data) from 21 locations of documented *C. riparius* occurrence (including the five natural populations of this study, Oppold et al. 2016a). Red line marks the random distribution of Eigenvalues (broken stick analysis). Components under this curve are expected to be non-significant.



Supporting Figure S2.2:

PCA loading of the significant components: PC1 – cold temperatures (top), PC2 – precipitation (middle), PC3 – warm temperatures (bottom). See Supplementary Tab. S1 for a list of the climate variables.

Supporting Table S2.1:

Climate variables (WorldClim data) with highest PCA loadings of the three significant components.

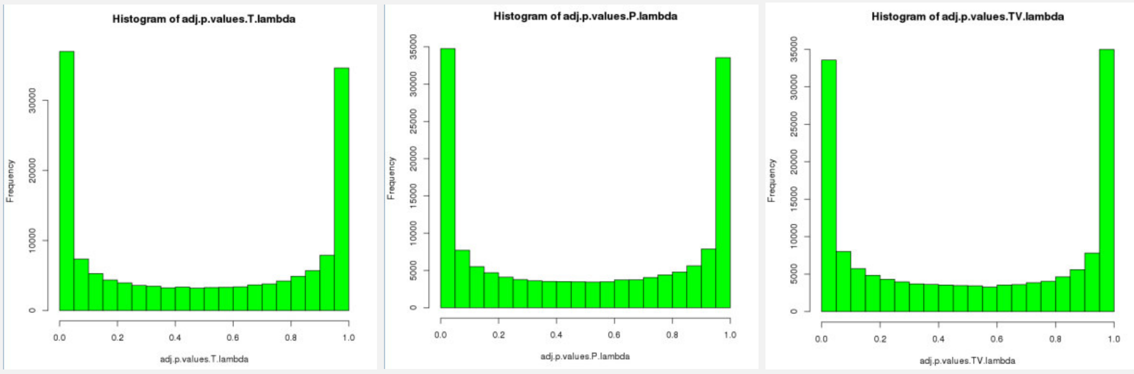
PC1	PC2	PC3
tmin1	prec9	tmin7
tmax1	bio12	tmin8
tmin2	prec10	prec8
tmax2		bio3
tmin3		bio10
tmax3		
tmin4		
tmax4		
tmin10		
tmax10		
tmin11		
tmax11		
tmin12		
tmax12		
bio1		
bio6		
bio11		

2.1 LFMM Workflow box

1. 149474 loci of 5 populations (extended to 20 individuals each), associated to 3 environmental variables
2. lfmm runs with 5 repetitions and emits a z.score per locus per env.variable, we take the median z.score of the separate lfmm runs
3. the genomic inflation factor lambda (should be close to 1) is estimated based on the z.scores with the chisq distribution as null model, this factor is meant to correct for non-neutral patterns that result from confounding factors (demography)
4. the z.scores are converted to p-values and corrected with lambda
5. fdr correction of adjusted p-values with Benjamini Hochberg algorithm

```
zs = z.scores(lfmm.extended, K=5, d=1)              #d=1 for first env.variable
zs.median = apply(zs, MARGIN=1, median)
lambda = median(zs..median^2)/qchisq(0.5, df=4)     #degrees of freedom = 4 (=n-1)
```

```
lambda.1 → 0.7076576                                #lambda “cold temperatures”
lambda.2 → 0.6743437                                #lambda “precipitation”
lambda.3 → 0.7067992                                #lambda “warm temperatures”
```



Box-Figure 2.1: Frequency distribution of adjusted p-values (adjusted with respective lambda) after association to three different environmental variables: warm temperatures, precipitation, cold temperatures (from left to right).

#these distribution are fine, since we enriched for fixed loci by extracting the 99% Fst-quantile (this explains the peak at 1), otherwise flat p-value distribution

#Benjamini-Hochberg correction with fdr level of 5%

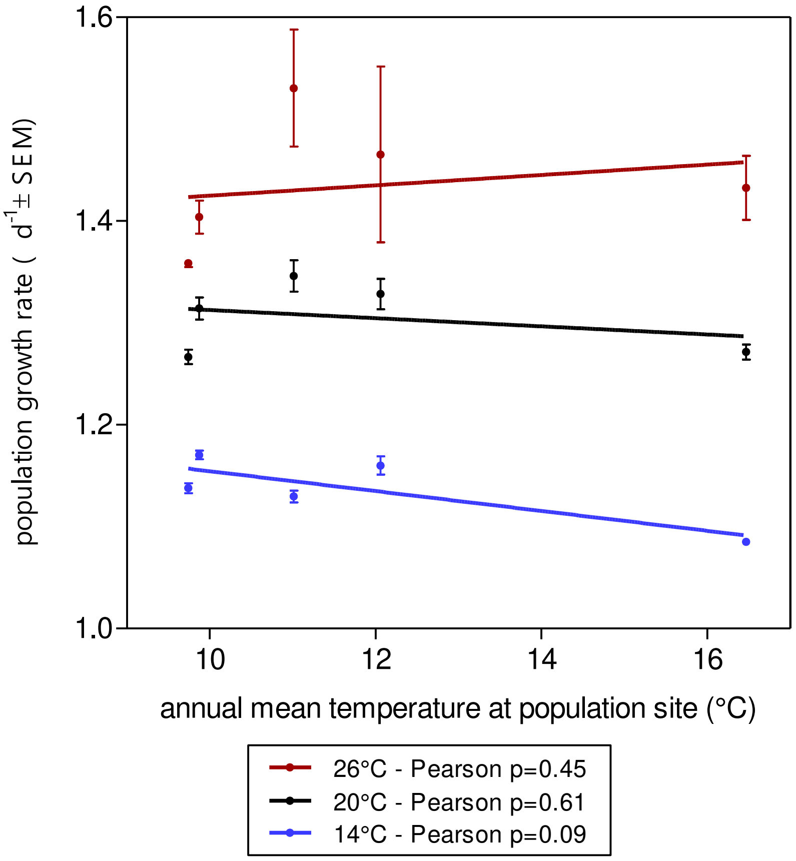
```
candidates.1 → 22959                   # ~23k loci associated to cold temperatures
candidates.2 → 19720                   # ~20k loci associated to precipitation
candidates.3 → 16956                   # ~17k loci associated to warm temperatures
```

3. Life-Cycle experiments

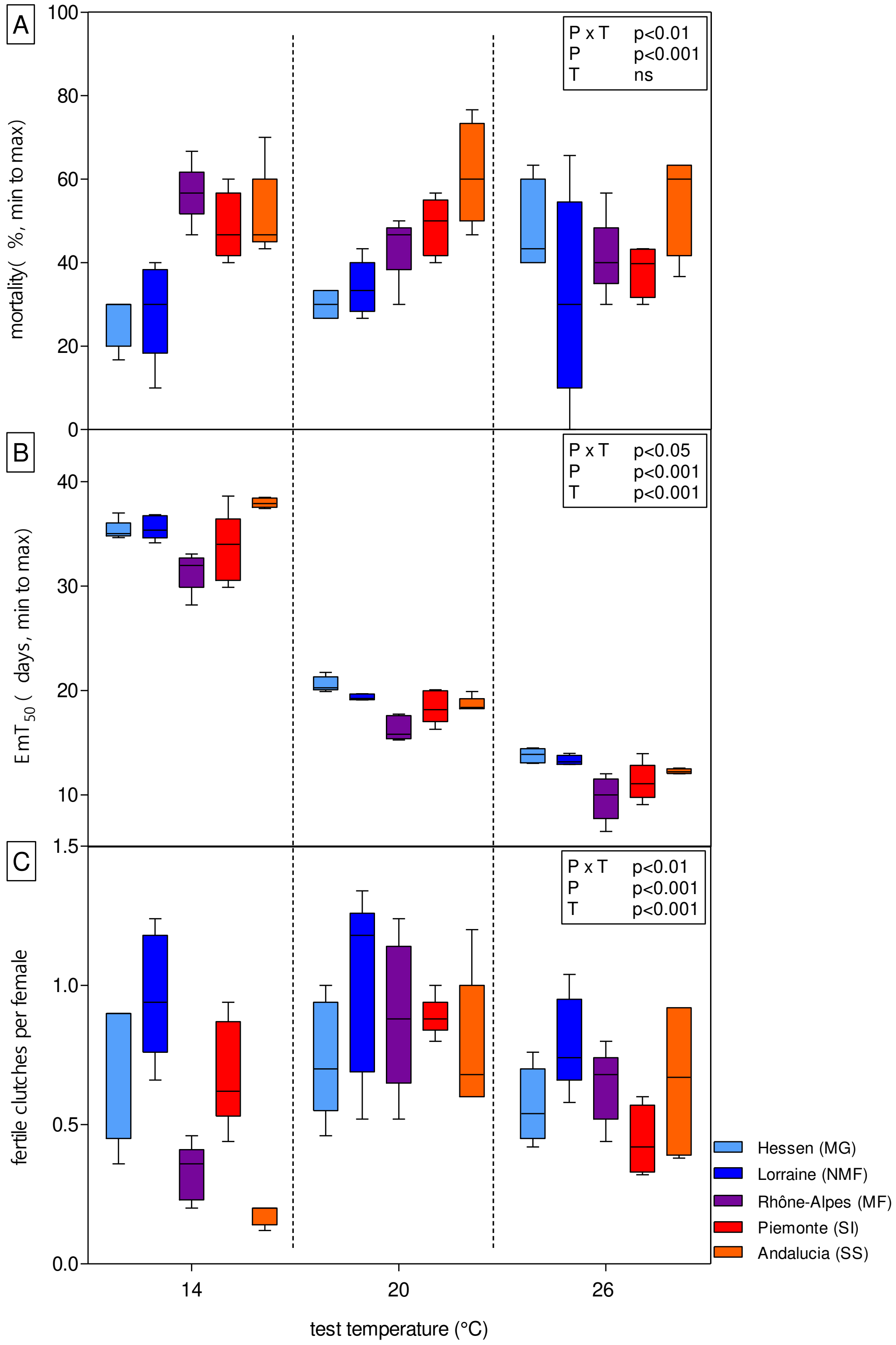
3.1. Extended Material & Methods

We recorded mortality, mean emergence time (EmT_{50}), sex ratio, number of clutches per female, number of eggs per clutch, and fertility of clutches (successful early embryonic development of at least half of the eggs per clutch) to finally calculate the population growth rate (PGR), as integrative fitness measure, based on a simplified Euler-Lotka calculation (Vogt *et al.* 2007b). Life-cycle parameters were analysed with two-way ANOVA to test for the effect of temperature, population, and the interaction of both in GraphPad Prism® (v5, GraphPad Software, San Diego, USA).

3.2. Extended Results



Supporting Figure S3.1:
Correlation of *C. riparius* population growth rates from life-cycle experiments at three different test temperatures (14, 20, 26 °C) with annual mean temperatures at respective population sites as proxy for the climate gradient across Europe. Population growth rates at 26 °C show large variation, especially in populations that experience strong seasonal temperature differences (Rhône-Alpes, Piemont), thus, correlation shows a weak tendency. Population growth rates at 14 °C show a strong tendency in correlation to annual mean temperatures.



Supporting Figure S3.2:
Life-cycle parameters of the natural *C. riparius* populations at the different test temperatures, shown as Box-Whiskers ranging from minimum to maximum: (A) mortality, (B) mean emergence time, (C) fertile clutches per female. P-value thresholds of two-way ANOVA shown in the box: effect of population (P), temperature (T), and interaction of both factors (P × T).

Supporting Table S3.3: Matrices with significance thresholds of two-way ANOVA with Bonferroni post-test for the different life-cycle parameters (A: mortality, B: mean emergence time, C: number of fertile clutches per female) at three test temperatures in five natural *C. riparius* populations. The population codes correspond to the legend in Supplemental Figure S1.

A:mortality	MG	NMF	MF	SI	SS
MG	**14°C**	ns	0.001	0.01	0.01
	20°C	ns	ns	0.05	0.001
	26°C	0.05	ns	ns	ns
NMF	ns	**14°C**	0.001	0.05	0.01
	ns	**20°C**	ns	ns	0.001
	0.05	**26°C**	ns	ns	0.01
MF	0.001	0.001	**14°C**	0.001	ns
	ns	ns	**20°C**	ns	0.05
	ns	ns	**26°C**	ns	ns
SI	0.01	0.05	0.001	**14°C**	ns
	0.05	ns	ns	**20°C**	ns
	ns	ns	ns	**26°C**	ns
SS	0.01	0.01	ns	ns	**14°C**
	0.001	0.001	0.05	ns	**20°C**
	ns	0.01	ns	ns	**26°C**
B:EmT50	MG	NMF	MF	SI	SS
MG	**14°C**	ns	0.001	ns	0.05
	20°C	ns	0.001	ns	ns
	26°C	ns	0.001	0.05	ns
NMF	ns	**14°C**	0.001	ns	0.05
	ns	**20°C**	0.01	ns	ns
	ns	**26°C**	0.001	ns	ns
MF	0.001	0.001	**14°C**	ns	0.001
	0.001	0.01	**20°C**	ns	0.05
	0.001	0.001	**26°C**	ns	0.05
SI	ns	ns	ns	**14°C**	0.001
	ns	ns	ns	**20°C**	ns
	0.05	ns	ns	**26°C**	ns
SS	0.05	0.05	0.001	0.001	**14°C**
	ns	ns	0.05	ns	**20°C**
	ns	ns	0.05	ns	**26°C**
C:clutches	MG	NMF	MF	SI	SS
MG	**14°C**	ns	0.05	ns	0.01
	20°C	ns	ns	ns	ns
	26°C	ns	ns	ns	ns
NMF	ns	**14°C**	0.001	ns	0.001
	ns	**20°C**	ns	ns	ns
	ns	**26°C**	ns	ns	ns
MF	0.05	0.001	**14°C**	0.05	ns
	ns	ns	**20°C**	ns	ns
	ns	ns	**26°C**	ns	ns
SI	ns	ns	0.05	**14°C**	0.01
	ns	ns	ns	**20°C**	ns
	ns	ns	ns	**26°C**	0.05
SS	0.01	0.001	ns	0.01	**14°C**
	ns	ns	ns	ns	**20°C**
	ns	ns	ns	0.05	**26°C**

4. MSMC analysis

4.1. Extended Material & Methods

Whole genome individual resequencing

DNA of adult midges was extracted using the DNeasy Blood & Tissue Kit (QIAGEN, Hilden, Germany). DNA concentration was measured with the Qubit® dsDNA BR Assay Kit in a Qubit® fluorometer and quality was assessed by gel-electrophoresis. As the total amount of DNA per individual was below 1 µg, preparation of 150 bp paired-end libraries was performed with the KAPA HyperPrep Kit (KR0961, KAPA Biosystems). Libraries were sequenced to an expected mean coverage of 25X on an Illumina HiSeq4000 (BGI sequencing facility, Hongkong).

Raw sequences were trimmed and clipped with TRIMMOMATIC (ILLUMINACLIP:adapters.fa:2:30:10:8 CROP:145 LEADING:10 TRAILING:10 SLIDINGWINDOW:4:20 MINLEN:50; v0.32, Bolger *et al.* 2014) and afterwards inspected with FASTQC (v0.11.2; http://www.bioinformatics.babraham.ac.uk/projects/fastqc/).

The phased data per scaffold of two populations and mappability mask of the respective scaffold were combined into one MSMC2 input per scaffold (comprising 2 populations x 4 individuals, i.e. 16 haplotypes). Therefore, ten alternative population-pairs were independently analysed concerning their respective cross coalescence.

Approach to decrease uncertainties in coalescence estimates

Since the script to generate the MSMC input (*generate_multihetsep.py*) cannot deal with missing data, there is a slight variation in the combination of sites between different pairs. This also slightly affects the N_e estimation of each population in a respective pair. To overcome this potential bias, the estimations were averaged per time index over the four independent runs per population.

Contrasting to studies with human or *Drosophila* genome sequences, there is no high-quality haplotype data available for *C. riparius*. It is hence not possible to estimate the actual phasing error in terms of the switch error rate (SER). To alternatively decrease uncertainties in the coalescence estimates, we only used time slices with a minimum number of ten coalescence events for downstream analyses. Since *C. riparius* and *Drosophila melanogaster* share similar genetic properties (µ, N_e (Oppold & Pfenninger 2017), chromosome number), we additionally used the conservative SER of 2.1 % from *Drosophila* (Bukowicki *et al.* 2016), corresponding to our sequencing coverage of approximately 20X in a data set with 20 unrelated individuals. To infer the expected mean haplotype length (MHL), genome-wide heterozygosity was estimated as an average of all individuals (number of diallelic SNPs per callable site of the genome). The MHL together with the *Drosophila* autosomal recombination rate of r = 2.1 cM Mb^{-1} (Mackay *et al.* 2012) enabled the calculation of an approximated time horizon (as time to the most recent common ancestor – tMRCA) below which phasing error precludes coalescence rate estimates:

$$tMRCA = \frac{1}{2 \cdot r \cdot MHL}$$

5. Simulation Study

5.1 Adjustment of Effective Population Size

To account for different generation times in our populations, we adjusted population sizes for recent epochs: $N_E^{adjust} = N_E \frac{G_a}{G_m}$,

where N_E^{adjust} is the adjusted population size, G_a is the number of generations per year and G_m is the mean number of Generations per year over all populations (Table S5.1, (Oppold et al. 2016)).

We refrained from adjusting population sizes in the distant past, as additional information on local climate and the spacial distribution of the population is not readily available (or even possible to obtain).

Table S5.1: Populations with generations per year

Population	Abbreviation	Generations per year	θ (Migrate Analysis)
Hessen (G)	MG	7.85	0.0316
Metz (F)	NMF	7.7	0.197
Lyon (F)	MF	9.07	0.396
Piemont (I)	SI	10.57	0.031
Andalucia (S)	SS	14.86	0.025
mean		10.01	

5.2 Models in Detail

General settings, consistent in all models:

Number of simulations per model: 200,000 Number of populations: 5
Number of samples: 20 per population
Length of sequence: 1,000 base pairs
Mutation rate per site and generation μ: 4.1×10^{-9}
Recombination rate: 0
Transition bias: 0.595

Simulations were performed using *fastsimcoal* v. 2.5.2 (Excoffier and Foll 2011).

Migration pathways

In all models migration between neighboring populations is allowed (Fig. S5.1 and Table S5.2).

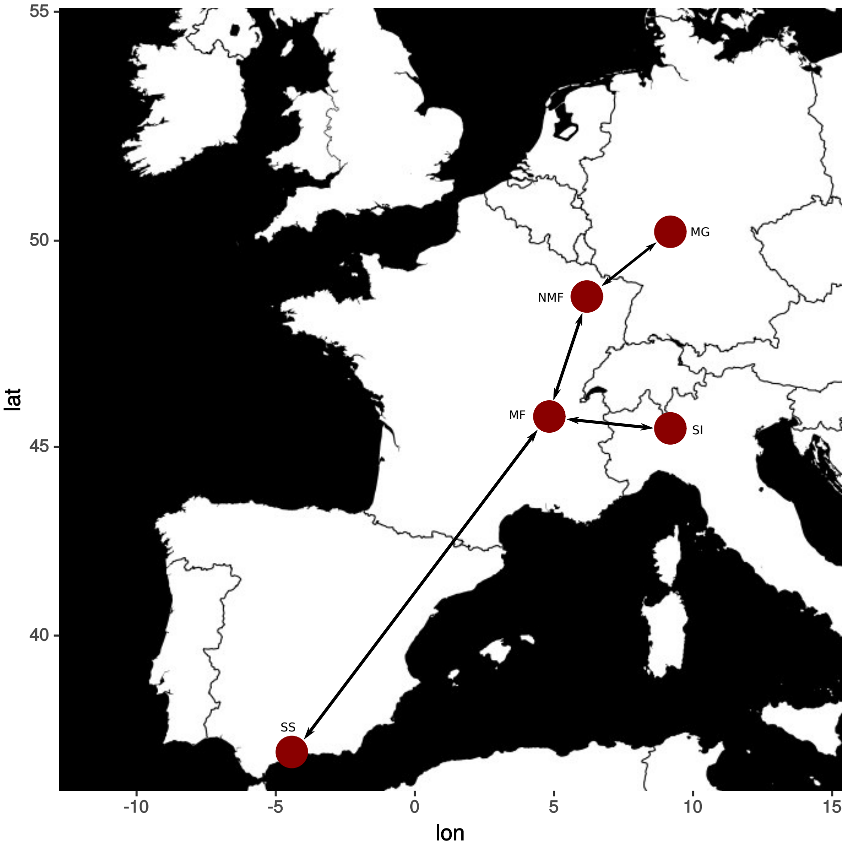


Figure S5.1: Locations of populations and possible migration routes between them (Kahle and Wickham 2013).

Table S5.2: Matrix of possible migration between neighboring populations.

	MG	NMF	MF	SI	SS
MG	0	possible	0	0	0
NMF	possible	0	possible	0	0
MF	0	possible	0	possible	possible
SI	0	0	possible	0	0
SS	0	0	possible	0	0

Constant Demography Model

As the simplest option we chose a population split model of constant population sizes and migration rates constant over time (Fig. S5.2, Table S5.3). Migration rates and population sizes are based on the results of the Migrate-n analysis (Beerli and Felsenstein 2001; Beerli 2006).

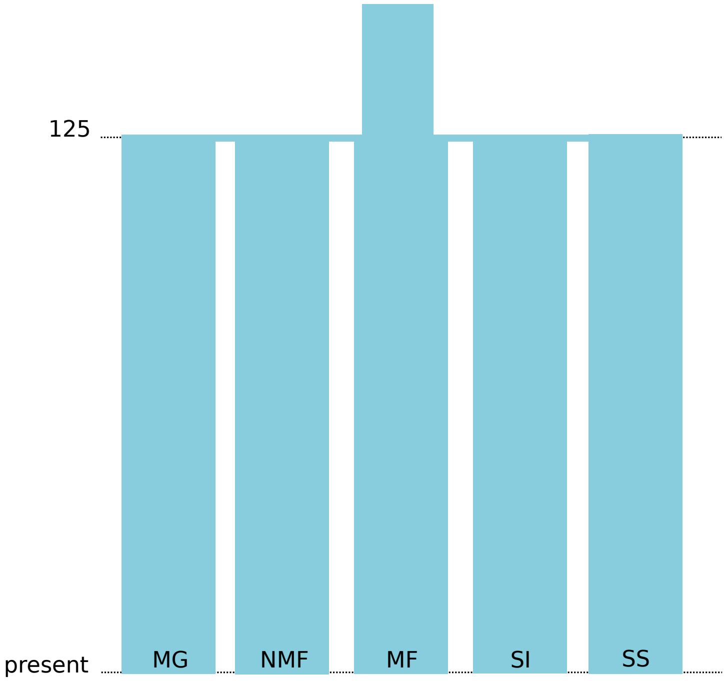


Figure S5.2: Graphical representation of demographic model: constant demography model (time scales in thousand generations).

Table S5.3: Population sizes in constant demography model. N_E^{adjust} and $N_E^{initial}$ give number of individuals at present and 125,000 generations ago.

Population	Abbreviation	N_E^{adjust}	$N_E^{initial}$
Hessen (G)	MG	1000857	0
Metz (F)	NMF	6120294	0
Lyon (F)	MF	14491648	28983
Piemont (I)	SI	1322063	0
Andalucia (S)	SS	1498905	0

Table S5.4: Migration matrix in constant demography model.

	MG	NMF	MF	SI	SS
MG	0	9.1×10^{-4}	0	0	0
NMF	9.1×10^{-4}	0	2.5×10^{-5}	0	0
MF	0	9.1×10^{-4}	0	3.0×10^{-5}	9.1×10^{-4}
SI	0	0	1.9×10^{-5}	0	0
SS	0	0	1.5×10^{-5}	0	0

Population Growth Model

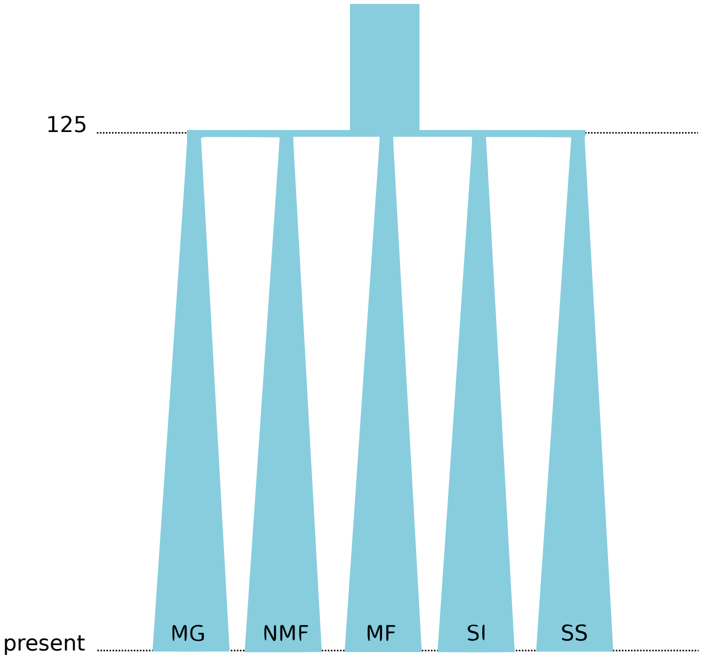


Figure S5.3: Graphical representation of demographic model: population growth model (time scales in thousand generations).

All parameters of this model are the same as in the Constant Demography Model, except for the addition of a population expansion (Fig. S5.3). The growth rate is $r = 1.0 \times 10^{-5}$ and population growth is given by:

$N_t = N_0 e^{rt}$,

where N_t equals population size in generation t and N_0 is the initial population size (Excoffier and Foll 2011).

Approximated Demographic Model

Based on the results of our MSMC2 analysis (Schiffels and Durbin 2014) we developed an approximated demographic model (Fig. S5.2 of main article) of population split, shrinkage and following expansion. Migration rates change over time, first decreasing to near isolation and then rising again, mirroring inferences on cross-coalescence rate from the MSMC2 analysis (Fig. S5.4, Table S5.4).

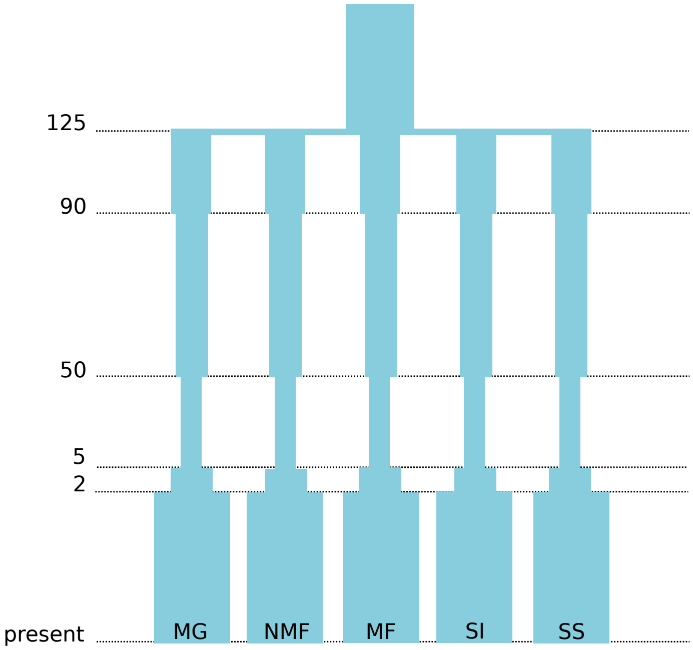


Figure S5.4: Graphical representation of demographic model: approximated population model (time scales in thousand generations).

Table S5.4: Population sizes at different time points (in generations) and migration rates (MIG) in these epochs.

	present	2000	5000	20000	50000	90000	125000
MG	282318	39211	18037	16000	20000	27000	0
NMF	615385	38462	16923	12000	19000	25000	0
MF	308072	47117	18122	13000	20000	26000	29000
SI	253427	26399	15839	18000	25000	30000	0
SS	504735	37113	14845	8000	15000	18000	0
MIG	1.02×10^{-5}	2.7×10^{-4}	3.0×10^{-4}	3.7×10^{-3}	5.1×10^{-3}	2.9×10^{-3}	0

5.3 Calculation of F_{ST} values

Pairwise F_{ST} values are used to detect short term genetic distances between populations (Excoffier and Lischer 2010; Reynolds, Weir, and Cockerham 1983; Slatkin 1995). We calculated these for all models as well as the empirical data, generated density functions and compared them (Fig. S5.5). Computation was performed with arlsumstat, the command-line version of Arlequin 3.5 (Excoffier and Lischer 2010).

Kruskal-Wallis tests showed significant differences in all pairings (Hollander, Wolfe, and Chicken 2013).

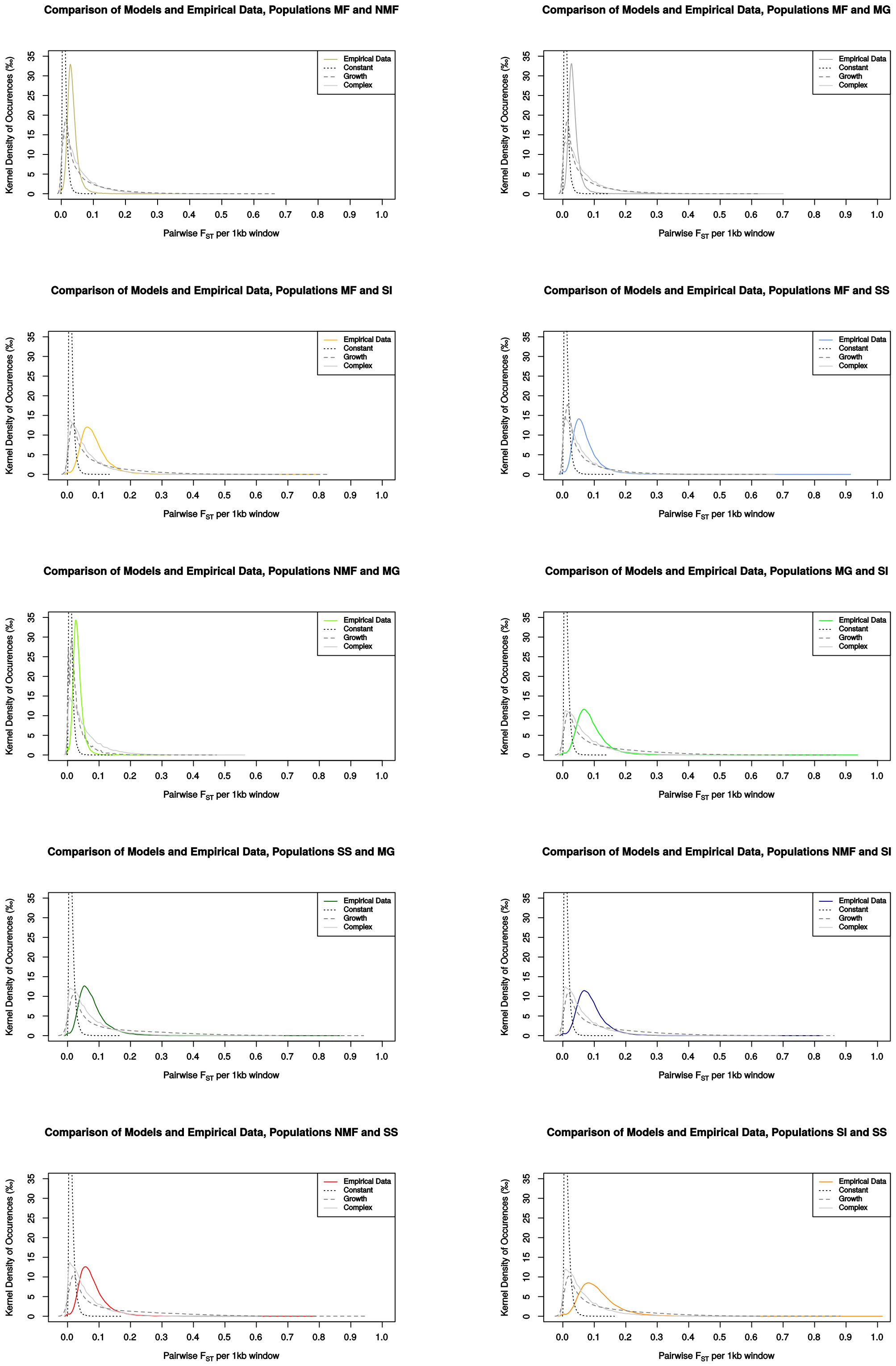


Figure S5.5: Comparisons of density functions of pairwise F_{ST} values between all pairs of populations

5.4 References

Beerli, Peter. 2006. "Comparison of Bayesian and Maximum-Likelihood Inference of Population Genetic Parameters." *Bioinformatics* 22 (3). Oxford Univ Press: 341–45.

Beerli, Peter, and Joseph Felsenstein. 2001. "Maximum Likelihood Estimation of a Migration Matrix and Effective Population Sizes in N Subpopulations by Using a Coalescent Approach." *Proceedings of the National Academy of Sciences* 98 (8). National Acad Sciences: 4563–8.

Excoffier, Laurent, and Matthieu Foll. 2011. "Fastsimcoal: A Continuous-Time Coalescent Simulator of Genomic Diversity Under Arbitrarily Complex Evolutionary Scenarios." *Bioinformatics* 27 (9): 1332–4. doi:10.1093/bioinformatics/btr124.

Excoffier, Laurent, and Heidi EL Lischer. 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). Wiley Online Library: 564–67.

Hollander, Myles, Douglas A Wolfe, and Eric Chicken. 2013. *Nonparametric Statistical Methods*. John Wiley & Sons.

Kahle, David, and Hadley Wickham. 2013. "Ggmap: Spatial Visualization with Ggplot2." *The R Journal* 5 (1): 144–61. http://journal.r-project.org/archive/2013-1/kahle-wickham.pdf.

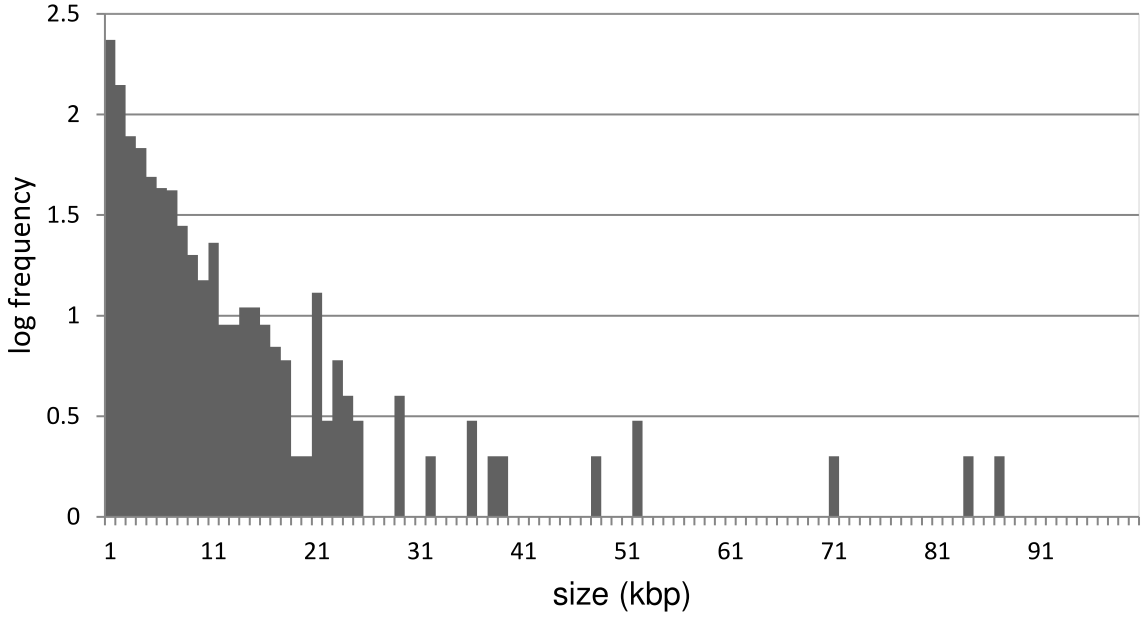
Oppold, Ann-Marie, João AM Pedrosa, Miklós Bálint, João B Diogo, Julia Ilkova, João LT Pestana, and Markus Pfenninger. 2016. "Support for the Evolutionary Speed Hypothesis from Intraspecific Population Genetic Data in the Non-Biting Midge Chironomus Riparius." In *Proc. R. Soc. B*, 283:20152413. 1825. The Royal Society.

Reynolds, John, Bruce S Weir, and C Clark Cockerham. 1983. "Estimation of the Coancestry Coefficient: Basis for a Short-Term Genetic Distance." *Genetics* 105 (3). Genetics Soc America: 767–79.

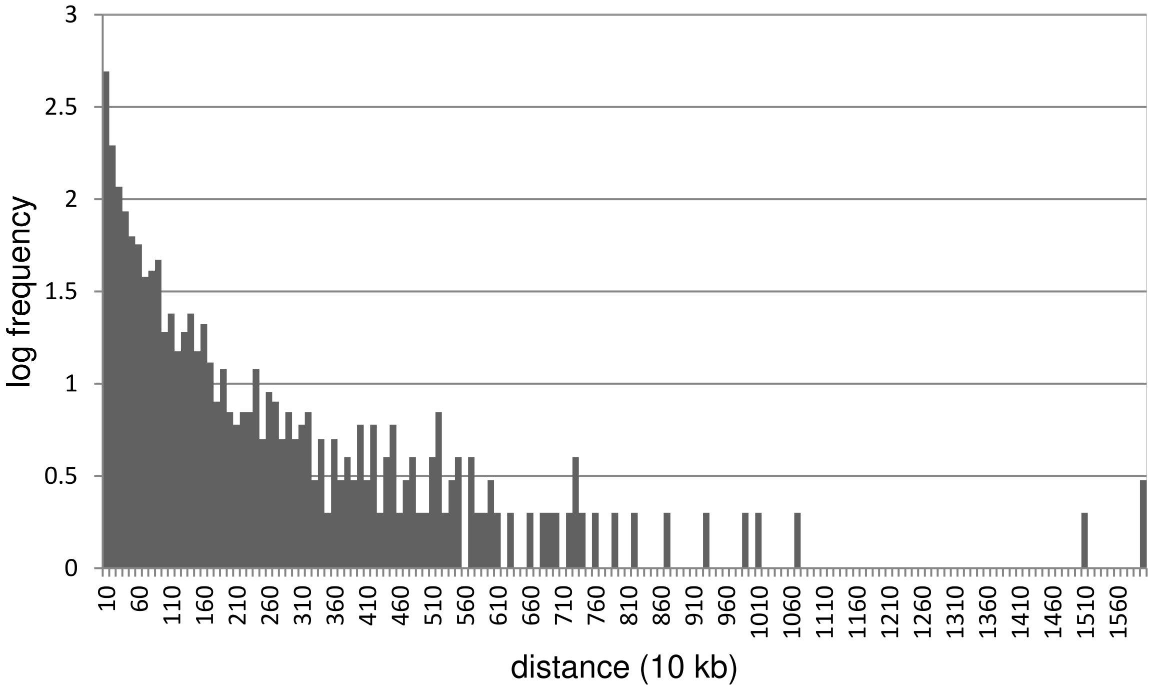
Schiffels, Stephan, and Richard Durbin. 2014. "Inferring Human Population Size and Separation History from Multiple Genome Sequences." *Nature Genetics* 46 (8). Nature Publishing Group: 919–25.

Slatkin, Montgomery. 1995. "A Measure of Population Subdivision Based on Microsatellite Allele Frequencies." *Genetics* 139 (1). Genetics Soc America: 457–62.

6. Analyses of population differentiation



Supporting Figure S6.1:
Size distribution of divergence regions (joined adjacent 1 kb outlier windows above 5 % F_{ST} threshold) for all pairwise comparisons.



Supporting Figure S6.2:
Distribution of distances among divergence regions on the same scaffold.

Supporting Table S6.1: Estimated migration rates between *C. riparius* populations across Europe.

direction	geographic distance(km)	migration rate per generation
MG→NMF	233.73	9×10^{-4}
MG←NMF	233.73	9×10^{-4}
NMF→MF	380.84	2×10^{-5}
NMF←MF	380.84	9×10^{-4}
MF→SI	274.25	3×10^{-5}
MF←SI	274.25	2×10^{-5}
MF→SS	1224.41	9×10^{-4}
MF←SS	1224.41	2×10^{-5}

Supporting Table S6.2: Statistics of pairwise F_{ST} from empirical Pool-Seq data of *C. riparius* populations.

population pair	geographic distance	median FST	mean FST	max FST
MF:MG	572.25	0.030	0.034	0.643
MF:NMF	380.84	0.031	0.035	0.551
MF:SI	274.25	0.074	0.083	0.862
MF:SS	1224.41	0.060	0.071	0.905
MG:NMF	233.73	0.029	0.032	0.415
MG:SI	532.58	0.078	0.088	0.926
MG:SS	1824.44	0.066	0.078	0.918
NMF:SI	1390.2	0.079	0.089	1.000
NMF:SS	1532.18	0.067	0.078	0.905
SI:SS	1387.39	0.097	0.111	1.000

Supporting Table S6.3: Comparisons of F_{ST} from Pool-Seq data (empirical) and simulation data (three different models). F_{ST} above 99 % threshold from empirical data was taken as threshold (highlighted in grey), above which we exclude the effect of drift. Numbers of highly diverged windows above this threshold before and after error correction are given.

population pair	99 % F_{ST}-threshold				number of windows above empirical 99 % F_{ST} threshold	number of windows after FDR correction
	empirical data	constant model	growth model	approximated model		
MF:MG	0.118	0.028	0.267	0.232	428	402
MF:NMF	0.115	0.028	0.264	0.204	437	399
MF:SI	0.260	0.034	0.397	0.224	407	407
MF:SS	0.250	0.034	0.269	0.211	519	519
MG:NMF	0.100	0.027	0.109	0.211	287	235
MG:SI	0.283	0.035	0.463	0.274	426	426
MG:SS	0.274	0.039	0.533	0.258	519	519
NMF:SI	0.278	0.034	0.461	0.248	444	444
NMF:SS	0.269	0.039	0.532	0.230	533	533
SI:SS	0.360	0.043	0.479	0.259	476	476

7. Tajima’s D analysis

Box 7: Molecular signatures of selection in divergent outlier windows

To analyse evolutionary forces acting on the identified outlier windows, we estimated Tajima’s D (T_D) in 1 kb windows in each of the five Pool-Seq data sets. Relative deviations from the mutation-drift equilibrium (measured as T_D) are expected to result from non-neutral evolution. With focus explicitly on highly divergent outlier regions, selection can be expected to be the major process contributing to divergence, whereas demographic effects can be neglected.

Method

We used the PoPoolation tool package (Kofler *et al.* 2011a) with high stringency settings for T_D estimation. T_D per population for significant 1kb outlier windows (according to the upper 1 % of the F_{ST} distribution) were extracted and compared to T_D per population for a random subset of the remaining 1kb windows that fell below the statistical threshold of neutral divergence (hereafter named “genome-wide average 1kb windows”). Summary statistics were calculated in GraphPad Prismv5.

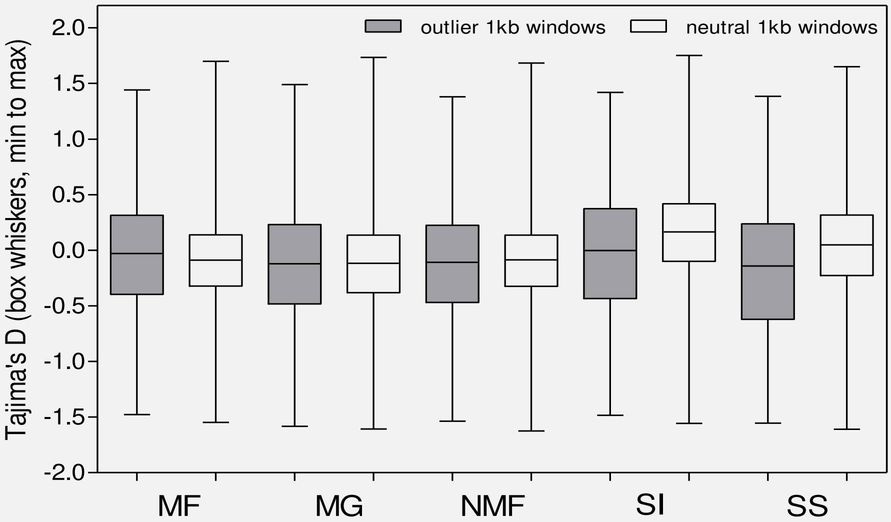
T_D values ±1 were defined as threshold above/below which we expected selective processes. This strict threshold simplifies the complex situation of genome-wide divergence processes; however, for our data it was more conservative (see upper 5 % values in Box-Table 6.1) than the upper 5 % T_D distribution threshold suggested in Feulner *et al.* (2015). We considered the following scenarios for a comparison with *n* populations (modified from Pfenninger *et al.* 2015): (i) negative T_D in one to *n*-1 populations indicate that the site has evolved under positive selection in the respective population, (ii) negative T_D in all populations is indicative for strong purifying selection, (iii) positive T_D is indicative for balancing selection in the respective population. With Chi^2 tests in R we compared the occurrences highly divergent outlier windows with signatures of positive or balancing selection in population-pairs and afterwards applied the Benjamini-Hochberg correction for multiple testing (*p.adjust* in R).

Results & Discussion

T_D of genome-wide average 1 kb windows of all populations levelled around zero (medians in a range of -0.116 and 0.165, Box-Table 6.1), indicating the major influence of neutral processes in shaping the genome. While the overall range of the T_D distribution was similar in 1 kb windows of divergent outliers and the genome-wide average (Box-Figure 6.1), median (as well as mean) T_D values in outlier regions were shifted towards negative values except for the MF population (Rhône-Alpes). These negative shifts are consistent with selection as major mechanism driving divergence in these genomic regions (Feulner *et al.* 2015).

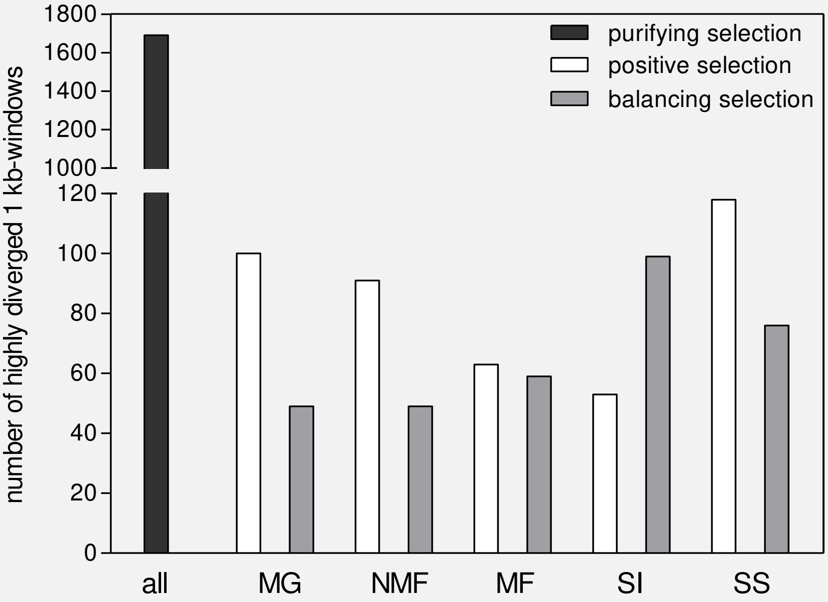
Box-Table 7.1: Summary statistics of T_D in outlier and genome-wide average 1 kb windows. Population-means of the two categories are given in the first two columns.

	outliers mean	neutral mean	outliers MF	neutral MF	outliers MG	neutral MG	outliers NMF	neutral NMF	outlier SI	neutral SI	outlier SS	neutral SS
Minimum	-1.527	-1.589	-1.476	-1.548	-1.583	-1.606	-1.537	-1.626	-1.485	-1.557	-1.554	-1.610
Maximum	1.423	1.704	1.441	1.700	1.490	1.734	1.379	1.684	1.419	1.752	1.385	1.652
Median	-0.080	-0.015	-0.029	-0.0872	-0.121	-0.116	-0.107	-0.087	-0.003	0.165	-0.140	0.0484
Mean	-0.095	-0.028	-0.0424	-0.0960	-0.1193	-0.1254	-0.1160	-0.0989	-0.027	0.1453	-0.168	0.0357
lower 5%	-1.020	-0.710	-0.9018	-0.7154	-1.028	-0.8197	-0.9937	-0.7268	-0.960	-0.564	-1.218	-0.722
upper 5%	0.835	0.603	0.8154	0.4881	0.8090	0.5352	0.7979	0.4855	0.8734	0.7825	0.8785	0.7260



Box-Figure 7.1: Distribution of T_D in 1 kb windows of highly divergent outlier windows and all remaining neutral windows.

Applying the three mutually exclusive scenarios to TD values ±1, we were able to quantify the relative contribution of different selection mechanisms in the divergent outlier 1 kb windows. Purifying selection was found to act on the majority of outlier windows (note that this value cannot be inferred population-wise). Number of highly divergent 1 kb windows evolved by positive and balancing selection differed among populations (Box-Figure 6.2). Northernmost and southernmost populations (MG, NMF, SS) showed major impact of positive selection (significantly different to MF and SI, Box-Table 6.2). Balancing selection was significantly increased in the two Southern populations (SI, SS, see Box-Table 6.2 for p-values).



Box-Figure 7.2: Occurrences of molecular signatures of selection in divergent outlier 1 kb windows (statistical comparisons in Box-Table 6.2).

Box-Table 7.2: Statistical p-values (Chi^2-tests, Benjamini-Hochberg correction for multiple testing) of numbers of molecular signatures of selection displayed in Box-Figure 6.2: (A) occurrences of positive selection among populations, (B) occurrences of balancing selection among populations.

A: positive selection	MG	NMF	MF	SI	SS
MG		0.5543	0.0082	6.67E-04	0.2976
NMF	0.5543		0.0448	0.0043	0.0939
MF	0.0082	0.0448		0.4414	2.11E-04
SI	6.67E-04	0.0043	0.4414		6.11E-06
SS	0.2976	0.0939	2.11E-04	6.11E-06	

B: balancing selection	MG	NMF	MF	SI	SS
MG		1	0.4231	0.0002	0.0368
NMF	1		0.4231	0.0002	0.0368
MF	0.4231	0.4231		0.0053	0.2317
SI	0.0002	0.0002	0.0053		0.1499
SS	0.0368	0.0368	0.2317	0.1499	

Since we can exclude a significant difference in N_e between populations (Fig. 3A), this pattern suggests that positive selection has been playing a major role in populations at the outer margins of the investigated climatic gradient. Referring back to the hypothesis that populations expanded from central France, i.e. the centre of the thermal cline, might provide an explanation for this spatial pattern of positive selection. According to the surfing mutation phenomenon, mutations occurring at the edge of the range expansion are lost at a reduced rate and can more easily be driven to fixation (Klopfstein et al. 2006). Therefore, the time of population range expansion is an evolutionary important period, where mutations can accumulate and contribute to adaptation processes. However, the existence of spatial and even temporal heterogeneity in the intensity and direction of selection is known from other species (Bergland et al. 2014; Charbonnel & Pemberton 2005). This could furthermore explain the observed spatial difference in the proportion of balancing selection.

References

Bergland AO, Behrman EL, O'Brien KR, Schmidt PS, Petrov DA (2014) Genomic Evidence of Rapid and Stable Adaptive Oscillations over Seasonal Time Scales in *Drosophila*. *Plos Genet* 10(11), e1004775.

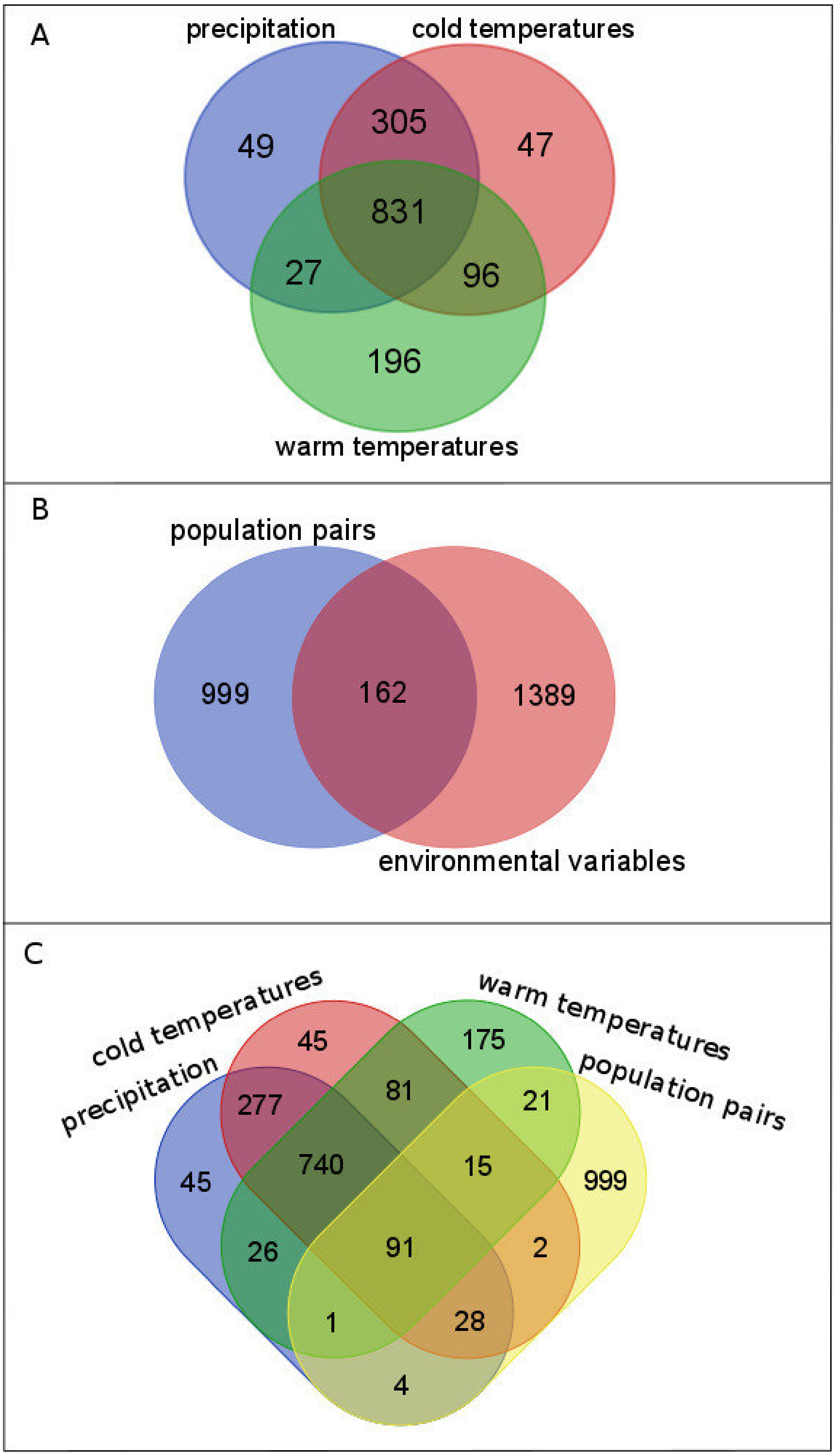
Charbonnel N, Pemberton J (2005) A long-term genetic survey of an ungulate population reveals balancing selection acting on MHC through spatial and temporal fluctuations in selection. *Heredity* 95, 377-388.

Feulner PGD, Chain FJJ, Panchal M, Huang Y, Eizaguirre C, Kalbe M, Lenz TL,Samonte IE, Stoll M, Bornberg-Bauer E, Reusch TBH, Milinski M (2015) Genomics of Divergence along a Continuum of Parapatric Population Differentiation. *PloS Genet* 11(7), e1004966.

Klopfstein S, Currat M, Excoffier L (2006) The fate of mutations surfing on the wave of a range expansion. *Mol Biol Evol* 23, 482-490.

Pfenninger M, Patel S, Arias-Rodriguez L, Feldmeyer B, Riesch R, Plath M (2015) Unique evolutionary trajectories in repeated adaptation to hydrogen sulphide-toxic habitats of a neotropical fish (*Poecilia mexicana*). *Mol Ecol* 24, 5446-5459

8. Functional enrichment analysis



Supporting Figure S8.1:
Venn diagrams (produced online at http://bioinformatics.psb.ugent.be/webtools/Venn/) of intersected candidate gene lists: (A) candidate genes for clinal adaptation correlated to the three environmental variables, (B) gene hits annotated to the significant outlier 1 kb-windows from pairwise population comparisons (99 % F_{ST} threshold) and candidate genes for clinal adaptation, and (C) detailed comparison of gene hits from population comparisons with the three environmental variables separately.

Supporting Table S8.1:

Results of enrichment analysis on the level of biological functions (GO terms) and molecular pathways (KEGG pathways). The amount of genes involved in the respective adaptation pattern is given against the complete annotation of 13,093 protein coding genes. Gene hits for populations integrate all hits that result from comparison of the respective population with the others (gene hits in significant outlier 1 kb-windows). Gene hits correlated to environmental variables result from the locus-specific environmental association study with LFMM. Note that the number of significantly enriched GO terms and KEGG pathways is relative to input genes. Therefore, there can be less significant hits on the superior level compared to subgroups (e.g. 9 GO terms among overall candidates for local adaptation against 19 GO terms among local candidates of SS).

	gene hits from comparisons with all other populations	% of all genes	enriched GO terms	KEGG pathways
MG	669	5.1	16	
NMF	728	5.6	7	
MF	708	5.4	6	
SI	603	4.6	14	
SS	656	5.0	19	
candidates for local adaptation	999	7.6	9	77
significant clinal candidates	162	1.2	10	23
all environmental candidates	1389	10.6	6	87
"cold temperatures" candidates	47	0.4	4	20
"precipitation" candidates	49	0.4	4	2
"warm temperatures" candidates	196	1.5	6	114