# The genomic footprint of climate adaptation in *Chironomus riparius*

Ann-Marie Oppold[1,2], Andreas Wieser[1,2], Tilman Schell[1,2], Simit Patel[2], Hanno Schmidt[2], Thomas Hankeln[3], Barbara Feldmeyer[2], Markus Pfenninger[1,2]

1 Molecular Ecology Group, Institute for Ecology, Evolution & Diversity, Goethe-University, Frankfurt am Main, Hessen, Germany

2 Senckenberg Biodiversity and Climate Research Centre, Frankfurt am Main, Hessen, Germany

3 Institute of Organismic and Molecular Evolution, Molecular Genetics and Genome Analysis, Johannes Gutenberg-University, Mainz, Rheinland-Pfalz, Germany

*Corresponding author: Ann-Marie.Oppold@senckenberg.de
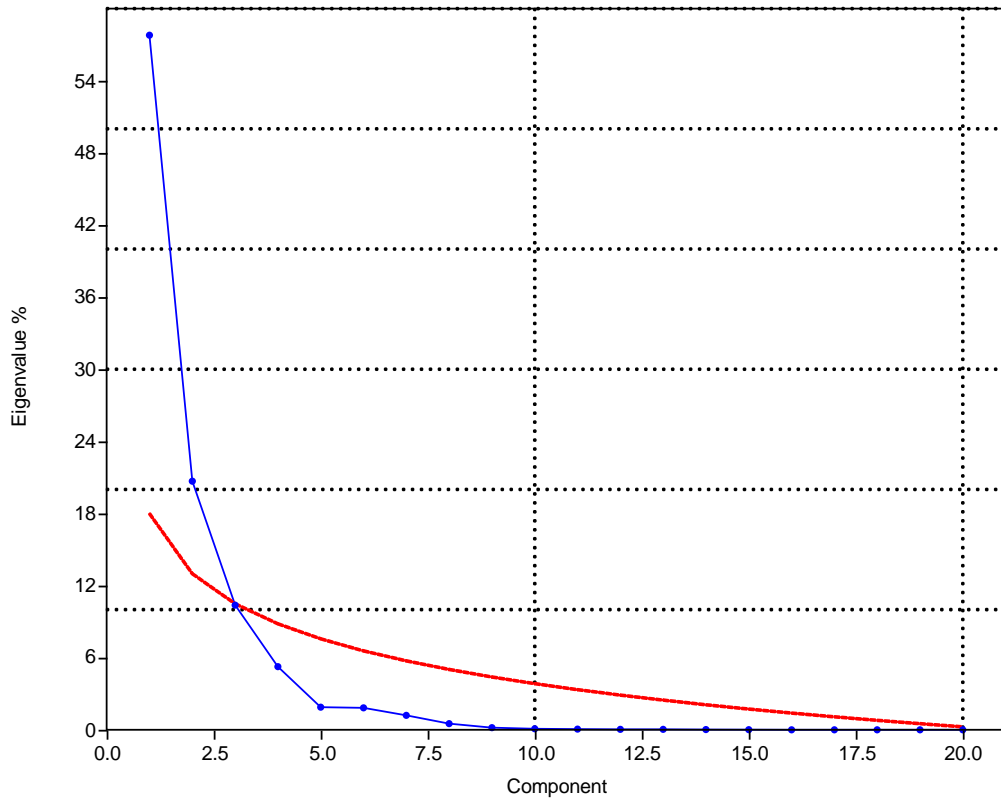
Supplementary Figure S1:
Disribution of Eigenvalues (%) of components (blue line) after principal component analysis (PCA) with 57 climatic variables (WorldClim data) from 21 locations of documented *C. riparius* occurrence (including the five natural populations of this study, Oppold et al. 2016a). Red line marks the random distribution of Eigenvalues (broken stick analysis). Components under this curve are expected to be non-significant.
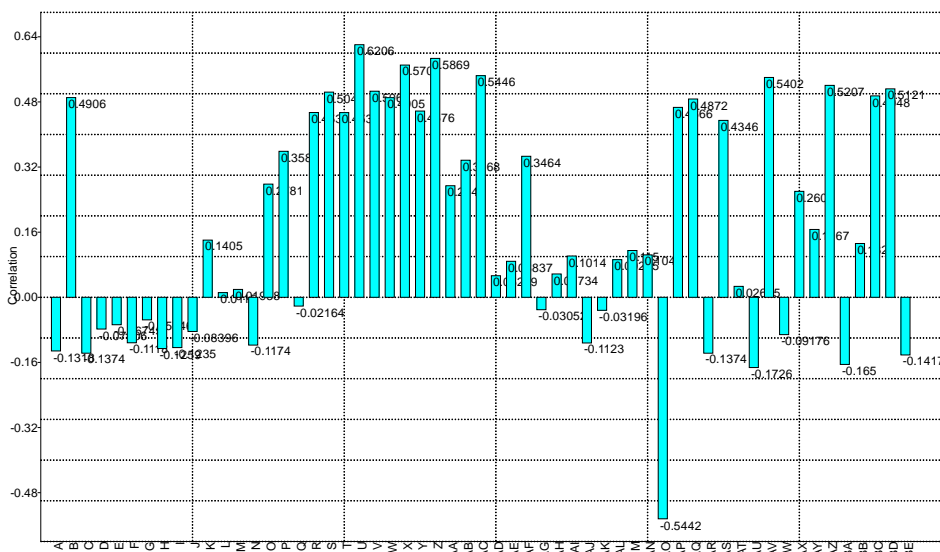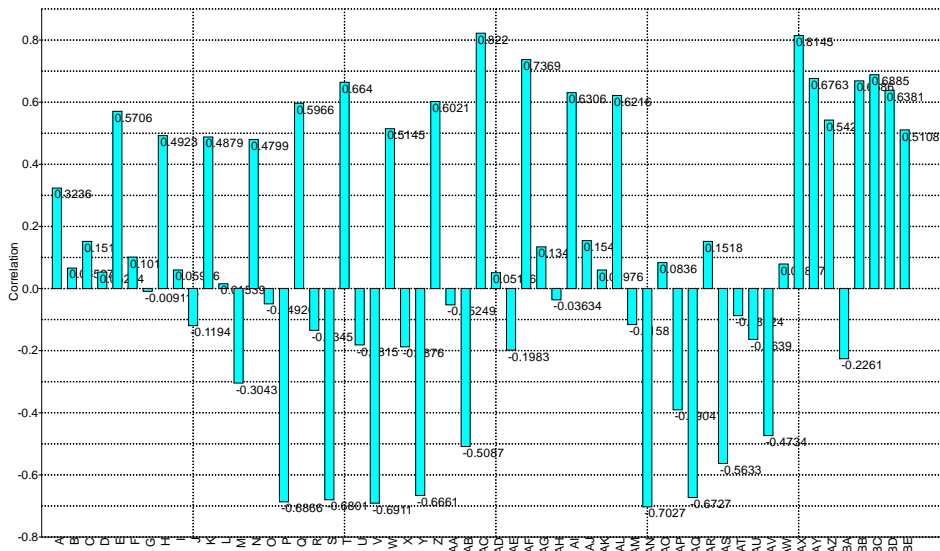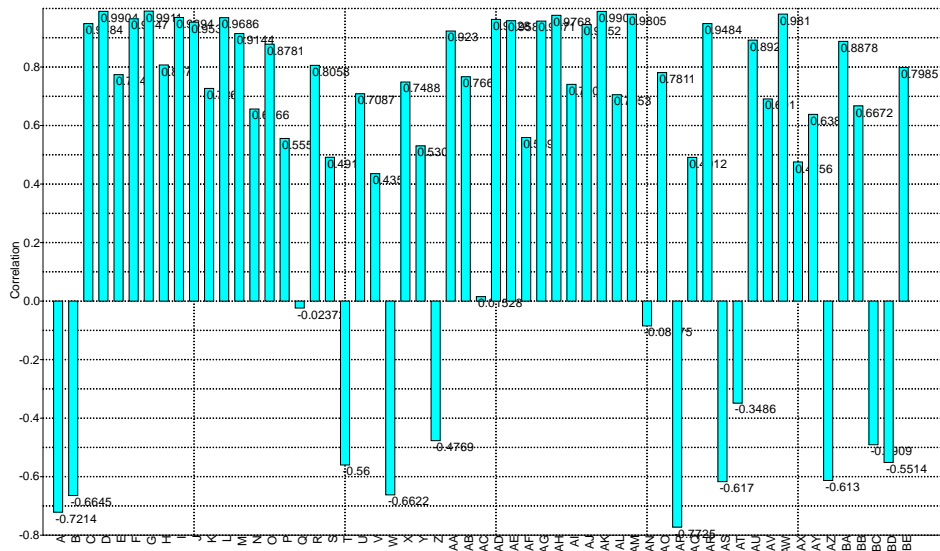
Supplementary Figure S2:

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.

3
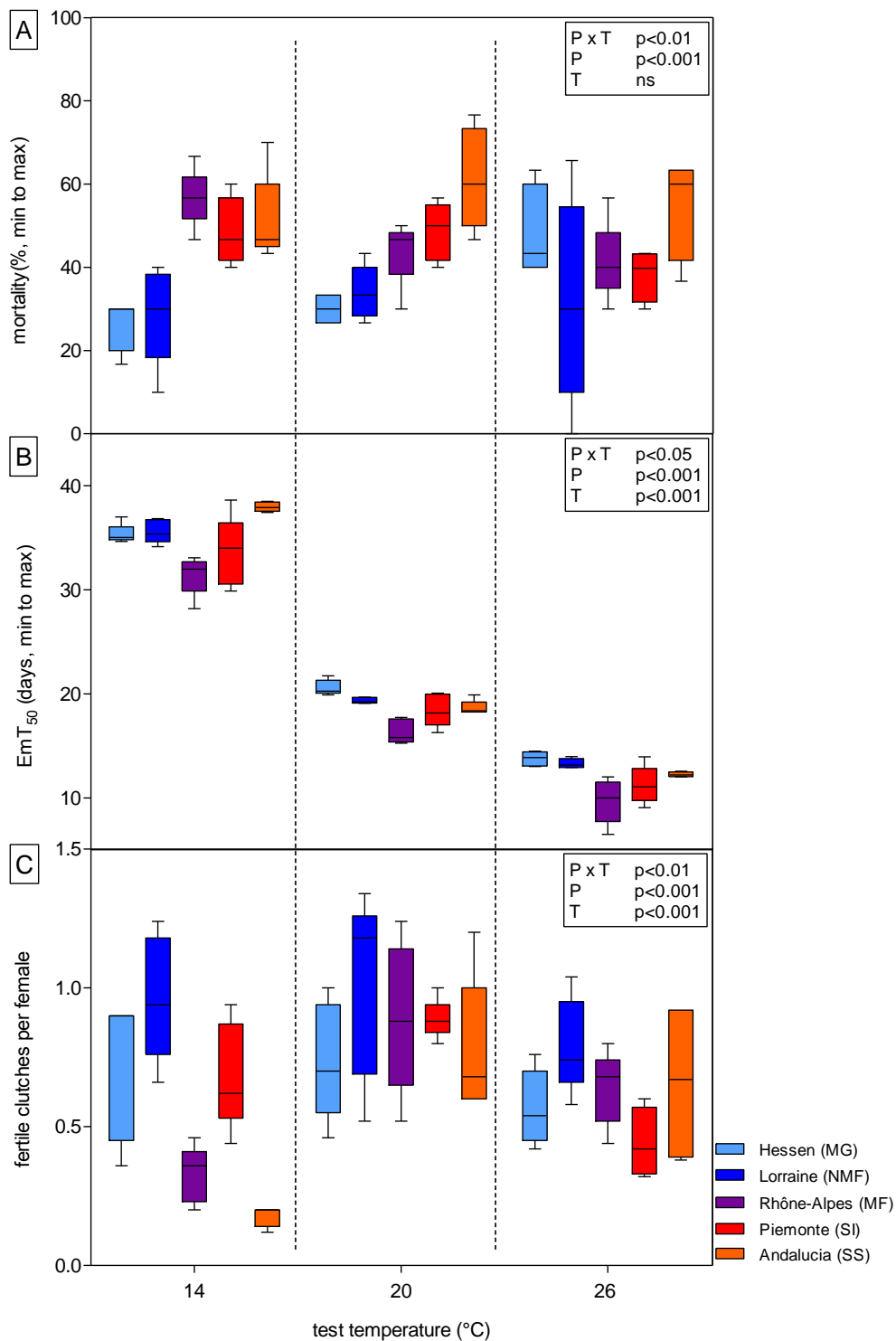
Supplementary Figure S3:

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).
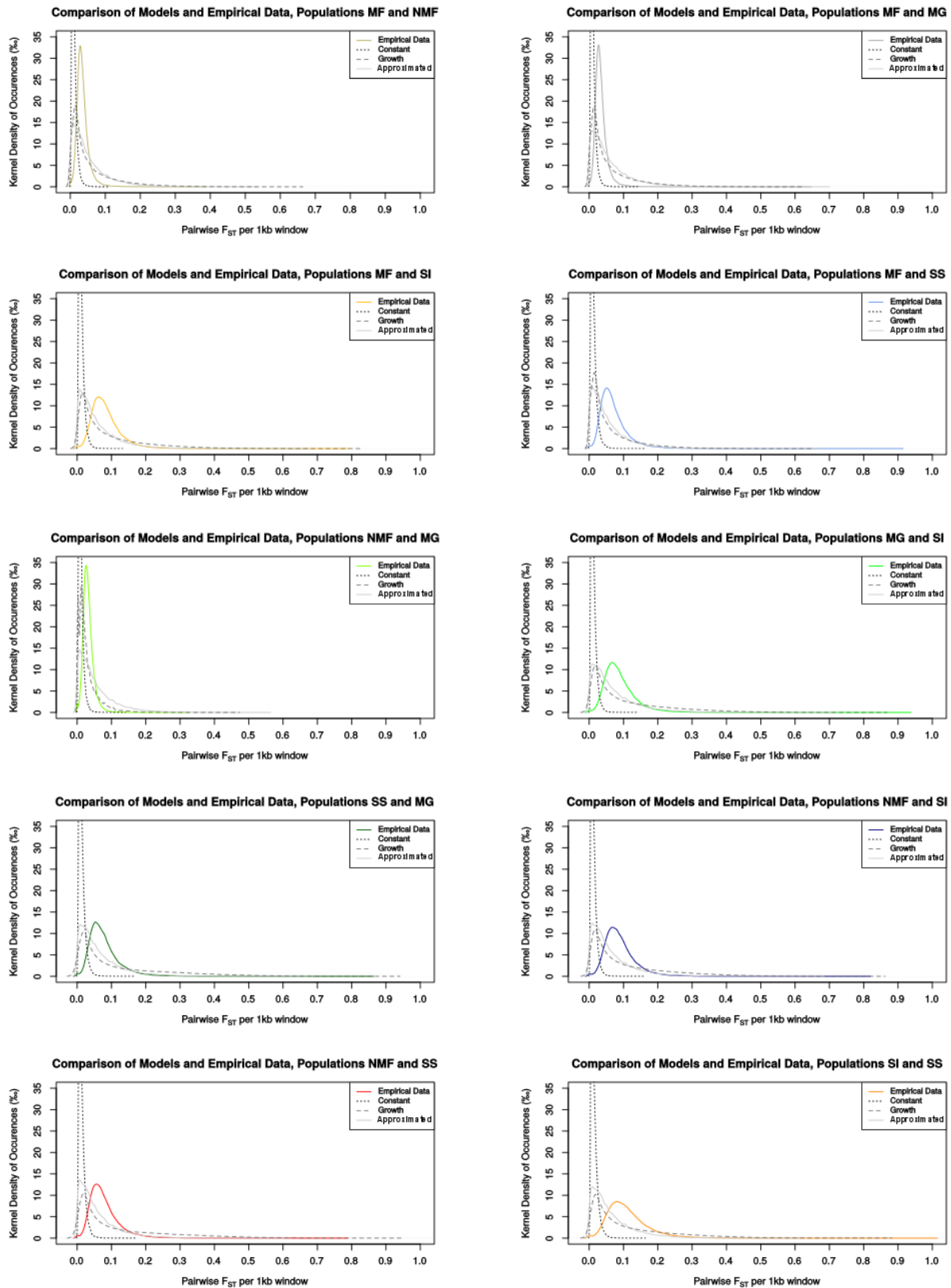
Supplementary Figure S4:
Density functions of pairwise F_{ST} per 1 kb-windows of empirical data (*C. riparius* Pool-Seq data) and simulated data from three different models (*constant*, *growth*, *approximated*). Each population comparison is plotted separately.
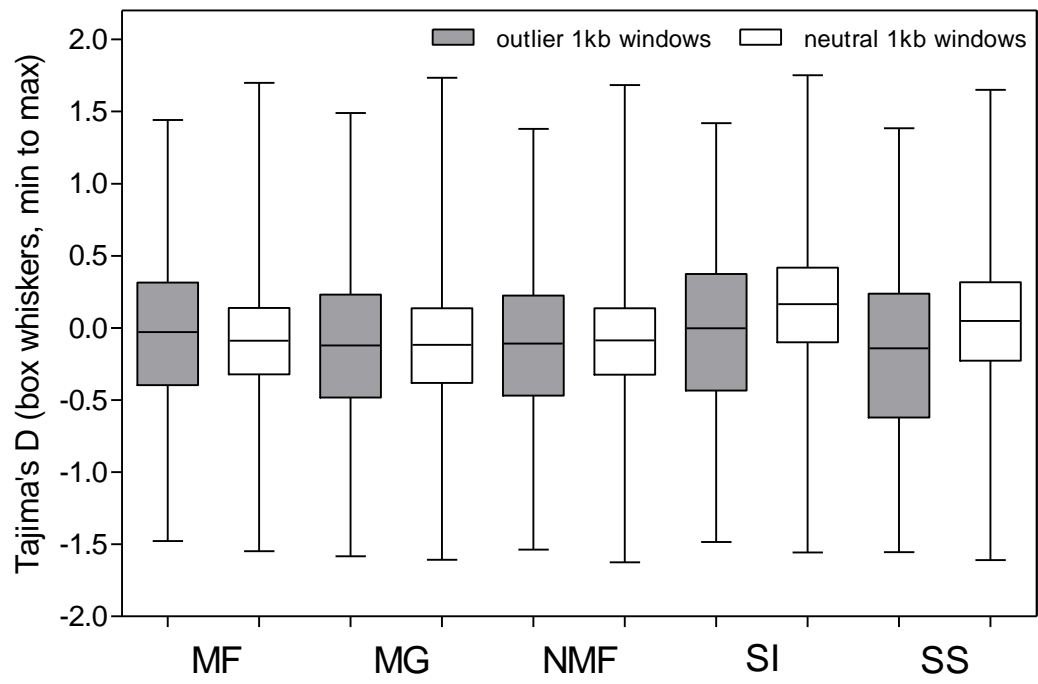
Supplementary Figure S5:
Distribution of Tajima's D in 1 kb-windows of highly diverged (outlier) and remaining neutral windows.
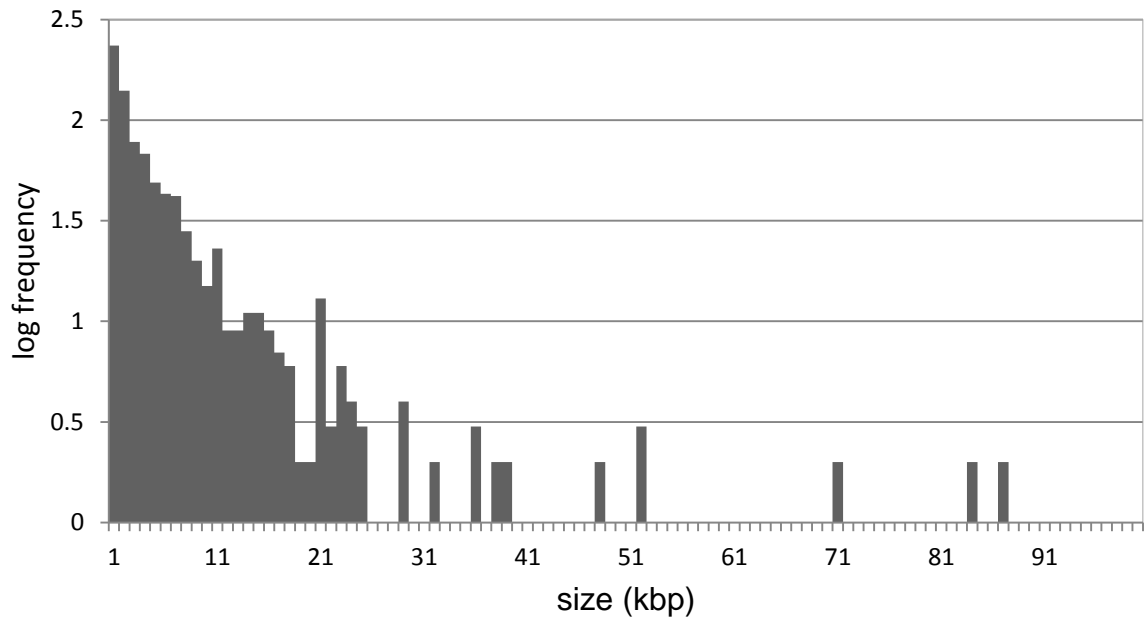
Supplementary Figure S6:
Size distribution of divergence regions (joined adjacent 1 kb outlier windows above 5 % F_{ST} threshold) for all pairwise comparisons.
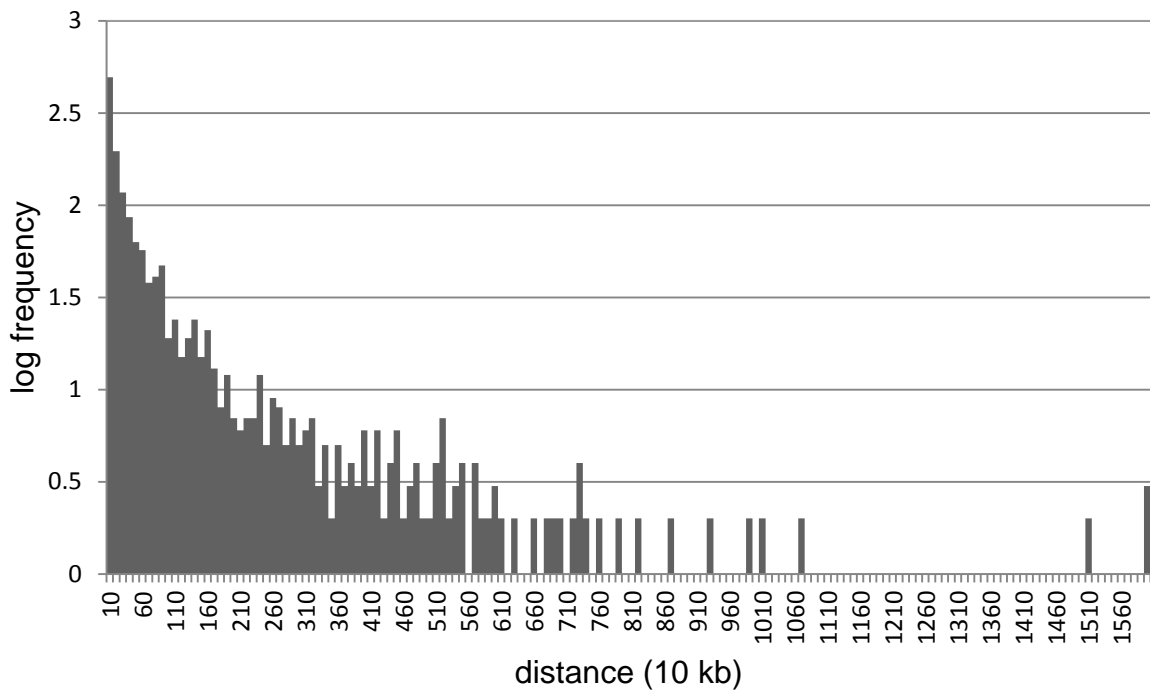
Supplementary Figure S7:
Distribution of distances among divergence regions on the same scaffold.
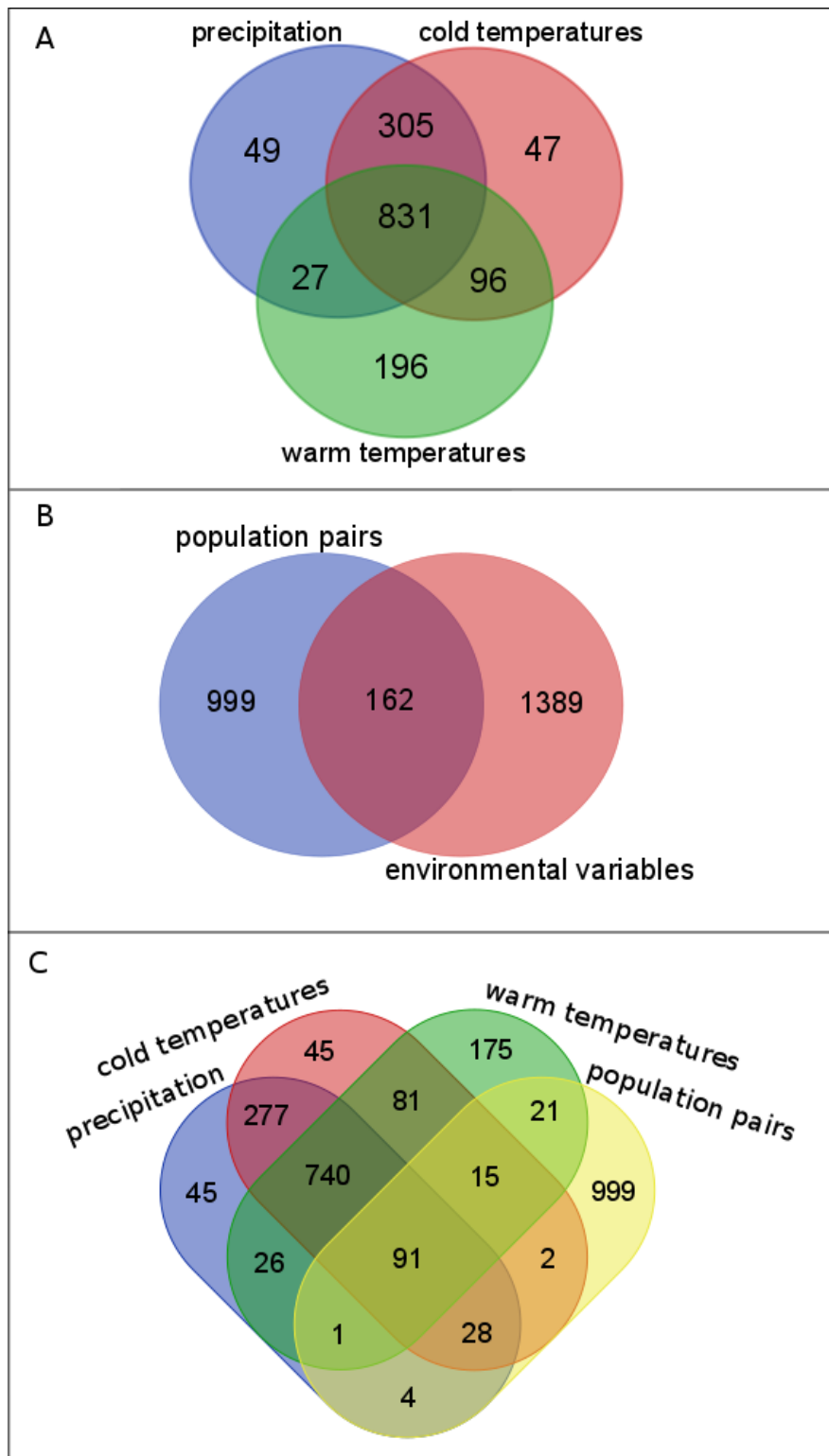
Supplementary Figure S8:
Venn diagrams (produced online at http://bioinformatics.psb.ugent.be/webtools/Venn/) of intersected candidate gene lists: (A) candidate genes for clinal adaptation correlated to the three environmental variables, (B) gene hits annotated to the significant outlier 1 kb-windows 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.

Supplementary Table S1:

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

population	mean coverage	% mapped reads
MG	26.7	81.09
NMF	55.3	77.57
MF	41.0	78.03
SI	40.2	77.88
SS	36.9	80.92

Supplementary Table S2:

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		

Supplemental Table S3: 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**

Supplementary Table S4: 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}

Supplementary Table S5: Statistics of pairwise F_{ST} from empirical Pool-Seq data of *C. riparius* populations.

population pair	geographic distance	median F_{ST}	mean F_{ST}	max F_{ST}
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

Supplementary Table S6: 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

Supplementary Table S7: Statistical p-values (Chi^2-tests, Benjamini-Hochberg correction for multiple testing) of numbers of molecular signatures of selection displayed in Figure 4:
(A) occurrences of selective sweeps among populations, (B) occurrences of balancing selection among populations.

A: selective sweeps	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	

Supplementary Table S8:
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