Supplementary Methods S1

Sequencing, assembly, scaffolding and annotation of the *C. riparius* draft genome

Approximately 50 larvae of a long standing laboratory culture of *C. riparius* were used for extraction of genomic DNA with the Wizard® SV Genomic DNA Purification System (Promega). DNA was quality checked by gel electrophoresis and quantified using the Qubit® dsDNA HS Assay Kit (ThermoFisher Scientific). Five sequencing libraries (Illumina) were constructed, including two paired-end libraries (100 and 200 bp insert size, see Supplementary Table S3) and two mate-pair libraries (3,000 and 5,500 bp insert size). Libraries were sequenced on an Illumina HiSeq2500 (Institute of Molecular Genetics, Johannes Gutenberg University Mainz), except for the 200 bp insert size paired-end library, which was sequenced on an Illumina MiSeq (StarSEQ).

Reads were quality checked with FastQC v0.11.3 (http://www.bioinformatics.babraham.ac.uk/projects/fastqc/) and processed with fastx_toolkit v0.0.13 (http://hannonlab.cshl.edu/fastx_toolkit/). Nine nucleotides were trimmed from the 5'-end. Universal and index adapters as well as positions with quality < 20 were trimmed from 3'-ends. Reads with overall bad quality (> 20 % positions with quality < 20) were discarded. To get rid of bacterial contaminations, all reads were mapped with Bowtie (Langmead et al 2009) against RefSeq bacterial genomes (as of March 2013), sorting out all reads with perfect hits. Error correction of reads was conducted using the tools Quake (Kelley et al. 2010) and Musket (Liu et al. 2013), respectively. Since this did not significantly improve the data quality, we did not use the output and rather continued with the unaltered reads.

All reads were assembled using Platanus v1.2.1 (Kajitani et al. 2014) with stepwise (6 bp) increasing kmer-sizes between k=32 and k=84 and with a minimum coverage of 10. A first round of scaffolding of contigs was performed using Platanus, fusing contigs linked by at least 3 read pairs. Afterwards, all contigs < 500 bp were discarded. A second scaffolding step with SSPACE 3.0 (Boetzer et al. 2011) using the extension option was done. Contigs were joined if a minimum of 5 read pairs spanned a gap or if 100 bp of two contigs overlapped. All scaffolds < 1,000 bp were discarded from the final *draft genome* (for results see Supplementary Table S4).

A first quality check was done by mapping all reads used for obtaining the assembly as well as reads from another project (Vicoso & Bachtrog 2015) back onto the assembly with BWA v0.7.10 (Li & Durbin 2009) using the *bwa mem* algorithm (for results see Supplementary Table S5). The tool BUSCO v1.1b1 (Simão et al. 2015) with the arthropod database was used to verify completeness of core genes in the assembly (for results see Supplementary Table S6). Quality of the scaffolding was checked with REAPR 1.0.18 (Hunt et al. 2013). As suggested by those authors, the largest dataset (Lib2) was mapped with *perfectmap* and the dataset with the biggest insert size (Lib5) with the *smaltmap* algorithm (both part of REAPR) onto the genome. Then the REAPR pipeline was used to evaluate the scaffolding based upon the correct pairing and orientation of the reads across N-stretches.

The whole annotation process was performed with the MAKER2 v2.31.8 (Cantarel *et al.* 2008, Holt *et al.* 2011) pipeline and affiliated programs. Before running MAKER2, all data available for annotation of the genome has to be prepared accordingly and handed to MAKER2 in form of input files. To ensure discovery of most repeat sequences in the *draft genome* we built a custom repeat library from all repeats from *Chironomus* species available in NCBIs databases plus unpublished *C. riparius* specific repeat sequences (Supplementary material 2). For integration of EST-evidence we assembled a reference transcriptome from RNA-Seq data from several life cycle stages of *C. riparius* (see below for workflow and Supplementary Table S6 for used datasets) and inputted this transcriptome together with the draft genome to the Augustus v3.2.1 training (Stanke *et al.* 2008) at the university of Greifswald's webserver (http://bioinf.uni-greifswald.de/webaugustus/training/create). This training's output then served as input for the first round of MAKER2. MAKER2 can work much more accurate when provided with genome-specific gene models at the beginning. Therefore we ran CEGMA v2.5 (Parra *et al.* 2007) on CRIP_Laufer and converted the output using scripts from the SNAP gene finder v2006-07-28 (Korf 2004) to a hidden Markov model. Additionally, we created another hidden Markov model by running GeneMark v4.32 (Lomsadze et al. 2014) with min_contig set to 20,000 on CRIP_Laufer.

The first round of MAKER2 annotation was then run using MPICH2 v3.2 (https://www.mpich.org/) parallelisation with the described transcriptome, SNAP, GeneMark and Augustus models, our custom

repeat library plus the SwissProt database (as at 13.1.2016) as input. The MAKER2 pipeline was run with the programs Augustus v3.2.1, BLAST v2.2.28+ (Altschul et al. 1990), Repeatmasker v4.0.6 (http://www.repeatmasker.org), SNAP v2006-07-28, GeneMark v4.32 and Exonerate v2.2.0 (Slater et al. 2005). We applied default parameters with only max_dna_len set to 500,000 to prevent loss of gene parts from genes with larger introns, min_protein set to 10 to receive as much potential protein sequences as possible and fix_nucleotides set as flag to allow for non-ACGT-characters in the genome file. The gff files of MAKER2's output were merged to a single file using gff3_merge from the MAKER2 distribution. Afterwards this gff file was converted to a hidden Markov model using SNAP scripts as described above for CEGMA with only the cegma2zff script being replaced by maker2zff. The information from the first MAKER2 run was also used for retraining the gene model in Augustus. The genome.ann file was first retransformed to gff and modified to match the gff format and then fed into the autoAug.pl script from Augustus v3.2.1 for retraining the EST-based gene models. The second round of the MAKER2 pipeline was then started with the same settings and input files as described above but with the updated Augustus and SNAP gene models and the parameters min_protein set to 30 and alt_splice on. Afterwards, a third round of the MAKER2 pipeline was run exactly as the second one, including another re-training of the Augustus gene model with autoAug.pl and again updating the SNAP gene model. From the resulting output of MAKER2's third round we merged all gff files with gff3_merge as described above and renamed the included gene tracks to ensure easier handling. To allow for assigning putative gene functions to the annotated gene tracks we gathered all predicted proteins from the MAKER2 output folder, performed BLASTP searches against the SwissProt database (as at 13.1.2016) and then added the BLAST hits to the accordant gene tracks. This output file (Supplementary Table S6) was then used for visualisation in JBrowse v1.12.0 (Skinner et al. 2009) and for the analyses involving gene annotations.

Preparation of the reference transcriptome used for the annotation of the genome

Overall 14 cDNA datasets (Supplementary Table S6) were iteratively assembled to include as much transcriptomic information as possible. All Illumina datasets were assembled using the CLC Genomics Workbench v6.5.2 (https://www.qiagenbioinformatics.com/) with word size 20 and bubble size 50.

Reads were mapped back to the contigs and "contig update" was performed. The Roche 454 data was assembled using SeqMan NGen® v13.0.1 (DNASTAR, Madison, WI) with k-mer size 31, minimum match percentage of 99 % and minimum sequences set to 2. The resulting contigs from the CLC and the NGen assemblies, as well as reads from the 454 data that were previously unassembled, were meta-assembled using NGen with k-mer size 31 and minimum match percentage of 100 %. Both assembled and unassembled contigs were then combined and resulted in the final transcriptome with 158,017 sequences with N50 of 539 nt.

References

Altschul, S. F. et al., Basic local alignment search tool. J. Mol. Biol. 215 (3), 403 (1990).

Behura SK, Lobo NF, Haas B, et al. (2011) Complete sequences of mitochondria genomes of *Aedes aegypti* and *Culex quinquefasciatus* and comparative analysis of mitochondrial DNA fragments inserted in the nuclear genomes. Insect Biochemistry and Molecular Biology 41, 770-777.

Boetzer M, Henkel CV, Jansen HJ, Butler D, Pirovano W (2011) Scaffolding pre-assembled contigs using SSPACE. Bioinformatics 27, 578-579.

Cantarel, Brandi L. et al., MAKER: an easy-to-use annotation pipeline designed for emerging model organism genomes. Genome Res. 18 (1), 188 (2008).

Holt RA, Subramanian GM, Halpern A, et al. (2002) The genome sequence of the malaria mosquito *Anopheles gambiae*. Science 298, 129-+.

Holt, Carson and Yandell, Mark, MAKER2: an annotation pipeline and genome-database management tool for second-generation genome projects. BMC Bioinformatics 12 (1), 491 (2011).

Hunt M, Kikuchi T, Sanders M, et al. (2013) REAPR: a universal tool for genome assembly evaluation. Genome Biology 14, R47.

Kajitani R, Toshimoto K, Noguchi H, et al. (2014) Efficient de novo assembly of highly heterozygous genomes from whole-genome shotgun short reads. Genome Research 24, 1384-1395.

Kelley DR, Schatz MC, Salzberg SL (2010) Quake: quality-aware detection and correction of sequencing errors. Genome Biology 11, R116.

Kelley JL, Peyton JT, Fiston-Lavier AS, et al. (2014) Compact genome of the Antarctic midge is likely an adaptation to an extreme environment. Nat Commun 5.

Klomp J, Athy D, Kwan CW, et al. (2015) A cysteine-clamp gene drives embryo polarity in the midge *Chironomus*. Science 348, 1040-1042.

Korf, Ian, Gene finding in novel genomes. BMC Bioinformatics 5 (1), 1 (2004).

Langmead B, Trapnell C, Pop M, Salzberg SL. (2009) Ultrafast and memory-efficient alignment of short DNA sequences to the human genome. Genome Biology 10:R25.

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

Liu Y, Schröder J, Schmidt B (2013) Musket: a multistage k-mer spectrum-based error corrector for Illumina sequence data. Bioinformatics 29, 308-315.

Lomsadze, Alexandre, Burns, Paul D., and Borodovsky, Mark, Integration of mapped RNA-Seq reads into automatic training of eukaryotic gene finding algorithm. Nucleic Acids Res. 42 (15), e119 (2014).

Marinkovic M, de Leeuw WC, de Jong M, et al. (2012) Combining Next-Generation Sequencing and Microarray Technology into a Transcriptomics Approach for the Non-Model Organism *Chironomus riparius*. Plos One 7.

Nene V, Wortman JR, Lawson D, et al. (2007) Genome sequence of *Aedes aegypti*, a major arbovirus vector. Science 316, 1718-1723.

Parra, Genis, Bradnam, Keith, and Korf, Ian, CEGMA: a pipeline to accurately annotate core genes in eukaryotic genomes. Bioinformatics 23 (9), 1061 (2007).

Simão FA, Waterhouse RM, Ioannidis P, Kriventseva EV, Zdobnov EM (2015) BUSCO: assessing genome assembly and annotation completeness with single-copy orthologs. Bioinformatics 31, 3210-3212.

Skinner, Mitchell E. et al., JBrowse: a next-generation genome browser. Genome Res. 19 (9), 1630 (2009).

Slater, Guy S. C. and Birney, Ewan, Automated generation of heuristics for biological sequence comparison. BMC Bioinformatics 6 (1), 31 (2005).

Stanke, Mario, Diekhans, Mark, Baertsch, Robert, and Haussler, David, Using native and syntenically mapped cDNA alignments to improve de novo gene finding. Bioinformatics 24 (5), 637 (2008).

Vicoso B, Bachtrog D (2015) Numerous transitions of sex chromosomes in Diptera. PLoS Biology DOI: 10.1371/journal.pbio.1002078.

Supplementary Tables

Supplementary Table S1: Information about *C. riparius* populations that were used for Pool-Seq genome scans and all experiments of this study. Mapping statistics are given for the mapping of 50 Mio reads for PoPoolationTE analyses.

Population (code)	Coordinates of location	Number of pooled individuals	Number of reads sequenced	Mean coverage and % mapped reads (mapping PoPTE)
Hesse (MG)	50.1680610, 9.0819270	112	63,506,436	26.18; 86.38
Lorraine (NMF)	49.1765430, 6.2156670	105	128,327894	24.47; 83.25
Rhône-Alpes (MF)	45.8616760, 4.8865000	168	84,747,040	25.31; 83.33
Piemont (SI)	45.4036180, 8.3473320	155	82,265,386	26.19; 82.82
Andalusia (SS)	37.3990800, -4.5267980	118	81,667,612	25.52; 85.99

Supplementary Table S2: Sequence data used for genome assembly and scaffolding

	Library kit	Library type	Sequencing technology	Insert size	Read length	Number of raw reads	Number of clean reads
Lib1	TruSeq	Paired end	HiSeq2500	100 bp	101 bp	29,136,088	17,504,534
Lib2	TruSeq	Paired end	HiSeq2500	200 bp	100 bp	167,264,372	104,627,969
Lib3	TruSeq LT	Paired end	MiSeq	200 bp	301 bp	58,447,092	53,336,377
Lib4	Mate Pair Library v2	Mate pair	HiSeq2500	3,000 bp	101 bp	48,853,530	31,793,337
Lib5	Nextera Mate Pair	Mate pair	HiSeq2500	5,500 bp	108 bp	47,978,148	44,252,069

Supplementary Table S3: Assembly statistics. NG50 and LG50 were calculated using an estimated genome size for *C. riparius* of 200 Mb (Schmidt-Ott et al. 2009).

Total sequence length	180,652,019
Total assembly gap length	28,829,127
Number of scaffolds	5,292
Longest scaffold	2,056,324
Scaffold N50	272,065
Scaffold L50	166
Scaffold NG50	227,750
Scaffold LG50	204
Number of contigs	41,974
Longest contig	146,150
Contig N50	6,791
Contig L50	5,682
Contig NG50	4,283
Contig LG50	10,177

Supplementary Table S4: Back-mapping of reads used to obtain the assembly onto the *draft genome*. Additionally, the reads used in the study of Vicoso & Bachtrog 2015 were mapped to the *draft genome* (marked with asterisks).

	Lib1	Lib2	Lib3	Lib4	Lib5	SRR1738173*	SRR1738174*
mapped reads [%]	88.2	94.9	98.9	84.5	82.9	93.1	93.4

Supplementary Table S5: Gene content evaluation. The *draft genome* was scanned for a core set of genes present in all arthropods using the tool BUSCO.

	count	fraction
total BUSCOs searched (arthropod database)	2,675	
complete BUSCOs found	2,101	79 %
BUSCOs partially recovered	381	14 %
BUSCOs not found	193	7 %

Supplementary Table S6: Datasets used for the assembly of a *C. riparius* reference transcriptome

sample	accession number	source
L4 larvae 454 Roche GS FLX sequences	SRR834592	Schmidt *et al.* 2013
L4 larvae, stress treated 454 Roche GS FLX sequences	SRR834593	Schmidt *et al.* 2013
different life cycle stages 454 Roche GS FLX sequences	SRR496839	Marinković *et al.* 2012
embryo Illumina HiSeq 2000	SRR1028867	Klomp *et al.* 2015
embryo Illumina HiSeq 2000	SRR1032319	Klomp *et al.* 2015
embryo Illumina HiSeq 2000	SRR1032320	Klomp *et al.* 2015
embryo Illumina HiSeq 2000	SRR1032321	Klomp *et al.* 2015
embryo Illumina HiSeq 2000	SRR1032322	Klomp *et al.* 2015
embryo Illumina HiSeq 2000	SRR1032323	Klomp *et al.* 2015
adult midges Illumina HiSeq	not published	1kite project
larvae Illumina HiSeq 2000	SRR pending	sequenced for the present study
larvae Illumina HiSeq 2000	SRR pending	sequenced for the present study
larvae Illumina HiSeq 2000	SRR pending	sequenced for the present study
larvae Illumina HiSeq 2000	SRR pending	sequenced for the present study

Supplementary Table S7: Content of protein-coding genes in *C. riparius* draft genome compared to *C. tentans* and *D. melanogaster*.

	gene count	average number of exons per gene	average exon length	protein coding part of the genome
CRIP_Laufer	13,093	4.7	355 bp	10.9 %
C. tentans (Kutsenko *et al.* 2014)	15,120	3.8	312 bp	9 %
D. melanogaster (Release 6)	13,907	5.5	538 bp	18.3 % (Release 5)

Supplementary Table S8: TEs inserted close to or into protein-coding regions.

population	scaffold	position	TE	frequency	protein hit	comment
SS	scaffold684	321	Balbiani ring repeat	1	Balbiani ring protein 2	expected (Kutsenko et al. 2014)
SS	scaffold475	16820	*Cla-element*	0.005	Cytochrome c oxidase subunit 1	pseudogene
SI	scaffold684	325	Balbianiringrepeat	1	Balbiani ring protein 2	expected (Kutsenko et al. 2014)
SI	scaffold88	100532	ChiroAlu element	0.23	Nitric oxide-associated protein 1	low frequency
NMF	scaffold1156	11060	CTRT1 element	0.5	Protein of unknown function	?
MF	scaffold684	327	Balbiani ring repeat	1	Balbiani ring protein 2	expected (Kutsenko et al. 2014)
MF	scaffold966	11048	ChiroAlu element	0.64	Retrovirus-related Polymerase poly-protein from trans-poson TNT 1-94	insertion into different TE
MF	scaffold798	15034	CTRT1 element	0.5	Deoxycytidylate deaminase	upstream

Supplementary Table S9: Statistics of TE distribution in the context of genome annotation. Chi squared test of differences between TE insertions in gene regions or exons and intergenic regions. 2969 random positions were drawn from the genome for the null distribution.

	% in genes	outside gene	in genes	in exons	chi²	p-value
SS	0.0207	2752	57	2	3.881	0.0488
SI	0.0167	3291	55	2	10.27	0.0014
NMF	0.0158	2840	45	1	10.941	0.0009
MF	0.0169	2372	40	3	8.104	0.0044
MG	0.0168	2982	50	0	9.57	0.002
random position	0.0291	2885	84	11		

Supplementary Table S10: Bioinformatic analysis of *Cla-element* transcription. Two publicly available transcriptome datasets of *C. riparius* (generated by Roche 454 sequencing) were screened for *Cla-element*-containing transcripts. We used the bwasw algorithm with default settings to map all reads against the dimeric *Cla-element* sequence described in the methods section of the main text. In dataset SRR834592 we found 29 out of 363,381 sequence reads with *Cla-element* homology, representing 0.01%. In dataset SRR834593 we found 27 out of 379,308 sequence reads with *Cla-element* homology, representing 0.01 % as well. Multiple sequence alignments showed the at least dimeric character of the *Cla-element* in most reads. We cannot distinguish reads with monomeric *Cla-element*s from reads that end at that position, so reads with no clear multimeric character of *Cla-element*s are marked with an interrogation point.

SRR834592			**SRR834593**		
read ID	alignment length	dimeric character	read ID	alignment length	dimeric character
GBCIR3W02FZBR1	236	y	GCKT4WV02FOSZ1	131	y
GBCIR3W02JMT2X	215	y	GCKT4WV02G6HFB	224	y
GBCIR3W02GRQMJ	171	y	GCKT4WV02FWFGP	125	y
GBCIR3W02JWW4M	119	y	GCKT4WV02JWBRQ	59	y
GBCIR3W02F5BA8	186	y	GCKT4WV02IA9I1	149	y
GBCIR3W02JCZ58	79	y	GCKT4WV02HM8FH	224	y
GBCIR3W02IAYP7	48	y	GCKT4WV02HPQ6J	106	y
GBCIR3W02HX11H	184	y	GCKT4WV02JQ0D3	113	y
GBCIR3W02JPXOH	140	y	GCKT4WV02FQGX7	94	?
GBCIR3W02H5M1S	122	y	GCKT4WV02HB1V1	198	y
GBCIR3W02IW29C	154	y	GCKT4WV02FHOSF	122	y
GBCIR3W02ILKBQ	116	y	GCKT4WV02G2DTB	69	?
GBCIR3W02G02NH	156	y	GCKT4WV02FTFMW	32	y
GBCIR3W02IJ48E	197	y	GCKT4WV02FJVDX	164	y
GBCIR3W02IAGR8	230	y	GCKT4WV02H6JFF	56	?
GBCIR3W02HWKDA	100	y	GCKT4WV02FKWN2	79	y
GBCIR3W02H8Z8J	97	y	GCKT4WV02HSXZQ	196	y
GBCIR3W02G4EEH	123	y	GCKT4WV02JPA7V	227	y
GBCIR3W02ICDY1	116	?	GCKT4WV02ILHAO	202	y
GBCIR3W02H5W9U	60	y	GCKT4WV02HRC0Q	178	y
GBCIR3W02FTDXO	66	?	GCKT4WV02GSFI5	172	y
GBCIR3W02JYYA0	125	y	GCKT4WV02GUTQW	96	y
GBCIR3W02HDPKU	96	?	GCKT4WV02JYKGS	57	y
GBCIR3W02GOMQ4	232	y	GCKT4WV02HI0GP	210	y
GBCIR3W02GHBXO	187	y	GCKT4WV02GUBAK	113	y
GBCIR3W02IEON4	215	y	GCKT4WV02JG7D4	89	y
GBCIR3W02IBZ1D	70	y	GCKT4WV02JPZL0	162	y
GBCIR3W02FSU5E	66	y			
GBCIR3W02FQ2BJ	189	y			

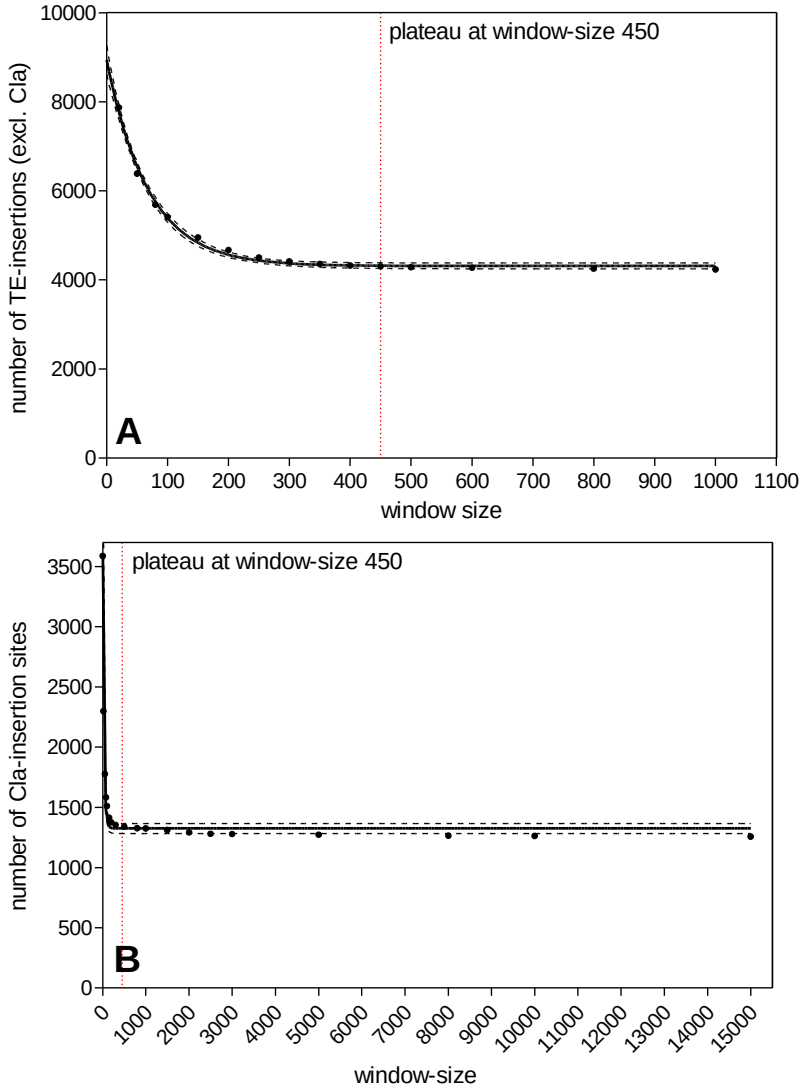
Supplementary Table S11: Number of successful FISH documentations for each chromosome of the different populations.

population	chromosome I	chromosome II	chromosome III	chromosome IV
MG×MG	6	5	5	4
SI×SI	2	1	3	1
SS×SS	4	3	1	5
MG♂×SI♀	2	3	3	3
MG♀×SI♂	5	9	6	5
SS♂×SI♀	2	3	3	2
SS♀×SI♂	3	2	4	1
SS♂×MG♀	2	4	3	6
SS♀×MG♂	2	1	1	2

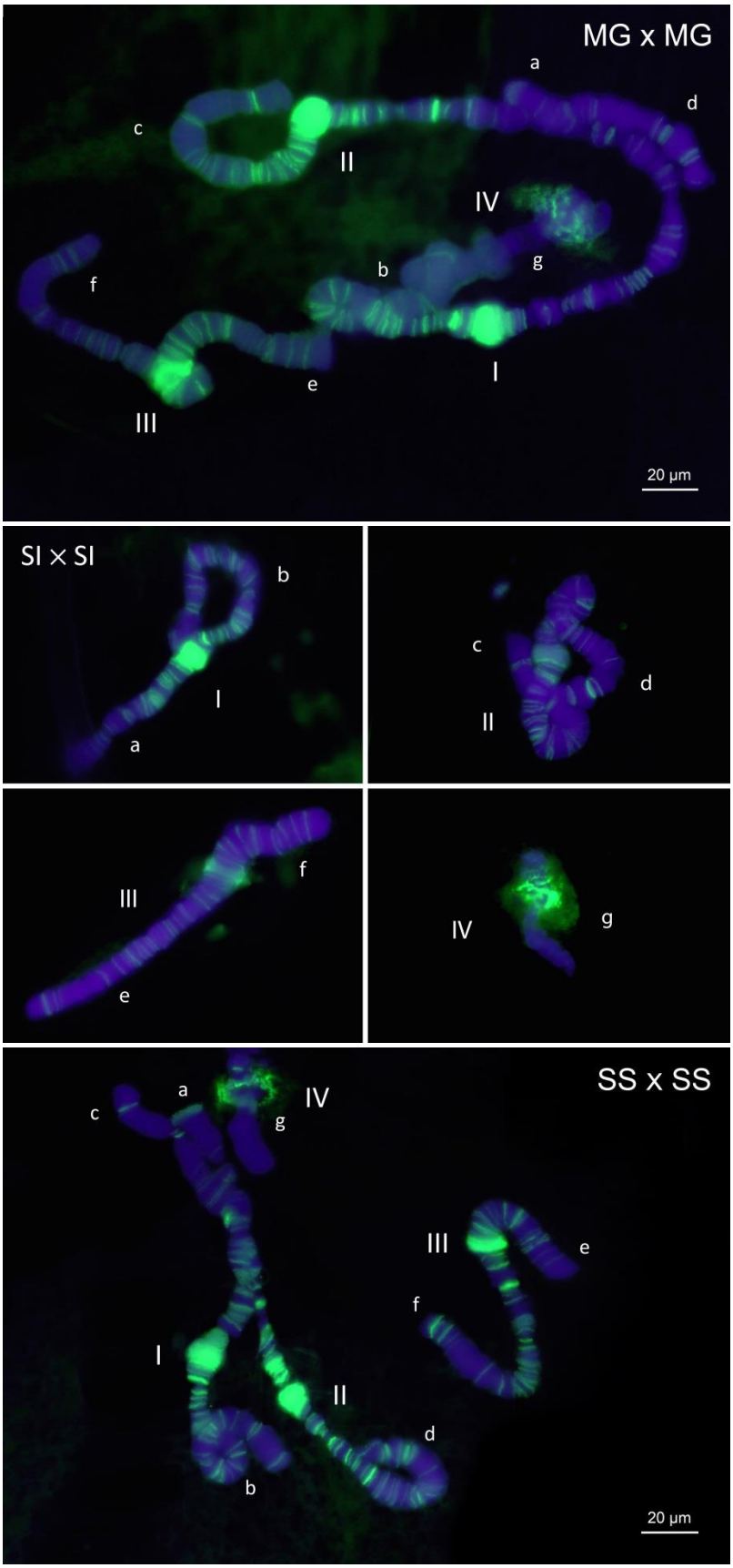
Supplementary Table S12: Comparison of the presented *draft genome* with other published *draft genome*s across the nematoceran infraorder Culicomorpha. All species with published *draft genome*s of the superfamily Chironomidae are included (marked with an asterisk). The remaining species are representatives from the only other culicomorphan superfamily Culicidae (mosquitoes). Hash keys mark genomes where updated genome metrics were available from vectorbase.org and used instead of the values from the cited publication.

	Total length (bp)	Number of scaffolds	N50	Publication
CRIP_Laufer*	180,652,019	5,292	272,065	Present study
*C. riparius**	154,534,000	29,677	7,097	Vicoso & Bachtrog 2015
*C. tentans**	213,463,000	26,025	7,697	Kutsenko et al. 2014
*Belgica antarctica**	89,583,700	4,997	13,687	Kelley et al. 2014
Culex quinquefasciatus[#]	579,057,705	3,172	486,756	Behura et al. 2011
Anopheles gambiae[#]	273,109,044	8	49,364,325	Holt et al. 2002
Aedes aegypti	1,383,957,531	4,756	1,547,048	Nene et al. 2007

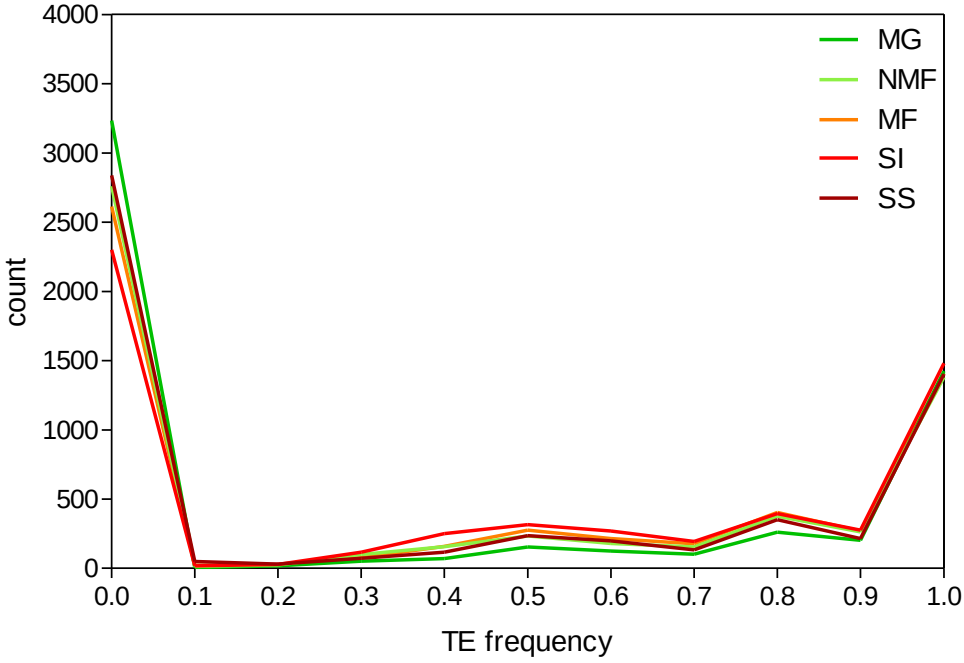
Supplementary Figures



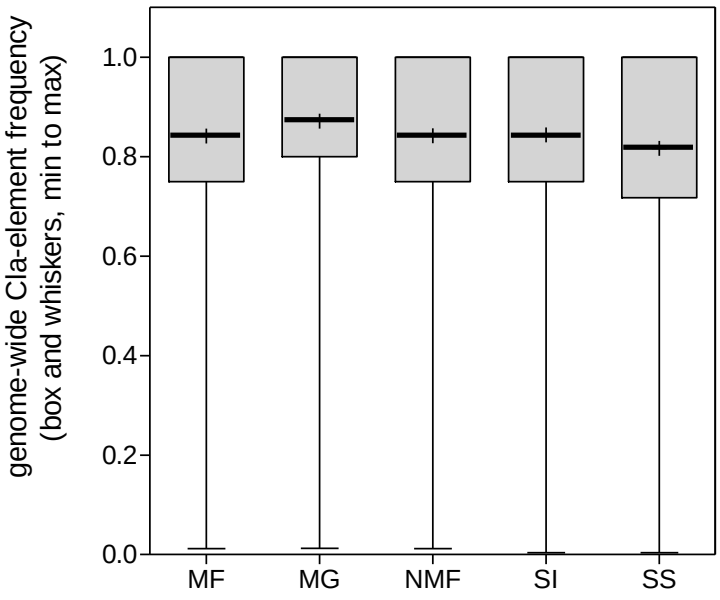
Supplementary Figure S1: Exponential one-phase decay model of increasing window-size with the number of (A) TE-insertions (exclusively *Cla-element* insertions) and (B) *Cla-element* insertions. In order to correct for variation between populations (insert size differences or indels), insertions of respective TEs in the different populations were merged to one single insertion (custom script temerger.py, Supplementary Material SM4). Plateau of both associations is reached at an approximate window-size of 450b.



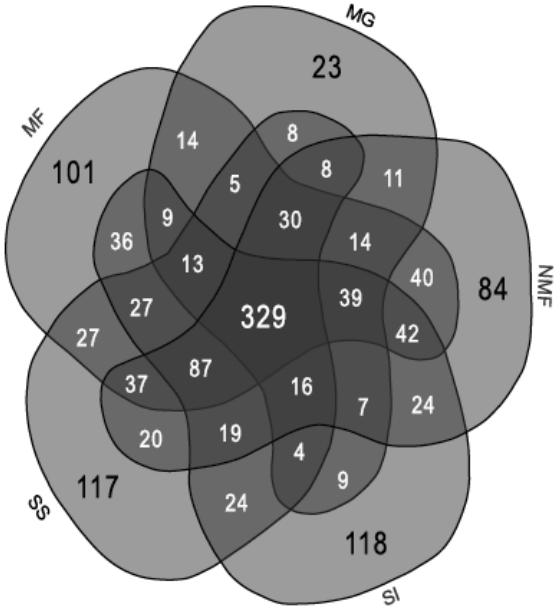
Supplementary Figure S2: Fluorescent *in-situ* hybridisation (FISH) micrographs of the *Cla-element* dimer clone *pAD* 2603 to polytene chromosomes of different *C. riparius* pure populations (MG – Hesse, SI – Piemont, SS – Andalusia). Pictures of DAPI-stained chromosomes were overlayed by the FISH documentation. Chromosomes are labelled with roman numbers, chromosomal arms with letters.



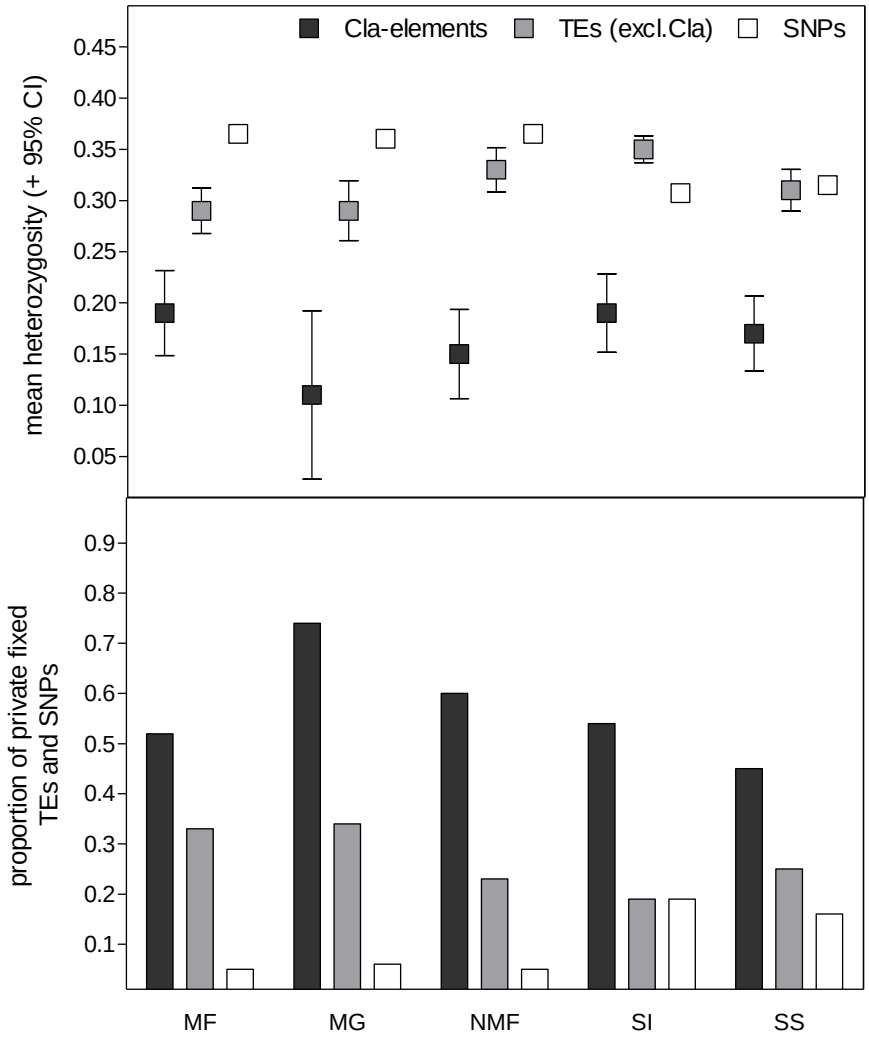
Supplementary Figure S3: Frequency distribution of all transposable elements (TE) identified in different *C. riparius* populations. Counts of absent insertions (frequency=0) result from comparisons between population (present insertion in at least one population defines a potential insertion site in other populations). In general, most TE-insertions are either fixed (frequency=1) within a population or absent.



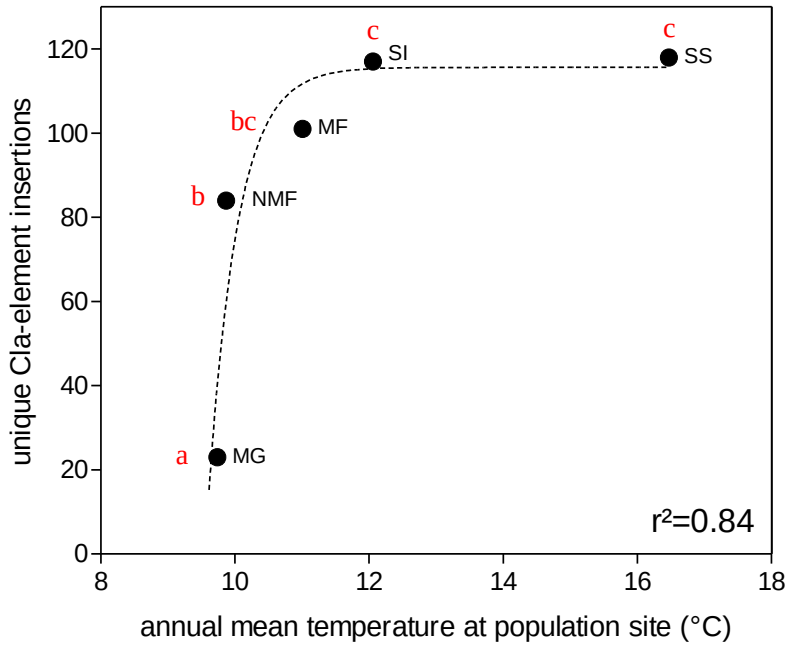
Supplementary Figure S4: Frequency of *Cla-element* insertions across the genome (absent insertions not taken into account) per population. Mean frequencies marked as horizontal bolt bars.



Supplementary Figure S5: Venn diagram of genome-wide *Cla-element* insertions in *C. riparius* including insertions at all frequencies. Sum of insertions present in all populations are in the centre, population-specific private insertions are peripheral.



Supplementary Figure S6: (**A**) Mean of heterozygosity on private *Cla-element* insertions, TE insertions excluding *Cla-elements*, and genome-wide SNPs, respectively in all *C. riparius* populations. (**B**) Proportion of fixed private to all private *Cla-element* insertions, other TE insertions, and SNPs in all *C. riparius* populations.



Supplementary Figure S7: Relation of the annual mean temperature at the population site with population-unique *Cla-element* insertions described as exponential one-phase decay function (given regression r^2). Pairwise comparisons of unique to total number of *Cla-element* insertions per population gave statistical support to the pattern (significance ($p < 0.05$) is indicated by different letters (a, b, c)).