### General information

This file explains the contents of Supplementary Files S2 – S11.

The numbering of baits and loci follows the numbering of contigs in the CLC assembly (ENA project accession PRJEB35099) from which they were derived. For example, CLC contig\_2892421 is the source of bait sequences Bv-contig2892421 \_0, Bv-contig2892421 \_50, Bv-contig2892421 \_100 and Bv-contig2892421 \_150 (File S4) that jointly capture the mapped locus 2892421 (FileS6). That particular CLC contig has BLAST+ matches to *Sox*3 genes in *Glandirana rugosa* and *Bombina orientalis* (File S7)*.*

In Files S3 and S5, the following names are applied to the F0 and F1 samples (ENA sample accession after the comma):

BvvGP = *B. v. variegata* F0 male, ERS3926742

BbGP = *B. bombina* F0 female, ERS3926743

F1M = F1 male (sire for families 6 and 7), ERS5083645

F1Fam6F = F1 female (family 6), ERS5083565

F1Fam7F = F1 female (family 7), ERS5083646

### **File S2 Fast Vector (FastVec) Mathematica script**

### The FastVec v2.2beta.cdf script can be opened using the (free) Wolfram Player: <https://www.wolfram.com/player/>

### File S3 Lep-MAP3 input file

The Lep-MAP3 input file contains 4799 loci that were suitable for mapping. Starting with line 7, each line represents a locus. The first column contains the name of the cluster to which the locus was assigned in an initial graph-based clustering analysis. The second column is the locus name. Typically, these columns give the scaffold name and sequence position, respectively. Both columns are ignored by Lep-MAP3.

### FileS4 bait sequences

The DNA sequences for 20,000 baits (100 base long each) in fasta format, derived from the *B. v. variegata* CLC genome assembly. There are four baits per target. See also the general information above.

### **File S5 Diplotype matrix from FastVec analysis**

This file contains the output from the FastVec Mathematica script (File S2) for each of the 5000 enrichment targets (= loci). Maximum likelihood diplotypes for the F0 (*n* = 2), F1 (*n* = 3) and F2 (*n* = 162) generations are listed in columns labeled <sample\_name>\_HHHest. Support estimates are listed in columns labeled <sample\_name>\_HHHsup. See the general information above regarding the naming of F0 and F1 individuals.

### File S6 Linkage map file

A list of the 4755 mapped loci, sorted by linkage group (LG) and map position. Segregation distortion was estimated separately for families 6 and 7: ‘chisqr\_fam#’ = χ2 estimate, ‘df\_fam#’ = degrees of freedom associated with χ2. ‘fdr\_fam#’ = FDR test significant (TRUE/FALSE). The critical value for the Bonferroni test (at experiment-wise α = 0.05, corrected for multiple comparisons, see publication) is 12.35. For the following three fields 0 = FALSE and 1 = TRUE: ‘rescored’ = locus was rescored based on manually selected variants; ‘js\_add’ = locus was added to the map via the Lep-MAP3 JoinSingles2All module, with reduced stringency; ‘synt\_stray’ = stray alignment in *X. tropicalis* synteny analysis. Annotations (0 if none exists): ‘ncbi’ = NCBI accessions (comma separated) with BLAST+ match in the CLC contig (BLAST+ data in FileS7). ‘ips’ = A comma-separated list of *B. v. variegata* transcripts that were used to scaffold the CLC contig associated with this locus and that also have InterProScan annotation (annotations in File S8).

### File S7 NCBI Blast+ records

BLAST+ matches to CLC contigs in standard tabular format, with the following added fields: gene description (column 2), query length (column 14) and subject length (column 15). It includes matches to three microsatellite markers that were used in previous *Bombina* hybrid zones analyses. Column ‘ncbi’ in File S6 links to these annotations.

### File S8 InterProScan records

The file provides Pfam annotations for *B. v. variegata* transcripts that are represented in the linkage map. The file has standard InterProScan tabular format:

1. Protein accession, 2. Sequence MD5 digest, 3. Sequence length, 4. Analysis, 5. Signature accession, 6. Signature description, 7. Start location, 8. Stop location, 9. Score (e-value), 10. Status of match (T: true), 10. Date of run, 11. InterPro annotations - accession, 12. InterPro annotations – Description.

Column ‘ips’ in File S6 links to these annotations.

**File S10 REPdenovo assembly**

The file in gzipped fasta format provides the assembly of *B. v. variegata* genomic repeats based on k-mers whose frequency in WGS reads exceeds the average k-mer frequency by a factor of 100 or more. We supply the unmerged version of this assembly that was used as a mapping reference to generate repeat-subtracted read sets.

**File S11 *B. variegata* mapping reference**

The file in gzipped fasta format contains the mapping reference for the 5000 target loci. Repeat-subtracted reads from bait-enriched libraries were mapped to this reference.