Supporting Information

Klasson et al. 10.1073/pnas.0810753106

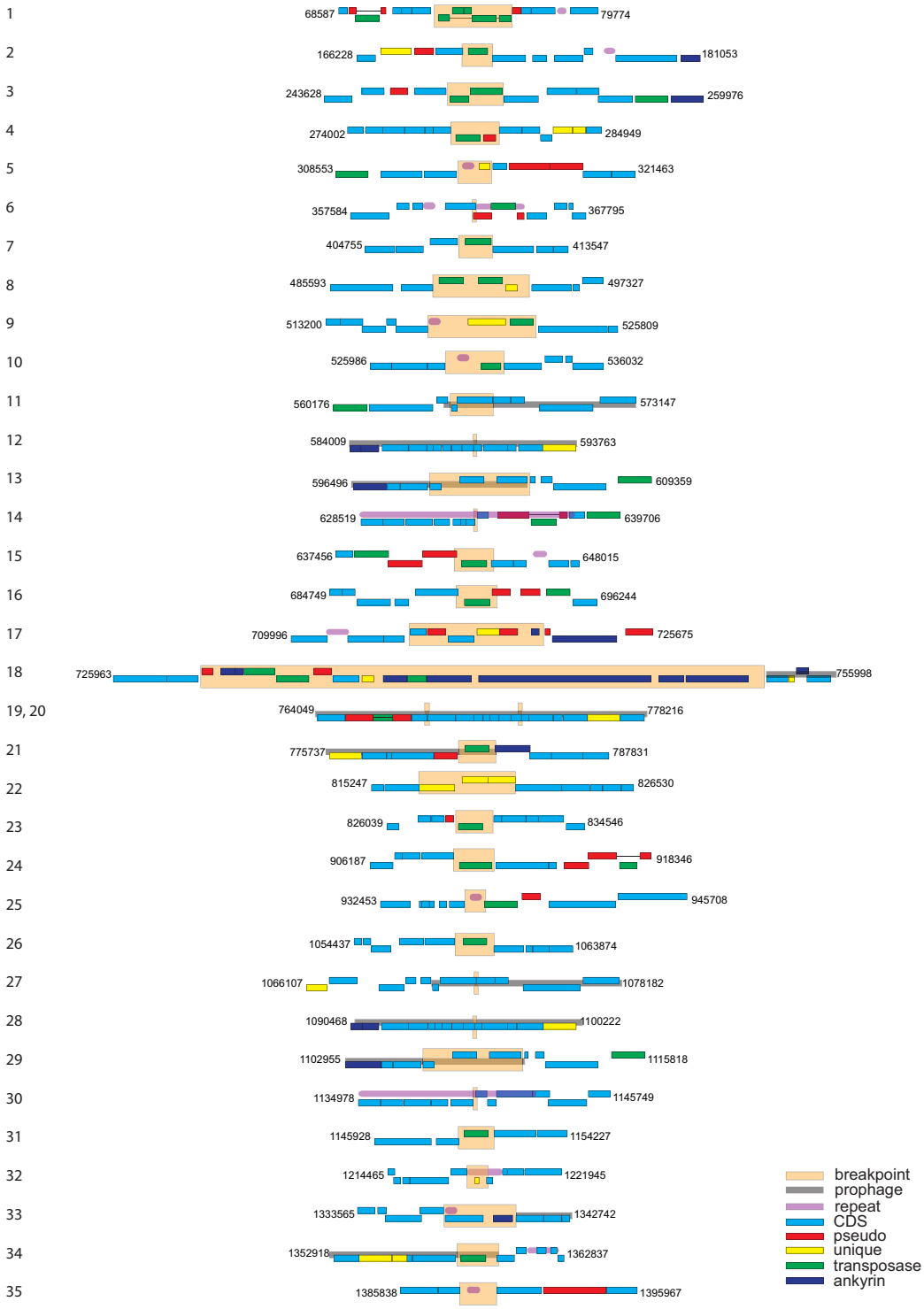

Fig. S1. Gene content at the 35 identified breakpoint positions (*light-orange boxed regions*) in the *w*Ri genome where the gene order differs relative to the *w*Mel genome. The numbers show the chromosomal position of the depicted region in the *w*Ri genome.

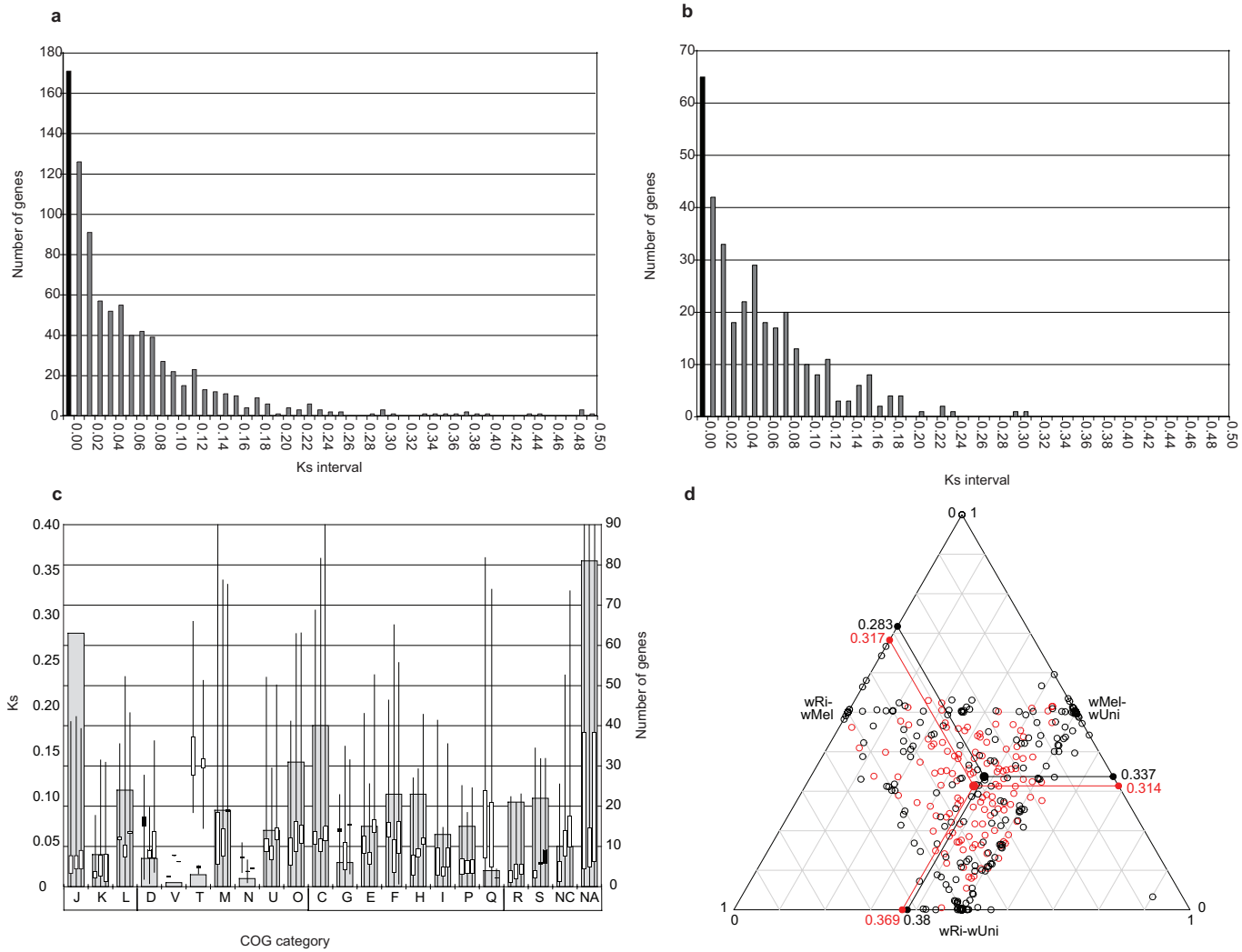

Fig. S2. Histogram showing the number of homologous gene pairs within a certain interval of substitution frequencies at synonymous sites (Ks) inferred from (*a*) 897 *w*Ri/*w*Mel gene pairs with Ks values between 0 and 0.5 and (*b*) 343 *w*Ri/*w*Mel core genes that have homologs in *Wolbachia w*Bm, *Orientia tsutsugamushi*, and 8 *Rickettsia* species [Fuxelius HH, Darby AC, Cho NH, Andersson SG (2008) Visualization of pseudogenes in intracellular bacteria reveals the different tracks to gene destruction. *Genome Biol* 9:R42]. The first bar in both plots contains genes with a Ks value of zero. (*c*) Substitution frequency variation in different functional categories estimated from 445 homologous gene triplets in *w*Ri, *w*Mel, and *w*Uni. The gene triplets were sorted according to COG categories, in which each group of 3 represents the pair-wise comparisons between first *w*Ri and *w*Mel, second *w*Mel and *w*Uni, and third *w*Ri and *w*Uni. The gray bar shows the number of genes, the black line shows the minimum and maximum Ks-value, and the black or white box shows the median and mean values of Ks within each functional category. The box is white if the median value is lower than the mean and black if the median is higher than the mean. (*d*) Ternary plot showing variation in substitution frequencies for *Wolbachia* genes that have been tested positive for recombination by RDP in red and genes where no recombination was detected by RDP in black. The mean relative Ks-values for each group are shown in the corresponding colors.

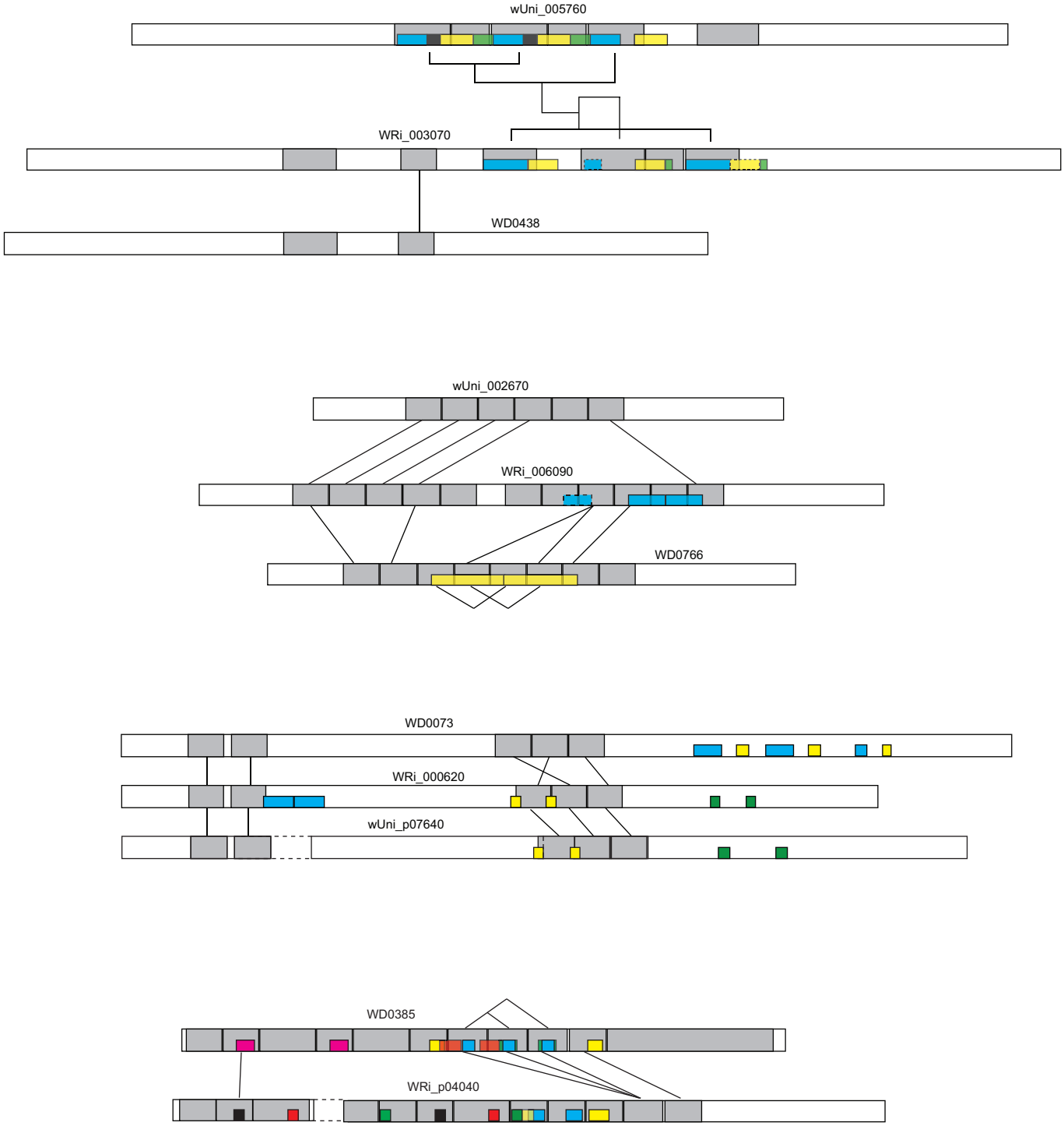

Fig. S3. Domain and repeat organization in 4 homologous ankyrin repeat genes in *w*Uni, *w*Ri, and *w*Mel. The gray boxes show the ankyrin repeats found using hmmsearch with the PFAM hmm for ankyrin repeat and the black lines show relationships across ankyrin repeats where the posterior probability from the Bayesian phylogenetic analysis of the ANK domains (Fig. 4) is greater than 0.9. Colored boxes with solid lines represent identical repeats, and colored boxes with dashed lines represent repeats with identity higher than 90%. Each color represents a different repeat inside a particular gene and colors do not necessarily correspond to the same repeat in the homologous gene. The dashed lines between 2 different parts of a gene show the position of an IS-element insertion in the gene.

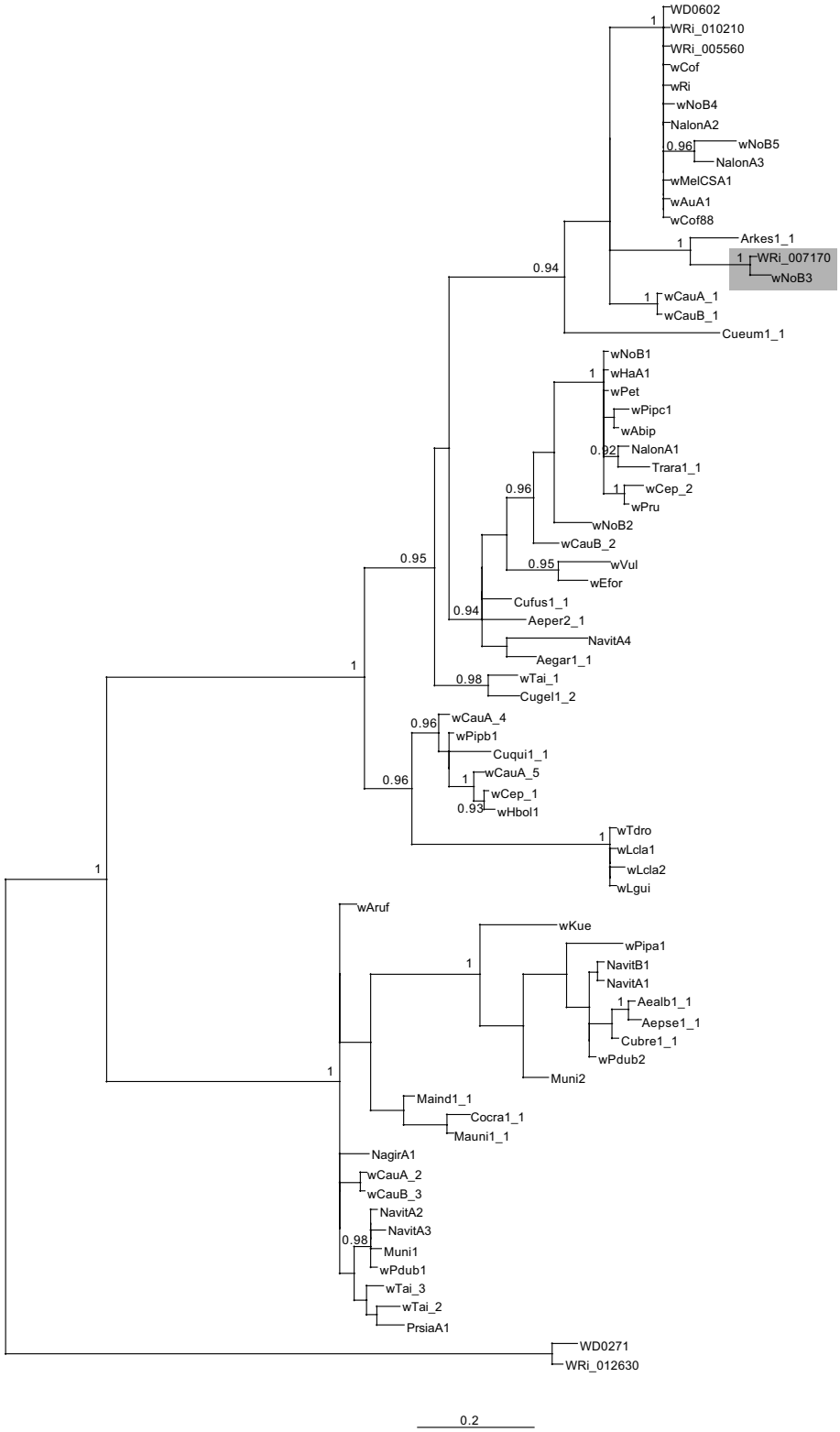

Fig. S4. Phylogenetic analysis based on the minor capsid protein using the Bayesian method. Numbers at nodes show posterior probabilities. The gray box highlights the horizontal transfer of the WOC prophage between *w*No and *w*Ri.

Table S1. Features of ankyrin repeat containing genes identified in the *w*Ri genome with highly similar positional homologs in the *w*Mel genome

*w*Ri			*w*Mel			*w*Uni			*w*Ri-*w*Mel		*w*Ri-*w*Uni		*w*Uni-*w*Mel	
Gene	Length (aa)	ANKs[a]	Gene	Length (aa)	ANKs[a]	Gene	Length (aa)	ANKs[a]	K_a	Ks	K_a	Ks	K_a	Ks
WRi_000280	721	0–3	WD00032/3/4	859	0–3	WUni_p06070	622	0–3	0.013	0.027	0.048	0.103	0.004	0.015
WRi_000290	288	6	WD0035	288	6	WUni_006060	288	6	0.000	0.000	0.003	0.000	0.003	0.000
WRi_001790	268	1	WD0191	268	1	WUni_004090[b]	214	1	0.021	0.030	0.008	0.000	0.020	0.029
WRi_002480	454	10–11	WD0498/9	454	9	WUni_009360	454	10–11	0.000	0.005	0.006	0.031	0.006	0.037
WRi_003040	400	1–2	WD0441	402	1–2	WUni_003380[b]	535	1–2	0.016	0.047	0.016	0.030	0.005	0.017
WRi_005390 WRi_010050	966	3–4	WD0633	966	3–4	-	-	-	0.006	0.006	-	-	-	-
WRi_005440 WRi_010100	152	3	WD0636	152	3	-	-	-	0.003	0.000	-	-	-	-
WRi_005450 WRi_010110	257	3	WD0637	208	3	-	-	-	0.002	0.006	-	-	-	-
WRi_005620 WRi_010280	493	9–10	WD0596	493	9–10	-	-	-	0.001	0.000	-	-	-	-

[a]Number of ANK domains found with Interproscan in searches against the PFAM and SMART databases. An interval is shown if the searches against the two databases yielded different results.

[b]Partial sequences, located at contig ends.

K_a, Nonsynonymous substitutions per site; Ks, Synonymous substitutions per site.

Table S2. Features of ankyrin repeat containing genes identified in the *w*Ri genome with highly divergent positional homologs in the *w*Mel genome

*w*Ri			*w*Mel			*w*Uni			*w*Ri-*w*Mel		*w*Ri-*w*Uni		*w*Uni-*w*Mel	
Gene	Length (aa)	ANKs[a]	Gene	Length (aa)	ANKs[a]	Gene	Length (aa)	ANKs[a]	K_a	Ks	K_a	Ks	K_a	Ks
WRi_000620	698	5–6	WD0073	800	5–6	WUni_p07640[b]	722	5–6	0.109	0.139	0.068	0.074	0.061	0.072
WRi_001060	1094	10	WD0147	954	11	-	-	-	0.429	0.990	-	-	-	-
WRi_003070	929	7–8	WD0438	632	2	WUni_005760	787	6–7	0.280	0.752	0.296	0.843	0.389	1.134
WRi_003100	237	0–2	WD0434	251	0	-	-	-	0.291	0.630	-	-	-	-
WRi_p03640[b]	371	6	WD0550	329	6	WUni_009500	309	5	0.556	1.283	0.210	0.477	0.319	1.118
WRi_p04040[b]	610	12	WD0385	542	10–11	WUni_004210	193	1	0.172	0.167	0.106	0.109	0.327	0.475
WRi_006090	615	11	WD0766	474	8	WUni_002670	427	6	0.071	0.122	0.028	0.024	0.125	0.112
WRi_006690	926	2–3	WD0633	966	3–4	-	-	-	0.416	1.914	-	-	-	-
WRi_006900	178	4	WD0566	173	2–3	-	-	-	0.503	83.939	-	-	-	-
WRi_p07220[b]	374	7	WD0596	493	9–10	-	-	-	0.305	2.856	-	-	-	-
WRi_007240	498	2–3	WD0753/4	544	2–4	WUni_p04990[b]	497	2–3	0.231	0.550	0.390	1.112	0.335	0.604
WRi_p07650[b]	110	2	WD0596	493	9–10	-	-	-	0.140	3.557	-	-	-	-

[a]Number of ANK domains found with Interproscan when searching against the PFAM and SMART databases. An interval is shown if the searches against PFAM and SMART yielded different results.

[b]Genes disrupted by IS-element insertions.

K_a, Nonsynonymous substitutions per site; Ks, Synonymous substitutions per site.

Table S3. Features of ankyrin repeat containing genes uniquely identified in the *w*Ri genome without homologs in the *w*Mel genome

wRi

Gene	Length (aa)	ANKs[a]
WRi_002880	491	6–7
WRi_p06670	95	2
WRi_006740	205	3
WRi_006750	110	1–2
WRi_006810	336	3–4
WRi_p06840[b]	617	16–17
WRi_006850	2474	34
WRi_006860	359	4
WRi_006870	866	21–22
WRi_012460	270	5

[a]Number of ANK domains found with Interproscan when searching against the PFAM and SMART databases. An interval is shown if the searches against PFAM and SMART yielded different results.
[b]Genes disrupted by IS-element insertions.

Table S4. General features of the *w*Ri genome compared to *w*Mel

	*w*Ri	*w*Mel
Genome size (bp)	1,445,873	1,267,782
CDSs	1,150	1,196
Pseudogenes	114	113
tRNAs	34	34
rRNA	One of each	One of each
GC %	35.2	35.2
Average length of CDS	976	851
Coding density %	77.7	80.3

Table S5. IS-elements identified in the *w*Ri genome

Name[a]	Family[a]	Group[a]	Color[b]
ISWpi1	IS5	IS1031	Light brown
ISWpi2	IS481		Green
ISWpi3	IS3	IS3	Dark purple
ISWpi4	IS481		Purple
ISWpi5	ISNCY	ISBst12	Deep pink
ISWpi6	IS4	IS231	Dark orange
ISWpi7	IS110		Light orange
ISWpi8	IS110		Light pink
ISWpi9	IS5	IS903	Blue
ISWpi10	IS630		Yellow
ISWpi11	IS630		Black

[a]Names are given in accordance with the IS nomenclature rules as defined in the IS Finder database

[b]Color refers to the color-coding of the IS-elements in Fig. 1.

Table S6. Genes uniquely present in the *w*Ri and *w*Mel genomes, respectively

	*w*Ri[a]	*w*Mel[a]
Hypothetical	31 (9)	35 (19)
ANKs	9 (0)	5 (2)
Prophage	7 (4)	4 (2)
Other	8 (5)	8 (7)
IS-element	5 (3)	2 (1)

[a]Genes present in multiple copies are only counted once. The number of genes present as pseudogenes in the other genome is shown in parenthesis. Unique pseudogenes are not included.

Table S7. *Drosophila* lines used in the breakpoint PCR assay

Host	Infection	Lines	Year when line established
D. simulans	Natural	KB28	1986 (1)
D. simulans	Transinfected	KB55	1994 (2)
D. simulans	Transinfected, Natural *w*No, *w*Ha	KB16, KB18	1997 (3)
D. yakuba	Transinfected	KB82, KB83	1996 (4)
D. teissieri	Transinfected	KB114, KB115	1999 (4)
D. santomea	Transinfected	KB127, KB128	2000 (4)

1. Hoffmann AA, Turelli M, Simmons GM (1986) Unidirectional Incompatibility between Populations of Drosophila simulans. *Evolution* 40:692–701.
2. Giordano R, O'Neill SL, Robertson HM (1995) *Wolbachia* infections and the expression of cytoplasmic incompatibility in *Drosophila sechellia* and *D. mauritiana*. *Genetics* 140:1307–1317.
3. Rousset F, Braig HR, O'Neill SL (1999) A stable triple *Wolbachia* infection in *Drosophila* with nearly additive incompatibility effects. *Heredity* 82 (Pt 6):620–627.
4. Zabalou S, et al. (2004) Natural *Wolbachia* infections in the *Drosophila yakuba* species complex do not induce cytoplasmic incompatibility but fully rescue the *w*Ri modification. *Genetics* 167:827–834.

Table S8. Median synonymous substitution-frequency per site (Ks), with strains and number of genes included detailed in Fig. 2

Strain or species	Median Ks
Wolbachia	
*w*Mel-*w*Uni	0.024
*w*Ri-*w*Uni	0.035
*w*Ri-*w*Mel	0.025
*Rickettsia**	
Rc-Rm	0.039
Rr-Rm	0.043
Rr-Rc	0.023
Neisseria meningitidis	
FAM18-MC58	0.056
053442-FAM18	0.053
053442-MC58	0.060
Staphylococcus aureus subsp. aureus	
MRSA252-COL	0.047
MRSA151-MSSA476	0.048
MSSA476-COL	0.012

*Species: Rc, *Rickettsia conorii*; Rm, *R. massiliae*; Rr, *R. rickettsii*.

Table S9. Tissue and developmental stage-specific RNA expression of the ANK genes in *w*Ri

*w*Ri gene	*w*Mel gene	♂	♀	T	O	2 h	o/n
WRi_000280	WD0032/3/4				Not tested		
WRi_000290	WD0035	n^+	n^+	n^+	n^+	n^+	n^+
WRi_001790	WD0191	n^+	n^+	n^+	w^+	n^+	n^+
WRi_002480	WD0498	+	+	+	+	w^+	+
WRi_003040	WD0441	+	+	+	+	+	+
WRi_005390	WD0633	+	+	+	+	+	+
WRi_010050							
WRi_005440	WD0636	n^+	n^+	n^+	n^+	n^+	n^+
WRi_010100							
WRi_005620	WD0596	n^+	n^+	n^+	+	n^+	n^+
WRi_010280							
WRi_000620	WD0073	n^+	n^+	n^+	n^+	n^+	n^+
WRi_001060	WD0147	+	+	+	+	+	+
WRi_003070	WD0438	n^-	n^+	n^-	n^+	n^+	n^+
WRi_003100	WD0434				Not tested		
WRi_p03640	WD0550	+	+	+	+	+	+
WRi_p04040[a]	WD0385	$n^{-/+}$	$n^{-/+}$	$n^{+/+}$	$n^{+/+}$	n^+/n^+	n^+/n^+
WRi_006090	WD0766	+	+	+	+	+	+
WRi_006690	WD0633	+	+	+	+	+	+
WRi_006900	WD0566	+	+	+	+	+	+
WRi_p07220	WD0596	+	+	+	+	+	+
WRi_007240	WD0754	+	+	+	+	+	+
WRi_p07650	WD0596	+	+	+	+	+	n^-
WRi_002880	Absent	+	+	+	+	+	+
WRi_p06670	Absent				Not tested		
WRi_006740	Absent	+	+	w^+	+	+	w^+
WRi_006750	Absent	+	+	+	+	+	+
WRi_006810	Absent	+	+	w^+	+	n^-	+
WRi_p06840	Absent	+	+	+	+	+	+
WRi_006850	Absent	+	+	+	+	w^+	+
WRi_006860	Absent	+	+	+	+	+	+
WRi_006870	Absent	+	+	w^+	+	n^-	n^-
WRi_012460	Absent	n^-	n^-	n^-	n^-	n^-	n^-

Absent, absence of a homologous gene in *w*Mel; +, positive using conventional RT-PCR; w^+, weakly positive using conventional RT-PCR; n^+, positive using nested RT-PCR; n^-, negative using both conventional and nested RT-PCR. ♂, male adult; ♀, female adult; T, testis; O, ovaries; 2 h, early embryos 2 h; o/n, late embryos overnight.

[a]WRi_p04040 has an IS-insertion. The two PCR results represent the 5′ and 3′ part of the gene respectively, not spanning the IS-insertion.

Table S10. Tissue and developmental stage-specific RNA expression of ANK genes in the B-group strains *w*No, *w*Mau, and *w*Ma

	*w*No						*W*Mau						*w*Ma					
Genes	♂	♀	T	O	2 h	o/n	♂	♀	T	O	2 h	o/n	♂	♀	T	O	2 h	o/n
WRi_006870	n^+	n^+	n^-	n^+	n^-	n^+	n^-	n^+	n^+	n^+	n^+	n^+	n^-	n^-	n^+	n^+	n^+	n^+
WRi_006900	+	+	+	+	n^+	n^-	+	+	+	+	+	+	+	+	+	+	n^+	+

Absent, absence of a homologous gene in *w*Mel; +, positive using conventional RT-PCR; w^+, weakly positive using conventional RT-PCR; n^+, positive using nested RT-PCR; n^-, negative using both conventional and nested RT-PCR. ♂, male adult; ♀, female adult; T, testis; O, ovaries; 2 h, early embryos 2 h; o/n, late embryos overnight.