Supplementary information

Evolutionary dynamics of enlarged sex chromosomes and novel pseudoautosomal regions in Sylvioidea songbirds

Hanna Sigeman & Bengt Hansson

This supplementary information includes **Tables S1-S6** and **Figures S1-S12**.

Table S1. Ancestral chromosomal positions, according to the zebra finch genome structure, of sex-linked genomic regions detected in previous studies across the Sylvioidea clade. Group numbers correspond to those in Figure 1. Two outermost edges are defined for each coherent non-recombining region ("NRR edges"), the proximal edge being the one closer to the beginning of the chromosome (0 Mb) and the distal edge being the one closer to the distal part of the chromosome. In each of the groups V-VII, there are two separate sex-linked regions ("R1" and "R2") on chromosome 3, and they have therefore one set of NRR edges for each region and group. The asterisks (* and **) mark the previously known fusion point between chromosomes Z and 4A (between the distal end of chromosome Z, position 72.9 Mb, and the distal NRR edge on chromosome 4A, position 9.6 Mb).

Chromosome		**NRR (Mb)**	**NRR edges (Mb)**		**NRR present (X) or absent (-) in Group nr.**						
ID	**Length (Mb)**		**Proximal**	**Distal**	**I**	**II**	**III**	**IV**	**V**	**VI**	**VII**
Z	72.9	0.0–72.9	0.0	72.9*	X	X	X	X	X	X	X
4A	20.7	0.0–9.6	0.0	9.6**	-	X	X	-	X	X	X
4A	20.7	5.3–9.6	5.3	9.6**	-	-	-	X	-	-	-
4	69.8	13.8–49.1	13.8	49.1	-	-	X	-	-	-	-
8	28.0	7.3–21.8	7.3	21.8	-	-	-	X	-	-	-
3	112.6	8.4–10.4 (R1); 18.1–24.1 (R2)	8.4 (R1); 18.1 (R2)	10.4 (R1); 24.1 (R2)	-	-	-	-	X	-	-
3	112.6	8.4–14.0 (R1); 18.1–24.1 (R2)	8.4 (R1); 18.1 (R2)	14.0 (R1); 24.1 (R2)	-	-	-	-	-	X	-
3	112.6	8.4–24.1 (R1); 29.8–88.0 (R2)	8.4 (R1); 29.8 (R2)	24.1 (R1); 88.0 (R2)	-	-	-	-	-	-	X
5	62.4	9.1–45.4	9.1	45.4	-	-	-	-	-	-	X

Table S2. Songbird PAR genes identified in the zebra finch annotation. The column "Gene name" corresponds to the gene names given in the Main text.

Gene name	Zebra finch gene ID (bTaeGut2.pat.W.v2)	Transcript ID	Scaffold	Start (bp)	End (bp)	protein coding
uncharacterized1	LOC116806781	XR_004365834.1	NC_045027.1	11005	11912	no
LMAN1	LOC100220399	XM_030258642.2	NC_045027.1	33731	58078	yes
uncharacterized2	LOC116806731	XR_004365759.1	NC_045027.1	52211	71108	no
RAX	RAX	NM_001243734.2	NC_045027.1	64707	66358	yes
GRP	LOC105760884	XM_030258647.2	NC_045027.1	71706	76475	yes
SEC11C	LOC116806596	XM_030257196.2	NC_045027.1	79268	84487	yes
ZNF532	LOC116806602	XM_032744080.1	NC_045027.1	90198	124057	yes
MALT1	LOC116806749	XM_032744259.1	NC_045027.1	135069	153662	yes
ALPK2	LOC116806817	XM_032744489.1	NC_045027.1	153718	182131	yes
uncharacterized3	LOC116806818	XR_004365896.1	NC_045027.1	182242	185760	no
MIR122	MIR122	NR_049051.1	NC_045027.1	182594	182665	no
NEDD4L	LOC116806603	XM_032744495.1	NC_045027.1	203780	290722	yes
uncharacterized4	LOC115491286	XR_003957248.2	NC_045027.1	209240	211463	no
uncharacterized5	LOC116806819	XR_004365897.1	NC_045027.1	215304	221401	no
ATP8B1	LOC116806743	XM_032744197.1	NC_045027.1	304375	325804	yes
NARS1	NARS1	XM_030259060.2	NC_045027.1	325874	333111	yes
FECH	LOC116806605	XM_030259064.2	NC_045027.1	334740	346420	yes
ONECUT2	LOC100226718	XR_003957244.2	NC_045027.1	350233	367344	yes
ST8SIA3	LOC116806742	XM_032744196.1	NC_045027.1	370724	378119	yes
uncharacterized6	LOC115491263	XR_003957222.2	NC_045027.1	379525	398137	no
WDR7	LOC116806851	XM_032744613.1	NC_045027.1	400103	447553	yes
TXNL1	LOC100218989	XM_032744615.1	NC_045027.1	444330	457645	yes
uncharacterized7	LOC116806852	XR_004365928.1	NC_045027.1	484714	493199	no

Table S3. Samples used for analyses. All samples were sequenced by us except for collared flycatcher (indicated with asterisks), which was downloaded from NCBI.

Species	**Common name**	**Female**	**Male**
Acrocephalus schoenobaenus	sedge warbler	H9-20_S16_L008	HB-03_S5_L003
Aegithalos caudatus	long-tailed tit	QF-1504-CP59475_S11_L004	QF-1504-BL37630_S12_L004
Alauda arvensis	eurasian skylark	QL-1681-19_S46_L006	QL-1681-21_S47_L006
Alauda razae	Raso lark	QL-1681-95694_S52_L008	QL-1681-246_S51_L008
Cecropis daurica	red-rumped swallow	QF-1504-P182141_S2_L001	QF-1504-P182142_S1_L001
Cettia cetti	Cetti's warbler	QF-1504-P182137_S9_L003	QF-1504-2L18122_S4_L002
Eremophila alpestris	horned lark	QF-1504-H-19_S8_L003	QF-1504-H-88_S7_L003
Cisticola juncidis	zitting cisticola	QF-1504-CISJUN-2_S6_L002	QF-1504-RA5680_S5_L002
Locustella luscinioides	Savi's warbler	QF-1504-LOCLUS-43_S1_L001	QF-1504-LOCLUS-24_S3_L001
Panurus biarmicus	bearded reedling	QF-1504-2KR32024_S2_L001	QF-1504-1ET92164_S3_L001
Phylloscopus collybita	common chiffchaff	QF-1504-R86159_S5_L002	QF-1504-Z81303_S4_L002
Pycnonotus barbatus	common bulbul	SJ-2333-Pbar-197_S24_L002	SJ-2333-Pbar-421_S22_L002
Sylvia atricapilla	eurasian blackcap	1EL38952_S2_L001	1EV02922_S4_L002
Sylvietta brachyura	northern crombec	SJ-2333-Sbra-553_S28_L002	SJ-2333-Sbra-878_S26_L002
Argya altirostris	Iraq babbler	SJ-2333-IB-2b_S32_L002	SJ-2333-IB-1a_S31_L002
Ficedula albicollis	Collared flycatcher	ERR637378*	ERR637360*

Table S4. Published reference genomes used in the study to identify fusion points and novel PARs, and what samples (from Table S3) was aligned to each reference genome (see Methods).

Reference genomes (downloaded)					**Analysed with (Table S3)**	
Species	**Common name**	**Assembly name**	**NCBI assembly accession**	**Sex**	**Female**	**Male**
Alauda arvensis	Eurasian skylark	skylark_genome	GCA_902810485.1	male	QL-1681-19_S46_L006	QL-1681-21_S47_L006
Alauda arvensis	Eurasian skylark	skylark_genome	GCA_902810485.1	male	QL-1681-95694_S52_L008	QL-1681-246_S51_L008
Cisticola juncidis	zitting cisticola	ASM1340021v1	GCA_013400215.1	male	QF-1504-CISJUN-2_S6_L002	QF-1504-RA5680_S5_L002
Eremophila alpestris	horned lark	CLO_EAlp_1.0	GCA_009792885.1	male	QF-1504-H-19_S8_L003	QF-1504-H-88_S7_L003
Sylvietta virens	green crombec	ASM1339951v1	GCA_013399515.1	male	SJ-2333-Sbra-553_S28_L002	SJ-2333-Sbra-878_S26_L002

Table S5. Scaffolds suggestive of fusion points between chromosome 4A and chromosome 3.

Species	Description	Scaffold ID	Scaffold range (bp)	Zebra finch range (bp)	Sex-linked
Eremophila alpestris	Fusion scaffold (Z;4A;3)	WMCF01000024.1	0-3269374	Z:69365659-72860660	sex-linked
Eremophila alpestris	Fusion scaffold (Z;4A;3)	WMCF01000024.1	3277173-12546504	4A:4588-9605363	sex-linked
Eremophila alpestris	Fusion scaffold (Z;4A;3)	WMCF01000024.1	12550475-12555560	3:12208-18513	sex-linked
Eremophila alpestris	Fusion scaffold (Z;4A;3)	WMCF01000024.1	12558560-12560541	3:24098751-24103322	sex-linked
Alauda arvensis	Fusion scaffold (4A;3)	CADDXX010000006.1	15000-5830000	4A:23654-5995271	sex-linked
Alauda arvensis	Fusion scaffold (4A;3)	CADDXX010000006.1	5000-15000	3:14664-16406	unsure
Alauda arvensis	Fusion scaffold (4A;3)	CADDXX010000006.1	0-5000	3:24099520-24099841	sex-linked

Table S6. Songbird PAR-scaffolds from reference genomes (in Table S4) used to generate Figure S10.

Species	**Chromosome copy**	**ID**
Taeniopygia guttata (3.2.4)	PAR-Z	Z_random
Taeniopygia guttata (bTaeGut2.pat.W.v2)	PAR-Z	NC_045027.1
Taeniopygia guttata (bTaeGut2.pat.W.v2)	PAR-W	NW_022611471.1
Ficedula albicollis	PAR-Z	N00298
Ficedula albicollis	PAR-Z	N00378
Ficedula albicollis	PAR-Z	N02597
Sylvietta virens	PAR-Z	scaffold_394
Alauda arvensis	PAR-Z	CADDXX010000025.1

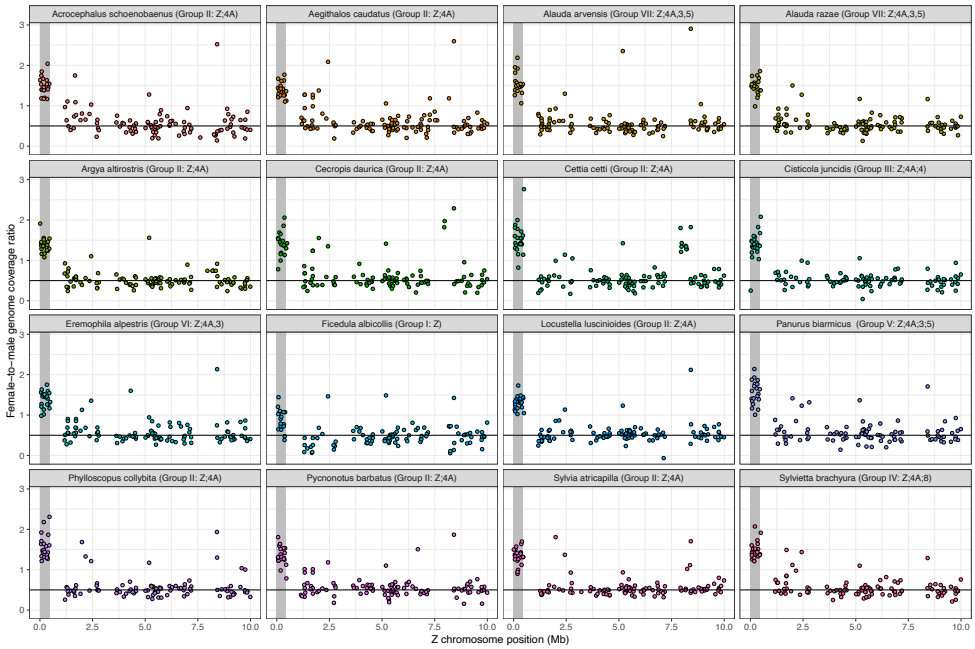

Figure S1. Female-to-male genome coverage values across 0-10 Mb on chromosome Z. The genes in the PAR (0-0.5 Mb) have higher genome coverage values compared to the genes in the fully sex-linked part (0.5-10 Mb). This suggests that the PAR genes have both a Z and W gametolog.

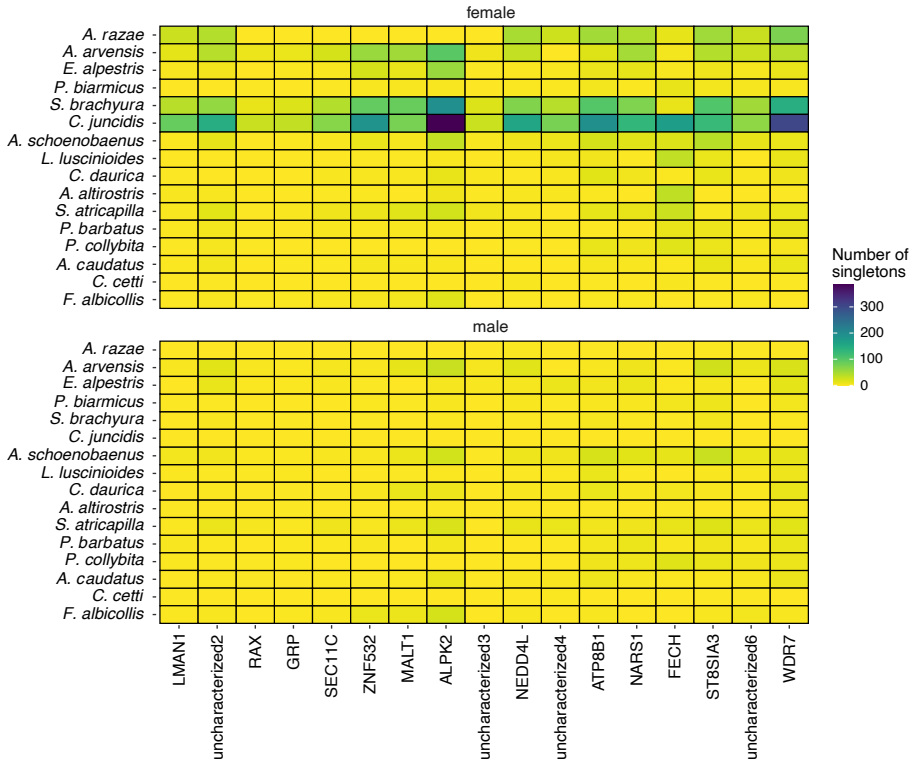

Figure S2. Number of singletons (genetic variants occuring in only the female or male sample, and in heterozygous form) in females and males separately. Females had more singleton variants than males, indicative of sex-linkage, in three lineages; Alauda, Cisticolidae and Sylvietta.

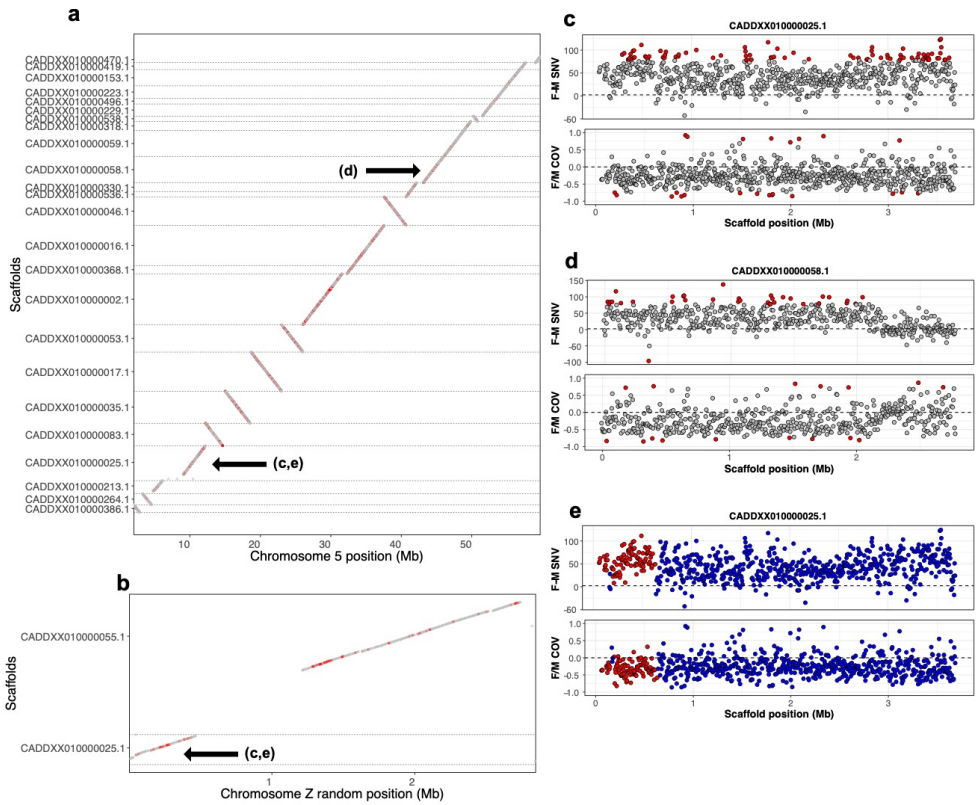

Figure S3. Synteny between *Alauda arvensis* scaffolds and zebra finch (a) chromosome 5 and the scaffold (b) "Z_random", which contain the zebra finch PAR within the first ~0.5 Mb. Genome coverage and single nucleotide variation (SNV) are based on aligned WGS data from one male and one female *Alauda arvensis* individual. Data points in (a) and (b) are colored red if either the female-to-male genome coverage or SNV exceeds the genome-wide 95% CI, and grey is they are not. Highlighted in (a) and (b) are scaffolds informative of fusion points (c,e) or a novel PAR boundary (d) shown in Figure 3a in the Main text. (c,d) Female-to-male SNV (upper) and genome coverage ratios (lower). Red data points exceed the genome-wide 95% CI, grey data points do not. (e) This panel is based on the same data as (c), but colored according to synteny to zebra finch chromosomes (chromosome 5 in blue; scaffold Z_random in red).

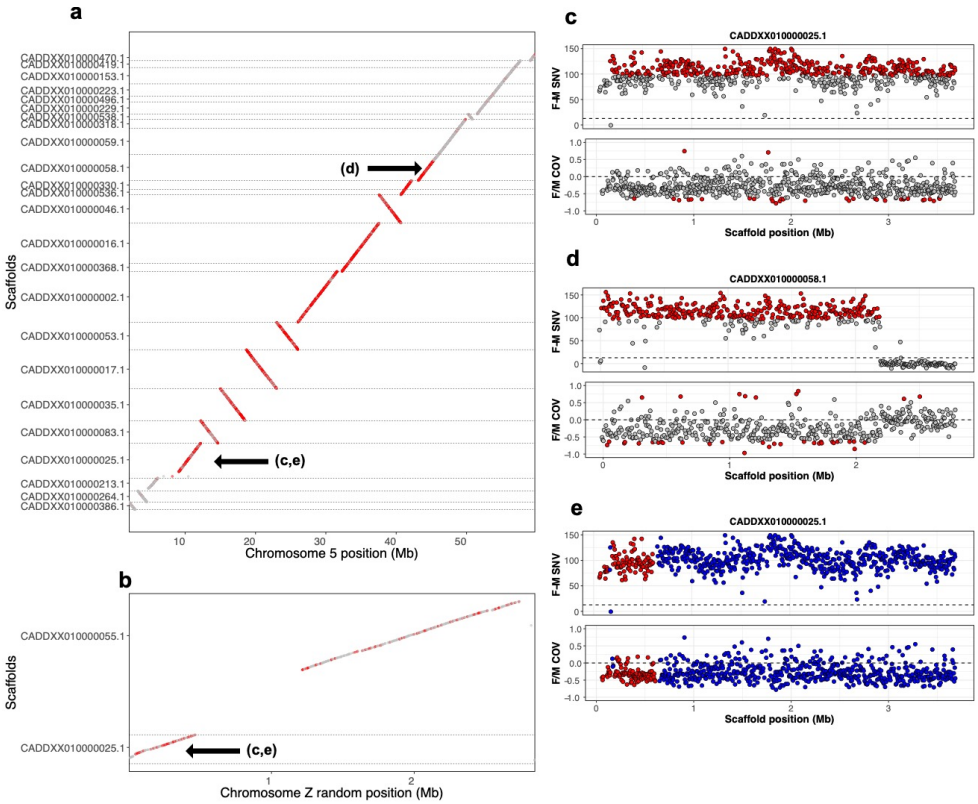

Figure S4. Same data as in Figure S3, except for that genome coverage and single nucleotide variation (SNV) are based on aligned WGS data from one male and one female *Alauda razae* individual.

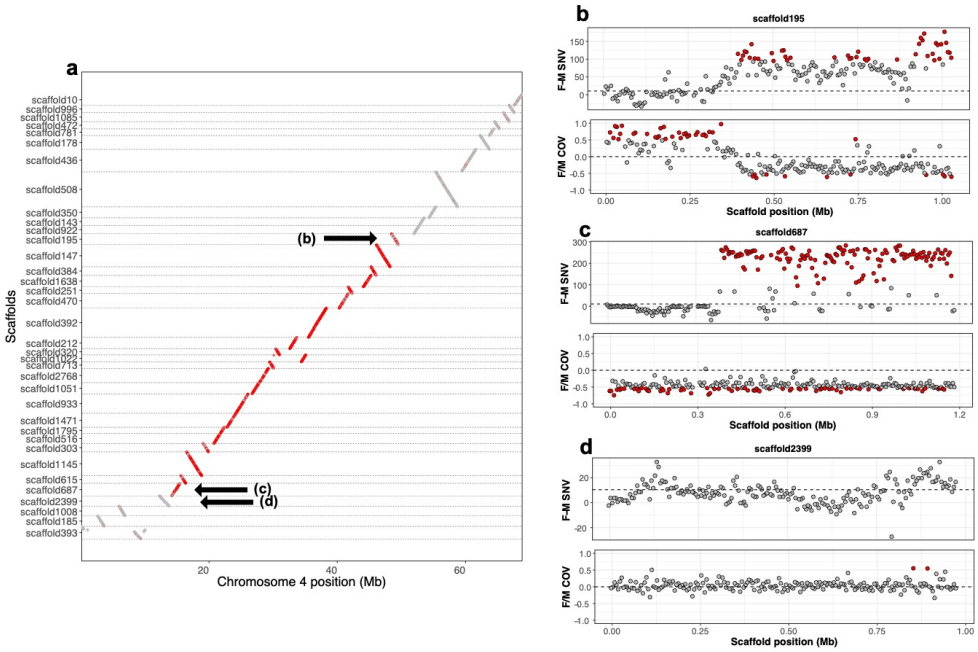

Figure S5. Synteny between *Cisticola juncidis* scaffolds and zebra finch (a) chromosome 4. Genome coverage and single nucleotide variation (SNV) are based on aligned WGS data from one male and one female *Cisticola juncidis* individual. Data points in (a) are colored red if either the female-to-male genome coverage or SNV exceeds the genome-wide 95% CI, and grey is they are not. Highlighted in (a) are scaffolds informative of (b) a novel PAR boundary or a pair of fission scaffolds (c,d). (b-d) Female-to-male SNV (upper) and genome coverage ratios (lower). Red data points exceed the genome-wide 95% CI, grey data points do not.

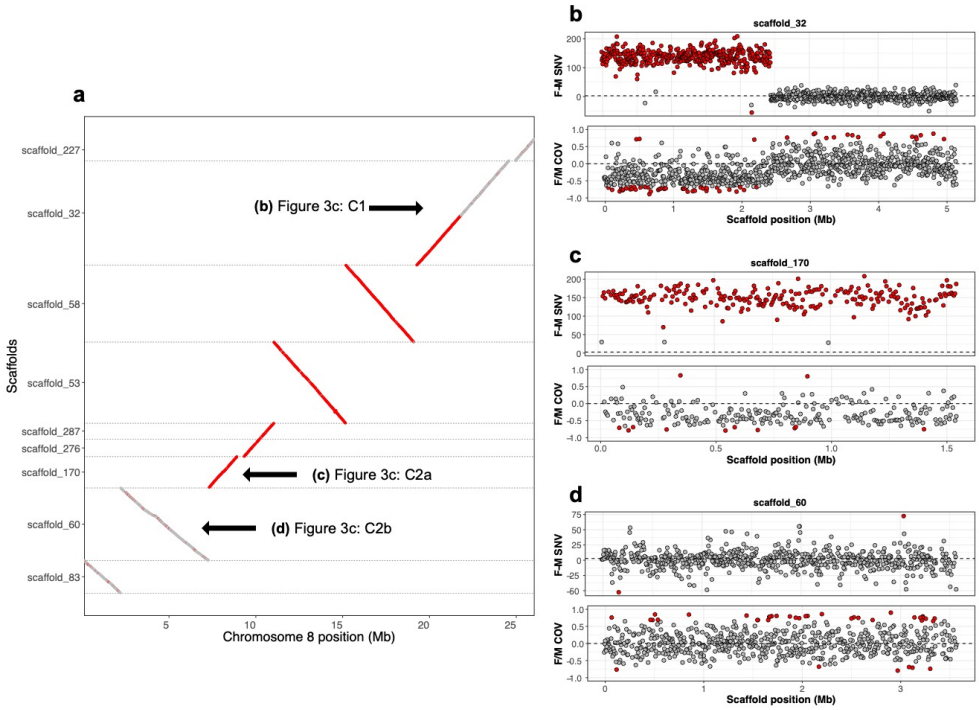

Figure S6. Synteny between *Sylvietta virens* scaffolds and zebra finch (a) chromosome 8. Genome coverage and single nucleotide variation (SNV) are based on aligned WGS data from one male and one female *Sylvietta brachyura* individual. Data points in (a) are colored red if either the female-to-male genome coverage or SNV exceeds the genome-wide 95% CI, and grey is they are not. Highlighted in (a) are scaffolds informative of a novel PAR boundary (b) or fusion points (c,d) shown in Figure 3c in the Main text. (b-d) Female-to-male SNV (upper) and genome coverage ratios (lower). Red data points exceed the genome-wide 95% CI, grey data points do not.

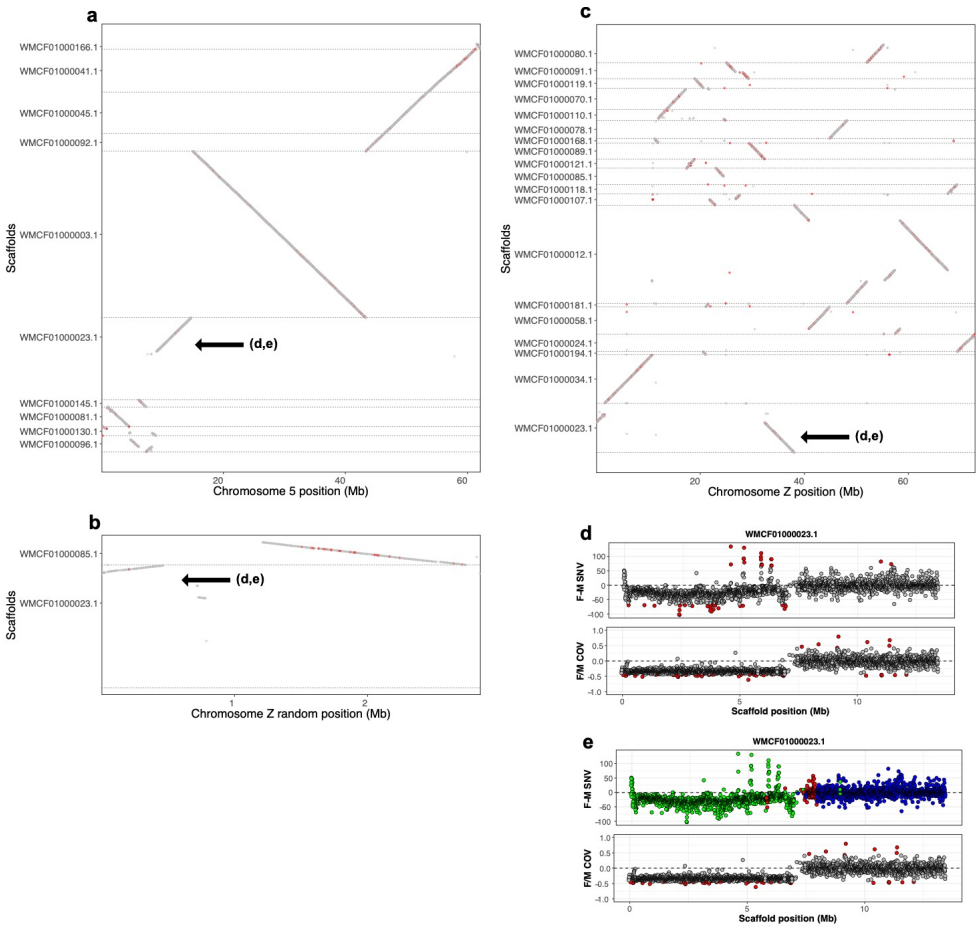

Figure S7. Synteny between *Eremophila alpestris* scaffolds and zebra finch (a) chromosome 5 and the scaffold (b) "Z_random", which contain the zebra finch PAR within the first ~0.5 Mb and the rest of (c) chromosome Z. Genome coverage and single nucleotide variation (SNV) are based on aligned WGS data from one male and one female *Eremophila alpestris* individual. Data points in (a-c) are colored red if either the female-to-male genome coverage or SNV exceeds the genome-wide 95% CI, and grey is they are not. Highlighted in (a-c) is the fusion scaffold shown in Figure 3b (d,e) in the Main text. (d) Female-to-male SNV (upper) and genome coverage ratios (lower). Red data points exceed the genome-wide 95% CI, grey data points do not. (e) This panel is based on the same data as (d), but colored according to synteny to zebra finch chromosomes (chromosome 5 in blue; scaffold Z_random in red, chromosome Z in green).

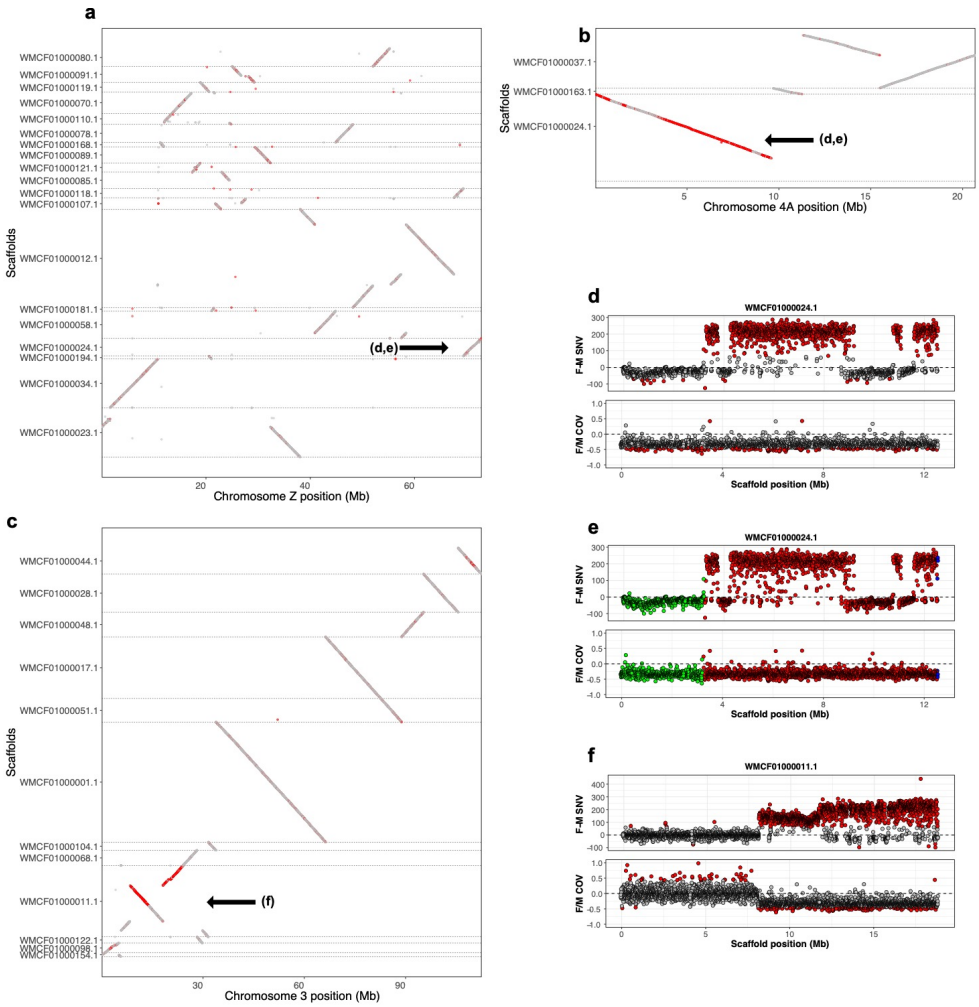

Figure S8. Synteny between *Eremophila alpestris* scaffolds and zebra finch (a) chromosome Z, (b) chromosome 4A and (c) chromosome 3. Genome coverage and single nucleotide variation (SNV) are based on aligned WGS data from one male and one female *Eremophila alpestris* individual. Data points in (a-c) are colored red if either the female-to-male genome coverage or SNV exceeds the genome-wide 95% CI, and grey is they are not. Highlighted in (a-c) are scaffolds informative of fusion sites (d,e) or (f) a novel PAR boundary. (d,f) Female-to-male SNV (upper) and genome coverage ratios (lower). Red data points exceed the genome-wide 95% CI, grey data points do not. (e) This panel is based on the same data as (d), but colored according to synteny to zebra finch chromosomes (chromosome 4A in red, chromosome Z in green and chromosome 3 in blue).

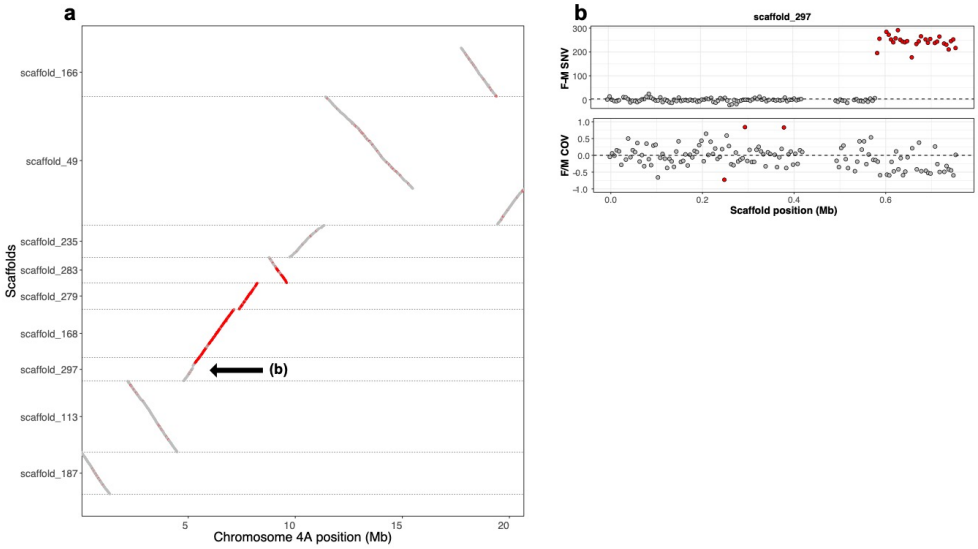

Figure S9. Synteny between *Sylvietta virens* scaffolds and zebra finch (a) chromosome 4A. Genome coverage and single nucleotide variation (SNV) are based on aligned WGS data from one male and one female *Sylvietta brachyura* individual. Data points in (a) are colored red if either the female-to-male genome coverage or SNV exceeds the genome-wide 95% CI, and grey is they are not. Highlighted in (a) is a scaffold informative of (b) a novel PAR boundary. (b) Female-to-male SNV (upper) and genome coverage ratios (lower). Red data points exceed the genome-wide 95% CI, grey data points do not.

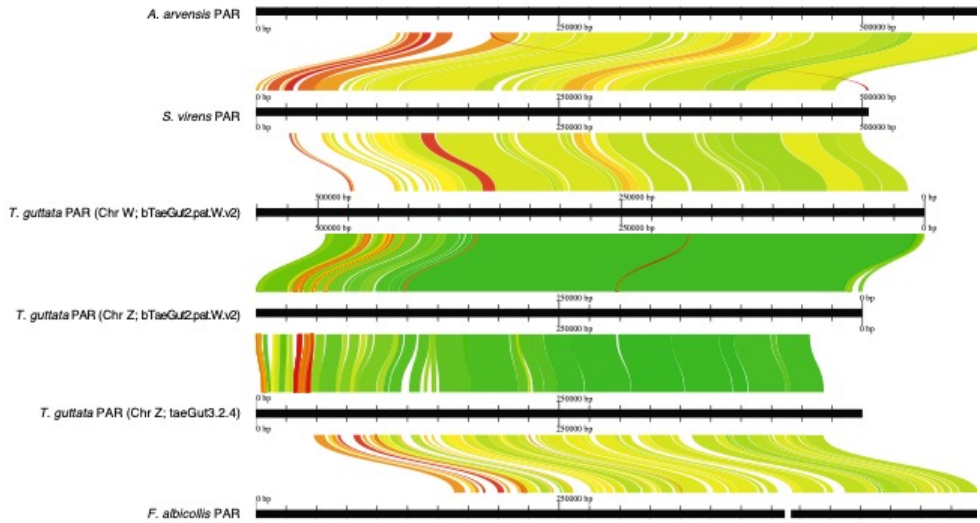

Figure S10. Synteny plot produced from PAR scaffolds of different species (Table S6), using the program AliTV (Ankenbrand et al. 2017)

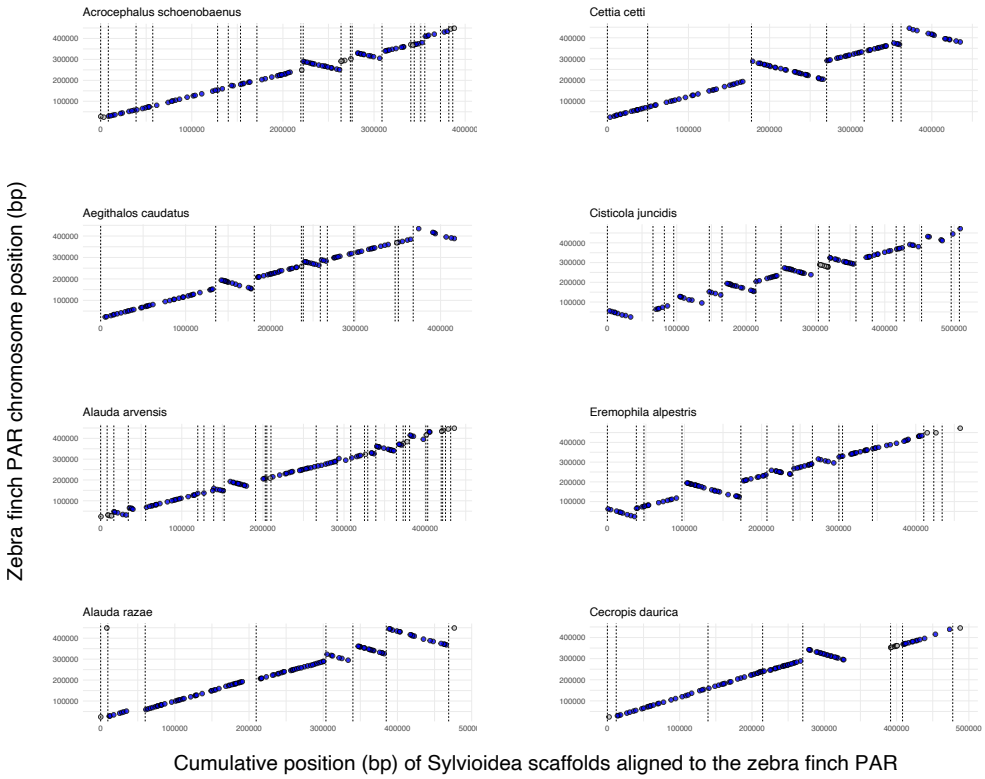

Figure S11. Synteny plots of scaffolds in the 15 Sylvioidea reference genomes (assembled from the male samples in Supplementary Table 1) and the zebra finch PAR. Scaffold boundaries are marked with dashed lines. Data points colored in blue are positioned on scaffolds with PAR genes. Grey data points are positioned on scaffolds that contain no genes.

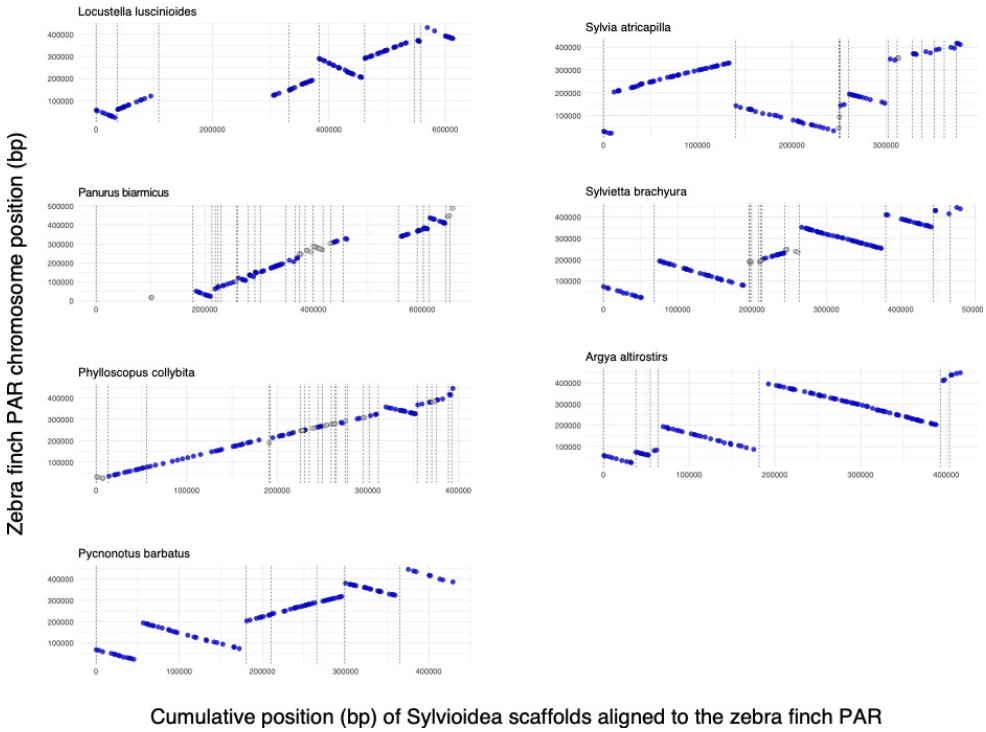

Figure S12. Synteny plots of scaffolds in the 15 Sylvioidea reference genomes (assembled from the male samples in Supplementary Table 1) and the zebra finch PAR. Scaffold boundaries are marked with dashed lines. Data points colored in blue are positioned on scaffolds with PAR genes. Grey data points are positioned on scaffolds that contain no genes.

References

Ankenbrand MJ, Hohlfeld S, Hackl T, Förster F. 2017. AliTV—interactive visualization of whole genome comparisons. PeerJ Computer Science 3:e116