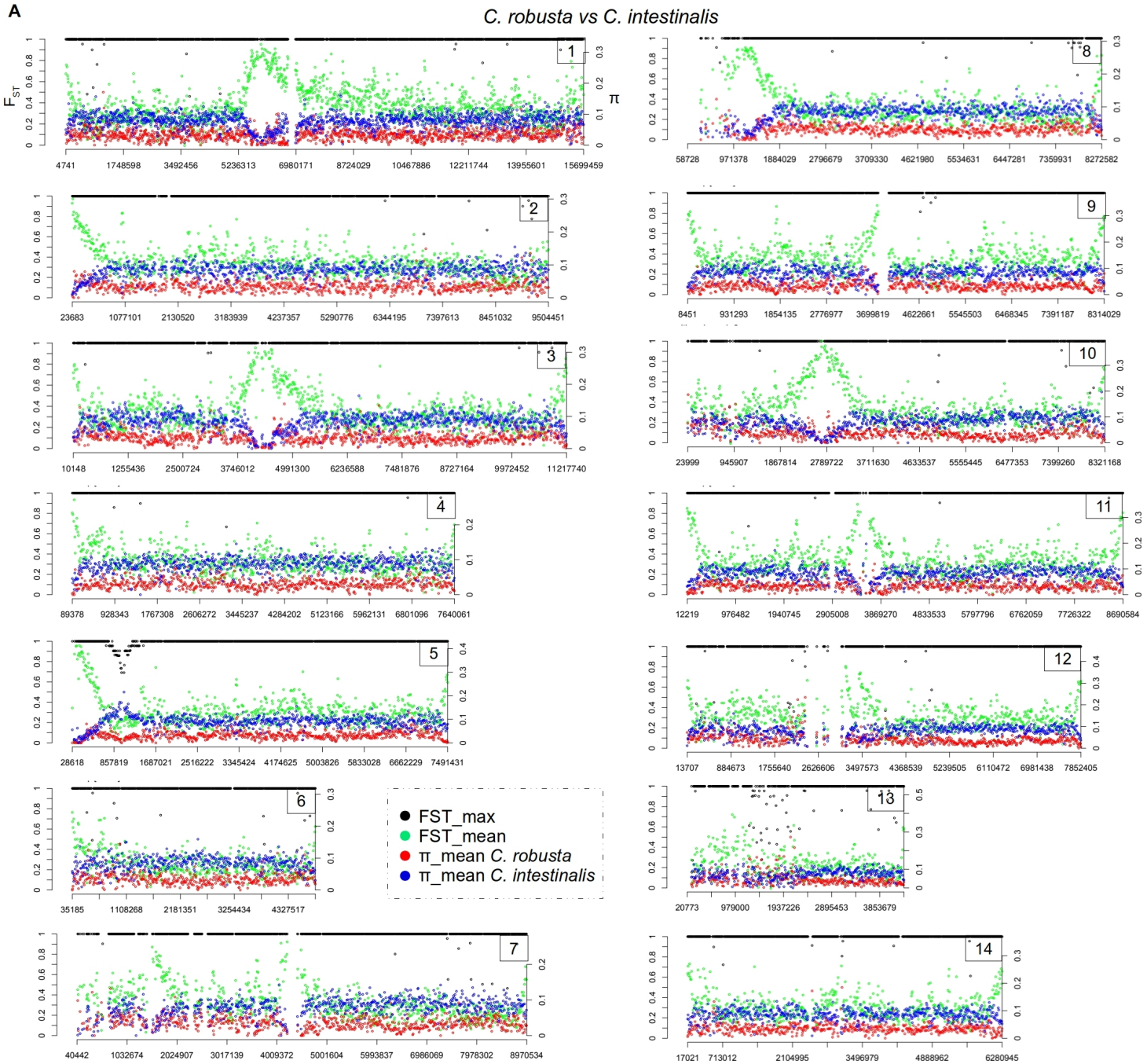
Figure S1 Population genetic statistics calculated in non-overlapping 10 Kb windows along the 14 chromosomes in the sea squirt genome. *C. robusta* is compared to *C. intestinalis* (**A**) and to *C. roulei* (**B**). Black points: maximal F_{ST} between the two populations compared; green points: average F_{ST} between the two populations compared; red points: average π in *C. robusta*; blue points: average π in *C. intestinalis* (**A**) or *C. roulei* (**B**). Windows with less than 10 SNPs were excluded. The x-axis is in bp. On the y-axis, the F_{ST} scale is depicted on the left, while the π scale is on the right, as shown on chromosome 1. Dataset **#2 "all SNPs with missing data"** was used.



B

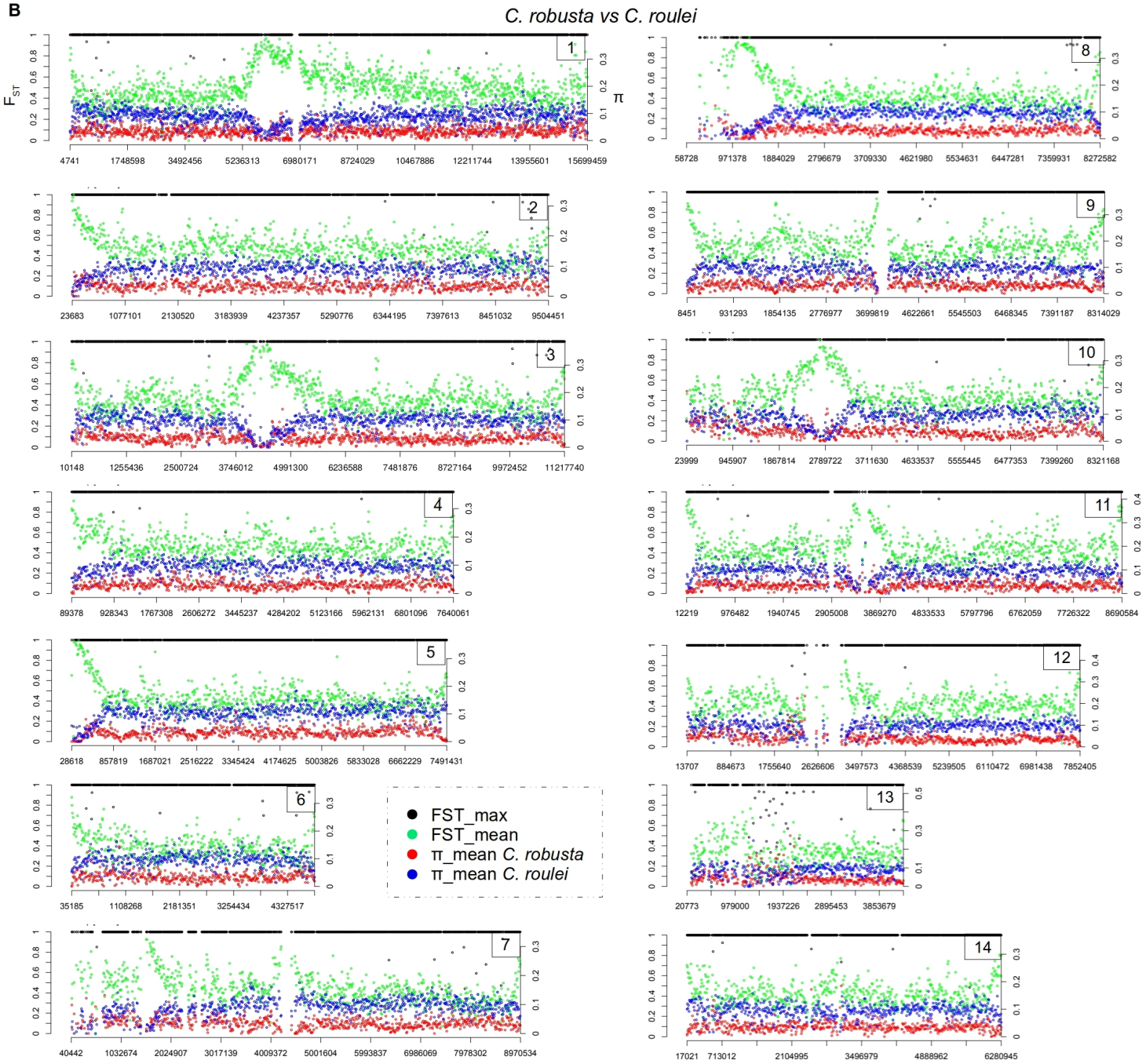

C. robusta vs C. roulei
F_ST
π
1
2
3
4
5
6
7
8
9
10
11
12
13
14
FST_max
FST_mean
π_mean C. robusta
π_mean C. roulei

Figure S2 *C. robusta* introgression into *C. intestinalis* shown across the 14 chromosomes. **A.** The haplotypes represent the allelic state at 111,951 SNPs fixed between *C. robusta* and *C. roulei*. Alleles of *C. roulei* origin are shown in yellow, while those of *C. robusta* origin are in red. SNPs are ordered on the x-axis following their physical position on chromosomes. Haplotypes were plotted using GPAT++ from vcflib. **B.** Local ancestry inference (Chromopainter) based on 640,044 phased SNPs. The frequency of the *C. robusta* alleles on the introgressed tracts is depicted with red bars along the chromosomes (x-axis in bp). Dataset **#3a "phased SNPs"** was used for Chromopainter and **#4 "ancestry informative phased SNPs"** for the raw haplotypes.

Figure S3 Population genetic statistics of the *C. robusta* introgressed coding sequences. CDS on non-introgressed tracts are depicted in blue (*n*=11,233); those on introgressed tracts are in yellow (*n*=69), among which 65 lie on chromosome 5 (in red; they mask the yellow points). A genomic tract is considered to be introgressed if at least one *C. intestinalis* individual carries the introgressed *C. robusta* allele. **A.** Correlation between diversity within (π_W) and between (d_{XY}) populations. The corresponding level of genetic differentiation (F_{ST}) is depicted with dashed lines. **B.** Comparison of the G_{min} value for migrant (red) and non-migrant (blue) tracts. d_{XY_min} is the minimal value of d_{XY}, and d_{XY_avg} is its average. Stars indicate that their difference is significant at 0.05 using a two-sided Student's test. **C.** The proportion of introgressed CDS on each chromosome. Numbers in brackets refer to their number. Dataset **#3b "CDS version of phased SNPs"** was used.

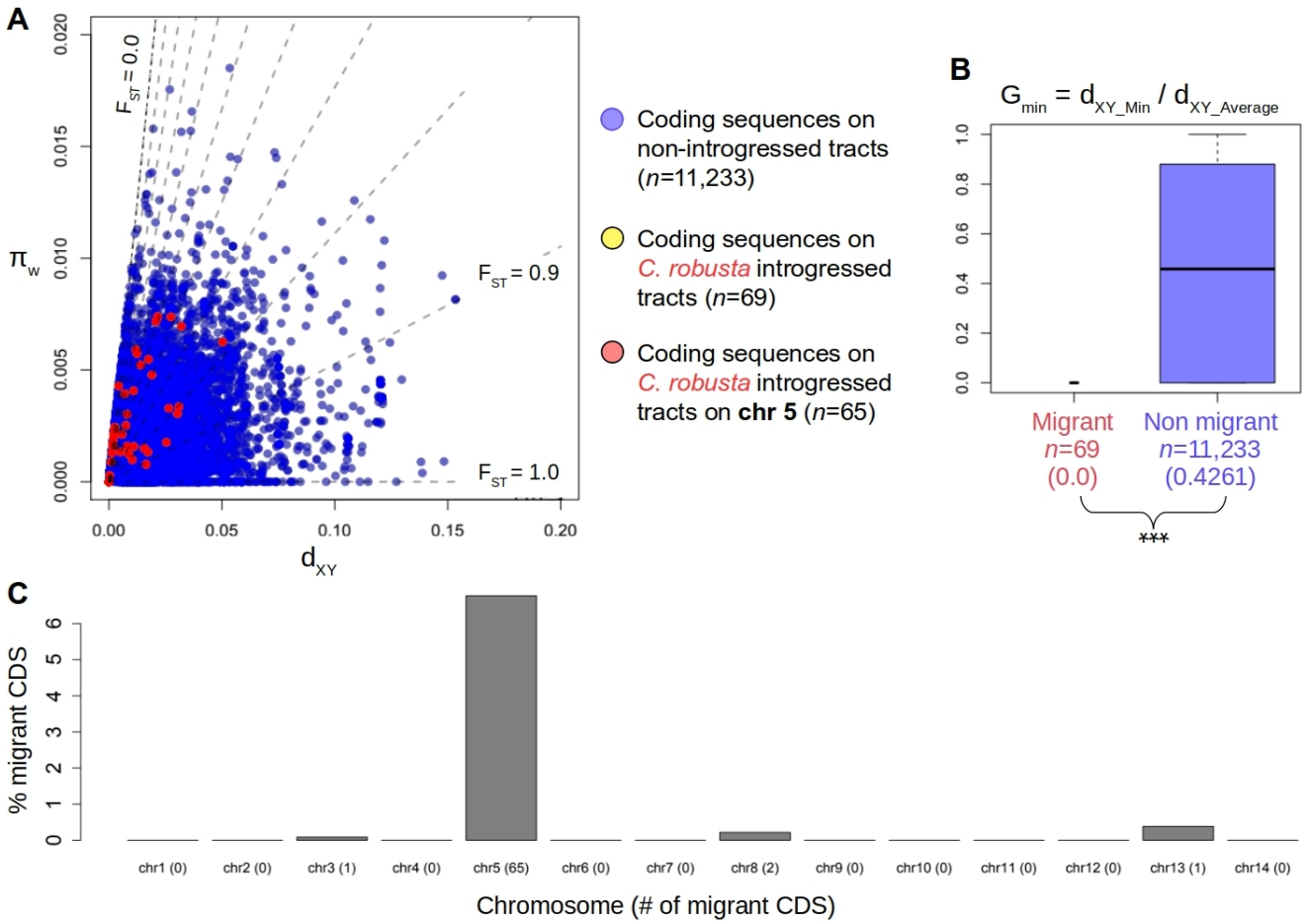


Figure S4 ABBA-BABA introgression patterns using *C. edwardsi* as an outgroup. Polarization was done using the monomorphic positions across four *C. edwardsi* samples (125,671 polarized SNPs). The following topology was used: (((P1 = *C. roulei* ; P2 = *C. intestinalis*) ; P3 = *C. robusta*) ; O = *C. edwardsi*). Therefore, a positive *D* indicates an excess of ABBA over BABA, i.e. excess shared ancestry between P2 and P3. **A.** *D* calculated for each chromosome. Its value for the genome is D=0.0543. **B**. Admixture proportion (*fd*) averaged per chromosome as a function of chromosome length in Mb. Spearman's rank correlation (after removing chromosome 5): ρ=-0.21, p-value=0.48. *fd* was calculated in non-overlapping windows of 100 SNPs. **C.** *fd* along each chromosome (position in bp). One line corresponds to one chromosome (overlapping), and chromosome 5 is depicted in red. The horizontal dashed line refers to *fd*=0. **D.** *fd* along chromosome 5. The introgression hotspot between 700 Kb and 1.5 Mb is shown in red with a maximum *fd* value located at 1,033,977 bp. Dataset **#6 "all polarized SNPs with missing data"** was used.

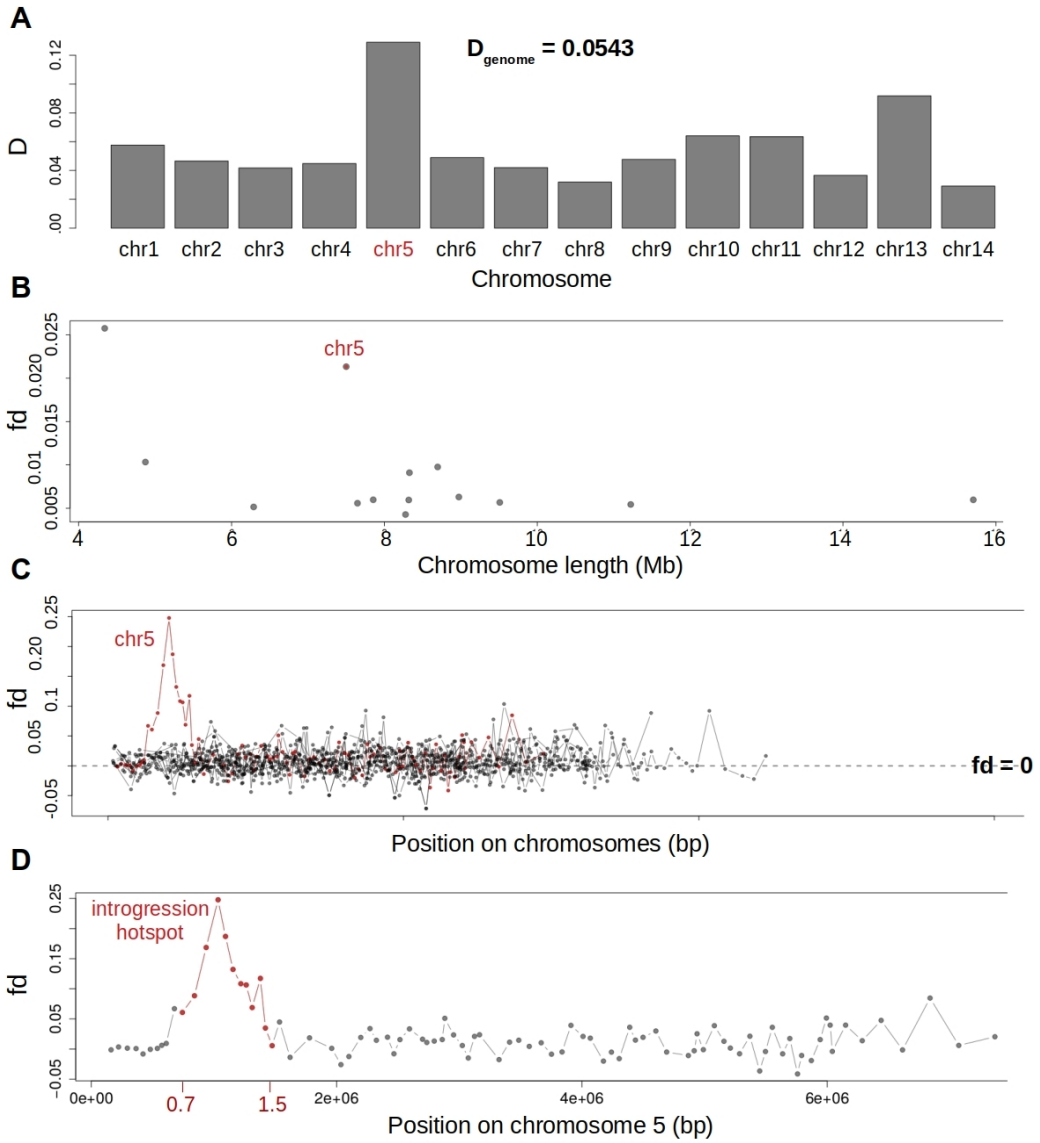
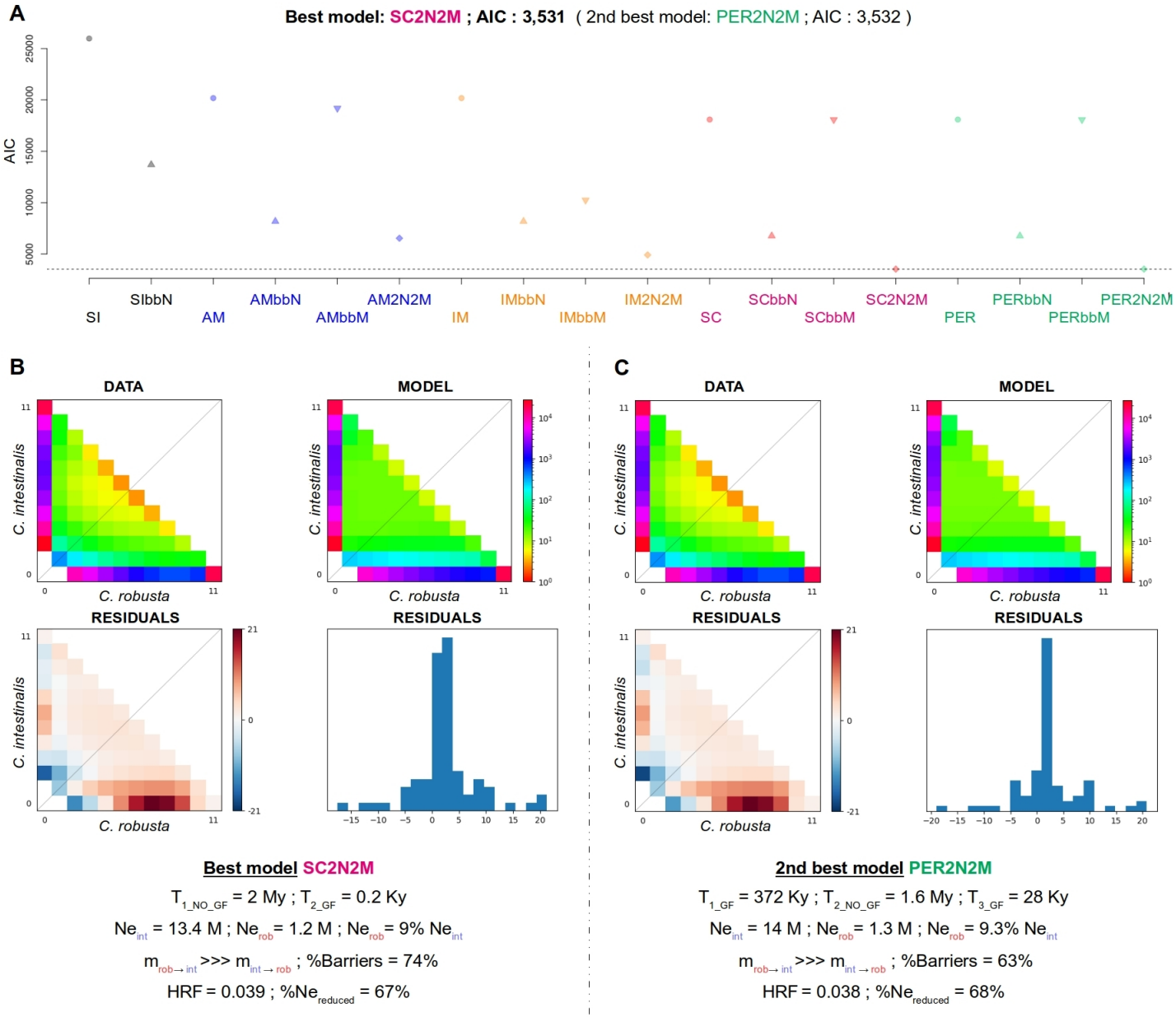


Figure S5 Inference of the divergence history between *C. robusta* and *C. intestinalis* with moments. **A.** AIC value of the best run for each model. **B.** Observed site frequency spectrum (SFS), modeled SFS and residuals of the best model. Maximum likelihood values of the parameters are provided. **C.** Same as in **B** but for the second-best model. Analyses were based on the folded SFS after LD-pruning the SNPs. Five demographic scenarios were modeled: SI = strict isolation, IM = isolation with continuous migration, SC = secondary contact, AM = ancient migration, PER = periodic connectivity. Different versions of these scenarios were tested: bbN = genomic heterogeneity of the effective population sizes, bbM = genomic heterogeneity of the effective migration rates, 2N2M = both types of heterogeneities, "" = no heterogeneities. Five replicates were run for each model. Parameters are as follows: T = times in years, assuming two generations per year in European waters (the "GF" label refers to gene flow), Ne = effective population sizes in numbers of individuals, m = migration rates (direction given by the arrow), %Barriers = proportion of the genome with null migration, $\%Ne_{reduced}$ = fraction of the genome experiencing reduced Ne, HRF = factor by which Ne is reduced. Full details are provided in **Table S4**. Dataset **#5 "all SNPs without missing data"** was used, including chromosome 5.



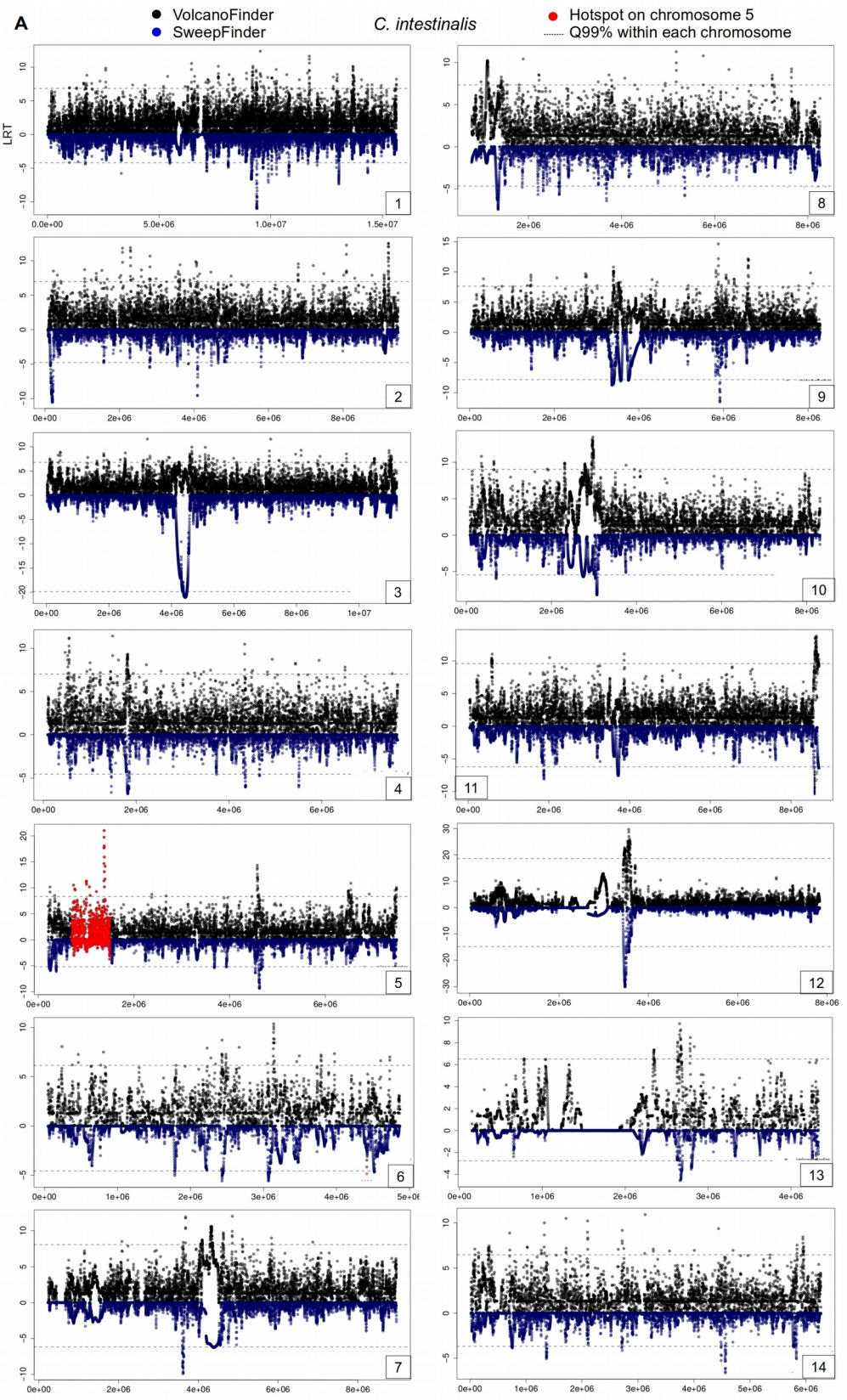


Figure S6 A. Selection tests in *C. intestinalis* using *C. edwardsi* as an outgroup. Polarization was done using the monomorphic positions across four *C. edwardsi* samples (125,671 polarized SNPs). The log-likelihood ratio test (LRT) for positive selection (SweepFinder, blue points) and adaptive introgression (VolcanoFinder, black points) is shown along each chromosome. No donor species needs to be specified in VolcanoFinder. LRT was calculated at test sites spaced by 1 Kb. The horizontal dashed lines (one for SweepFinder and one for VolcanoFinder) correspond to the 99th quantile of the LRT distribution for each chromosome. The 99th quantile overall chromosomes for SweepFinder is 6.2 and for VolcanoFinder is 8.0. **B.** Selection tests for chromosome 5 in *C. intestinalis* (VolcanoFinder, black points, Q99%=8.0) and *C. robusta* (SweepFinder, blue points, Q99%=10.9). Positive selection in *C. robusta* and adaptive introgression into *C. intestinalis* was detected in the hotspot on chromosome 5 (highlighted in red). Dataset **#6 "all polarized SNPs with missing data"** was used.

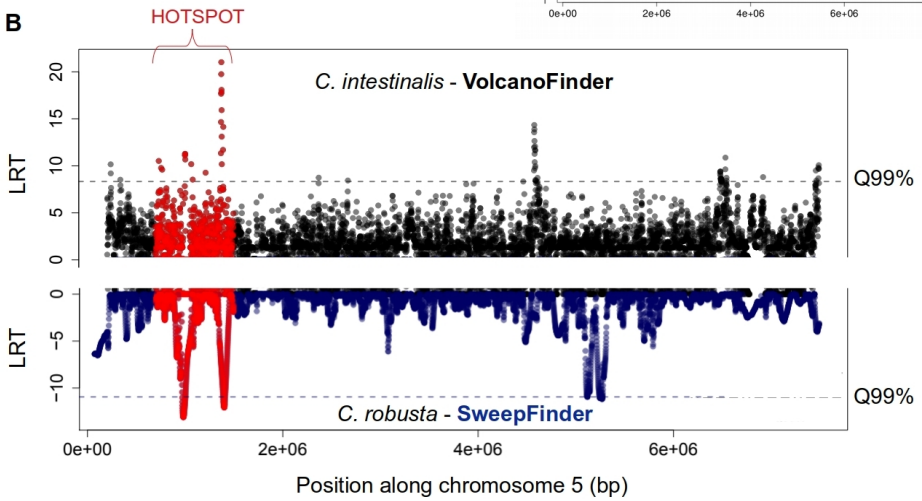


Figure S7 *C. robusta* ancestry along chromosome 5 in *C. intestinalis* individuals. Ancestry probabilities were inferred with Chromopainter and are shown before defining the haplotype bounds. For each individual, the probability of *C. robusta* is plotted for its two haplotypes: haplotype 1 in upper panels (red), haplotype 2 in lower panels (orange). The grey band delimits the "missing data region". The x-axis is in bp. Dataset **#3a "phased SNPs"** was used.

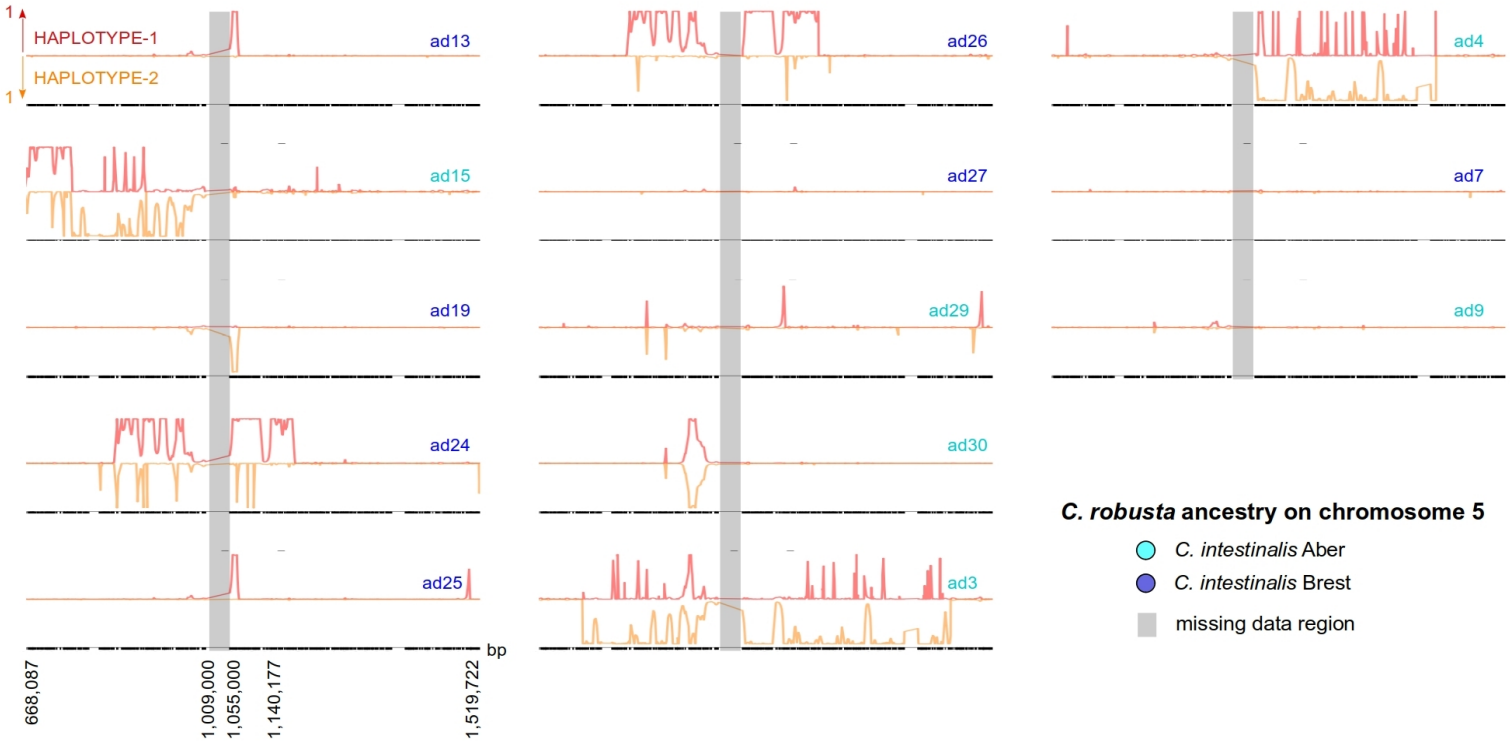


Figure S8 Neighbor-joining trees of 50 Kb windows framing the "missing data region" (grey band) at the center of the chromosome 5 hotspot. Dataset **#3c "FASTA version of phased SNPs"** was used.

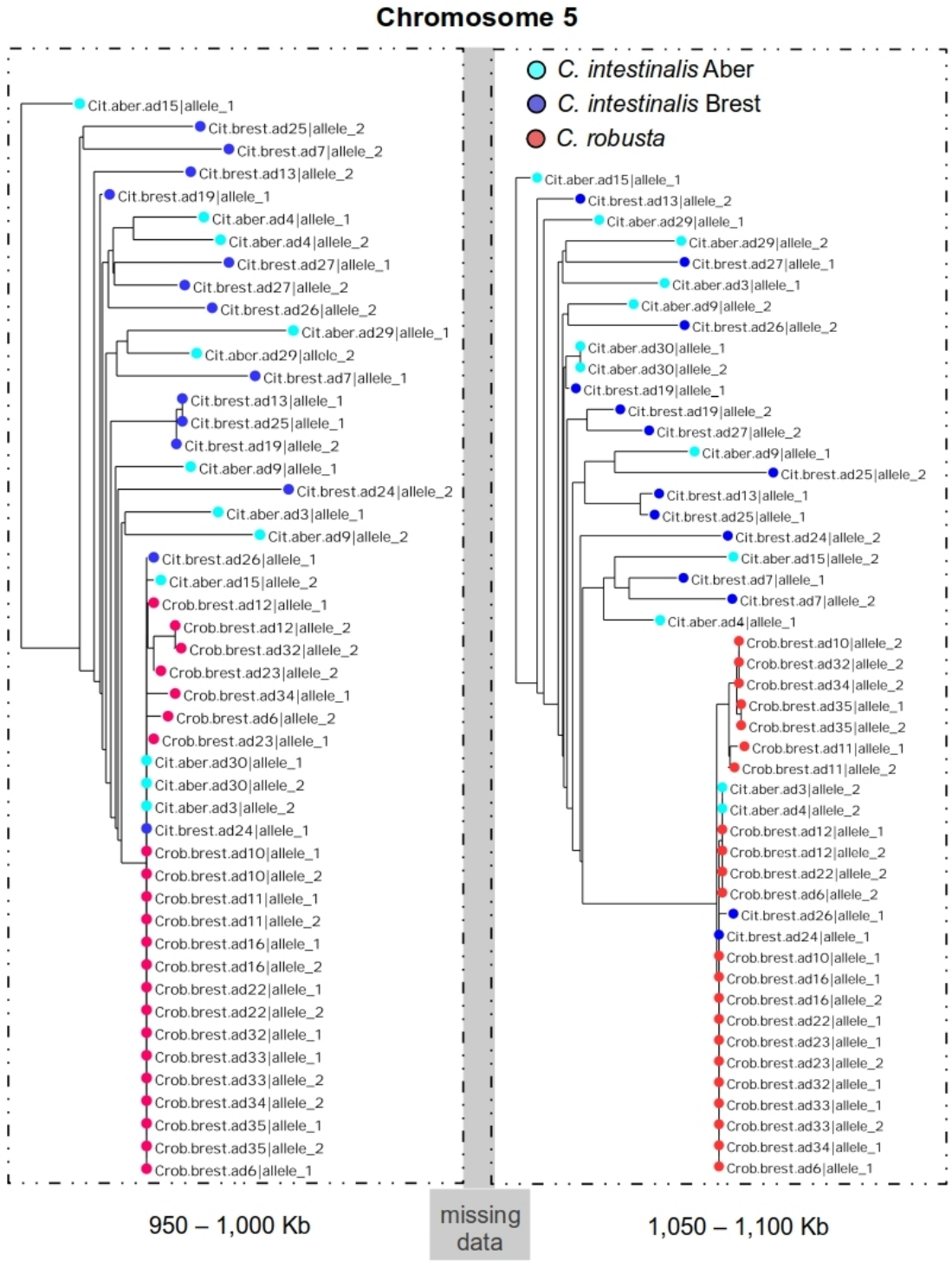


Figure S9 Copy number variation at candidate SNPs in the introgression hotspot on chromosome 5 (700 Kb - 1.5 Mb). Candidates are labeled (annotation and position in bp) on the top left of each panel, and those lying on the same locus are framed together with a specific color. SNPs located in the "missing data region" (due to high coverage) are indicated with a grey star. Candidates were defined as having a variant allele fraction, VAF <= 50% in *C. intestinalis* and, either **A**) VAF >= 90% in *C. roulei* and VAF <= 10% in *C. robusta*, or **B**) VAF >= 85% in *C. roulei* ; VAF <= 15% in *C. robusta* (only new candidate SNPs are shown in **B**). No candidates were found in the other direction (i.e. with the minor VAF in *C. roulei*). Copy number at each SNP was calculated as its allele read depth normalized by the per-site read depth averaged across all sites (excluding sites with less than ten reads) for each individual (labeled on the left). A copy number of one (vertical dashed line) means that the SNP lies on a single-copy locus. Values for the *C. robusta* allele (red) and the *C. intestinalis* allele (blue) are separately shown. Read depth was obtained from the bam files. Horizontal dashed lines separate the different species, and *C. intestinalis* individuals introgressed at the hotspot (see **Figure 4**) were labeled as "introgressed". Dataset **#7 "unfiltered mapping files"** was used.


A
Copy number at candidate SNPs
C. robusta allele
C. intestinalis allele
C. roulei
C. robusta
C. intestinalis introgressed
949,652 (non-coding)
949,653 (non-coding)
990,561 (aminoadipate-semialdehyde synthase)
990,571 (aminoadipate-semialdehyde synthase)
990,572 (aminoadipate-semialdehyde synthase)
995,239 (chitinase 1)
1,009,280 (adenosine deaminase RNA specific)
1,010,365 (adenosine deaminase RNA specific)
1,015,283 (cyclic nucleotide binding domain)
1,018,450 (cyclic nucleotide binding domain)
1,050,647 (cytochrome P450 family 2 subfamily U member 1)
1,050,917 (cytochrome P450 family 2 subfamily U member 1)
1,067,404 (non-coding)


B
Copy number at candidate SNPs
C. robusta allele
C. intestinalis allele
C. roulei
C. robusta
C. intestinalis introgressed
968,799 (cilia & flagella)
1,050,118 (cytochrome P450 family 2 subfamily U member 1)
1,050,355 (cytochrome P450 family 2 subfamily U member 1)
1,050,481 (cytochrome P450 family 2 subfamily U member 1)
1,050,816 (cytochrome P450 family 2 subfamily U member 1)
1,050,842 (cytochrome P450 family 2 subfamily U member 1)
1,050,922 (cytochrome P450 family 2 subfamily U member 1)
1,050,956 (cytochrome P450 family 2 subfamily U member 1)
1,050,957 (cytochrome P450 family 2 subfamily U member 1)
1,050,983 (cytochrome P450 family 2 subfamily U member 1)
1,051,868 (cytochrome P450 family 2 subfamily U member 1)
1,051,926 (cytochrome P450 family 2 subfamily U member 1)
1,051,983 (cytochrome P450 family 2 subfamily U member 1)
1,052,166 (cytochrome P450 family 2 subfamily U member 1)
1,052,545 (cytochrome P450 family 2 subfamily U member 1)

Figure S10 Structural analysis of the "missing data region" on chromosome 5 (from 1,009,000 to 1,055,000 bp). **A.** Synteny plot of this region with itself using the fasta sequence extracted from the reference genome (GCA_009617815.1). The black dashed square corresponds to the genomic region harboring the cytochromes P450 (four different genes were annotated). The orange dashed square corresponds to the candidate region on the cytochrome P450 2U1 gene. Alignment was done using YASS online https://bioinfo.lifl.fr/yass/index.php (Noé and Kucherov 2005). **B**. Snapshot of the aligned reads in the hotspot region of the *C. robusta* reference genome. A region carrying different types of variants is shown in four representative individuals: 1) SNPs at 1,050,917 bp and 1,050,922 bp (orange) are differentially fixed between *C. robusta* and *C. roulei*, and the introgressed *C. intestinalis* individual carries a fraction of both alleles; 2) SNP at 1,050,862 bp (red) is segregating in *C. intestinalis*, 3) SNP at 1,050,881 bp (blue) is segregating in *C. intestinalis* and *C. roulei* (not visible). The Integrative Genomics Viewer (IGV) was used for visualization (Robinson et al. 2011). Dataset **#7 "unfiltered mapping files"** was used.