

Description of Additional Supplementary Files

File Name: Supplementary Data 1

Description: Estimates of GRC size from meiotic spreads for 6 nightingale individuals. Each estimate is based off a linear regression of chromosome lengths. The average R² value represents the average of all 10 cells linear regressions.

File Name: Supplementary Data 2

Description: Reference somatic genome statistics for *L. megarhynchos* and *L. luscinia*

File Name: Supplementary Data 3

Description: All 192 identified genes from nightingale GRCs. "Bp Found" refers to the number of different nucleotides identified from said gene across all scaffolds. "Adjusted Proportion" represents how much of the cds length was found in the GRC, divided by the proportion of the cds that was found in the whole genome (to a maximum of 0.75). "Normalised Coverage" is the estimated copy number in the GRC, taking recent duplications into account by normalising to the expected coverage using the respective somatic dataset. Expression data is expressed in FPKM, using the length of the gene in the GRC. Note that comparisons of FPKM between datasets is not advised since it is not known what proportion of sequenced cells contained a GRC.

File Name: Supplementary Data 4

Description: Sequencing depth across 12 libraries. LL represents *L. luscinia* and LM represents *L. megarhynchos*. "t" stands for testis and "k" stands for kidney.

File Name: Supplementary Data 5

Description: Description of PCR primers used to validate GRC sequences