Figure S1:

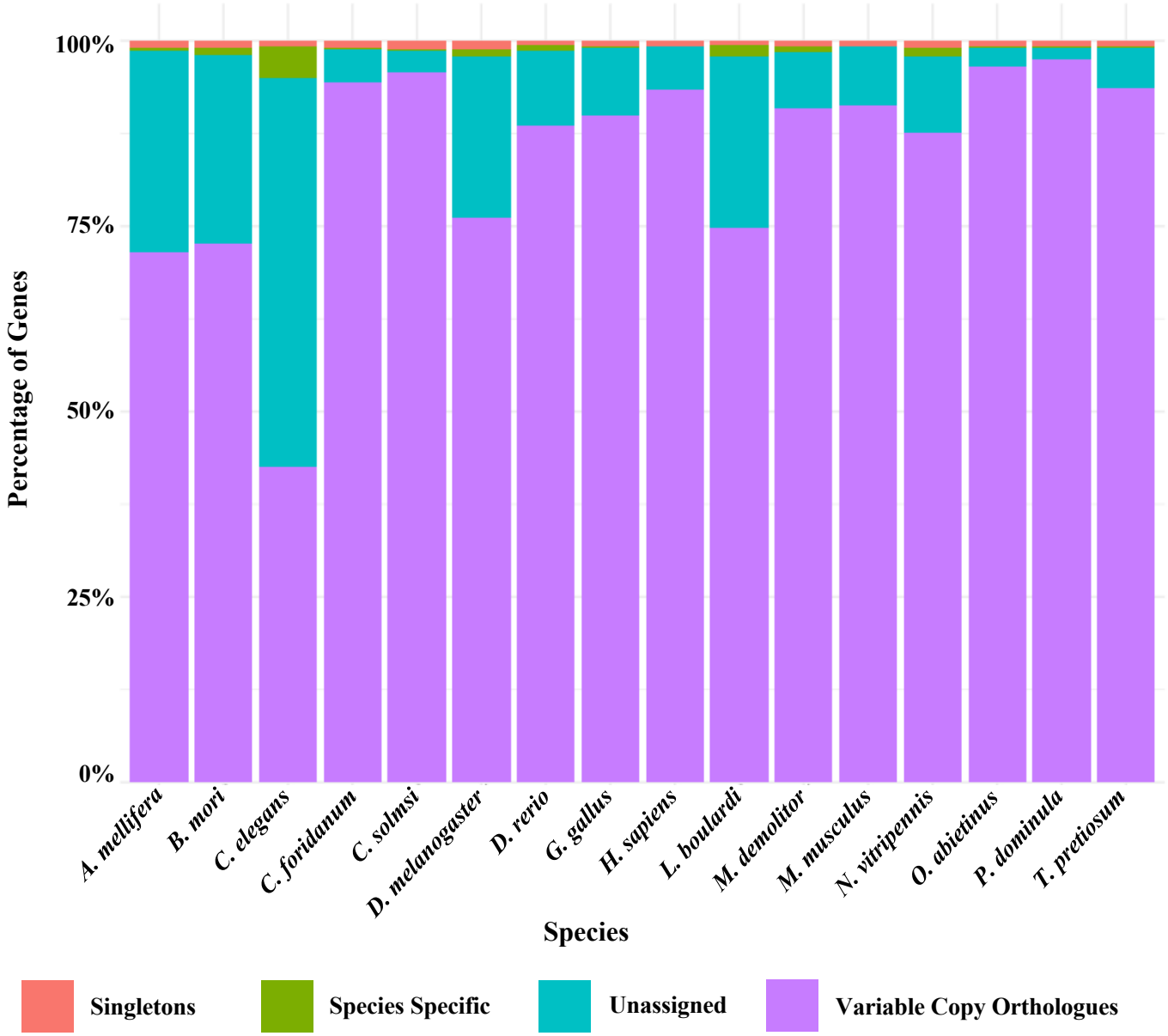
100%
75%
50%
25%
0%
Percentage of Genes
A. mellifera
B. mori
C. elegans
C. foridanum
C. solmsi
D. melanogaster
D. rerio
G. gallus
H. sapiens
L. boulardi
M. demolitor
M. musculus
N. vitripennis
O. abietinus
P. dominula
T. pretiosum
Species
Singletons
Species Specific
Unassigned
Variable Copy Orthologues

Figure S2:

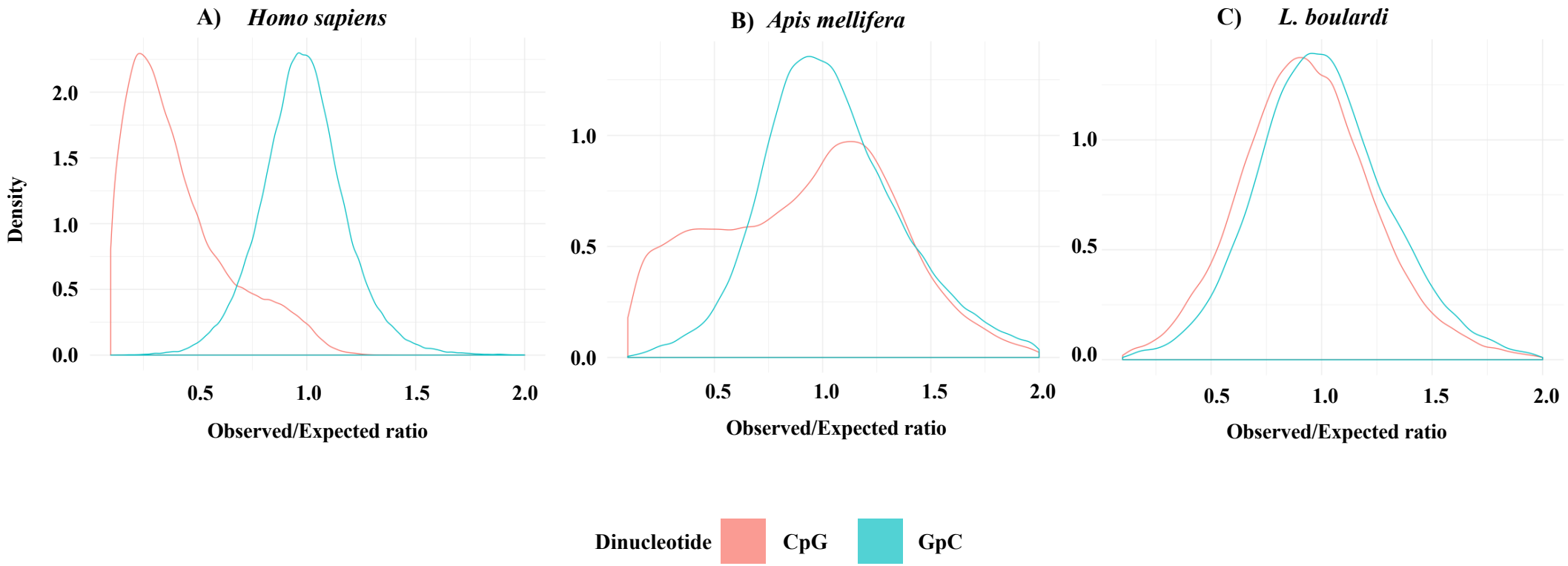
A) Homo sapiens
B) Apis mellifera
C) L. boulardi
Density
Observed/Expected ratio
Observed/Expected ratio
Observed/Expected ratio
Dinucleotide
CpG
GpC

Figure S3:

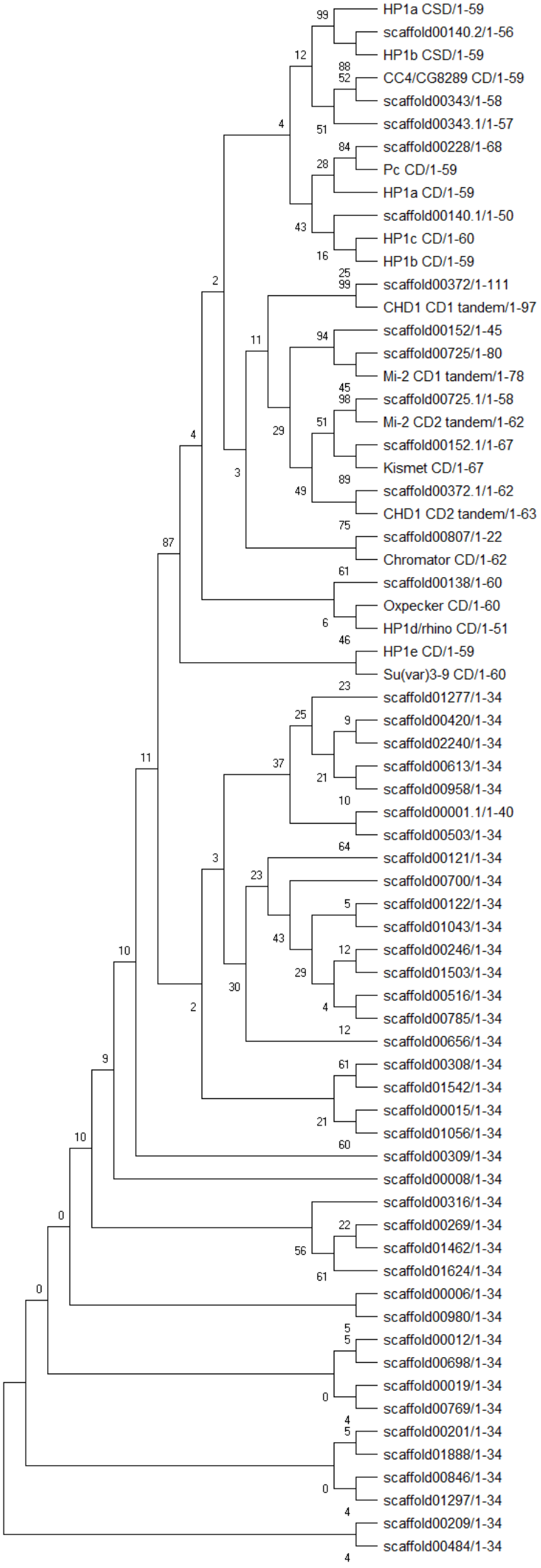
HP1a CSD/1-59
scaffold00140.2/1-56
HP1b CSD/1-59
CC4/CG8289 CD/1-59
scaffold00343/1-58
scaffold00343.1/1-57
scaffold00228/1-68
Pc CD/1-59
HP1a CD/1-59
scaffold00140.1/1-50
HP1c CD/1-60
HP1b CD/1-59
scaffold00372/1-111
CHD1 CD1 tandem/1-97
scaffold00152/1-45
scaffold00725/1-80
Mi-2 CD1 tandem/1-78
scaffold00725.1/1-58
Mi-2 CD2 tandem/1-62
scaffold00152.1/1-67
Kismet CD/1-67
scaffold00372.1/1-62
CHD1 CD2 tandem/1-63
scaffold00807/1-22
Chromator CD/1-62
scaffold00138/1-60
Oxpecker CD/1-60
HP1d/rhino CD/1-51
HP1e CD/1-59
Su(var)3-9 CD/1-60
scaffold01277/1-34
scaffold00420/1-34
scaffold02240/1-34
scaffold00613/1-34
scaffold00958/1-34
scaffold00001.1/1-40
scaffold00503/1-34
scaffold00121/1-34
scaffold00700/1-34
scaffold00122/1-34
scaffold01043/1-34
scaffold00246/1-34
scaffold01503/1-34
scaffold00516/1-34
scaffold00785/1-34
scaffold00656/1-34
scaffold00308/1-34
scaffold01542/1-34
scaffold00015/1-34
scaffold01056/1-34
scaffold00309/1-34
scaffold00008/1-34
scaffold00316/1-34
scaffold00269/1-34
scaffold01462/1-34
scaffold01624/1-34
scaffold00006/1-34
scaffold00980/1-34
scaffold00012/1-34
scaffold00698/1-34
scaffold00019/1-34
scaffold00769/1-34
scaffold00201/1-34
scaffold01888/1-34
scaffold00846/1-34
scaffold01297/1-34
scaffold00209/1-34
scaffold00484/1-34

Figure S4:

10
20
30
40
50
60
70
80
HP1a_CSD/1-41
scaffold00140.2/1-41
HP1b_CSD/1-41
CC4/CG8289_CD/1-41
scaffold00343.1/1-39
scaffold00228/1-55
scaffold00343/1-42
Pc_CD/1-41
scaffold00140.1/1-32
HP1c_CD/1-42
HP1a_CD/1-41
HP1b_CD/1-41
scaffold00372/44-90
CHD1_CD1_tandem/24-75
HP1e_CD/1-41
scaffold00138/1-42
Su(var)3-9_CD/1-42
scaffold01277/1-34
scaffold00308/1-34
scaffold01542/1-34
scaffold00001.1/1-38
scaffold00420/1-34
scaffold00613/1-34
scaffold00958/1-34
scaffold02240/1-34
scaffold00503/1-34
scaffold00316/1-34
scaffold00309/1-34
scaffold00269/1-34
scaffold01462/1-34
scaffold01624/1-34
scaffold00015/1-34
scaffold01056/1-34
scaffold00006/1-34
scaffold00012/1-34
scaffold00019/1-34
scaffold00201/1-34
scaffold00209/1-34
scaffold00484/1-34
scaffold00698/1-34
scaffold00769/1-34
scaffold00846/1-34
scaffold00980/1-34
scaffold01297/1-34
scaffold01888/1-34
scaffold00008/1-34
scaffold00121/1-34
scaffold00656/1-34
scaffold00700/1-34
scaffold00122/1-34
scaffold00246/1-34
scaffold00516/1-34
scaffold00785/1-34
scaffold01043/1-34
scaffold01503/1-34
Oxpecker_CD/1-42
HP1d/rhino_CD/1-42
scaffold00152/1-45
scaffold00725/1-61
Mi-2_CD1_tandem/1-59
scaffold00725.1/1-43
Mi-2_CD2_tandem/1-43
scaffold00152.1/1-48
Kismet_CD/1-48
scaffold00807/1-22
Chromator_CD/1-43
scaffold00372.1/1-43
CHD1_CD2_tandem/1-44
Consensus
YLVEKILRKRWA++T+++S++PSTSKTTRTPR+GGNKVLVKWLGFDNADLVFFQKIVYQIENDNKFFFLFPSHNTWEDKANV+DCP

Figure S1: Single copy orthologs, multi copy orthologs, and species-specific proteins identified by OrthoFinder analysis of 15 metazoan species.

Figure S2: Prediction of CpG methylation status based on observed/expected ratio of CpG dimers in exons of A) Human, B) Honey bee, and C) *L. boulardi.* Exons shorter than 100bp were excluded from the analysis. GpC content was used as control, which shows a unimodal distribution peaking at a ratio of 1 in all three species.

Figure S3: The phylogenetic tree constructed using the chromodomain/chromoshadow domain sequences of *D. melanogaster* and *L. boulardi*. The maximum likelihood tree was constructed using MEGA with 1000 bootstrap steps. The bootstrap values are shown at each branch. *Drosophila* sequences are named with protein and domain name. *Leptopilina* sequences are named after the scaffold number containing them.

Figure S4: The multiple sequence alignment of chromodomain/chromoshadow domain sequences of *D. melanogaster* and *L. boulardi*. *Drosophila* sequences are named with protein and domain name. *Leptopilina* sequences are named after the scaffold number containing them. The amino acids are highlighted in different colors based on their properties and degree of conservation. The consensus at each residue is represented in a bar graph at the bottom.