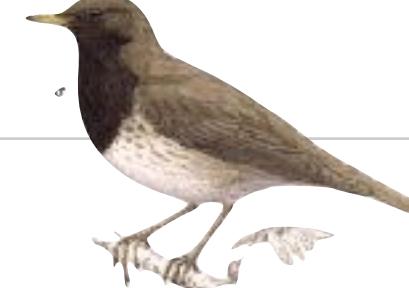
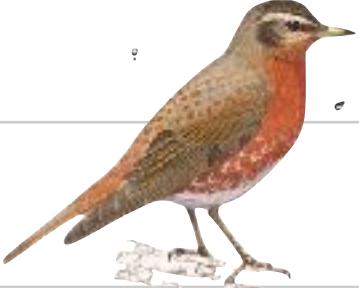
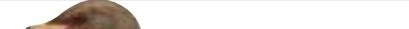


DUSKY THRUSH GENOME ASSEMBLY:

SEARCHING FOR PARALLEL PATTERNS IN BIRD COLORATION EVOLUTION

INPUT DATA

Species	library	notes	
Turdus eunomus	HiSeq 2 x 100 nt (600nt inserts)	1 lane 124,936,491 read pairs	
Turdus eunomus	HiSeq 2 x 100 nt (600nt inserts)	2 lanes 358,681,193 read pairs	
	3KB mate pairs (?)	202,549,012 read pairs	
Turdus eunomus	5KB mate pairs (?)	193,090,668 read pairs	
Turdus eunomus	single read	152,794,351 reads	
Turdus eunomus	overlapping pairends	265,398,145 read pairs	
Turdus atrogularis		single read	157,619,162 reads
Turdus atrogularis		single read	145,519,822 reads
	Turdus naumani	single read	147,645,002 reads
	Turdus naumani	single read	147,519,247 reads
	Turdus ruficollis	single read	145,715,239 reads
	Turdus ruficollis	single read	154,097,224 reads

GENOME ASSEMBLY

Assembly
contigs (>= 0 bp)
contigs (>= 10000 bp)
contigs (>= 50000 bp)
Total length (>= 0 bp)
Total length (>= 10000 bp)
Total length (>= 50000 bp)
contigs
Largest contig
Total length
GC (%)
N50
N75
L50
L75
N's per 100 kbp
Assembly2.scafSeq

1787669

17029

7049

1464699541

1100436484

848436212

80926

1439321

1196599872

41.19

94431

42565

3429

8117

32451.46



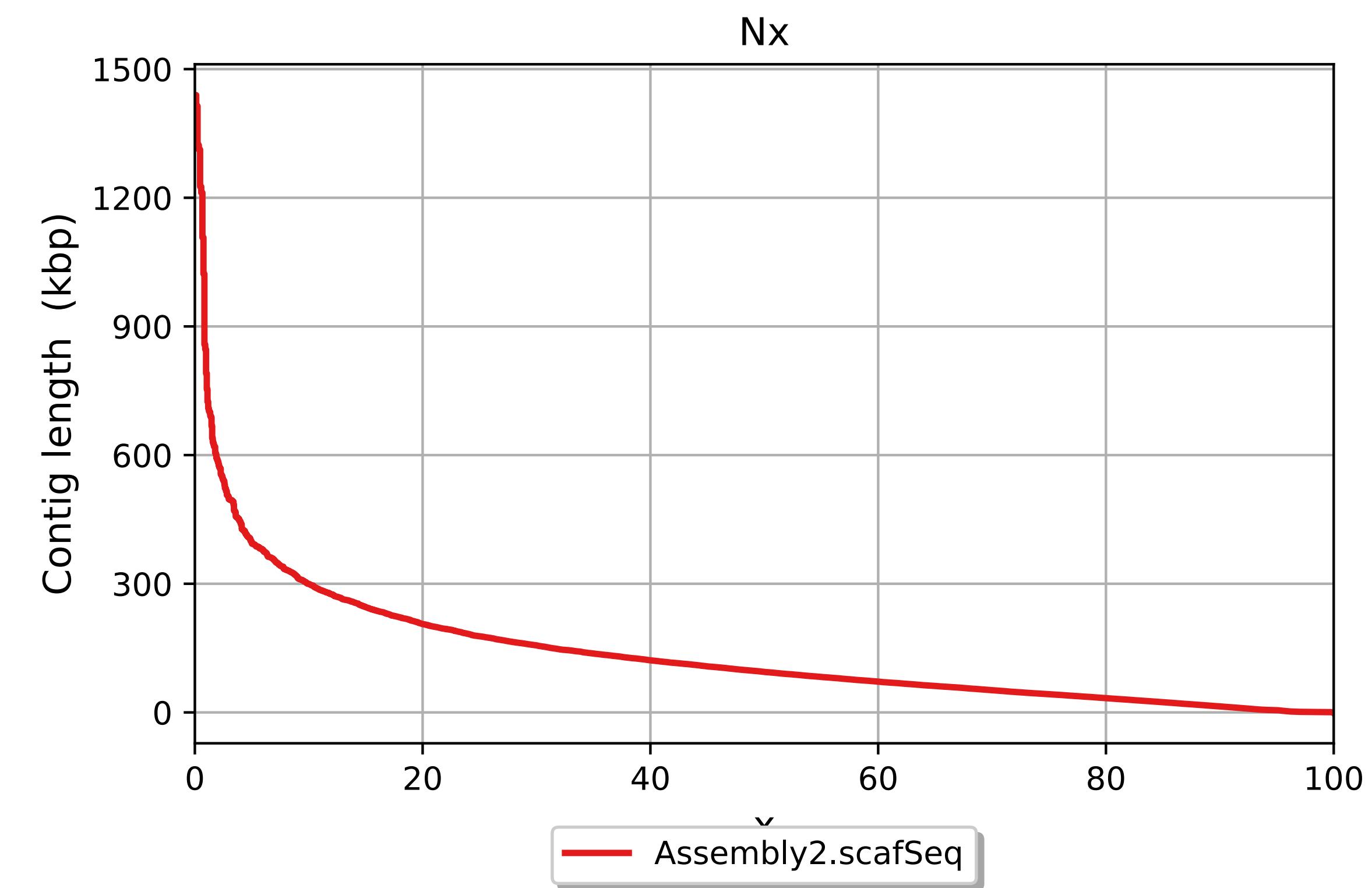
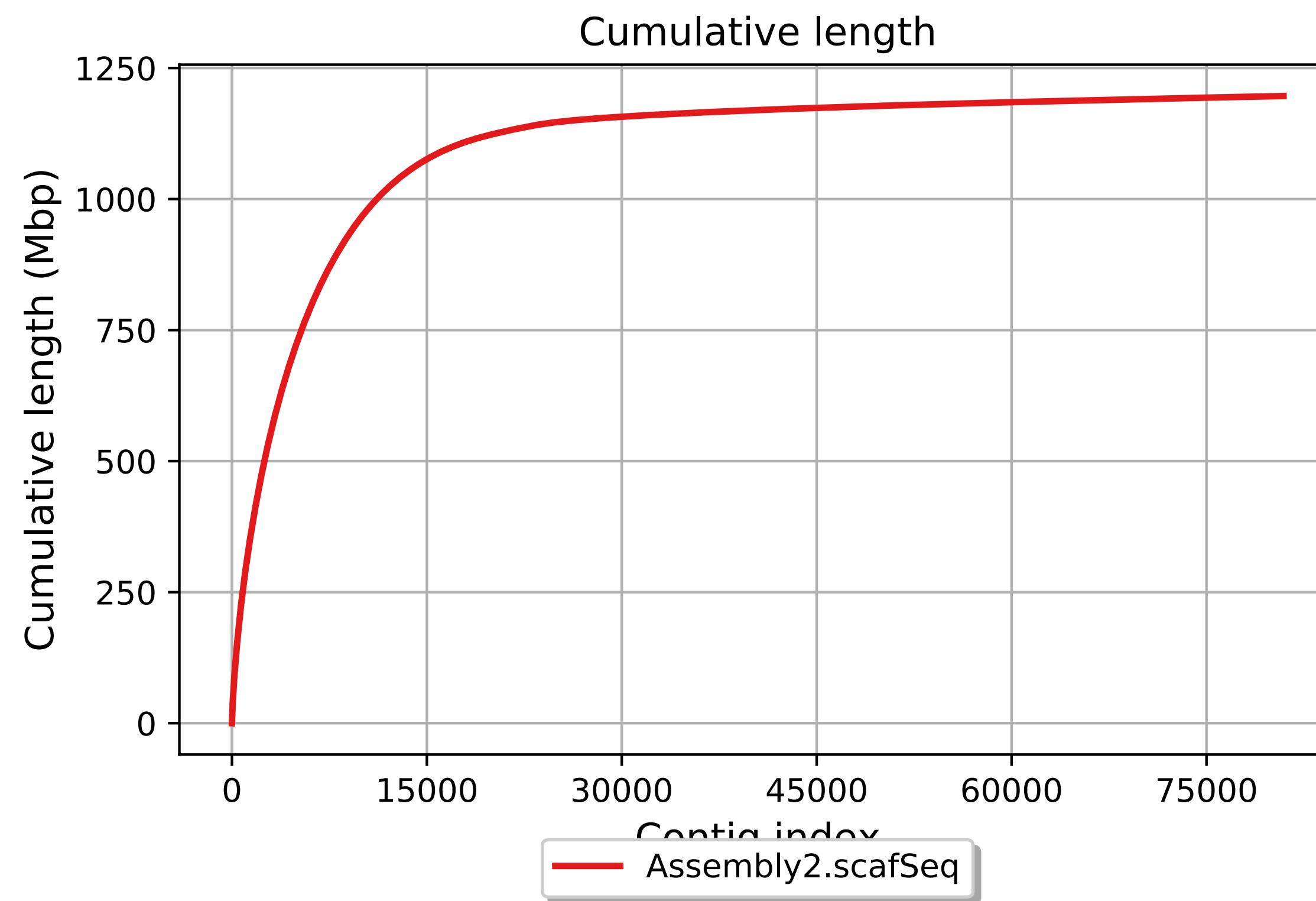
METHODS:
SOAP DE NOVO ASSEMBLY

DUSKY THRUSH GENOME ASSEMBLY



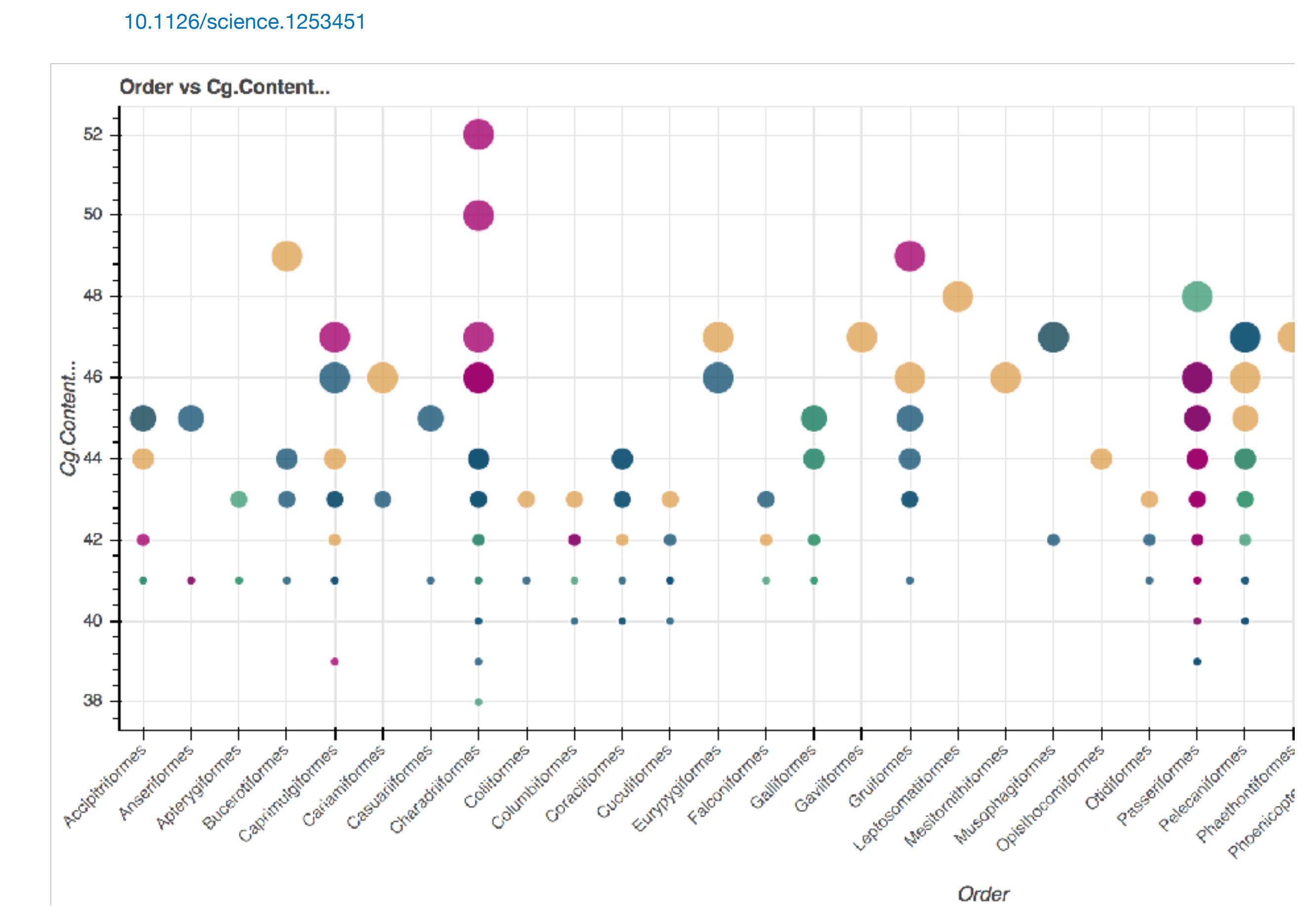
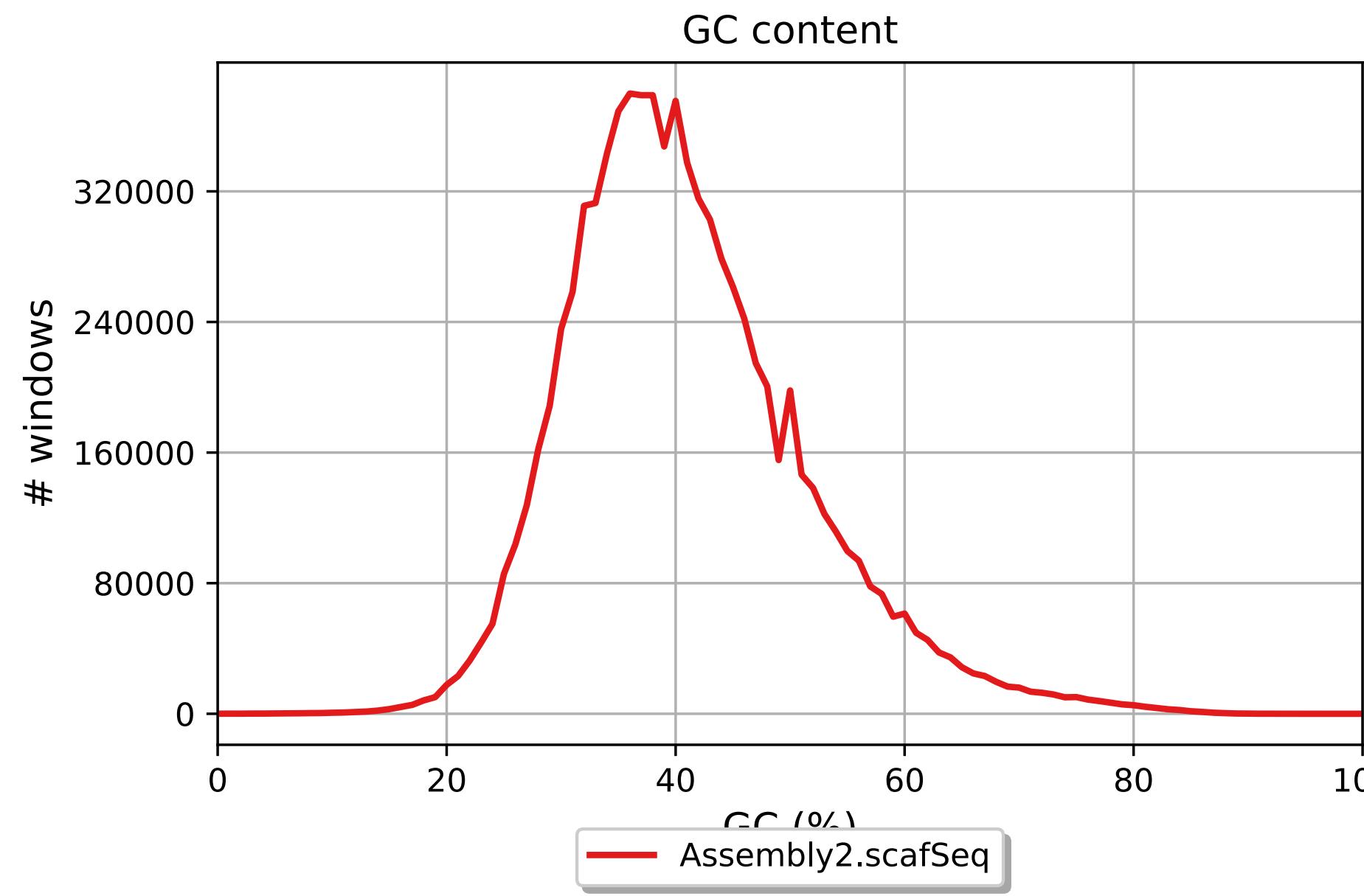
GENOME ASSEMBLY

Genome length and contiguity



DUSKY THRUSH GENOME ASSEMBLY

GC CONTENT EVALUATION



BUSCO

C:**40.8%**[S:40.8%,D:0.0%],F:23.9%,M:35.3%,n:255

104 Complete BUSCOs (C)

104 Complete and single-copy BUSCOs (S)

0 Complete and duplicated BUSCOs (D)

61 Fragmented BUSCOs (F)

90 Missing BUSCOs (M)

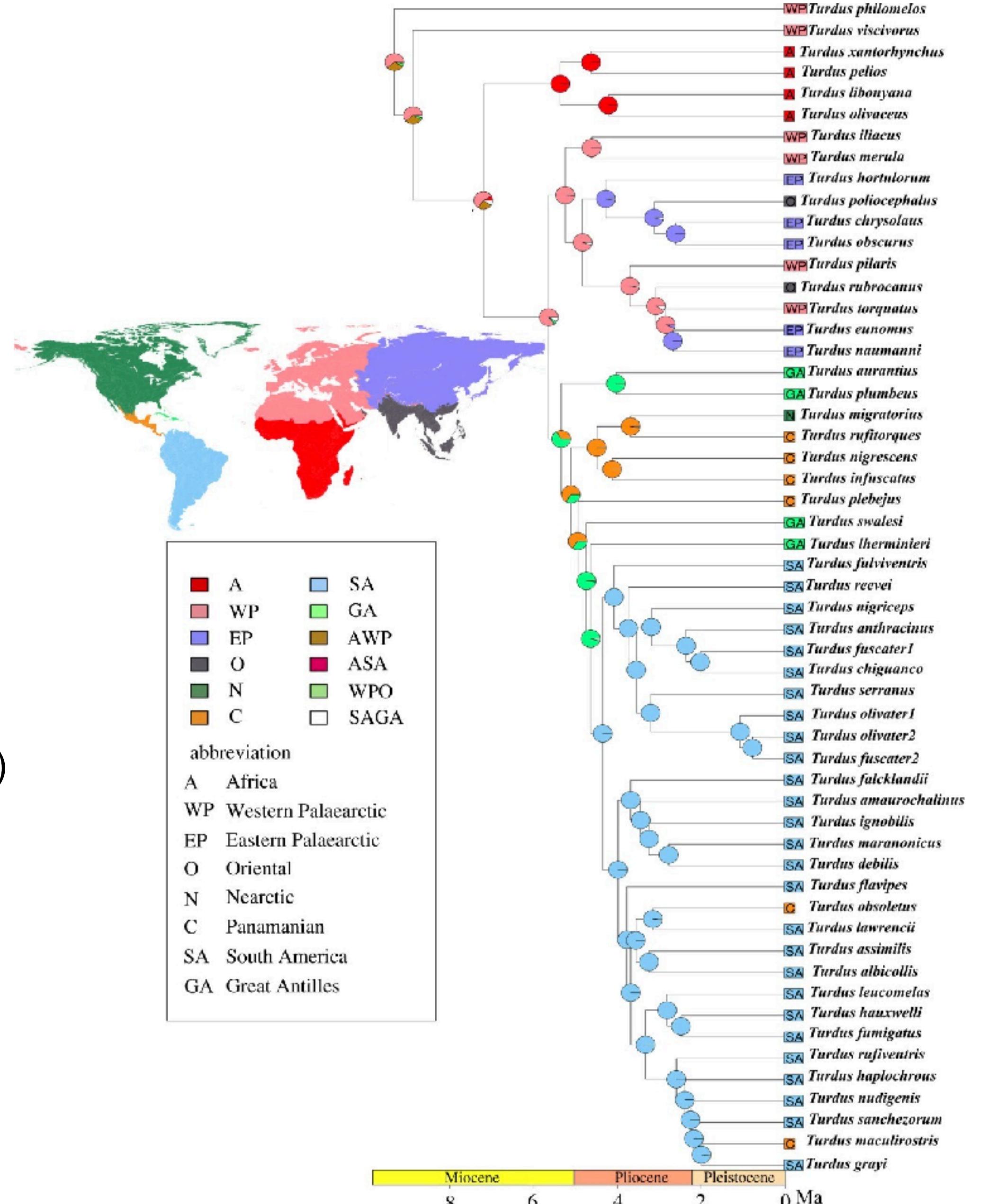
255 Total BUSCO groups searched

VARIANT CALLING: RELATEDNESS



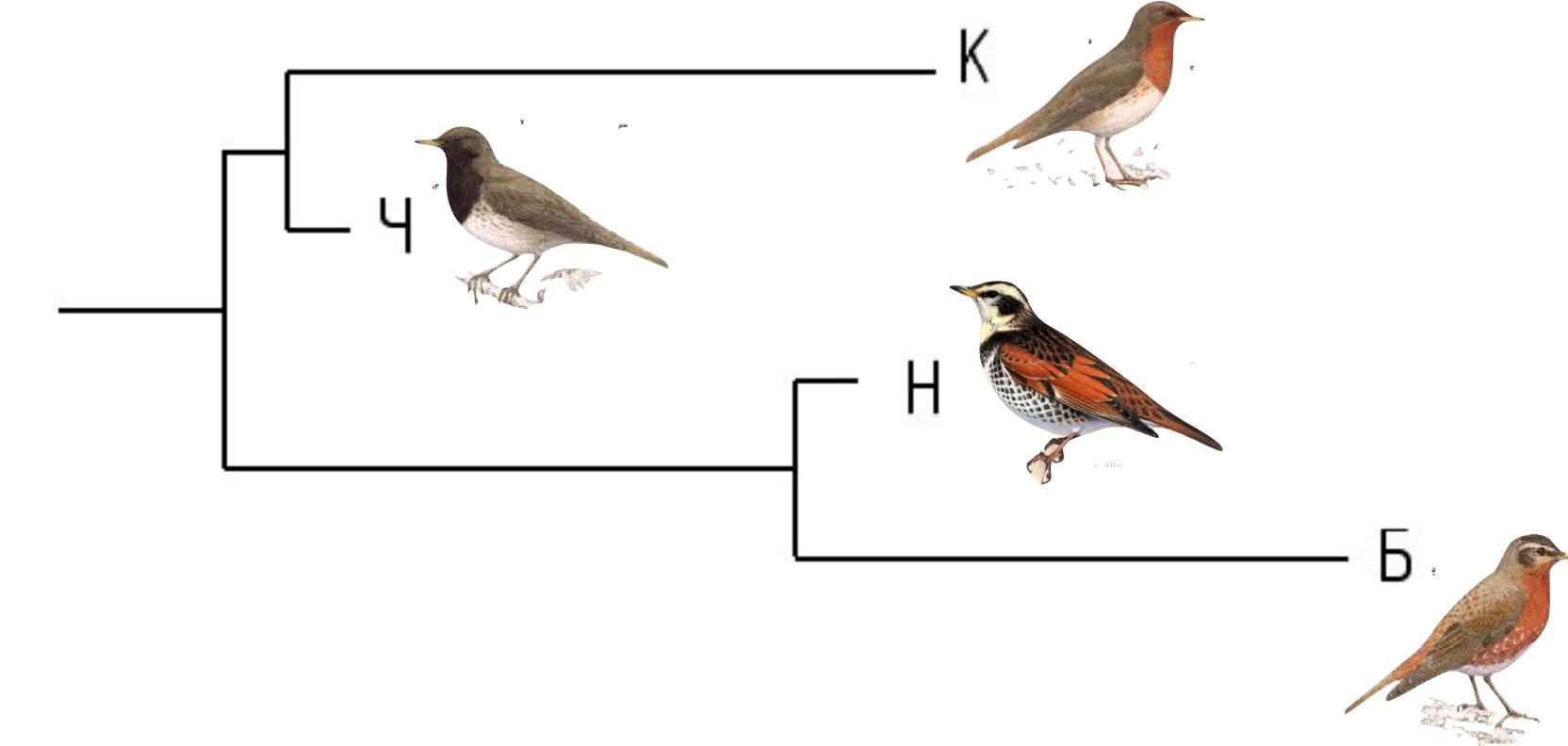
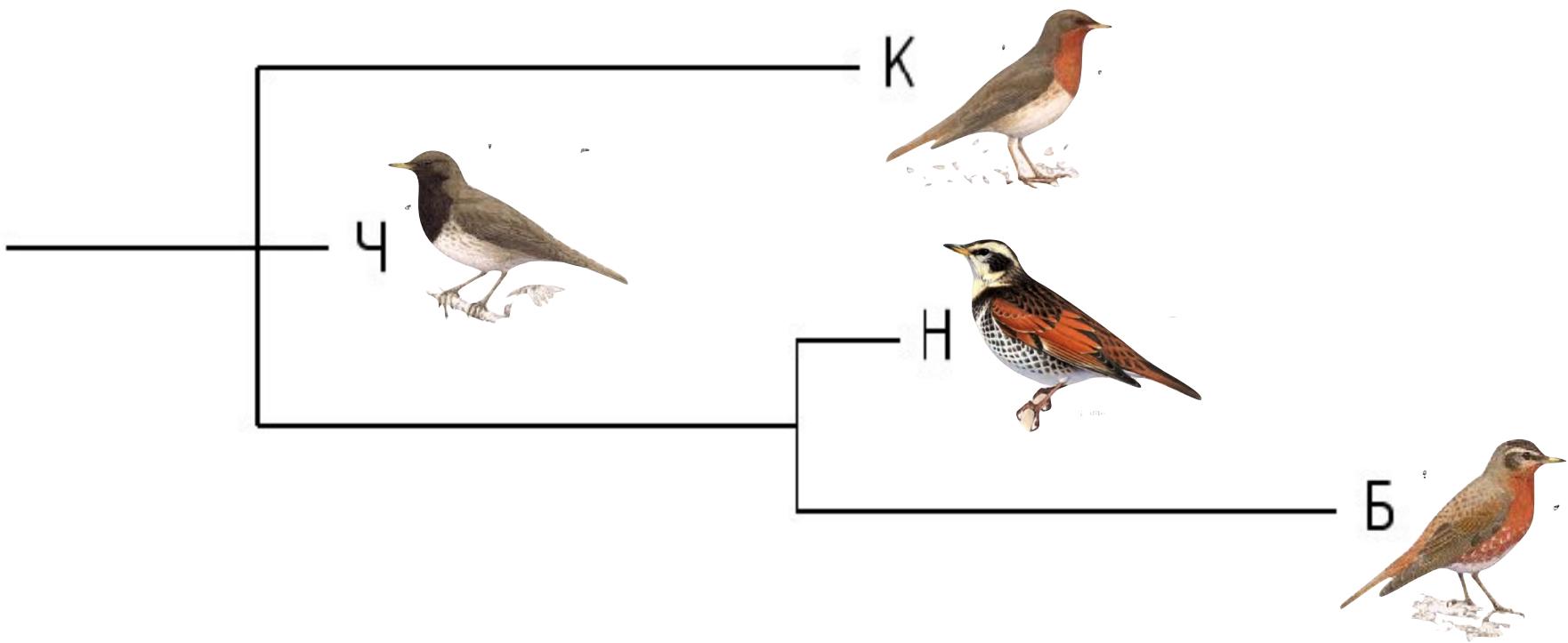
Fixed SNPs		90228 (1%)
Heterozygous	35537 (0.4%)	6781 (0.07%)
Total		9,077,325

METHODS: BWA > Samtools > Vcftools > Custom Scripts (Python)



DUSKY THRUSH GENOME ASSEMBLY

PHYLOGENY PROXY



CANDIDATE GENES

Total number of “parallel” SNPs: **647** (hits in unassembled contigs excluded)

Summary of BLAST hits:

Genes: **136** (including multiple hits)

Non-genic: **81**

Genomes (*BLAST shows hits to fasta, annotation information can be added): **302**

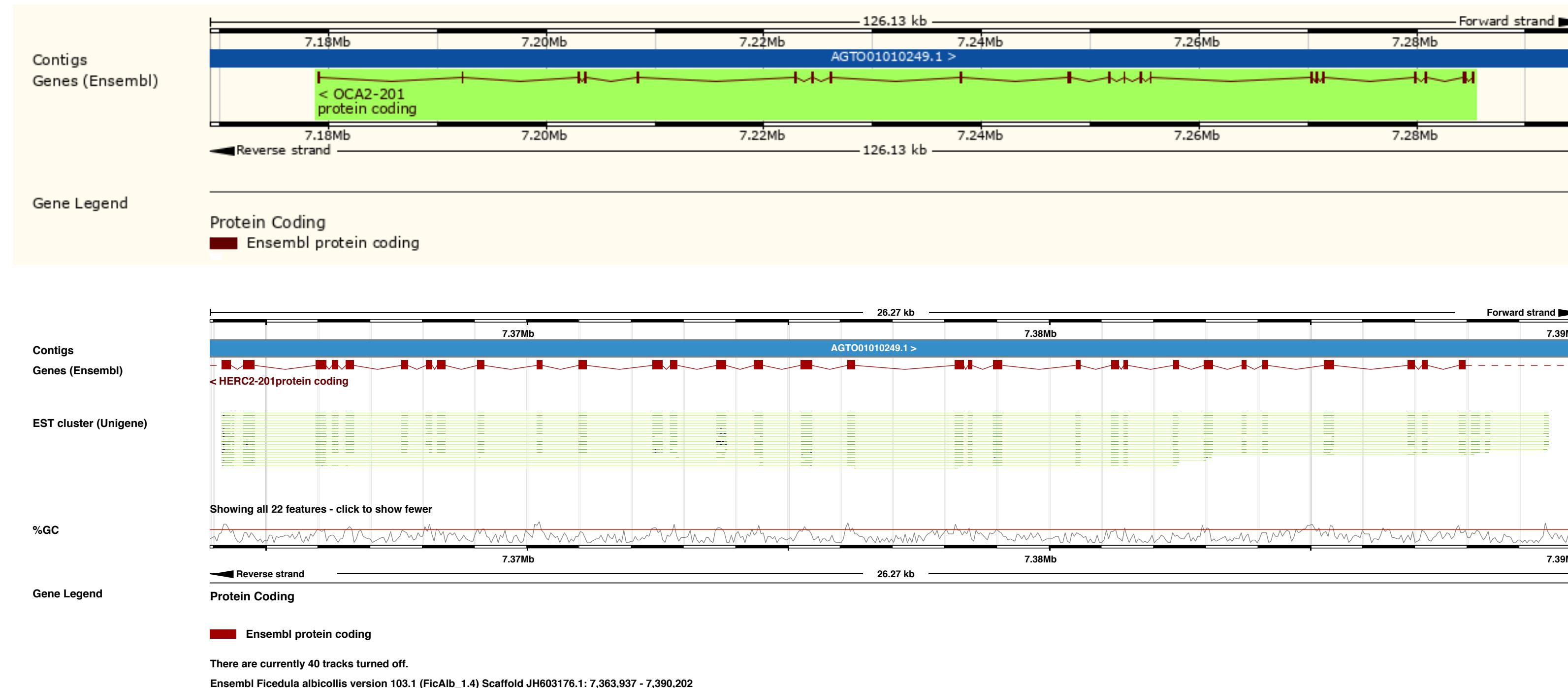
No hit: **128**

CANDIDATE GENES - BINGO!

scaffold730	45181-46181	NULL	NULL	NULL
scaffold730	67150-68150	NULL	NULL	NULL
scaffold730	68220-69220	Aquila chrysaetos chrysaetos genome assembly, chromosome: 23	LR606203.1	10357719 to 10357941
		PREDICTED: Catharus ustulatus OCA2 melanosomal transmembrane protein (OCA2), transcript variant X3, mRNA	XM_033053547.1	333 to 365
scaffold730	68471-69471	PREDICTED: Catharus ustulatus OCA2 melanosomal transmembrane protein (OCA2), transcript variant X3, mRNA	XM_033053547.1	333 to 365
scaffold730	68956-69956	Aquila chrysaetos chrysaetos genome assembly, chromosome: 23	LR606203.1	10358636 to 10359400
		PREDICTED: Catharus ustulatus OCA2 melanosomal transmembrane protein (OCA2), transcript variant X3, mRNA	XM_033053547.1	333 to 365
scaffold730	68982-69982	Aquila chrysaetos chrysaetos genome assembly, chromosome: 23	LR606203.1	10358636 to 10359406
		PREDICTED: Catharus ustulatus OCA2 melanosomal transmembrane protein (OCA2), transcript variant X3, mRNA	XM_033053547.1	333 to 365
scaffold730	70114-71114	PREDICTED: Corvus brachyrhynchos OCA2 melanosomal transmembrane protein (OCA2), transcript variant X12, mRNA	XM_017728600.1	1 to 319
scaffold730	70295-71295	PREDICTED: Corvus brachyrhynchos OCA2 melanosomal transmembrane protein (OCA2), transcript variant X12, mRNA	XM_017728600.1	1 to 319
scaffold730	71201-72201	Aquila chrysaetos chrysaetos genome assembly, chromosome: 23	LR606203.1	10360337 to 10360407
scaffold730	72020-73020	NULL	NULL	NULL
scaffold730	72324-73324	NULL	NULL	NULL
scaffold730	73045-74045	NULL	NULL	NULL
scaffold730	73067-74067	NULL	NULL	NULL
scaffold730	74321-75321	Aquila chrysaetos chrysaetos genome assembly, chromosome: 23	LR606203.1	10371358 to 10371657
scaffold730	75793-76793	NULL	NULL	NULL
scaffold730	75832-76832	NULL	NULL	NULL
scaffold730	75861-76861	NULL	NULL	NULL
scaffold730	75933-76933	NULL	NULL	NULL
scaffold730	77109-78109	NULL	NULL	NULL
scaffold730	78275-79275	Aquila chrysaetos chrysaetos genome assembly, chromosome: 18	LR606198.1	14133647 to 14133730
scaffold730	79820-80820	NULL	NULL	NULL
scaffold730	85480-86480	PREDICTED: Catharus ustulatus HECT and RLD domain containing E3 ubiquitin protein ligase 2 (HERC2), transcript variant X3, mRNA	XM_033053064.1	14581 to 14876
scaffold730	85498-86498	PREDICTED: Catharus ustulatus HECT and RLD domain containing E3 ubiquitin protein ligase 2 (HERC2), transcript variant X3, mRNA	XM_033053064.1	14581 to 14858
scaffold730	87768-88768	Aquila chrysaetos chrysaetos genome assembly, chromosome: 23	LR606203.1	10376271 to 10376739
		PREDICTED: Catharus ustulatus HECT and RLD domain containing E3 ubiquitin protein ligase 2 (HERC2), transcript variant X3, mRNA	XM_033053064.1	14026 to 14140
scaffold730	89571-90571	Aquila chrysaetos chrysaetos genome assembly, chromosome: 23	LR606203.1	10377760 to 10378303
		PREDICTED: Lonchura striata domestica HECT and RLD domain containing E3 ubiquitin protein ligase 2 (HERC2), transcript variant X3, mRNA	XM_031504129.1	14126 to 14249
scaffold730	89598-90598	Aquila chrysaetos chrysaetos genome assembly, chromosome: 23	LR606203.1	10377760 to 10378303
		PREDICTED: Lonchura striata domestica HECT and RLD domain containing E3 ubiquitin protein ligase 2 (HERC2), transcript variant X3, mRNA	XM_031504129.1	14126 to 14249
scaffold730	92371-93371	Aquila chrysaetos chrysaetos genome assembly, chromosome: 23	LR606203.1	10380195 to 10380527
		PREDICTED: Catharus ustulatus HECT and RLD domain containing E3 ubiquitin protein ligase 2 (HERC2), transcript variant X3, mRNA	XM_033053064.1	13461 to 13533
scaffold730	92463-93463	Aquila chrysaetos chrysaetos genome assembly, chromosome: 23	LR606203.1	10380292 to 10380527
		PREDICTED: Catharus ustulatus HECT and RLD domain containing E3 ubiquitin protein ligase 2 (HERC2), transcript variant X3, mRNA	XM_033053064.1	13461 to 13533
scaffold730	93252-94252	NULL	NULL	NULL
scaffold730	93511-94511	NULL	NULL	NULL
scaffold730	93513-94513	NULL	NULL	NULL
scaffold730	93999-94999	NULL	NULL	NULL
scaffold730	94173-95173	NULL	NULL	NULL
scaffold730	96067-97067	NULL	NULL	NULL

CANDIDATE GENES - WHAT'S NEXT?

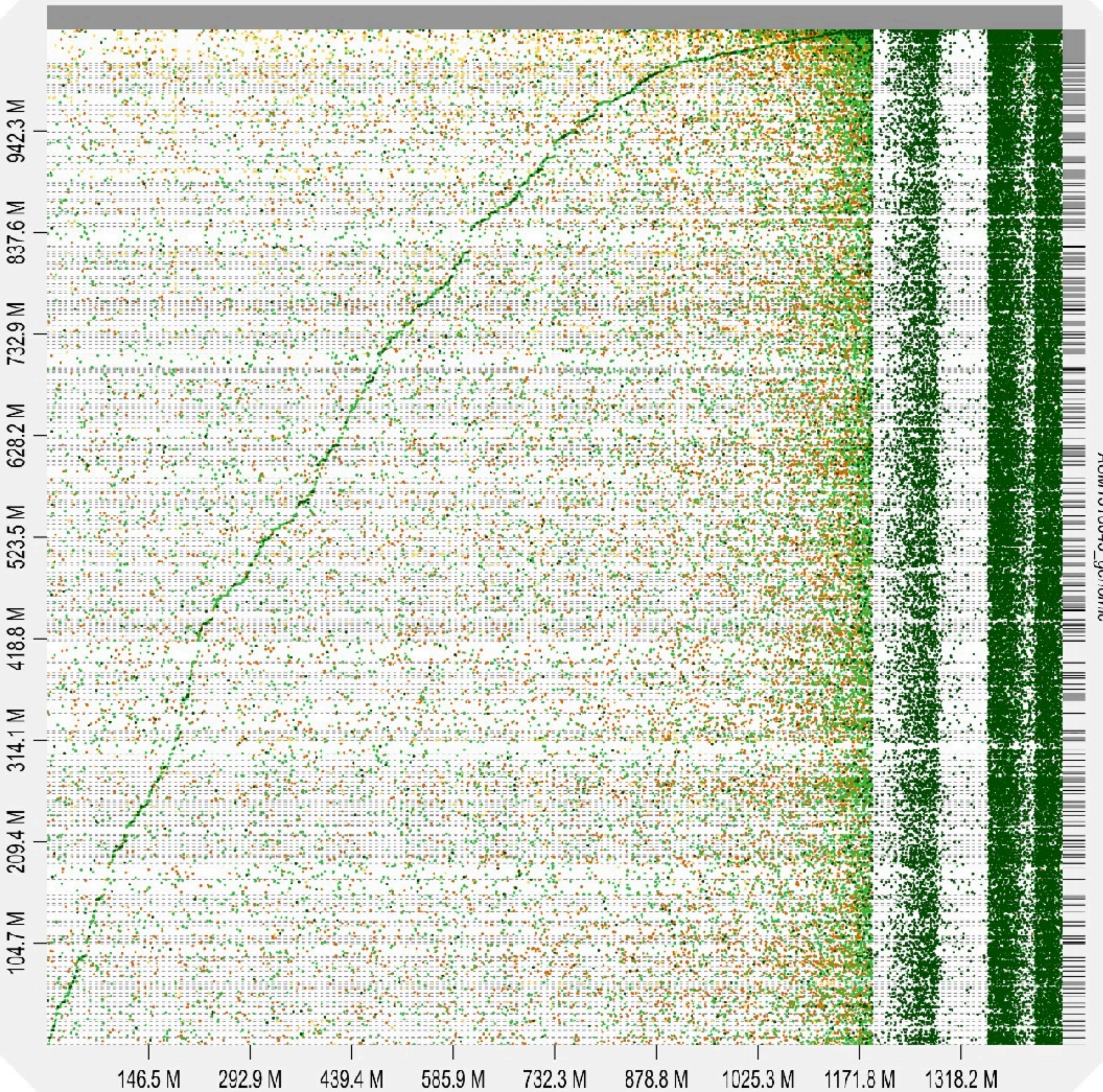
1. Closer look at the region. That's how it looks in flycatcher:



1. Visualize the result in a pseudo genome scan: proves, that this concentration of SNPs is real
2. Detecting selection: dN/dS, snpEFF
3. Reference-based scaffolding?

DUSKY THRUSH GENOME ASSEMBLY

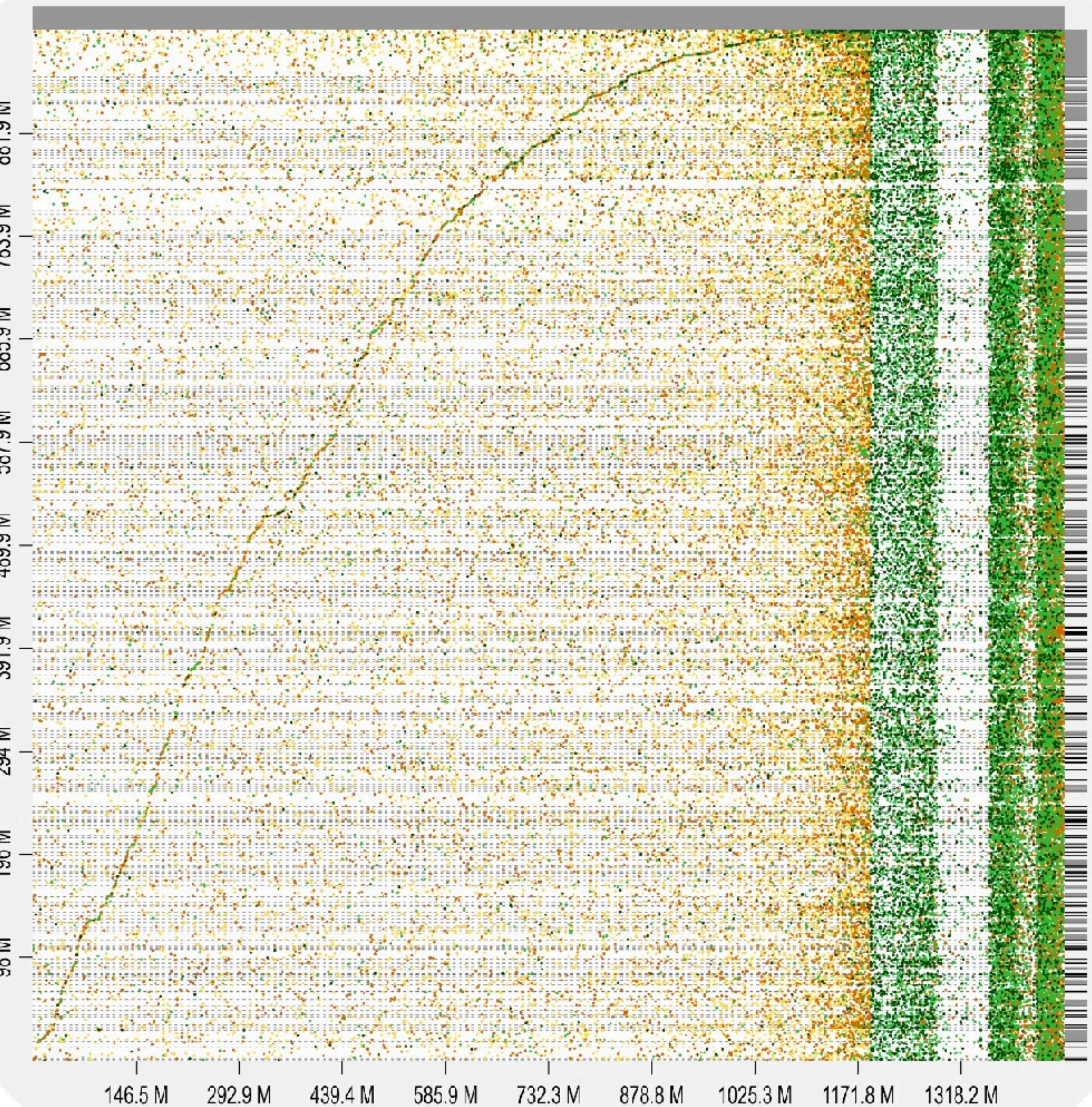
GENOME SYNTENY - THRUSH



DUSKY THRUSH GENOME ASSEMBLY

GENOME SYNTENY - ROBIN

Assembly2.scafSeq



GCA_013398395.1_ASM1339839v1_genomic