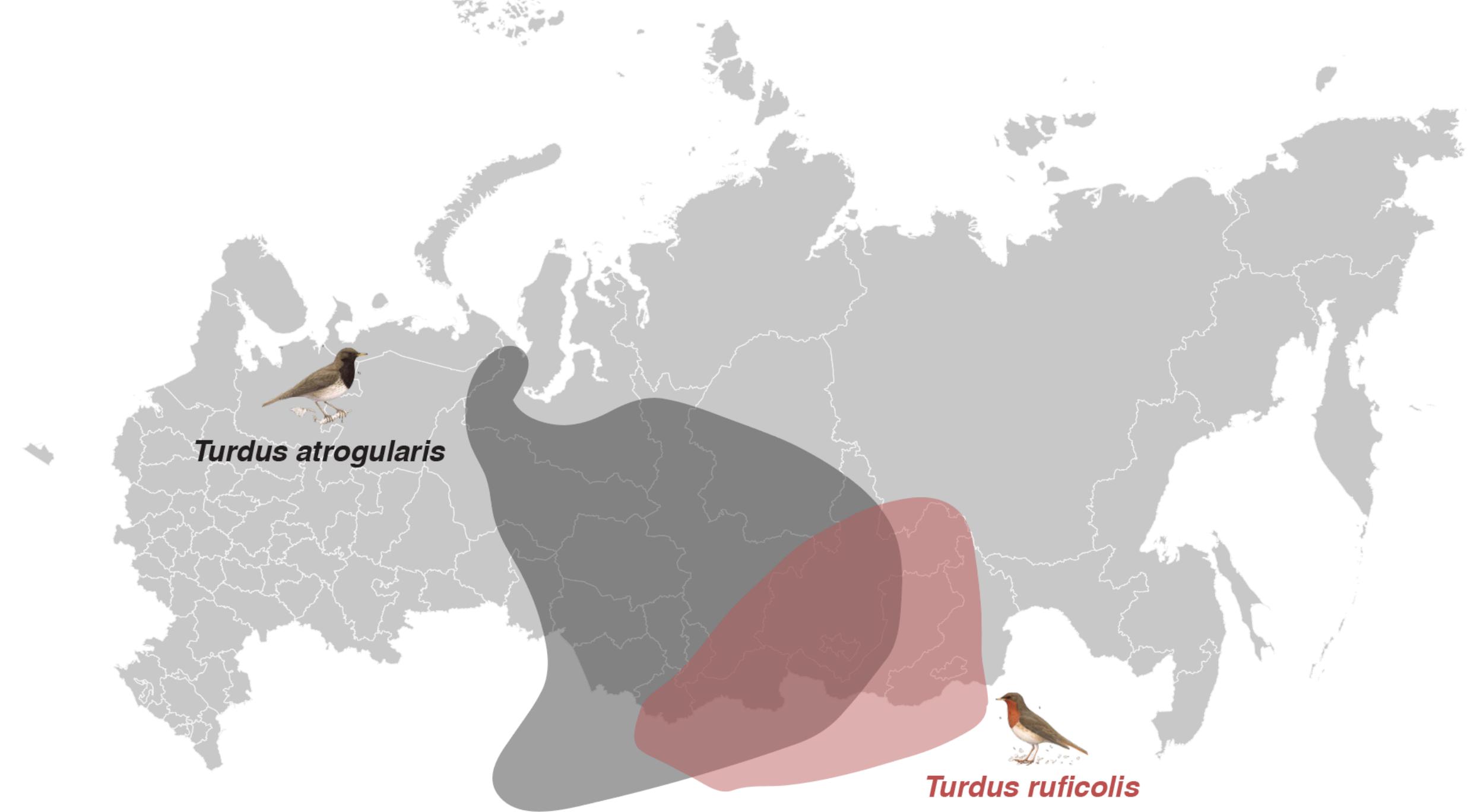


GENOMICS OF PARALLELISM IN BIRD COLORATION:

CASE STUDY IN TURDUS GENUS

GENOMICS OF PARALLELISM IN BIRD COLORATION

STUDY SYSTEM



GENOMICS OF PARALLELISM IN BIRD COLORATION

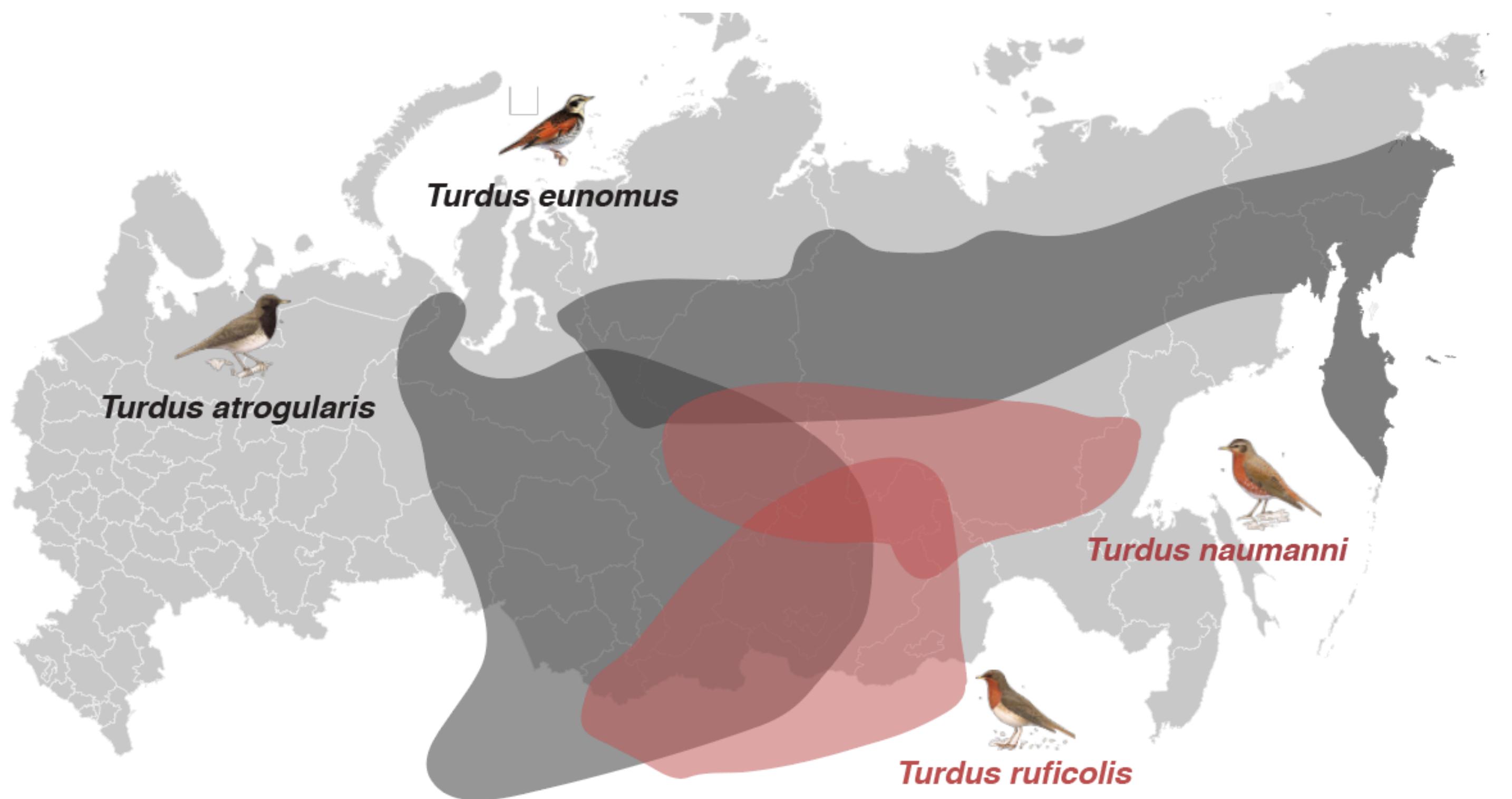
STUDY SYSTEM



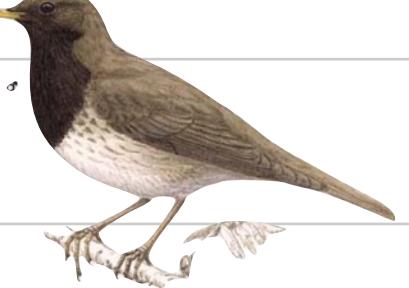
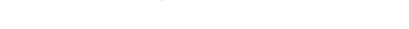
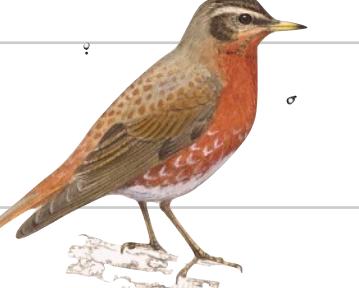
GOALS

Pilot: parallelism in coloration genes

1. Search for parallel substitutions on the genome level
 - Genome assembly
 - Search for the pattern and genome scans
2. Biology of the species, basis for reproductive isolation
 - Vocalization
 - Ecology

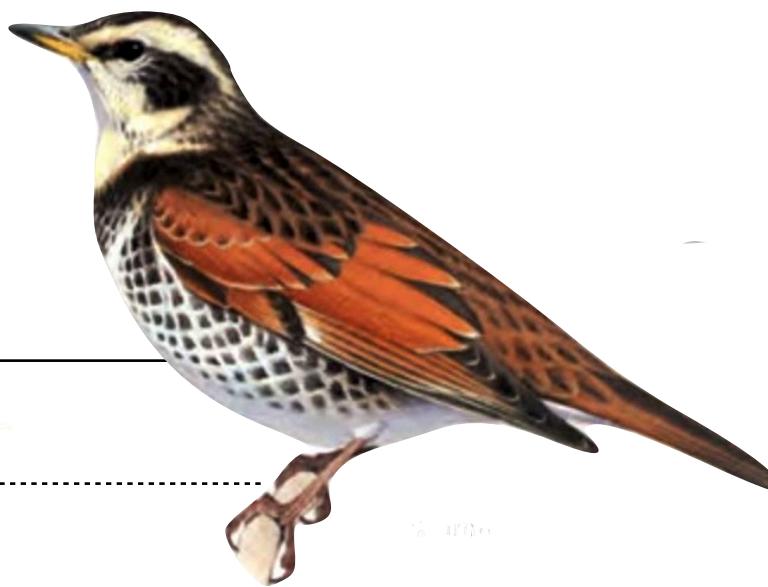


INPUT DATA

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

GENOME ASSEMBLY

Assembly	Assembly2.scafSeq
# contigs (>= 0 bp)	1787669
# contigs (>= 10000 bp)	17029
# contigs (>= 50000 bp)	7049
Total length (>= 0 bp)	1464699541
Total length (>= 10000 bp)	1100436484
Total length (>= 50000 bp)	848436212
# contigs	80926
Largest contig	1439321
Total length	1196599872
GC (%)	41.19
N50	94431
N75	42565
L50	3429
L75	8117
# N's per 100 kbp	32451.46

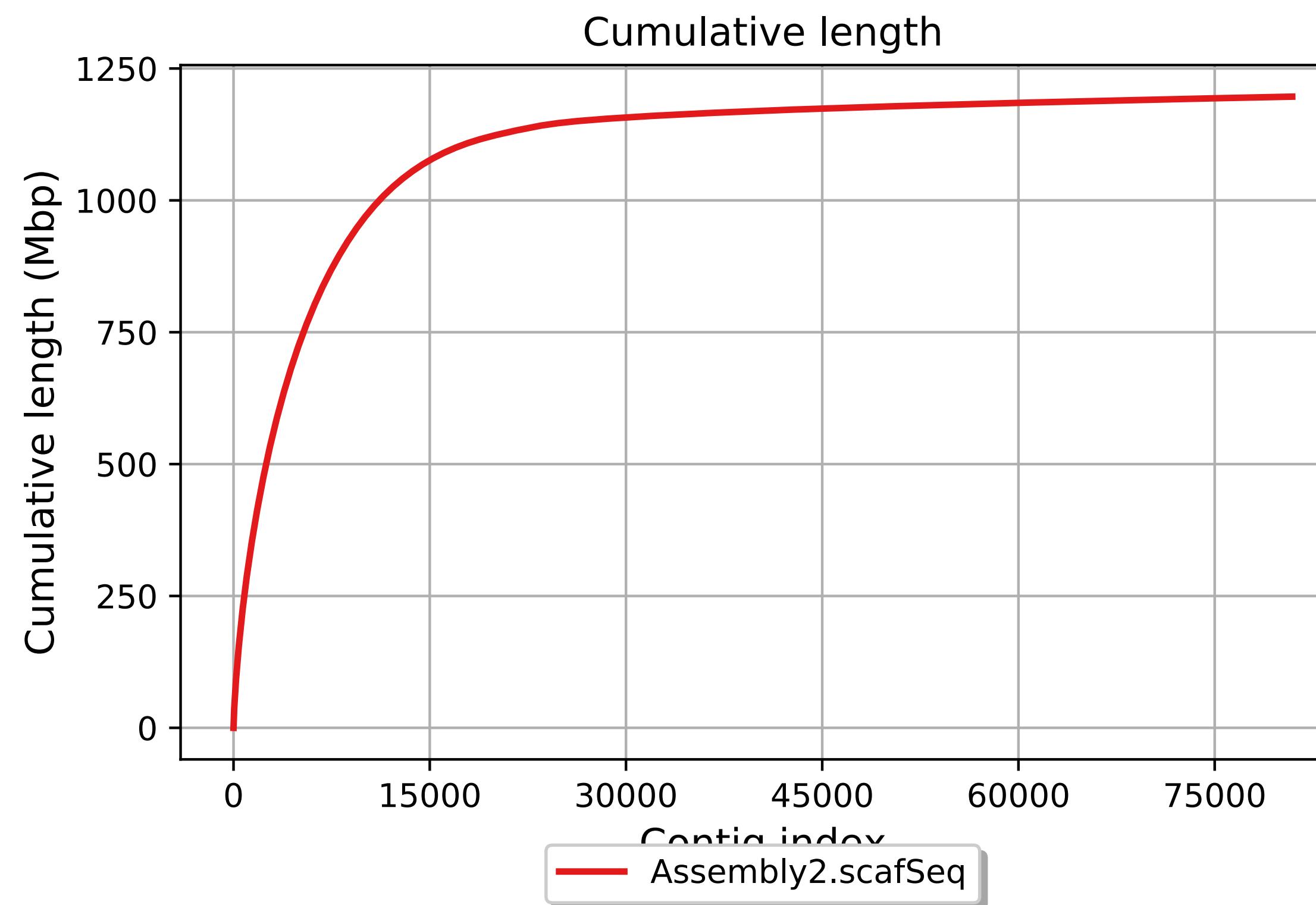


METHODS:
SOAP DE NOVO ASSEMBLY



GENOME ASSEMBLY

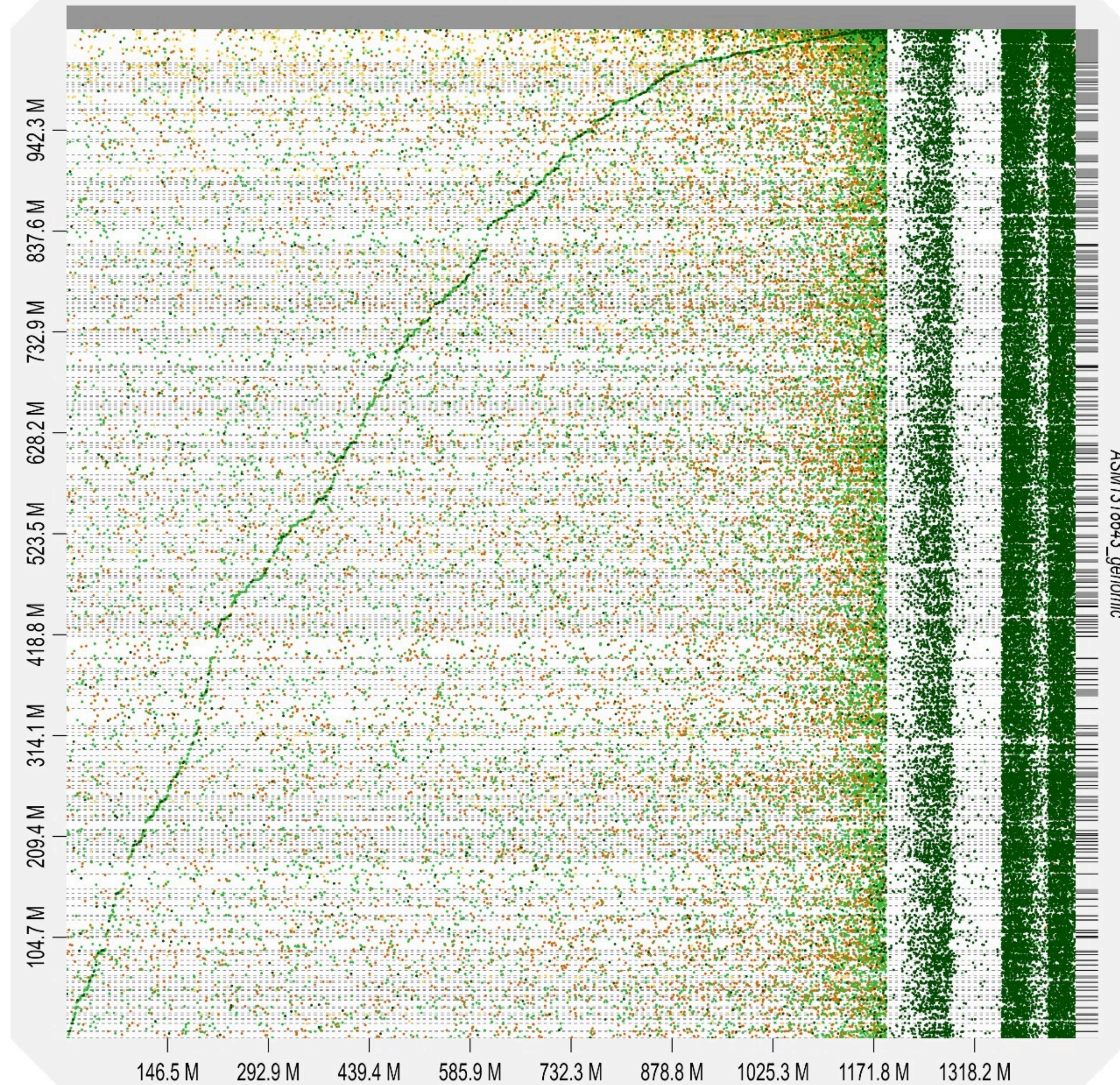
Genome length and contiguity



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

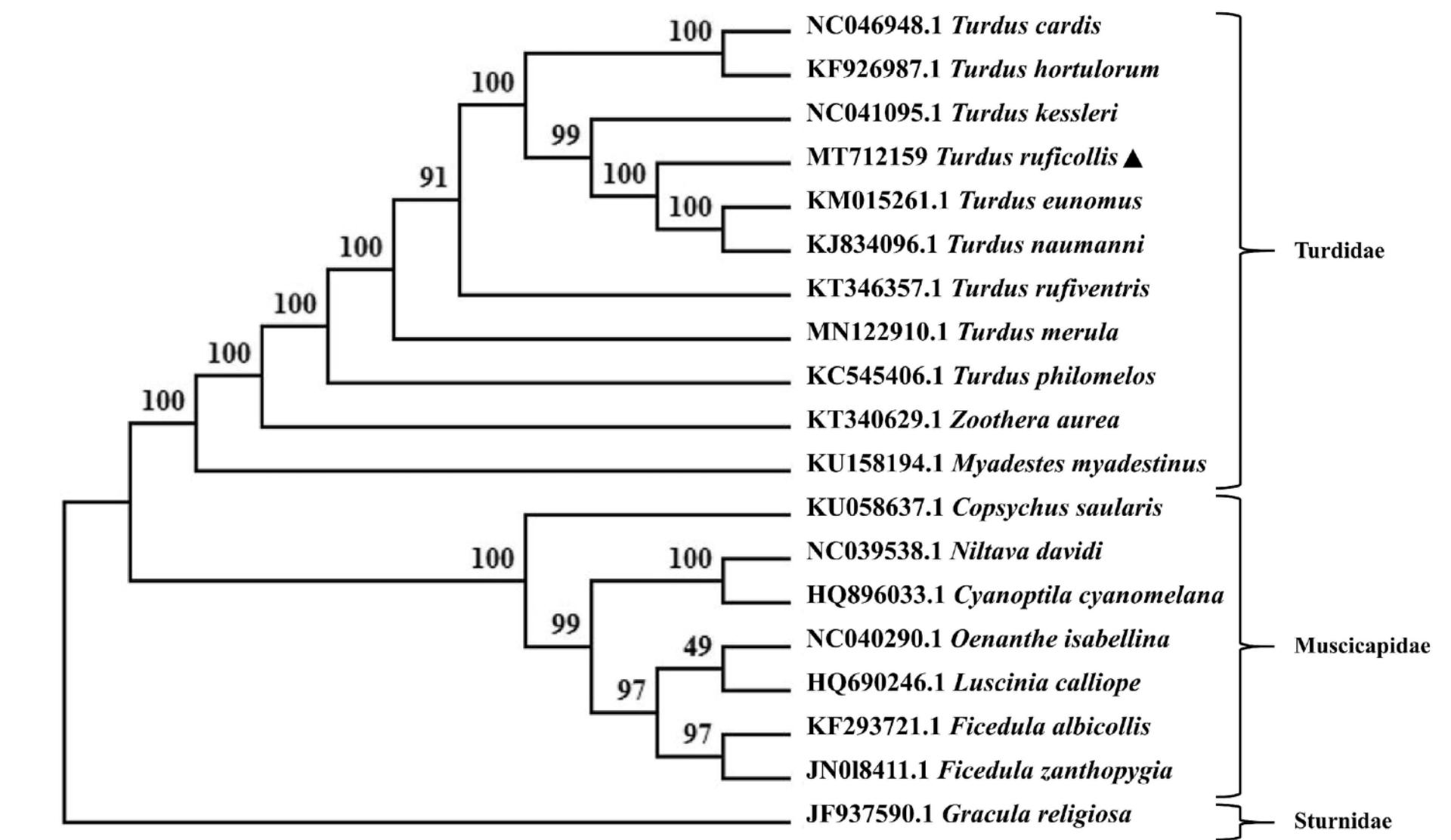
GENOMICS OF PARALLELISM IN BIRD COLORATION
**GENOME SYNTENY -
THRUSH**



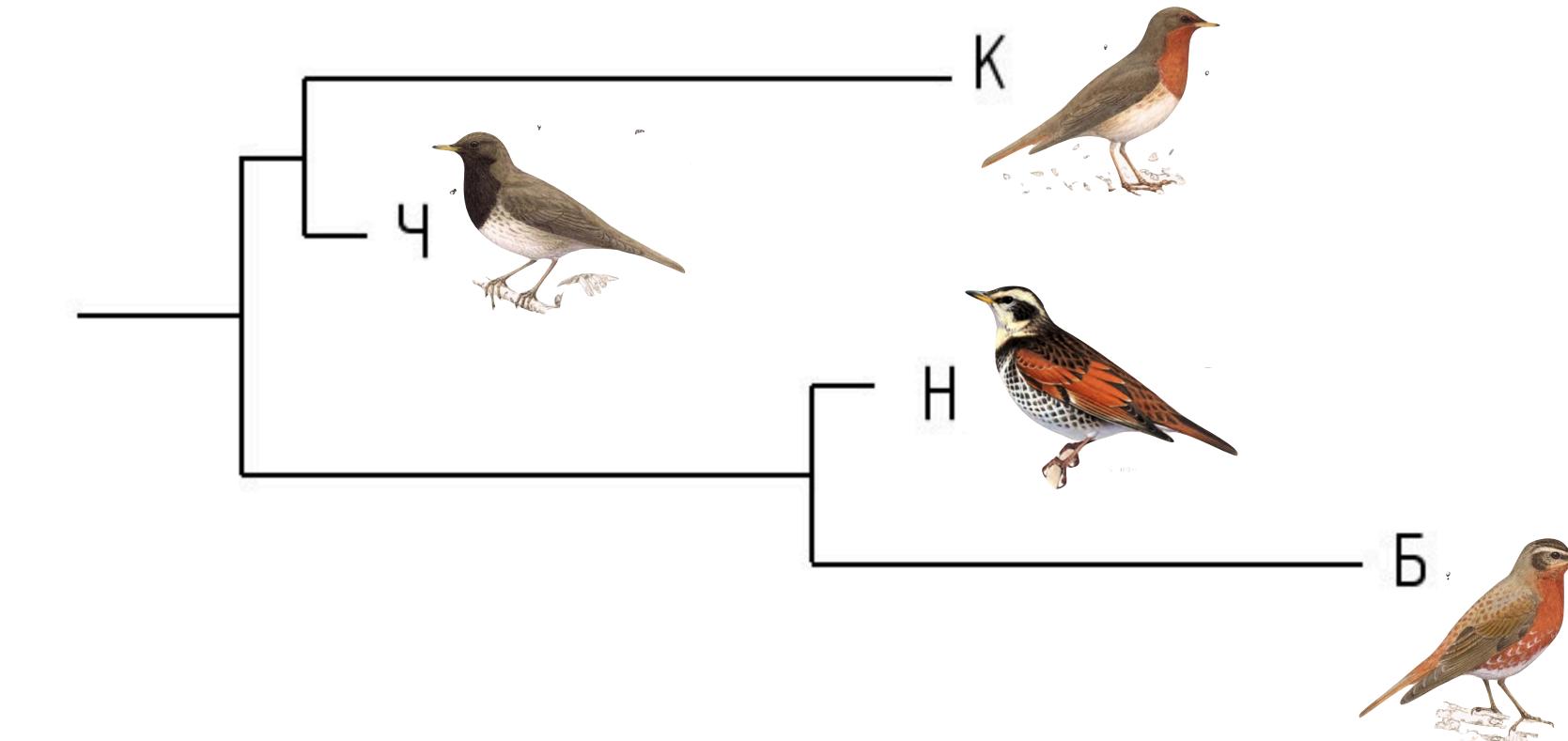
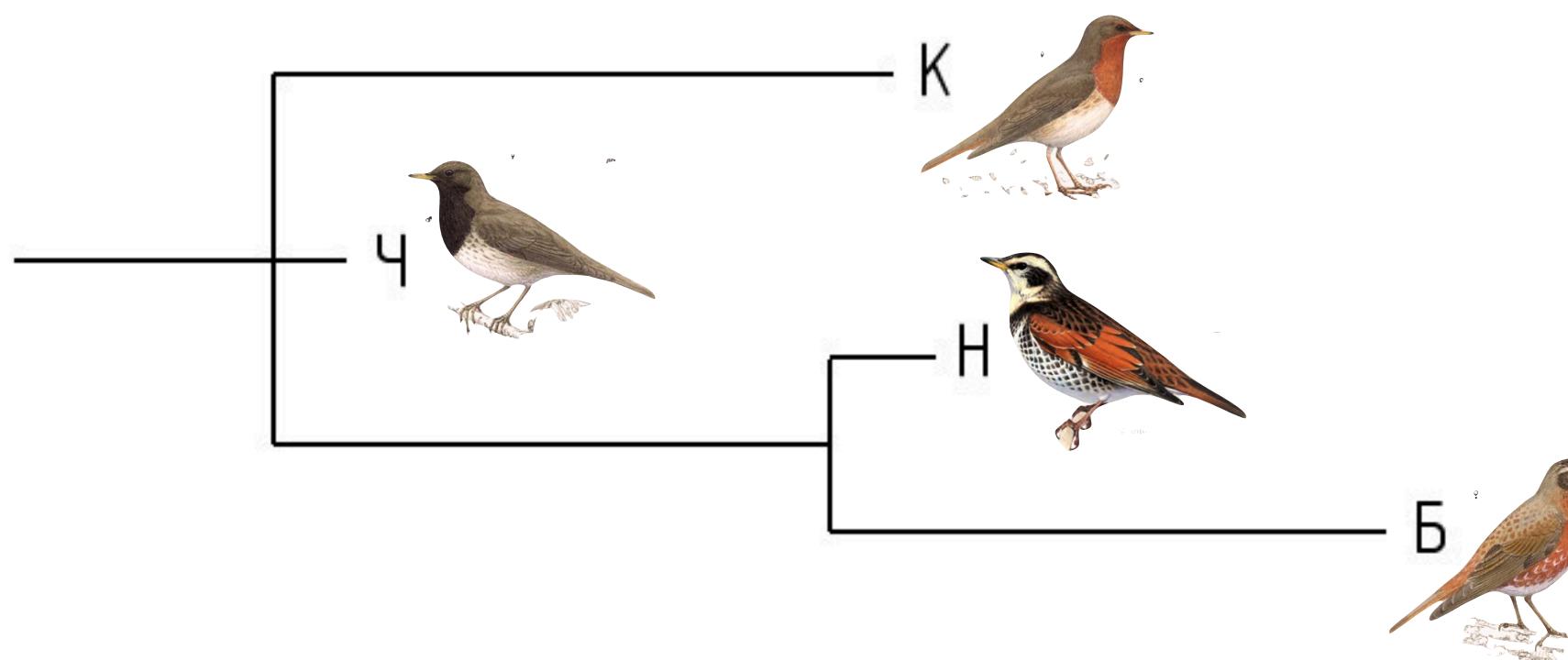
VARIANT CALLING



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



(Zhang, 2021)



SCAN FOR PARALLEL SUBSTITUTIONS

Total number of “parallel” SNPs: **647**



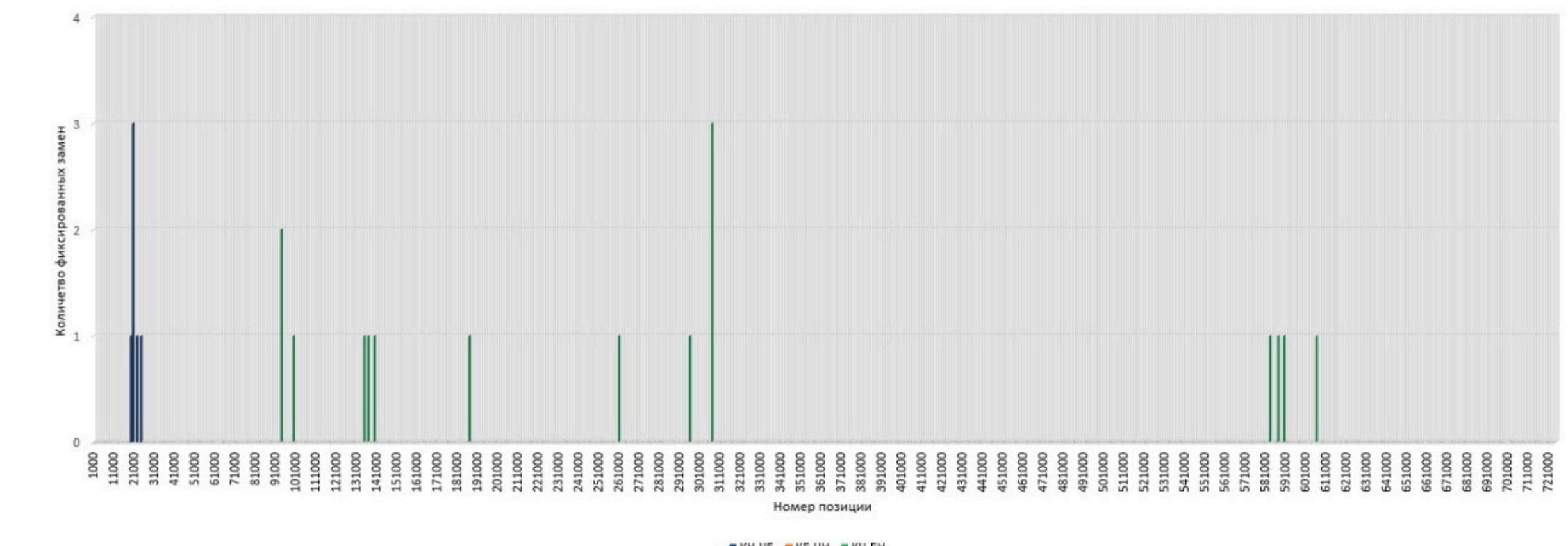
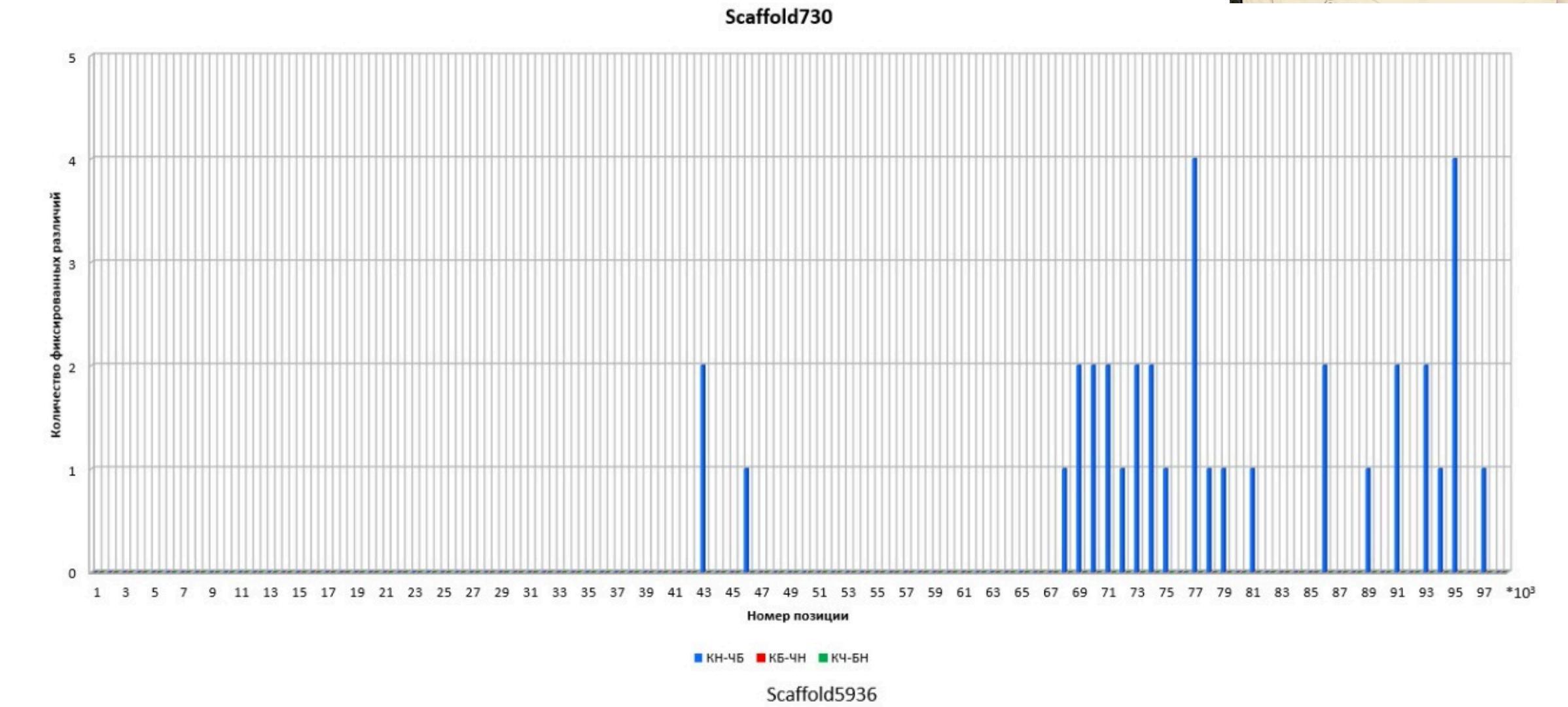
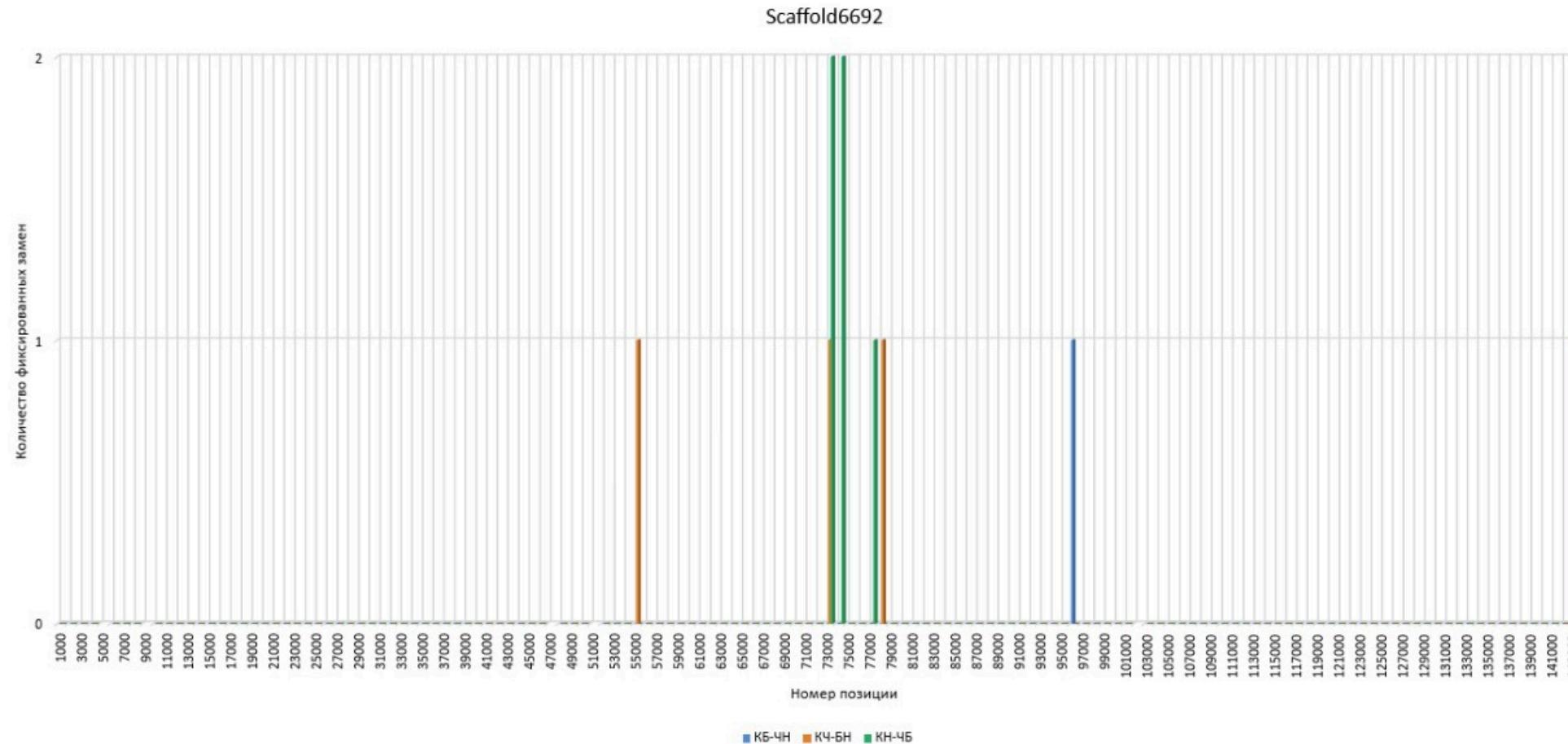
Summary of BLAST hits:

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

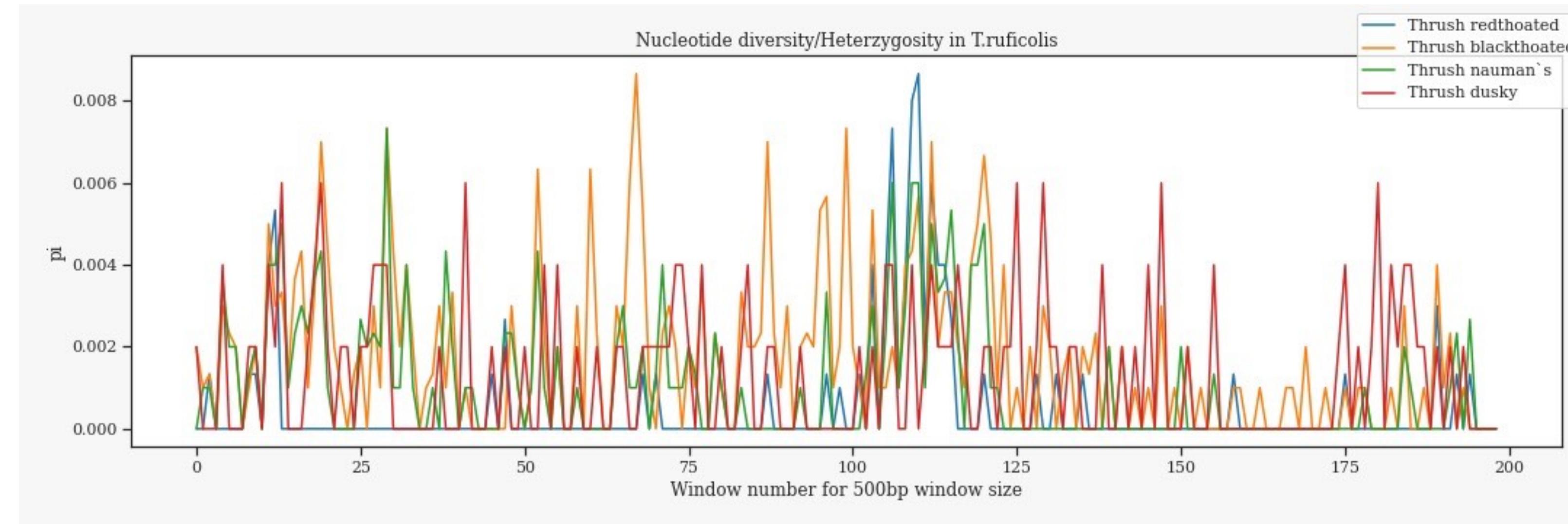
Non-genic: **81**

Genomes: **302**

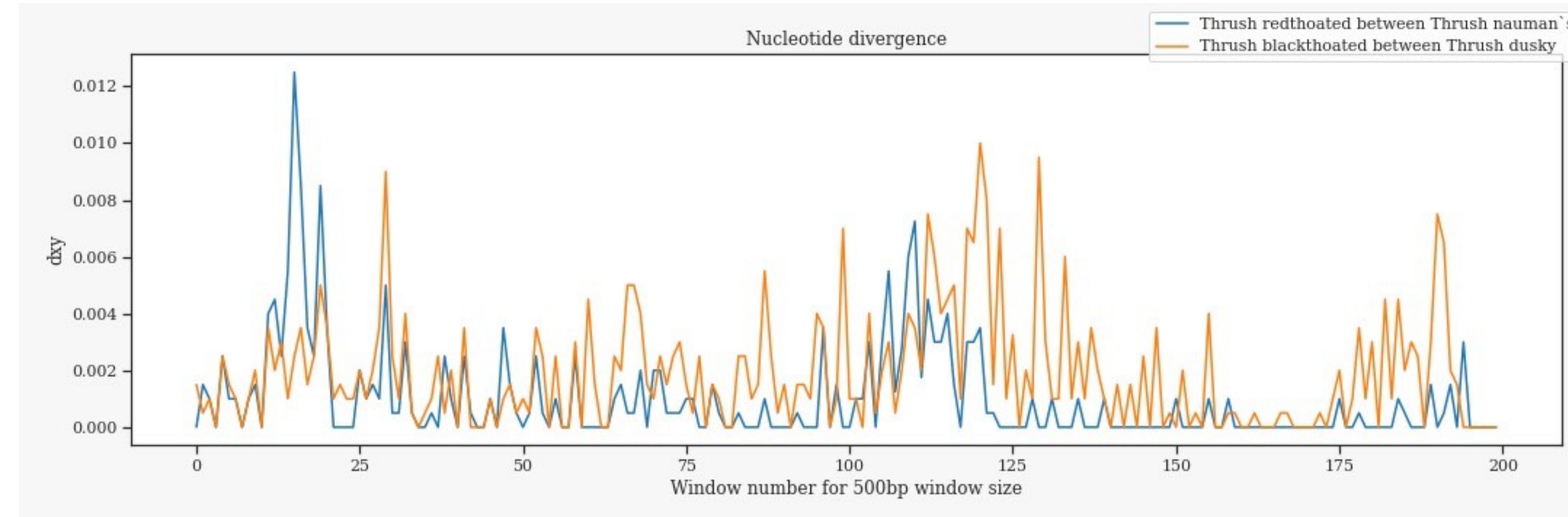
No hit: **128**



DIVERSITY



Grouped by Color

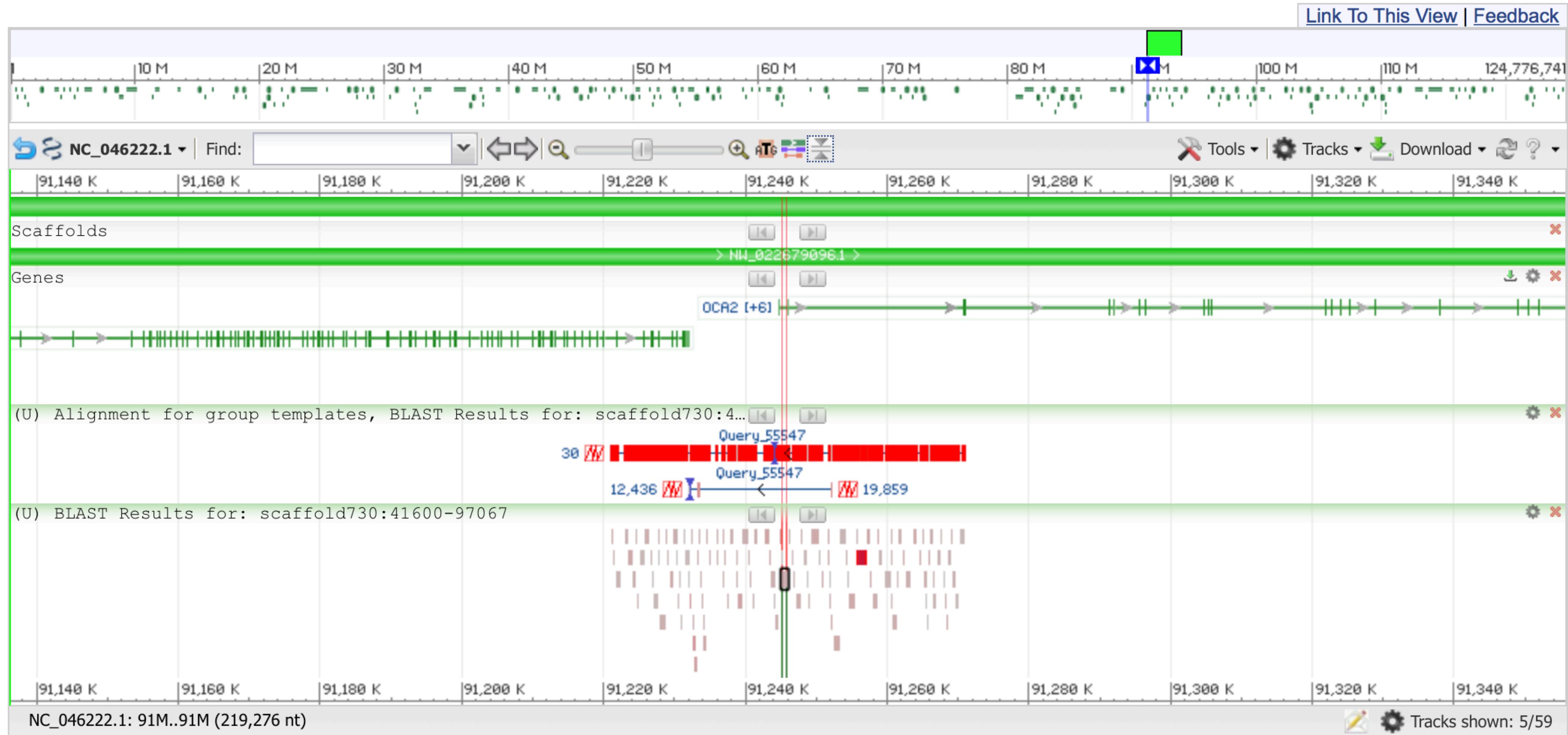


CANDIDATE GENES - BINGO!

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

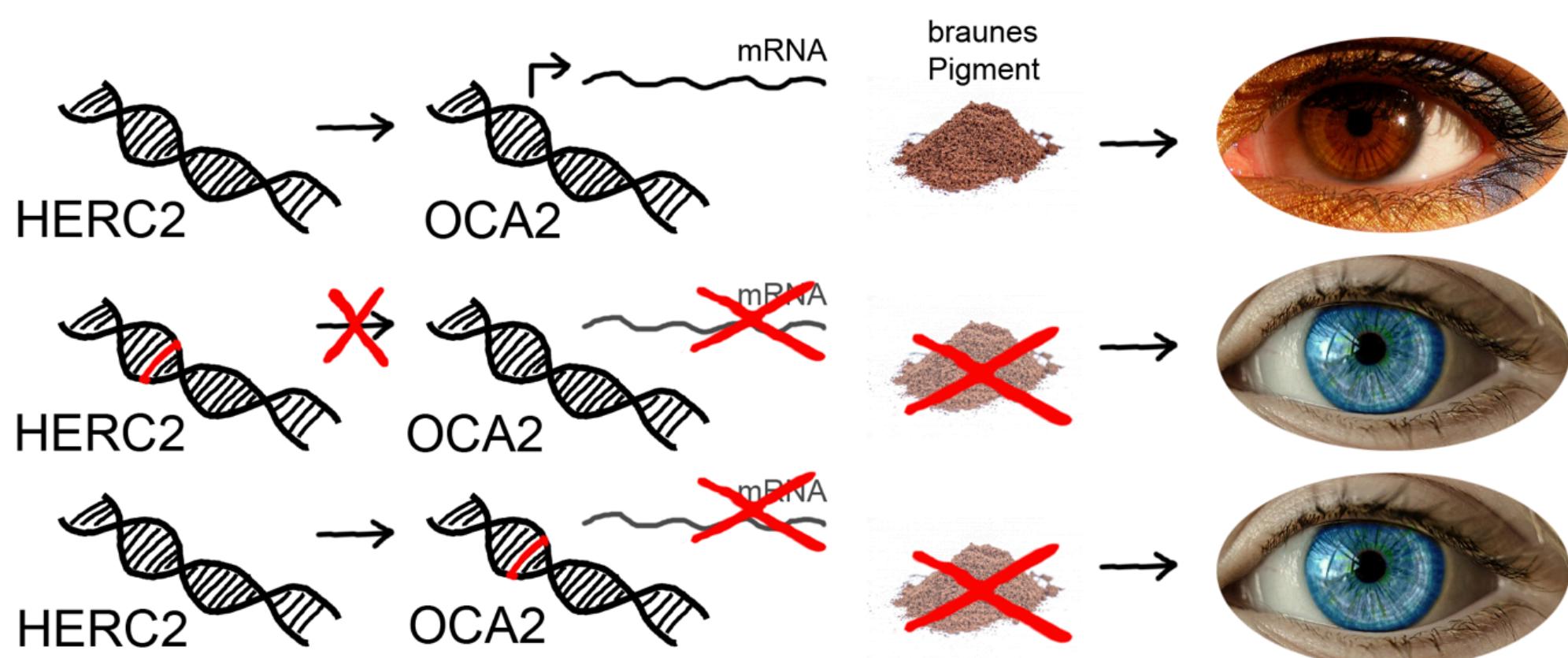
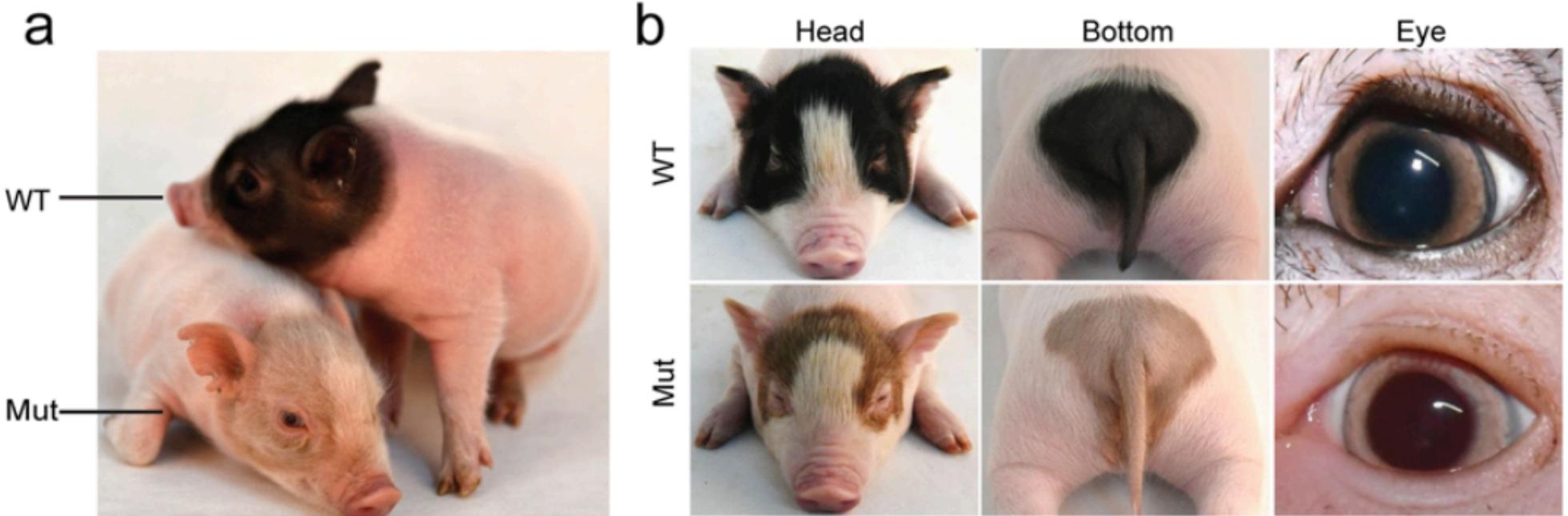
CANDIDATE GENES - WHAT'S NEXT?

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



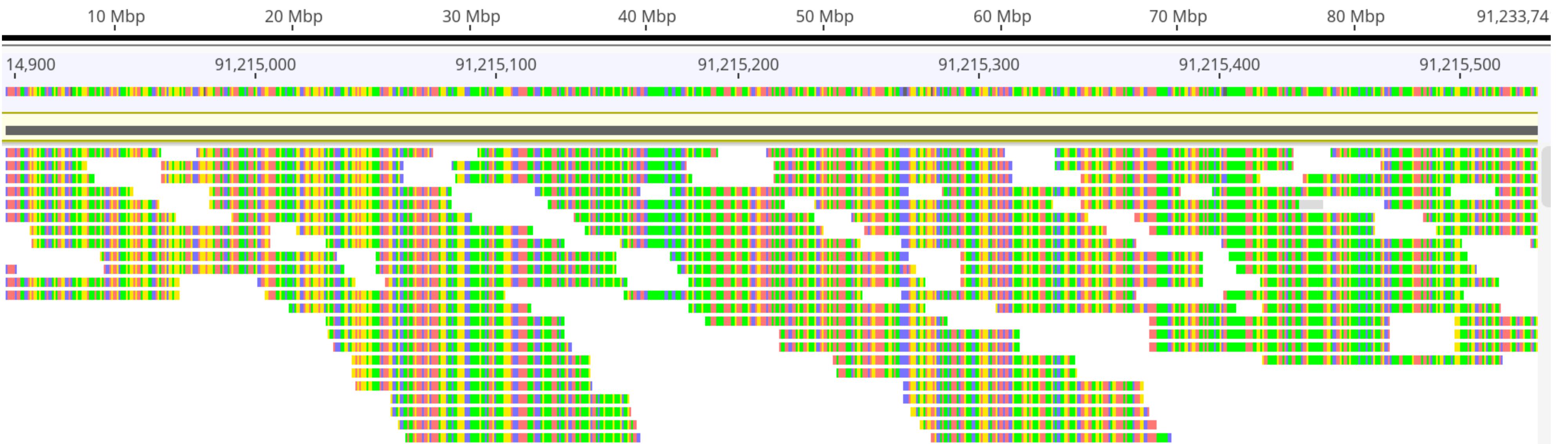
GENOMICS OF PARALLELISM IN BIRD COLORATION

OCA2 AND HERC2



WHAT'S NEXT?

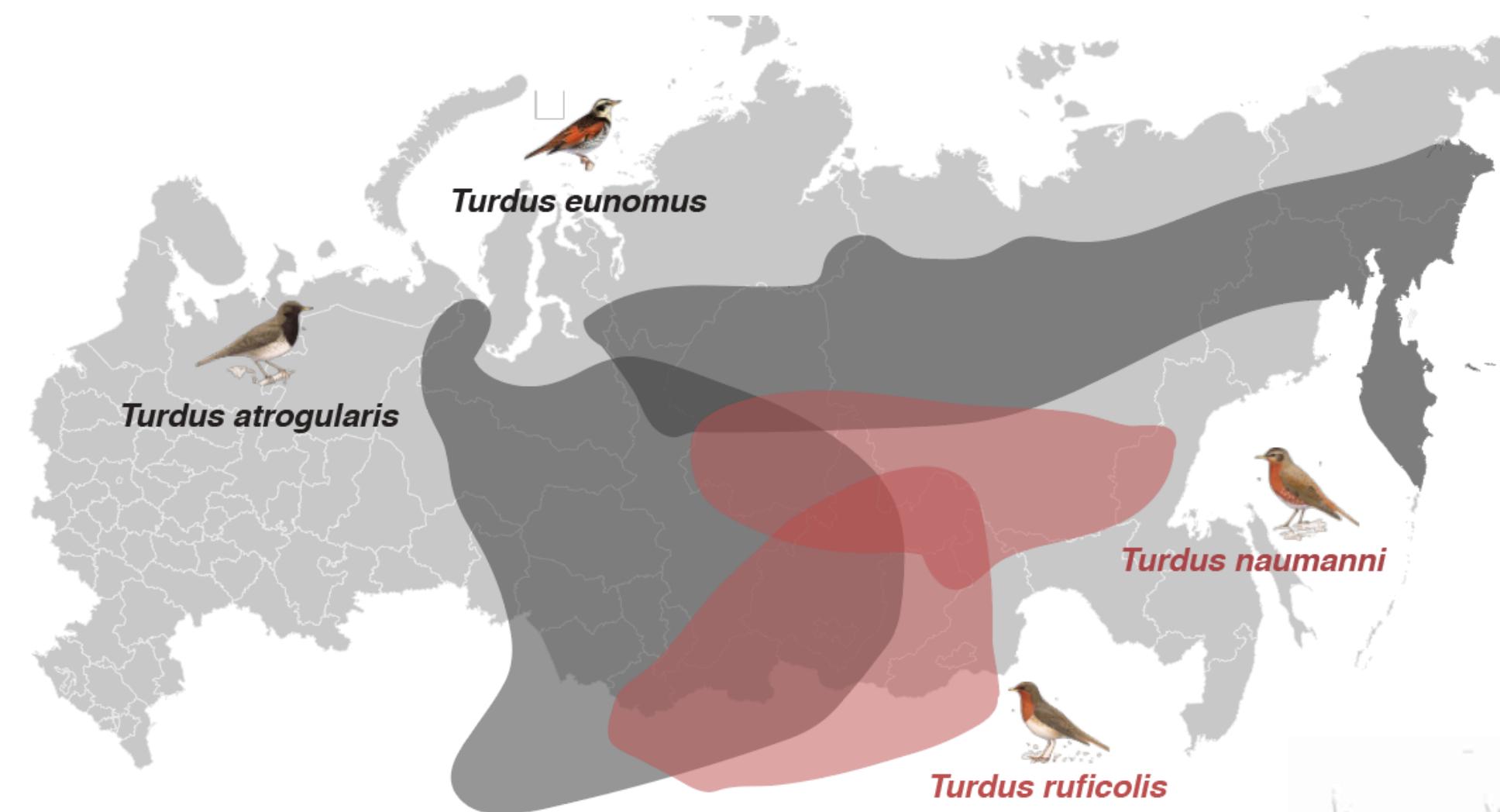
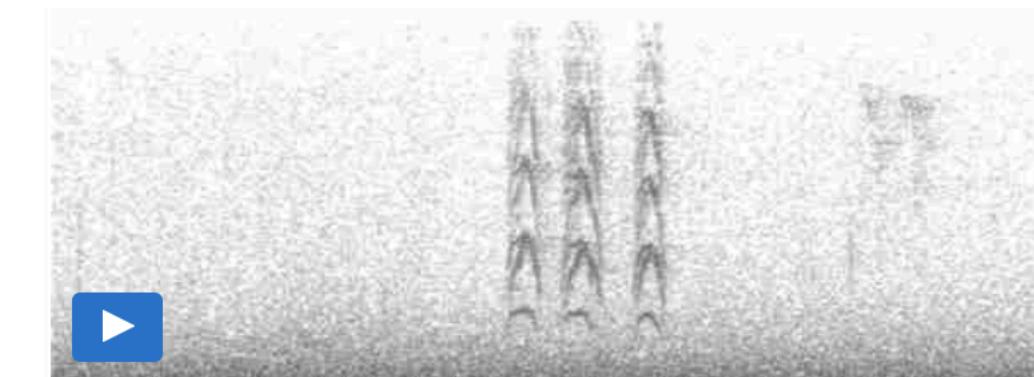
1. Constructing the new reference
2. Tests for selection
3. OCA2 evolution on a global scale



BEYOND COLORATION

Interaction in the sympatry zone

- Difference in coloration suggests reproductive isolation
- Ecology is different (latitudinal, habitat difference)
- SONG IS NOT FULLY DESCRIBED



THANKS

- Fedyo Kondrashov
- Natasha Kistkina
- Vladimir Arkhipov

