

atcgaatccggatccgcgatcgtcgttaccggtgacacagacgcggagcctcgttagcggcgtatgacag
acgtccgaacccggagcctcgtaccgttgcagtcacglatgacagacgtccggagcttagcgalcggatactcgatc
gatcctgmcnngtgcaccgttgongtcaccgatgacagacgiccggagcttagcgalcggatactcgatc
tgcctttcaccgintcccgintaccgaccgaccgatgacagacgiccggagcttagcgalcggatactcgatc
cgtatcttccggttagcttccgaccgttagcgalcggatactcgatcgtgacagacgiccggagcctcgttagc
acgtttccgttagcgttccgaccgttagcgalcggatactcgatcgtgacagacgiccggagcctcgttagc

THETA

Statistical Genetics Group
Institute of Animal Genetics



Group specific CNVs in scouts and recruits of honeybee

M. Mielczarek^{1,2}, S. L. Rodriguez-Zas³, J. Szyda^{1,2}, B. Southey³

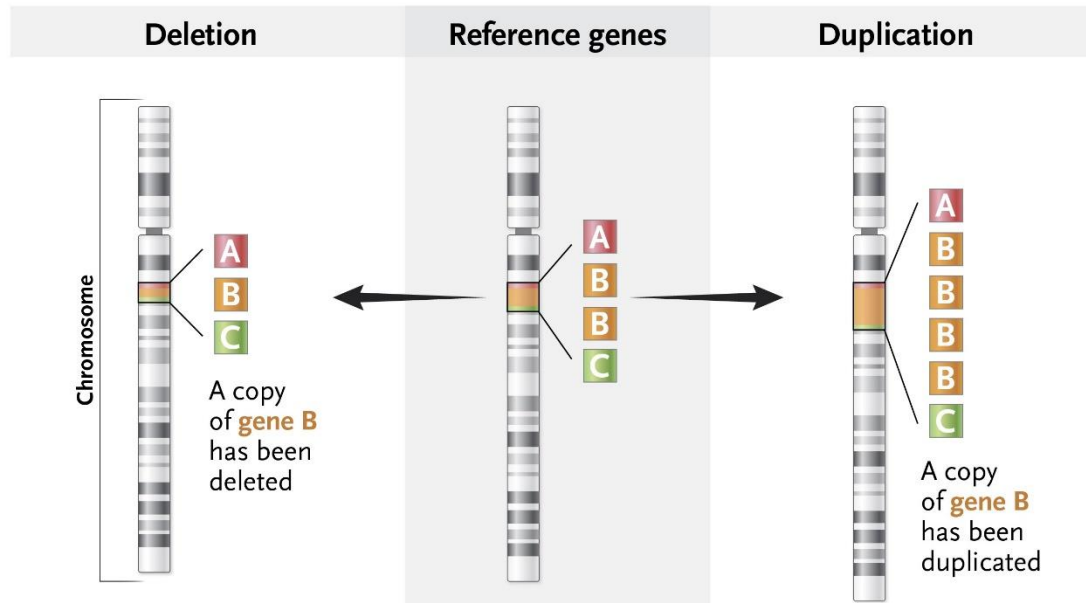
¹Wroclaw University of Environmental and Life Sciences; Kozuchowska 7, 51-631 Wroclaw, Poland

²National Research Institute of Animal Production, Krakowska 1, 32-083 Balice, Poland

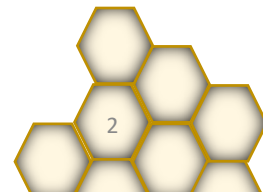
³University of Illinois at Urbana-Champaign; 1207 W. Gregory Dr. Urbana, IL 61801, the USA

Background & Objectives

- Genome-wide CNVs detection
- Determination of group specific CNVs



source: <https://illustrated-glossary.nejm.org>



Dataset



Queen

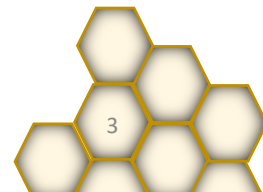


Drone

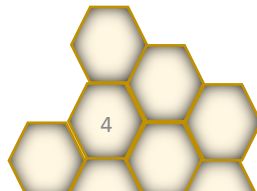
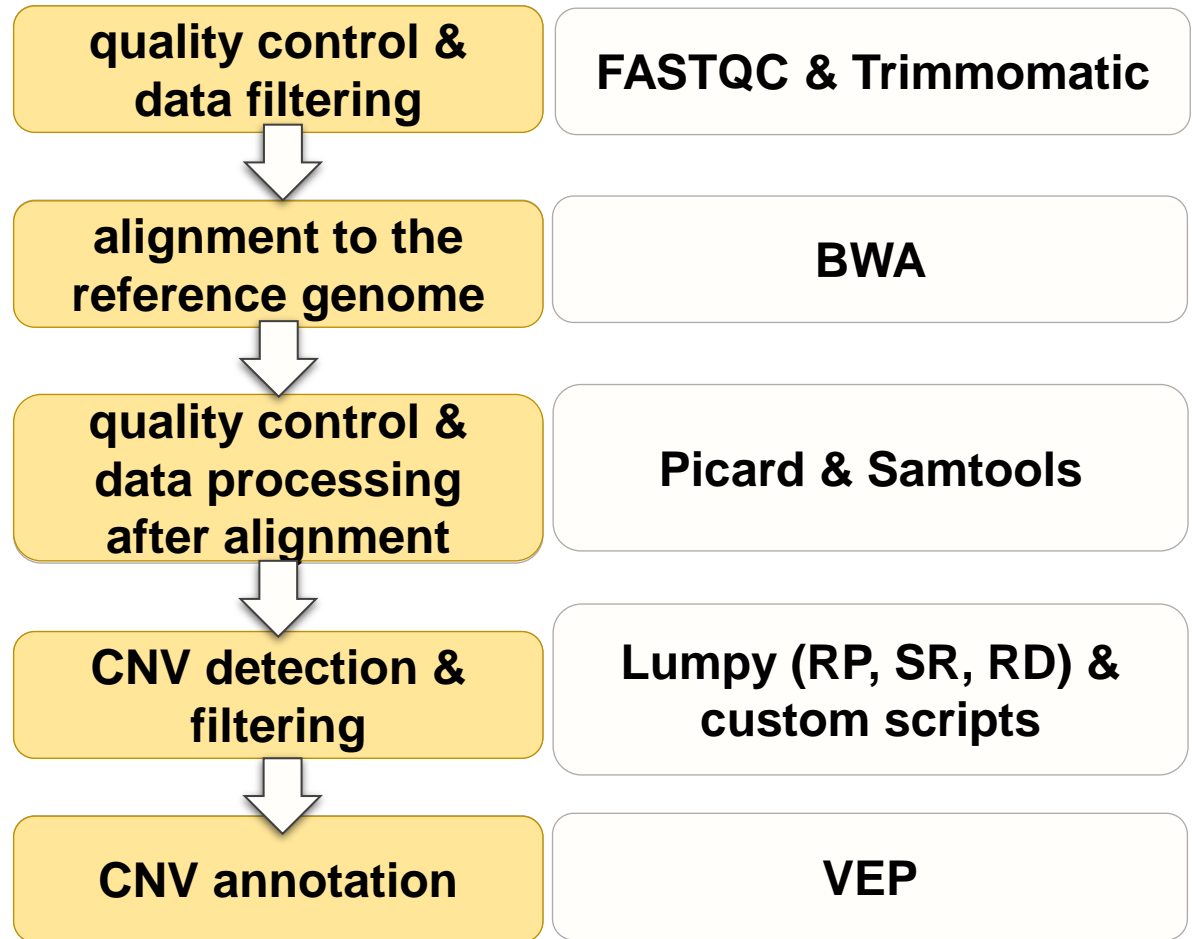


Worker

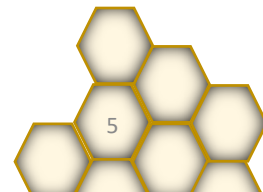
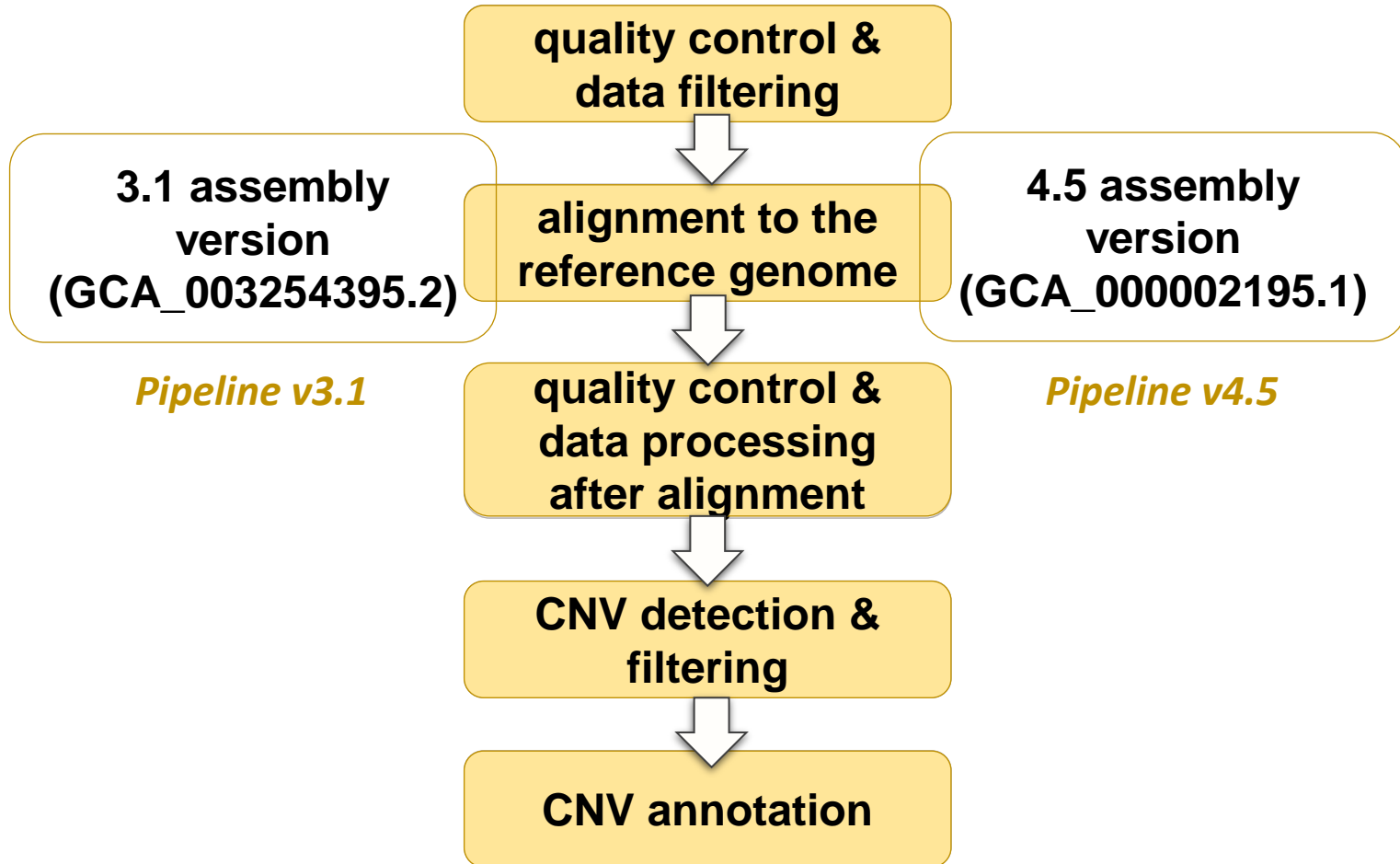
- Whole genome DNA sequences of *Apis mellifera* **foragers**:
 - 22 scouts (S)
 - 22 recruits (R)
- Scouts seek new resources and return to the colony, enlisting recruits to collect these resources
- Sequencing parameters:
 - Illumina HiSeq2000
 - PE 100 bp
 - ~ 10 x average coverage



Methods



Methods



Methods



**3.1 assembly
version
(GCA_003254395.2)**

Pipeline v3.1

**alignment to the
reference genome**

**4.5 assembly
version
(GCA_000002195.1)**

Pipeline v4.5

Global statistics

Total sequence length	225,250,884
Total ungapped length	223,937,270
Gaps between scaffolds	0
Number of scaffolds	177
Scaffold N50	13,619,445
Scaffold L50	7
Number of contigs	228
Contig N50	5,382,476
Contig L50	13
Total number of chromosomes and plasmids	17
Number of component sequences (WGS or clone)	228

Global statistics

Total sequence length	250,270,657
Total ungapped length	229,107,307
Gaps between scaffolds	324
Number of scaffolds	5,644
Scaffold N50	997,192
Scaffold L50	65
Number of contigs	16,501
Contig N50	45,688
Contig L50	1,390
Total number of chromosomes and plasmids	16
Number of component sequences (WGS or clone)	16,501

Results

CNVs number

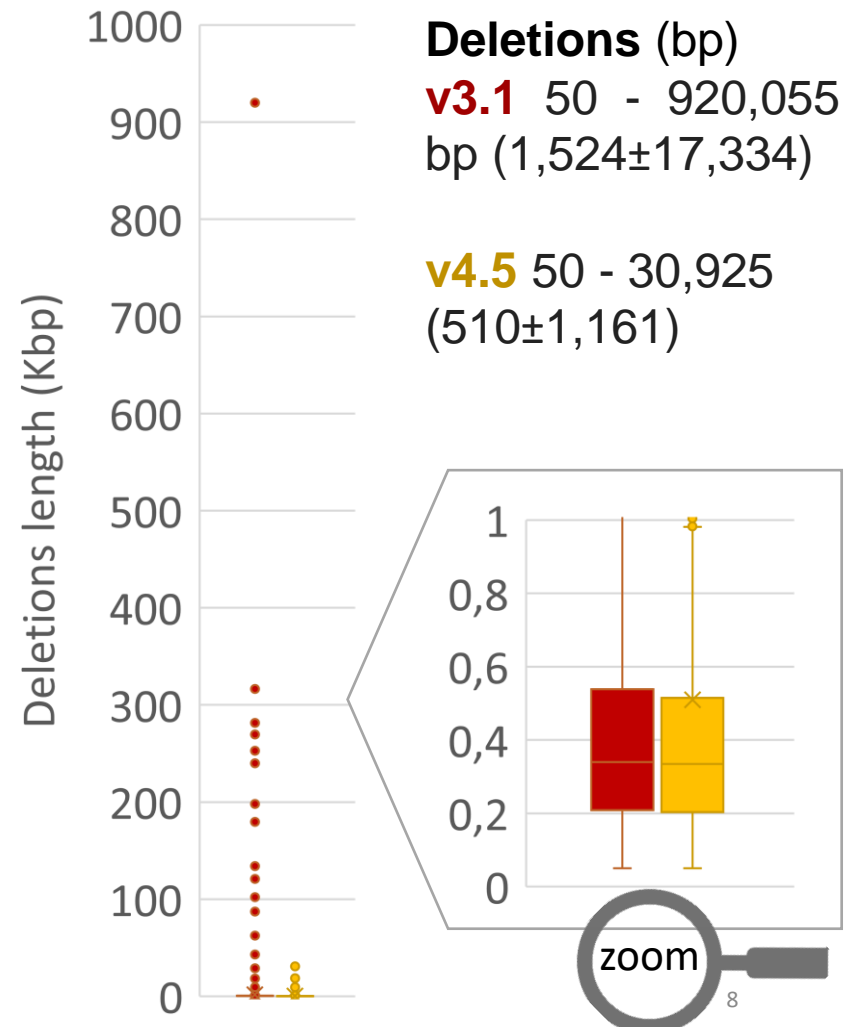
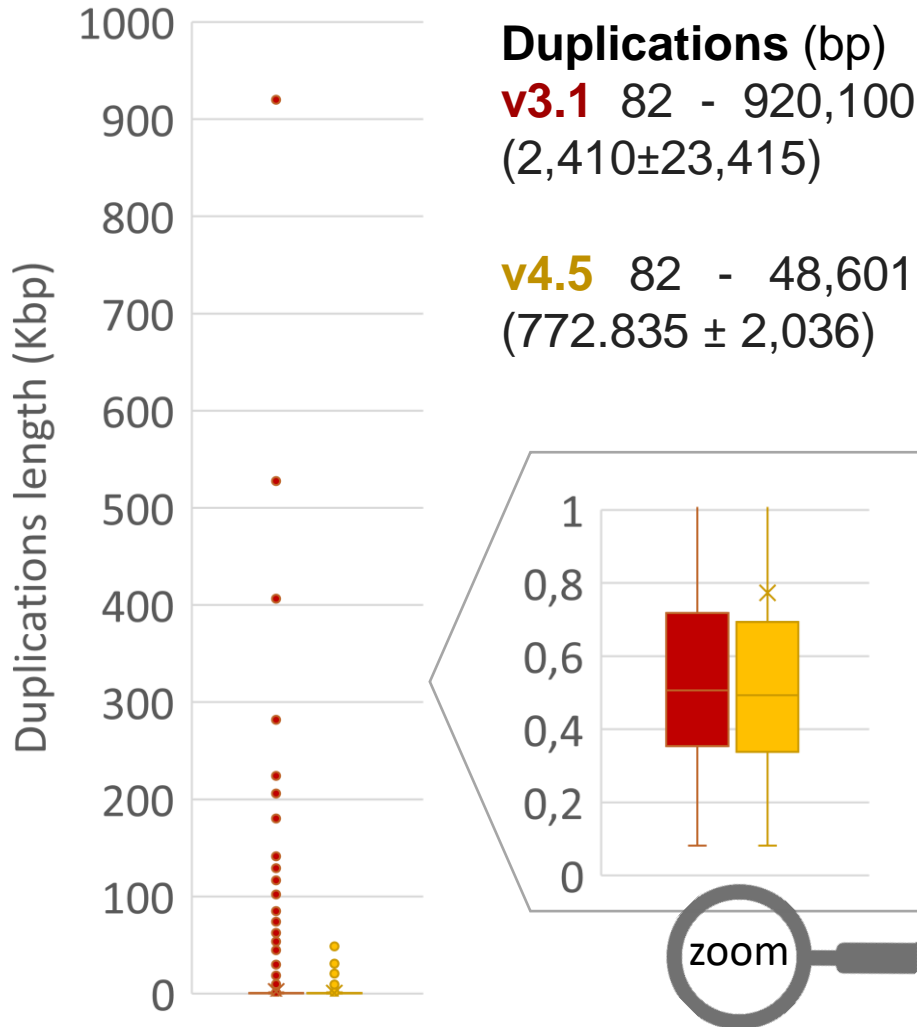
	v3.1	v4.5
# dup	3,225	3,093
# del	5,348	4,409



Results

CNVs length

■ v3.1 ■ v4.5

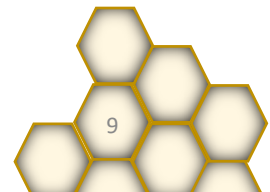


Results

Group specific CNVs

- No **scout**-specific CNVs
- **Recruit**-specific fragments deleted in both reference genome versions, duplication only in 3.1 assembly

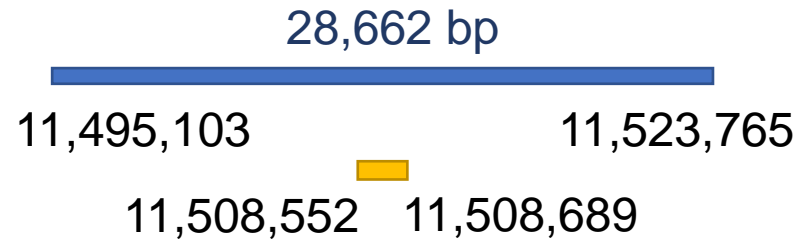
	v3.1	v4.5
dup	12:4684284-4684637 (353 bp)	-
del	5:9737471-9737608 (137 bp)	5:11508552-11508689 (137 bp)



Results

	v3.1	v4.5
dup	12:4684284-4684637 (353 bp)	-
del	5:9737471-9737608 (137 bp)	5:11508552-11508689 (137 bp)

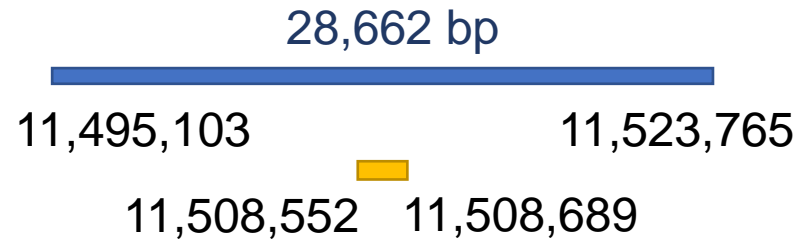
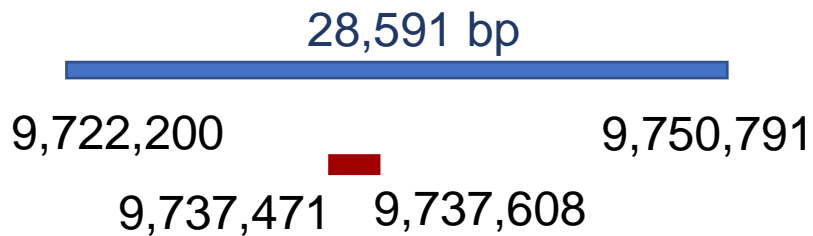
LOC408840 (5'-3'
exoribonuclease 2 homolog)



Annotation release	Status	Assembly	LG	Location
104	current	Amel_HAv3.1 (GCF_003254395.2)	LG5	NC_037642.1 (9722200..9750791)
103	previous assembly	Amel_4.5 (GCF_000002195.4)	LG5	NC_007074.3 (11495103..11523765, complement)

Results

	v3.1	v4.5
dup	12:4684284-4684637 (353 bp)	-
del	5:9737471-9737608 (137 bp)	5:11508552-11508689 (137 bp)



LOC408840 gene deletion in both genome assemblies.



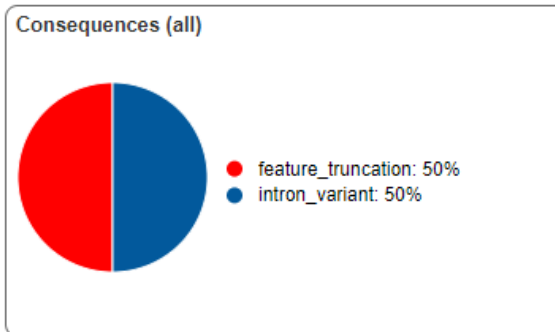
Annotation release	Status	Assembly	LG	Location
104	current	Amel_HAv3.1 (GCF_003254395.2)	LG5	NC_037642.1 (9722200..9750791)
103	previous assembly	Amel_4.5 (GCF_000002195.4)	LG5	NC_007074.3 (11495103..11523765, complement)

Results

LOC408840 (5'-3' exoribonuclease 2 homolog) function:

- 5'->3' exoribonuclease activity
- termination of transcription by RNA polymerase II

Category	Count
Variants processed	1
Variants filtered out	0
Novel / existing variants	-
Overlapped genes	1
Overlapped transcripts	1
Overlapped regulatory features	-



Results preview

Navigation (per variant) | Filters | Download | New job

Page: 1 of 1 | Show: 1 All variants | Uploaded variant is defined | Add | All: [VCF](#) [VEP](#) [TXT](#)

Show/hide columns (20 hidden)

Uploaded variant	Location	Allele	Consequence	Gene	Feature type	Feature	Biotype	Feature strand
5_11508552_deletion	5:11508551-11508689	deletion	intron_variant , feature_truncation	GB44445	Transcript	GB44445-RA.	protein_coding	-1

Page: 1 of 1 | Show: 1 All variants

A sequence variant that causes the reduction of a genomic feature, with regard to the reference sequence

Results

	v3.1	v4.5
dup	12:4684284-4684637 (353 bp)	-
del	5:9737471-9737608 (137 bp)	5:11508552-11508689 (137 bp)

- Intergenic

LOC113219153 | Exon Navigator: There are no genes in the region. | LOC413125

NC_037649.1 | Tools | Tracks | Download

200 | 4,684,250 | 4,684,300 | 4,684,350 | 4,684,400 | 4,684,450 | 4,684,500 | 4,684,550 | 4,684,600 | 4,684,650

Sequence

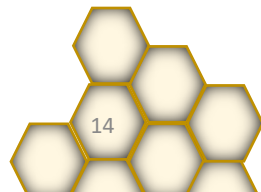
Scaffolds

> NH_020555783.1 >

- NCBI Genome Remapping Service (**v3.1** → **v4.5**):
 - 12:4684284 -4684637 → GroupUn5087:3954 -4179

Conclusions

- LOC408840 gene may play a role in behavioral differences between scouts and recruits
- Mapping of reads against different versions of the reference genomes provides validation of CNVs



Future work

Dataset:

- WGS of 100 honeybees (Illumina and Oxford Nanopore)
- Honeybees from different families

Goals:

- Describing the genomic landscape of honeybee genome
- Assessing inter-individual and inter-family variation.



Acknowledgements



POLISH NATIONAL AGENCY
FOR ACADEMIC EXCHANGE



Poznan Supercomputing and Networking Center
affiliated to the Institute of Bioorganic Chemistry, PAS

Thank you!

