ALMA MATER STUDIORUM
UNIVERSITÀ DI BOLOGNA

EAAP
Insect-IMP
Athens, Jan. 29-31, 2025

Population genomics from a food matrix: estimating *csd* variability using next generation sequencing on *Apis mellifera* DNA present in honey

Luca Fontanesi

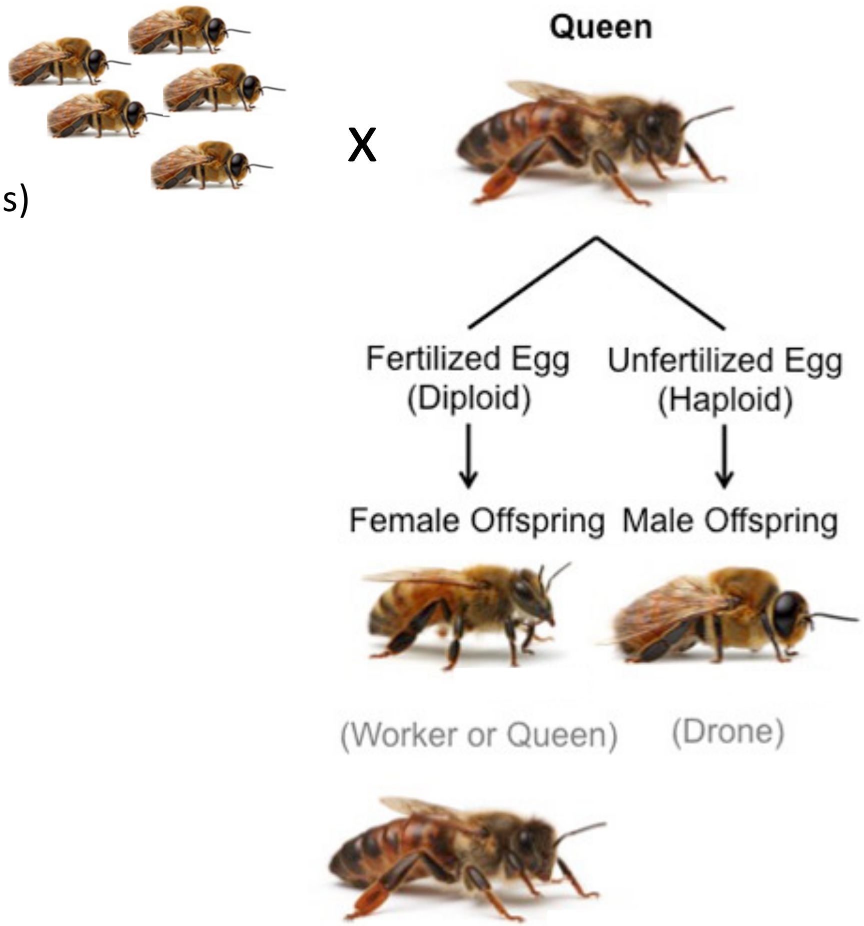
Animal and Food Genomics Group

Department of Agricultural and Food Sciences

University of Bologna, Bologna Italy

E-mail: luca.fontanesi@unibo.it

Apis mellifera sex determination: haplodiploidism



(Drones)

X

Queen

Fertilized Egg
(Diploid)

Unfertilized Egg
(Haploid)

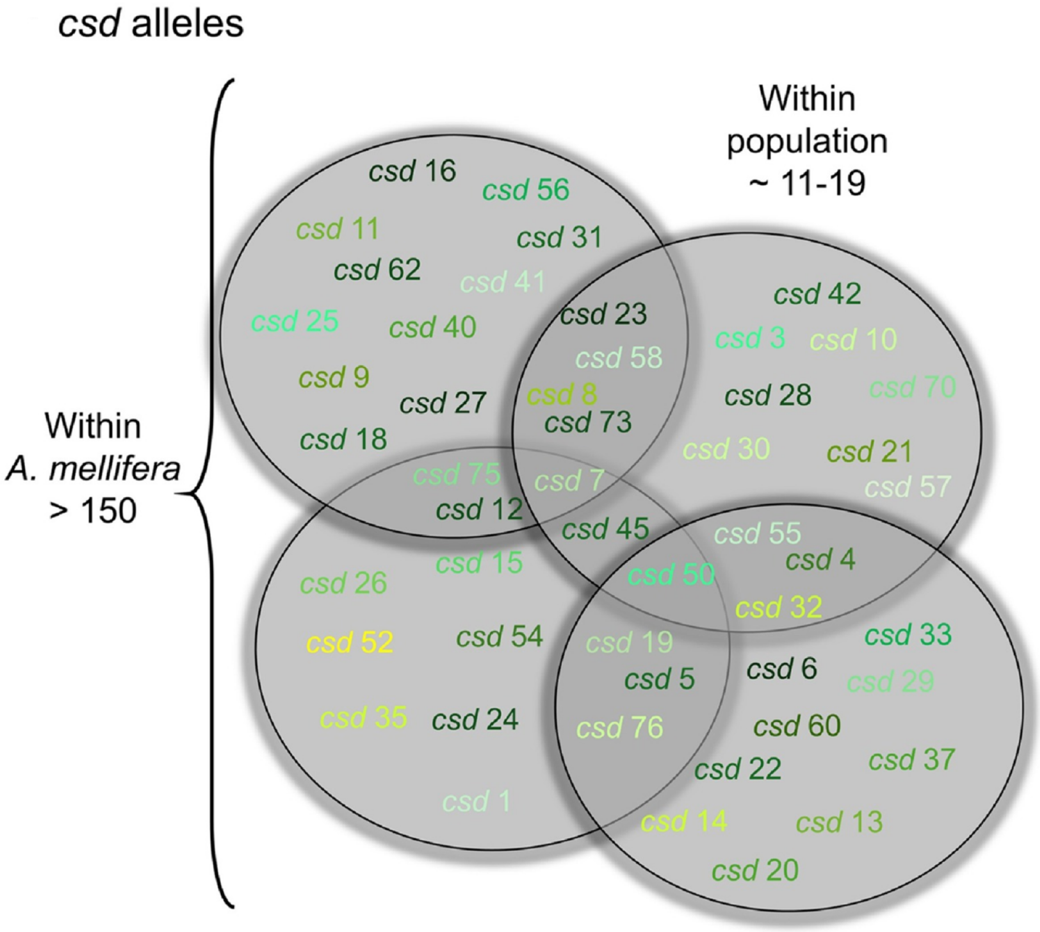
Female Offspring

Male Offspring

(Worker or Queen)

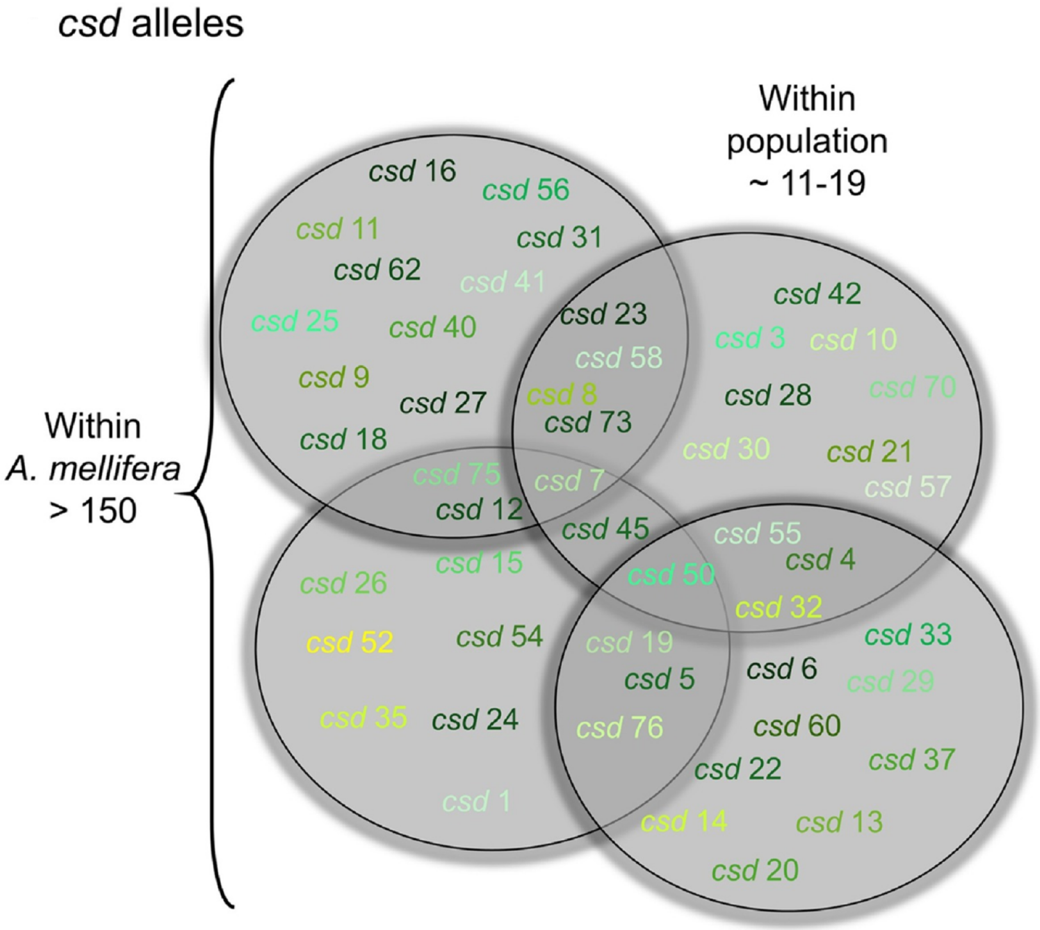
(Drone)

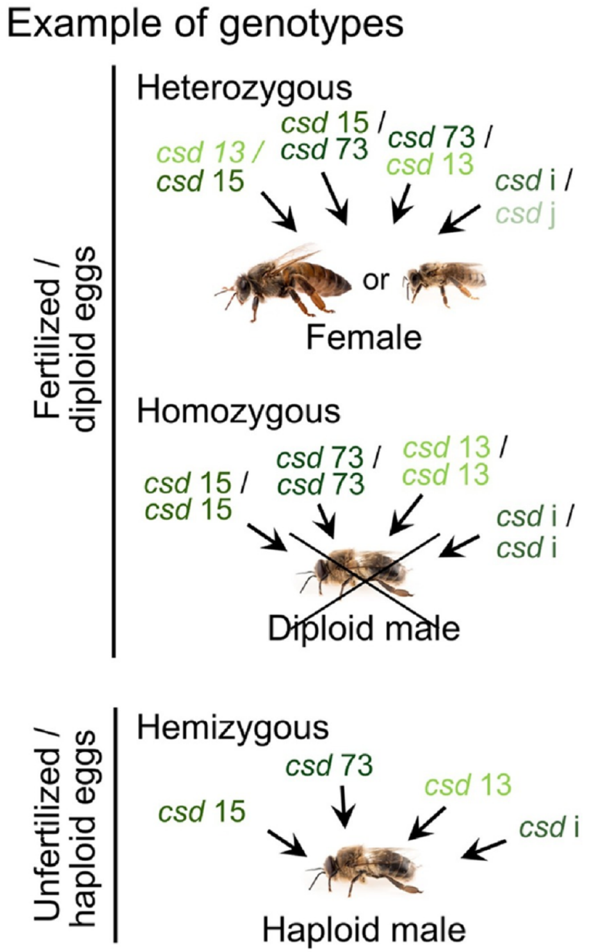
Apis mellifera complementary sex determiner (*csd*) gene



From Seiler & Beye (2024)

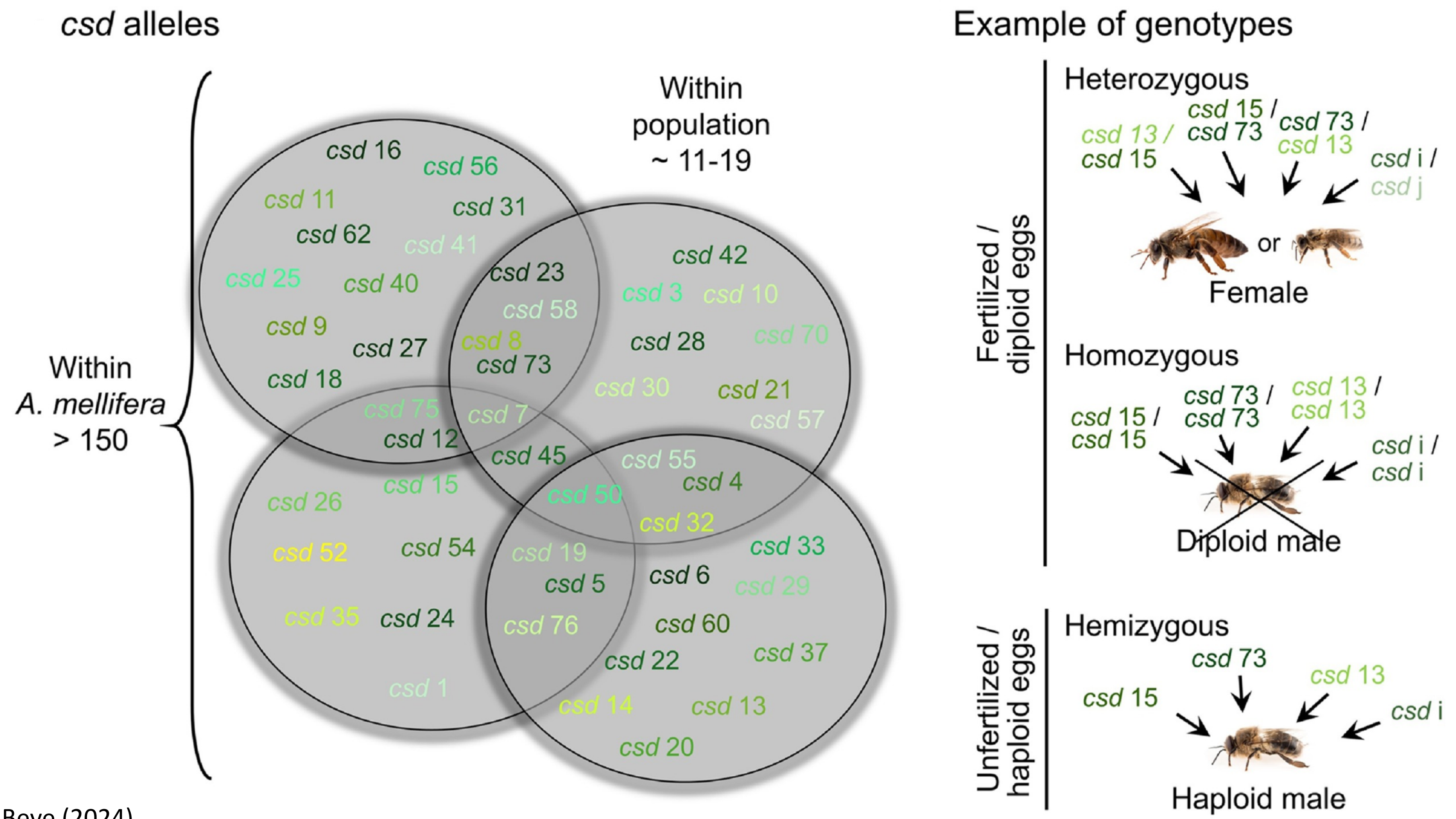
Apis mellifera complementary sex determiner (*csd*) gene

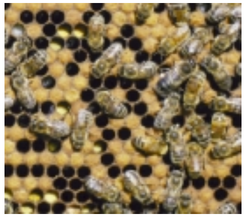




From Seiler & Beye (2024)

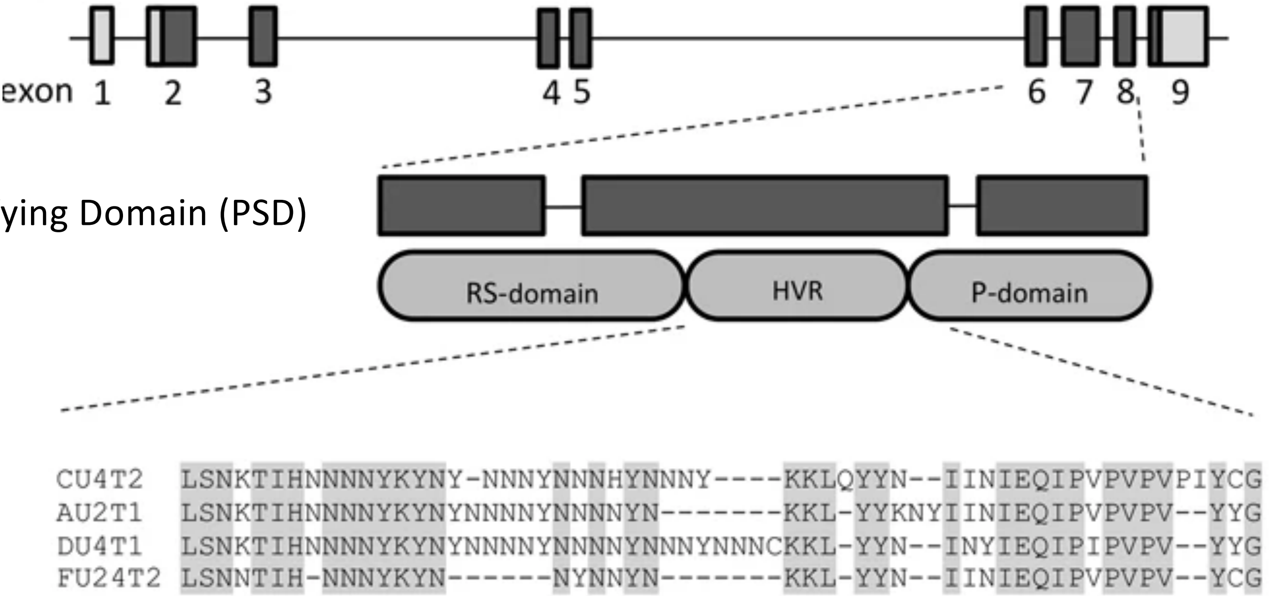
Apis mellifera complementary sex determiner (*csd*) gene





From Seiler & Beye (2024)

Apis mellifera complementary sex determiner (*csd*) gene

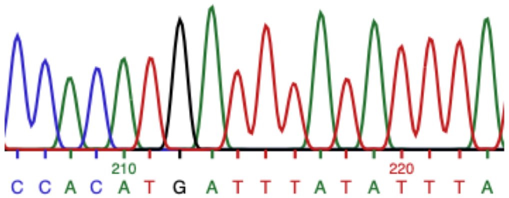


Potential Specifying Domain (PSD)

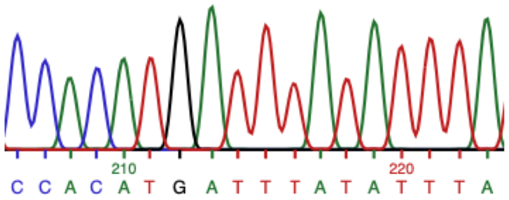
```
CU4T2  LSNKTIHNNNNYKYNY-NNNYNNNHYNNNY----KKLQYYN--IINIEQIPVPVPVPIYCG
AU2T1  LSNKTIHNNNNYKYNYNNNNYNNNNYN-------KKL-YYKNYIINIEQIPVPVPV--YYG
DU4T1  LSNKTIHNNNNYKYNYNNNNYNNNNYNNNYNNNCKKL-YYN--INYIEQIPIPVPV--YYG
FU24T2 LSNNTIH-NNNYKYN------NYNNYN-------KKL-YYN--IINIEQIPVPVPV--YCG
```

Zareba et al. (2017)

(Drones)



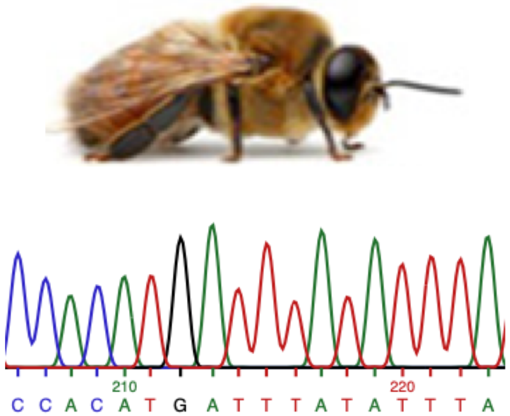
(Drones)

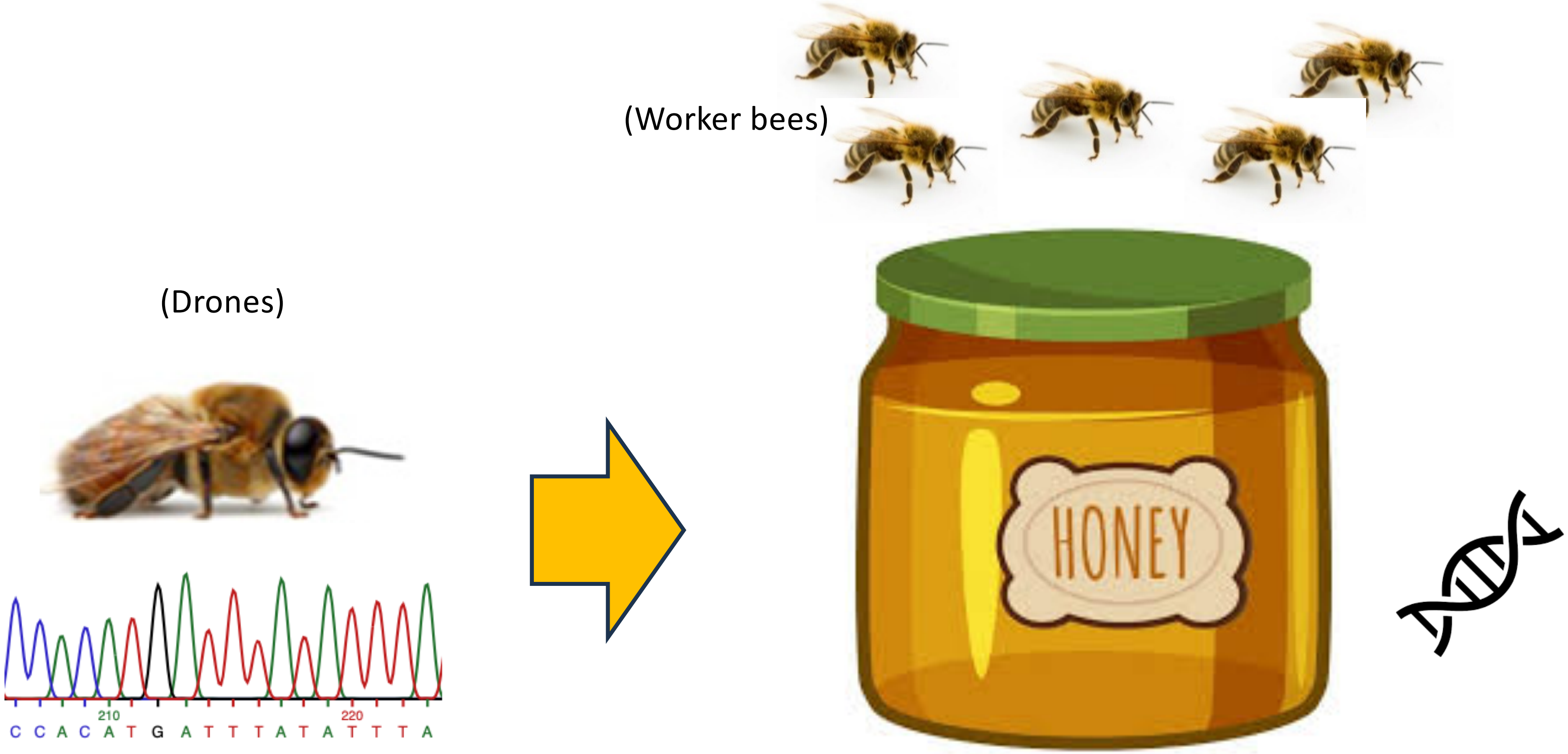


?

We aimed to transform a food matrix into a population genomic tool

(Drones)

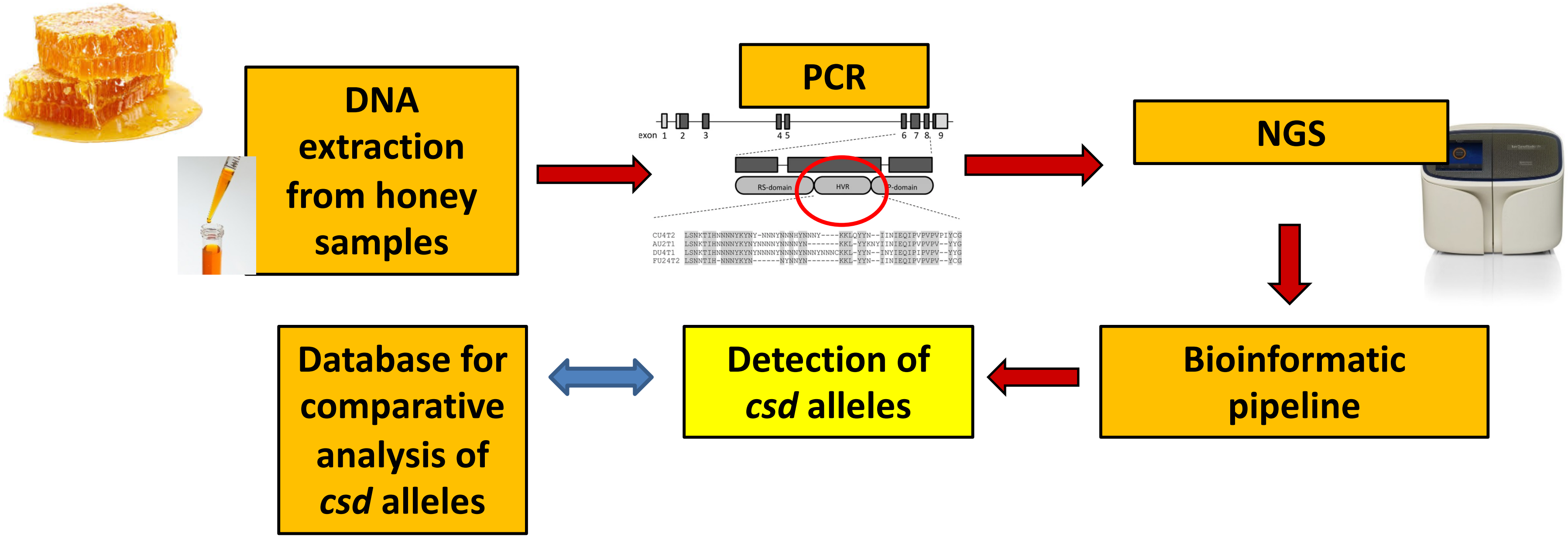




(Worker bees)

(Drones)

Analysis of *Apis mellifera complementary sex determiner* (*csd*) gene from honey DNA



DNA extraction from honey samples

PCR

NGS

Bioinformatic pipeline

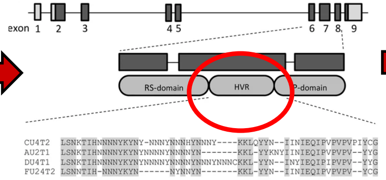
Detection of *csd* alleles

Database for comparative analysis of *csd* alleles

Analysis of *Apis mellifera complementary sex determiner* (*csd*) gene from honey DNA

DNA extraction from honey samples

PCR



NGS

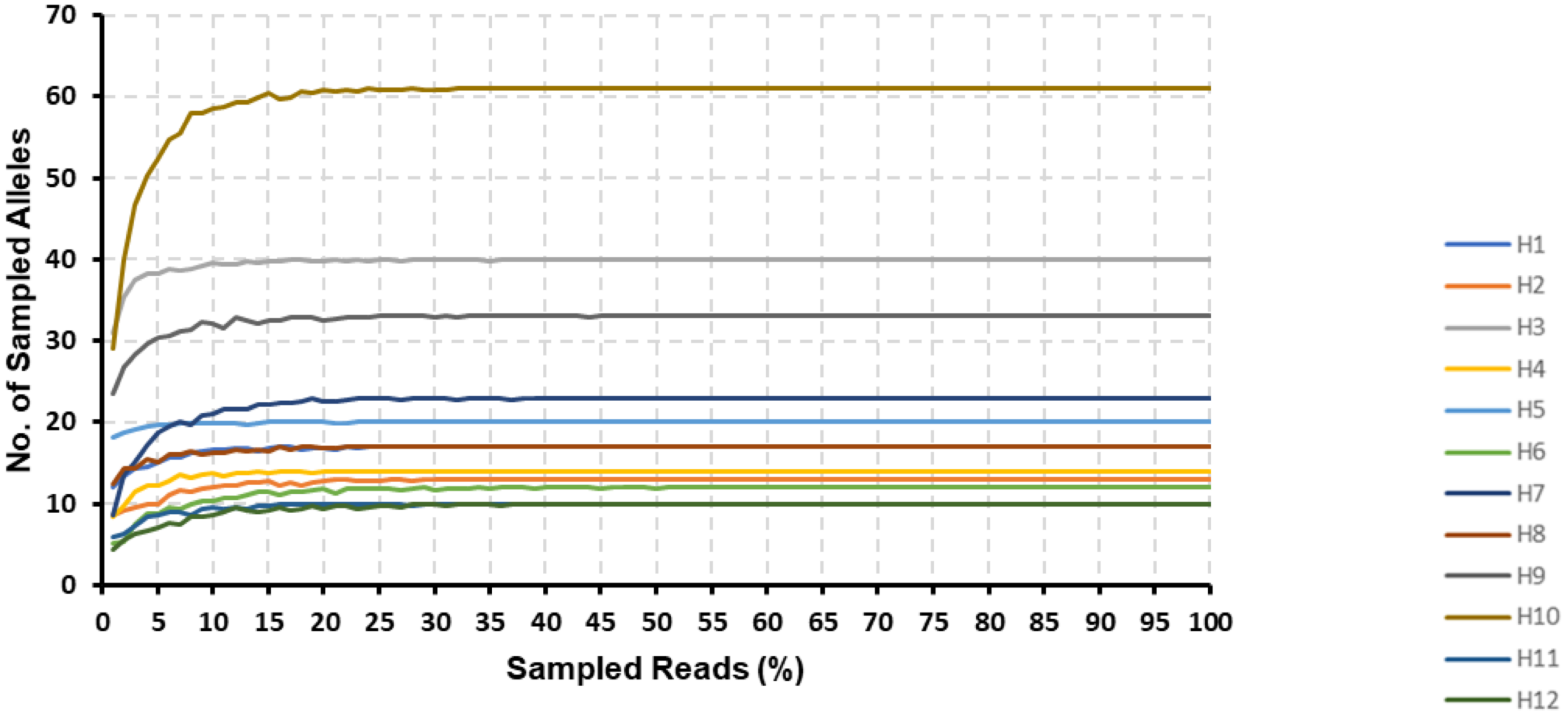
Bioinformatic pipeline

Detection of *csd* alleles

Database for comparative analysis of *csd* alleles

- Filtering low quality of reads and clustering
- Translation: protein residues (motif) upstream (KIIS) and downstream (IEQIP) – Protein motives in UniprotKB – absence of any stop codon
- Number of reads: Median minimum relative abundance of the alleles detected in more than one sample – alleles detected in at least two samples
- Rarefaction curve

Rarefaction curves obtained for the honey samples



ALMA MATER STUDIORUM
UNIVERSITÀ DI BOLOGNA

Results (1)

Sample ID [1]	Province	No. of Reads	No. of Retained Reads (%) [2]
H1	Bologna	173,064	149,678 (86.4)
H2 °	Reggio Emilia	75,454	72,027 (95.4)
H3	Piacenza	414,002	269,785 (64.8)
H4 °	Reggio Emilia	84,307	82,056 (97.3)
H5 °	Reggio Emilia	397,333	296,982 (74.7)
H6	Ferrara	34,350	33,503 (97.5)
H7	Piacenza	99,979	48,369 (48.3)
H8	Bologna	41,741	37,431 (89.7)
H9 *	Piacenza	97,735	73,994 (75.7)
H10	Rimini	70,307	53,728 (75.1)
H11 *	Piacenza	35,183	29,865 (84.9)
H12	Piacenza	1341	1220 (91.0)

Overall: 75.3%

Results (1)

Sample ID [1]	Province	No. of Reads	No. of Retained Reads (%) [2]	No. of csd Protein Alleles [3]	No. of Private csd Alleles (%) [4]	No. of New csd Alleles (Private) [5]
H1	Bologna	173,064	149,678 (86.4)	17	8 (47.1)	9 (7)
H2 °	Reggio Emilia	75,454	72,027 (95.4)	13	4 (30.8)	5 (3)
H3	Piacenza	414,002	269,785 (64.8)	40	16 (40.0)	13 (11)
H4 °	Reggio Emilia	84,307	82,056 (97.3)	14	2 (14.3)	5 (0)
H5 °	Reggio Emilia	397,333	296,982 (74.7)	20	7 (35.0)	5 (3)
H6	Ferrara	34,350	33,503 (97.5)	13	3 (23.1)	7 (3)
H7	Piacenza	99,979	48,369 (48.3)	23	11 (47.8)	12 (11)
H8	Bologna	41,741	37,431 (89.7)	17	4 (23.5)	3 (1)
H9 *	Piacenza	97,735	73,994 (75.7)	33	13 (39.4)	16 (11)
H10	Rimini	70,307	53,728 (75.1)	61	41 (67.2)	40 (37)
H11 *	Piacenza	35,183	29,865 (84.9)	10	2 (20.0)	3 (1)
H12	Piacenza	1341	1220 (91.0)	10	3 (30.0)	0 (0)

Overall: 75.3%

No correlation between the number of reads and the number of *csd* alleles

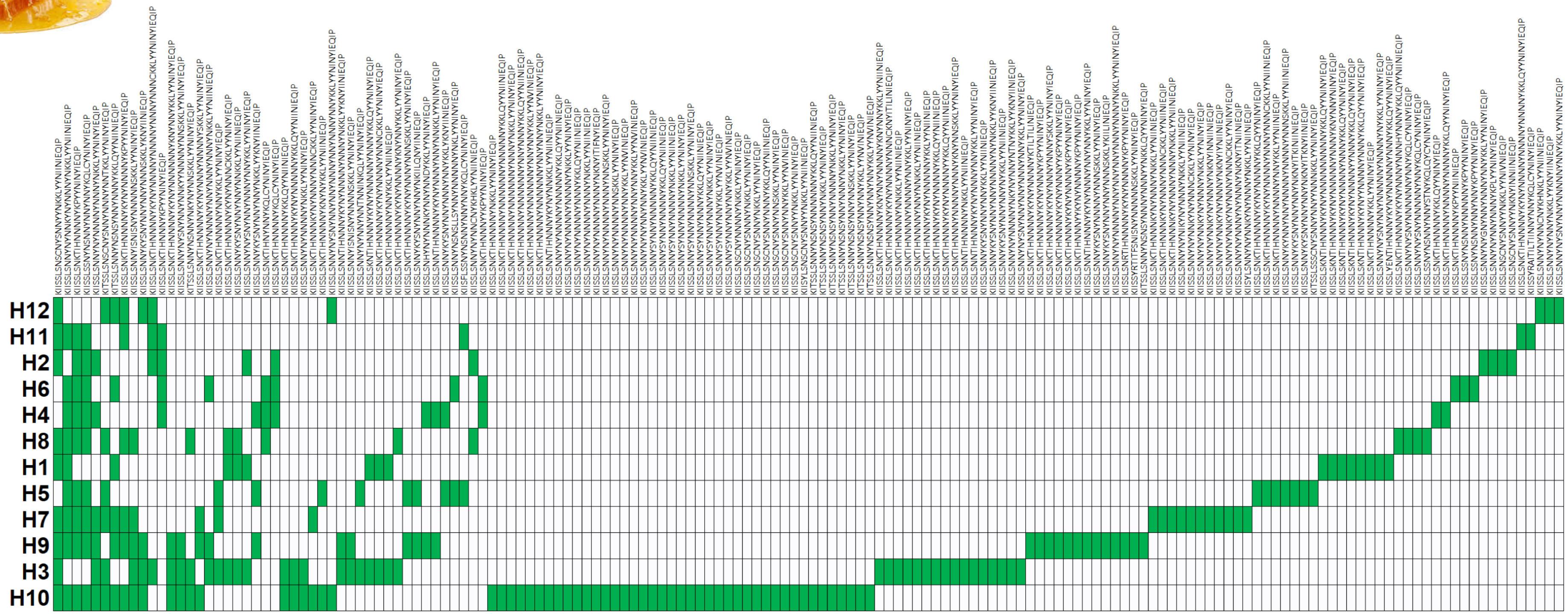
Results (1)

Sample ID [1]	Province	No. of Reads	No. of Retained Reads (%) [2]	No. of csd Protein Alleles [3]	No. of Private csd Alleles (%) [4]	No. of New csd Alleles (Private) [5]
H1	Bologna	173,064	149,678 (86.4)	17	8 (47.1)	9 (7)
H2 °	Reggio Emilia	75,454	72,027 (95.4)	13	4 (30.8)	5 (3)
H3	Piacenza	414,002	269,785 (64.8)	40	16 (40.0)	13 (11)
H4 °	Reggio Emilia	84,307	82,056 (97.3)	14	2 (14.3)	5 (0)
H5 °	Reggio Emilia	397,333	296,982 (74.7)	20	7 (35.0)	5 (3)
H6	Ferrara	34,350	33,503 (97.5)	13	3 (23.1)	7 (3)
H7	Piacenza	99,979	48,369 (48.3)	23	11 (47.8)	12 (11)
H8	Bologna	41,741	37,431 (89.7)	17	4 (23.5)	3 (1)
H9 *	Piacenza	97,735	73,994 (75.7)	33	13 (39.4)	16 (11)
H10	Rimini	70,307	53,728 (75.1)	61	41 (67.2)	40 (37)
H11 *	Piacenza	35,183	29,865 (84.9)	10	2 (20.0)	3 (1)
H12	Piacenza	1341	1220 (91.0)	10	3 (30.0)	0 (0)

Overall: 75.3%

Results (1)

Distribution of the 160 *csd* alleles across the honey samples



Results (2)

Most abundant *csd* alleles (>5%) identified in the analyzed honey samples

Sample ID [1]	***csd* Protein Allele Sequence** [2]	**Abundance %**	**Novel** [3]
H1	KIISSLSKNTIHNNNNYKYNYNNNNNYNNNYKKLQYYNINYIEQIP	53.4	
H1	KIISSLSNKTIHNNNNNYKKLYYNINYIEQIP	9.7	
H1	KIISSLSNNYNYSNYNNYNNNYNNYKKLYYNINYIEQIP ‡	8.9	
H2 °	KIISSLSNSCNYSNNYYNKKLYYNIINIEQIP †	53.7	
H2 °	KIISSLSNNYNYSNYNNYNNNYNNYKKLYYNINYIEQIP ‡	30.2	
H2 °	KIISSLSNKTIHNNNNNYKPYYNINYIEQIP **	7.7	
H3	KIISSLSNNYKYSNYNNYNNYNNKKLYYNIINIEQIP	40.0	Yes
H3	KIISSLSNKTIHNNNNNYNNYKKLYYNIINIEQIP	12.4	Yes
H3	KIISSLSNKTIHNNNNYKYNYNNNNNYKKLQYYNIINIEQIP	6.3	
H4 °	KIISSLSNKTIHNNNNNYKPYYNINYIEQIP **	83.6	
H5 °	KIISSLSSNYNSNNYNNYNNYKQLCYNINYIEQIP @	39.6	
H5 °	KIISSLSNNYKYSNYNNYNNYNKKLYYKNYIINIEQIP	12.0	
H5 °	KIISSLSNNYNYNNKYNYNNNYNKKLYYNIINIEQIP §	8.5	
H6	KIISSLSNKTIHNNNNNYKPYYNINYIEQIP **	93.1	
H7	KIISSLSNNYNYNNKYNYNNNYNKKLYYNIINIEQIP §	38.2	
H7	KIISSLSNKTIHNNNKYNYNNNYNNNCKKLYYNINYIEQIP	8.9	Yes
H8	KIISSLSNKTIHNNNNNYKPYYNINYIEQIP **	33.3	
H8	KIISSLSSNYNSNNYNNYNNYKQLCYNINYIEQIP @	25.9	
H8	KITSSLSNNYNSNNYNKYNYNNSKKLYYNINYIEQIP	13.2	
H8	KIISSLSNKTIHNNNNNYKYNYNNNNYKNYNNYKKLYYNINYIEQIP	5.8	
H9 *	KIISSLSNKTIHNNNNNYKYNYNNNNYKPYYNINYIEQIP	45.0	
H9 *	KIISSLSNKTIHNNNNNYKYNYNNNYNNNNYSKKLYYNINYIEQIP	10.3	Yes
H9 *	KIISSLSNNYISNISNYNNNNNSKKLYYNINYIEQIP	5.2	
H10	KIISSLSNSCNYSNNYYNKKLYYNIINIEQIP †	17.6	
H11 *	KIISSLSSNYNSNNYNNYNNYKQLCYNINYIEQIP @	43.8	
H11 *	KIISSLSNKTIHNNNNNYKPYYNINYIEQIP **	32.5	
H11 *	KIISSLSNNYNYNNKYNYNNNYNKKLYYNIINIEQIP §	6.7	
H12	KITSSLSNNYNSNSYNNYNNNYKKLQYYNIINIEQIP	37.5	
H12	KIISSLSNNYNYSNYNNYNNYNNNYNNYNNNYNNYKKLYYNINYIEQIP	31.5	
H12	KIISSLSNKTIHNNNNNYKYNYNNNNYNNNNYNNNYNNNCKKLYYNINYIEQIP	5.4	
H12	KIISSLSNNYKYSNYNNYNNYNNNSKKLYKNYIINIEQIP	5.3	
H12	KIISSLSNSCNYSNNYYNKKLYYNIINIEQIP †	5.1	

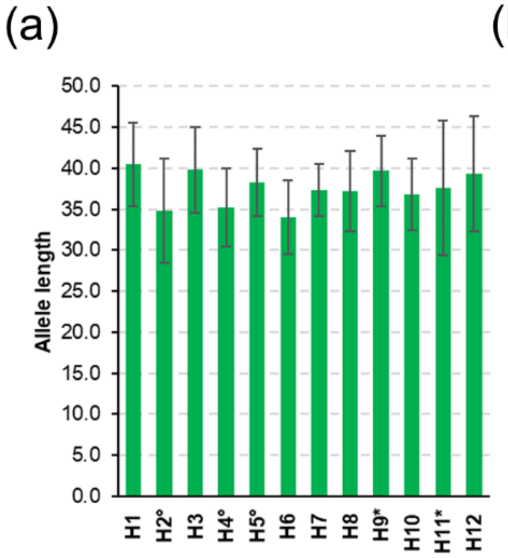
[1] Samples provided by the same beekeeper are marked with the same symbol (* or °).

[2] The same alleles identified in different samples are marked with the same symbol.

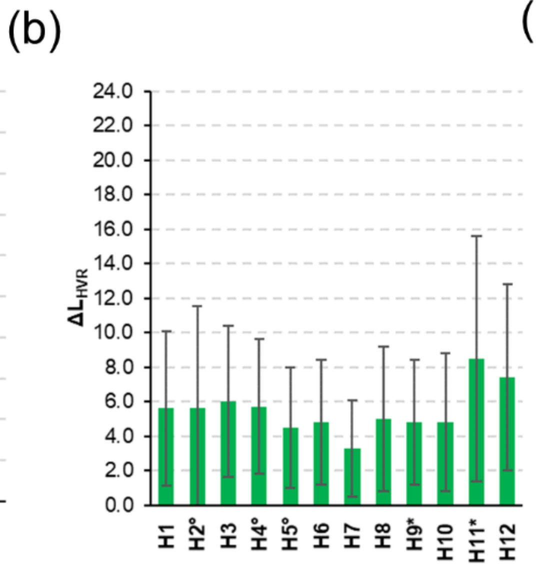
[3] Alleles that were not present in NCBI Database.

Results (3)

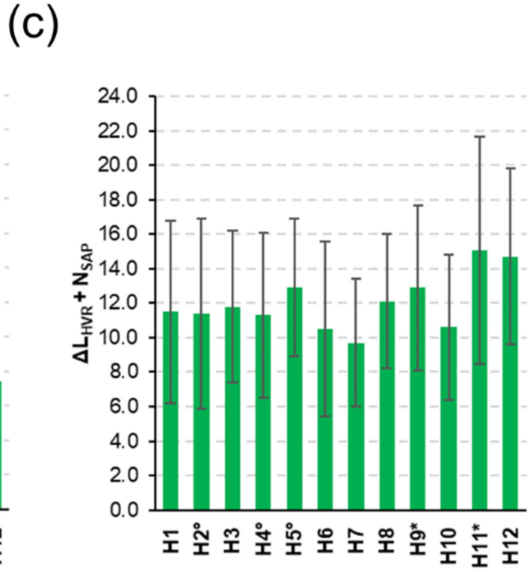
Study of the within sample HVR *csd* variability



Allele length (mean ± s.d.)



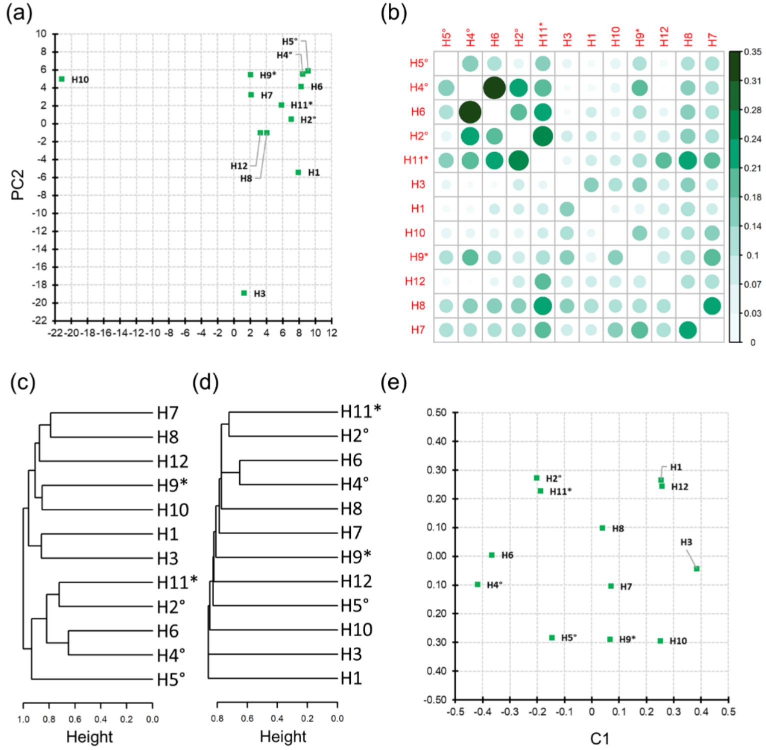
ΔL_{HVR} (mean ± s.d.)



Total differences
($\Delta L_{HVR} + N_{SAP}$; mean ± s.d.)

Results (4)

Analyses of honey sample similarity



Logistic
PCA

Jaccard similarity coefficient

Single-linkage
clustering based on
Jaccard distance

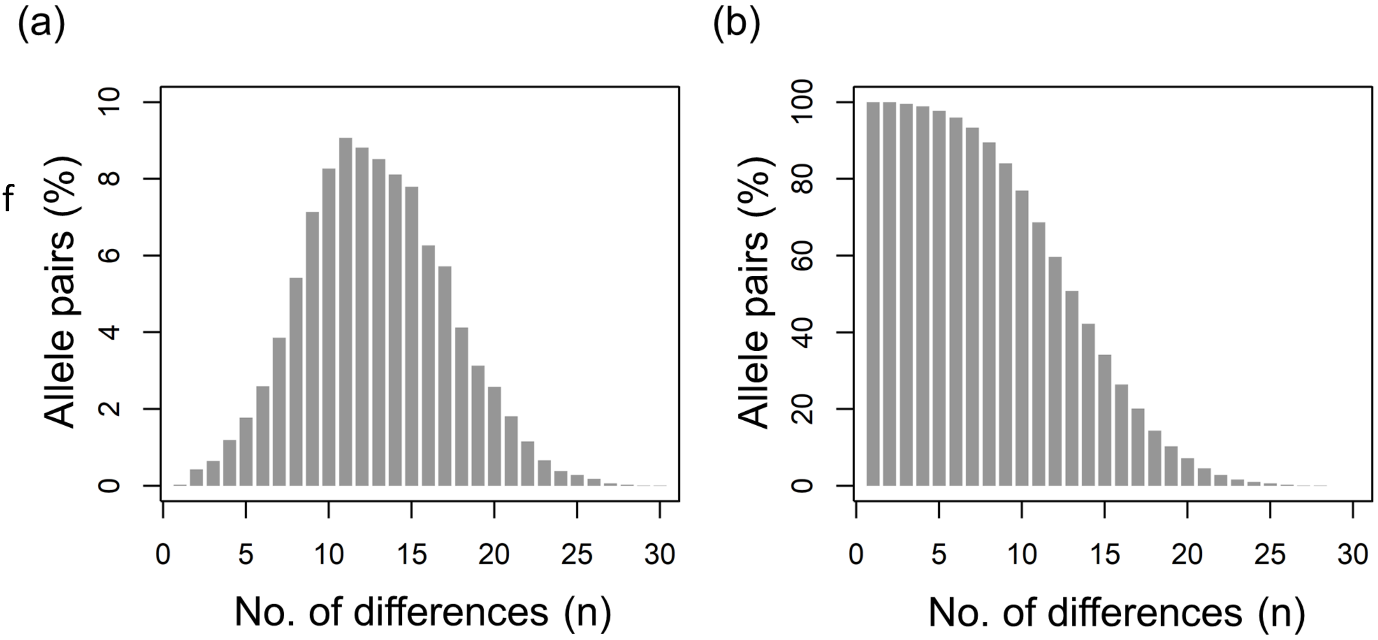
Multidimensional scaling
of the Jaccard distance
matrix

Complete-linkage
clustering based on
Jaccard distance

Results (5)

Study of the HVR *csd* variability via pairwise alignment of 160 non-identical alleles identified in the investigated honey samples.

Distribution of the HVR differences (alignments presenting n differences)



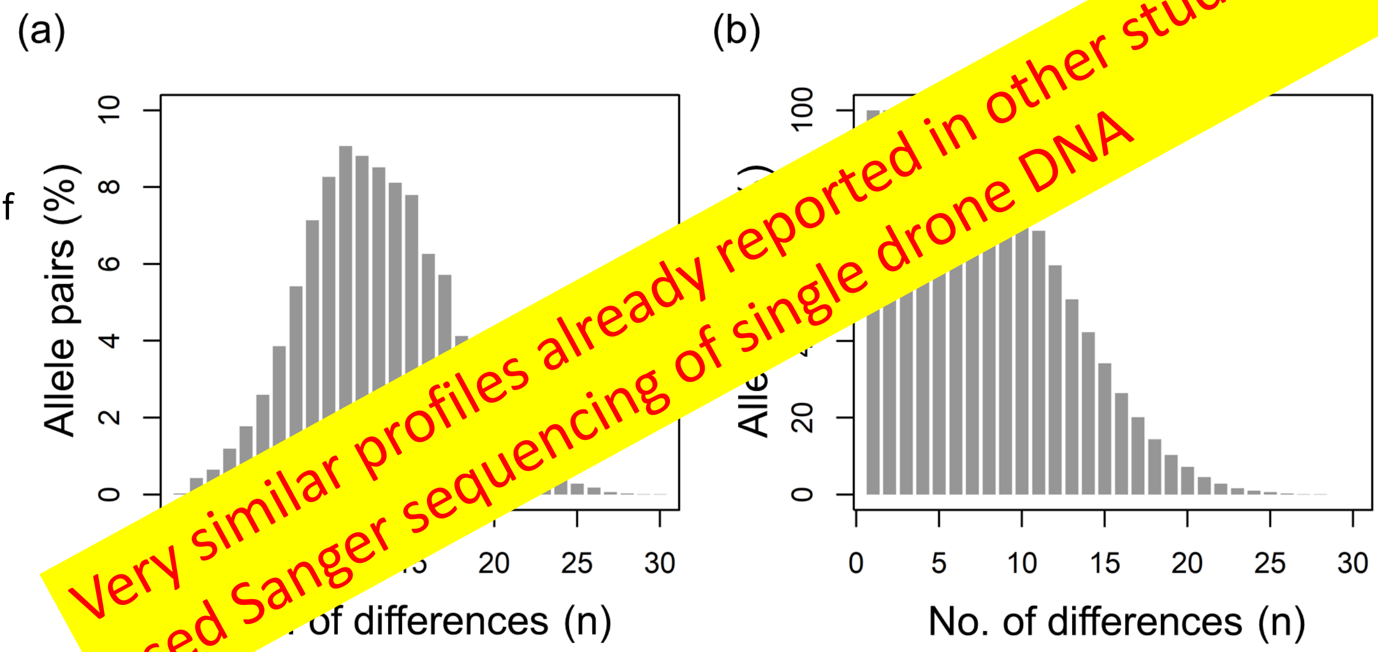
Cumulative distribution of the HVR differences (% of alignments presenting ≥ n differences).

Differences considered the $\Delta L_{HVR} + N_{SAP}$ values

Results (5)

Study of the HVR *csd* variability via pairwise alignment of 160 non-identical alleles identified in the investigated honey sa

Distribution of the HVR differences (alignments presenting n differences)



Cumulative distribution of the HVR differences (% of alignments presenting ≥ n differences).

Very similar profiles already reported in other studies that used Sanger sequencing of single drone DNA

Differences considered the $\Delta L_{HVR} + N_{SAP}$ values

ALMA MATER STUDIORUM
UNIVERSITÀ DI BOLOGNA

Conclusions

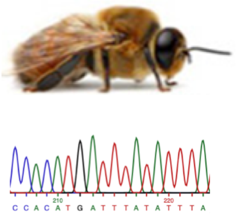
① Technical issues can be managed and considered to correctly interpret the final results

Conclusions

①

Technical issues can be managed and considered to correctly interpret the final results

②



≈

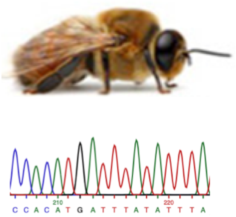
For honey bee population genetic studies, including analysis of the *csd* gene

Conclusions

①

Technical issues can be managed and considered to correctly interpret the final results

②



≈

For honey bee population genetic studies, including analysis of the csd gene

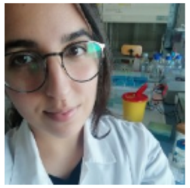
③

Can be more effective in capturing all *csd* alleles present in a colony

Co-authors of the Animal and Food Genomics Group of the University of Bologna

Samuele Bovo
Junior Assistant Professor

Anisa Ribani
Junior Assistant Profess



Valeria Taurisano
Post-doc

Giuseppina Schiavo
Junior Assistant Professor

Matteo Bolner
Post-doc

Francesca Bertolini
Associate Professor

Luca Fontanesi
Full Professor

Regione Emilia-Romagna

BEE-RER

ALMA MATER STUDIORUM
UNIVERSITÀ DI BOLOGNA

ALMA MATER STUDIORUM
UNIVERSITÀ DI BOLOGNA

EAAP
Insect-IMP
Athens, Jan. 29-31, 2025

Population genomics from a food matrix: estimating *csd* variability using next generation sequencing on *Apis mellifera* DNA present in honey

Luca Fontanesi

Animal and Food Genomics Group

Department of Agricultural and Food Sciences

University of Bologna, Bologna Italy

E-mail: luca.fontanesi@unibo.it