



HAL
open science

From workers to queen -Statistical imputation of individual genotype from pool sequencing

Sonia E Eynard, Christina Sann, Fanny Mondet, Benjamin Basso, Kamila Canake-Tabet, Rachid Mahla, Olivier Bouchez, Yves Le Conte, Yannick Poquet, Florence Phocas, et al.

► To cite this version:

Sonia E Eynard, Christina Sann, Fanny Mondet, Benjamin Basso, Kamila Canake-Tabet, et al.. From workers to queen -Statistical imputation of individual genotype from pool sequencing. *Approche Interdisciplinaire de l'Evolution Moléculaire*, Nov 2019, Toulouse, France. hal-04493790

HAL Id: hal-04493790

<https://hal.inrae.fr/hal-04493790v1>

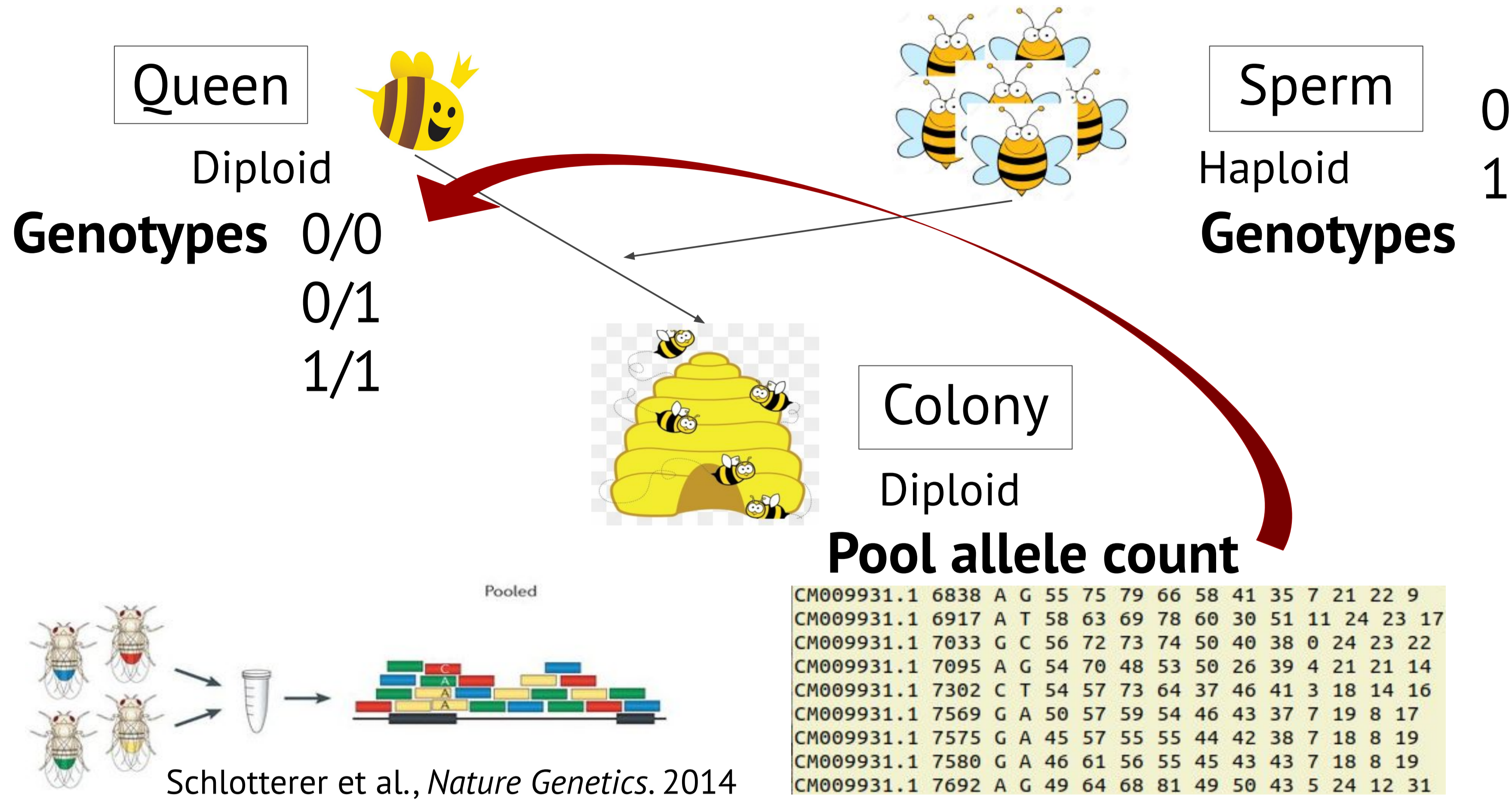
Submitted on 7 Mar 2024

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

From workers to queen - Statistical imputation of individual genotype from pool sequencing

Sonia E. Eynard¹, C. Sann⁴, F. Mondet³, B. Basso², K Canake-Tabet¹, R. Mahla⁴, O. Bouchez⁵, Y Le Conte³, Y Poquet⁴, F. Phocas⁶, F. Guillaume⁴, A. Decourtye³, L. Genestout⁴, A. Vignal¹, B. Servin¹
¹INRA GenPhySE; ²ITSAP; ³INRA Abeilles et Environnement; ⁴Labogena-Evolution; ⁵GetPlage; ⁶INRA GenAqua



Why is it useful to reconstruct queen genotype?

- Largest part of transmitted genetics
- Selection performed mainly on queens
- Run analysis with standard tools

Statistical model

Inference of Population Structure Using Multilocus Genotype Data
 Jonathan K. Pritchard, Matthew Stephens and Peter Donnelly
GENETICS, June 1, 2000 vol. 155 no. 2 945-959

$$\frac{g + f}{2} = \text{colony with queen genotype (g) and allele frequency in sperm (f)}$$

For each colony, let's consider

- 1) 1 colony queen + sperm = same population
- 2) Heterogeneous population: mixture of known genetic types + shared across markers (Q)

x_l : count referent allele at marker l

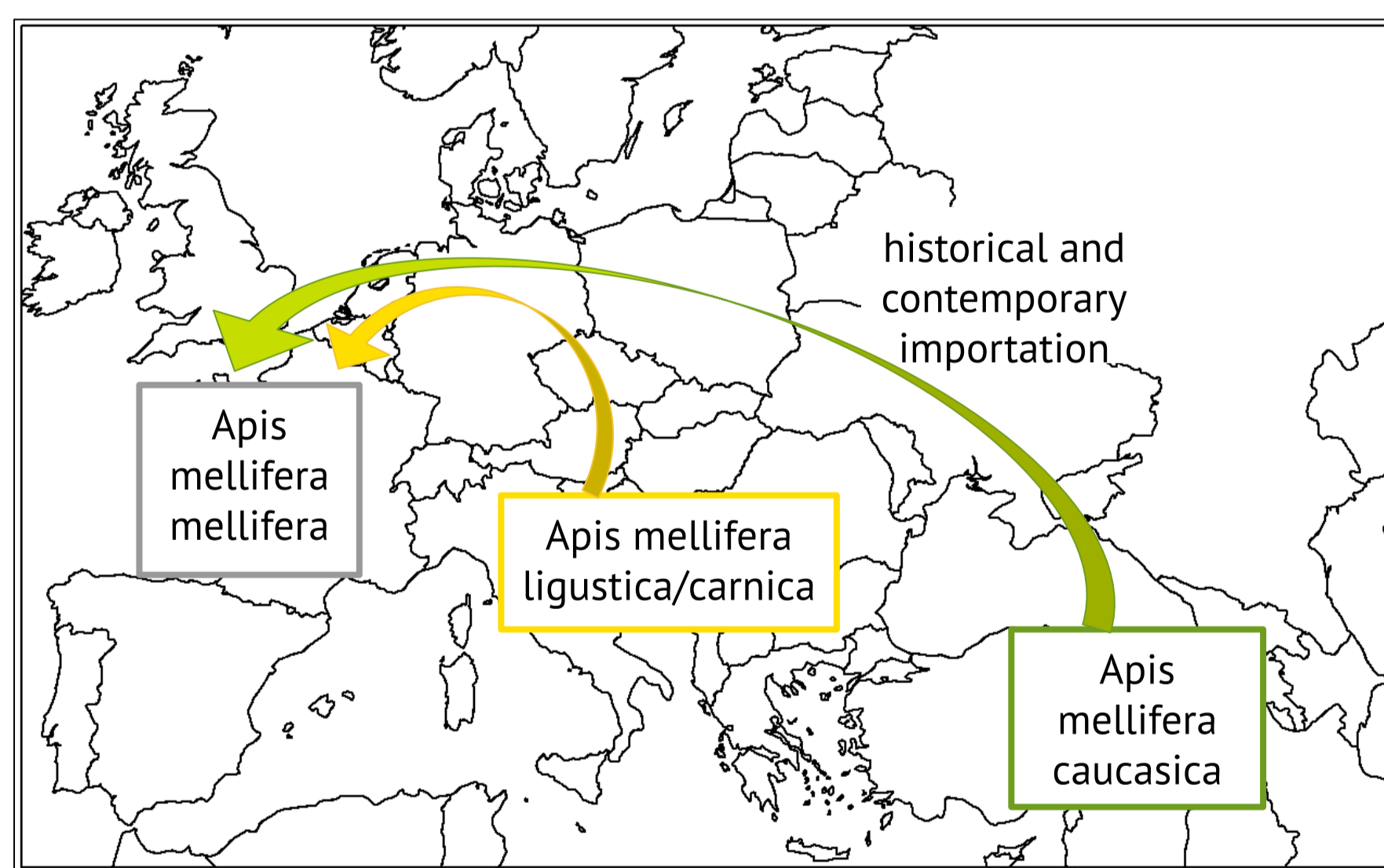
d_l : sequencing depth at marker l

To solve, obtain queen genotype: find Q maximising the likelihood

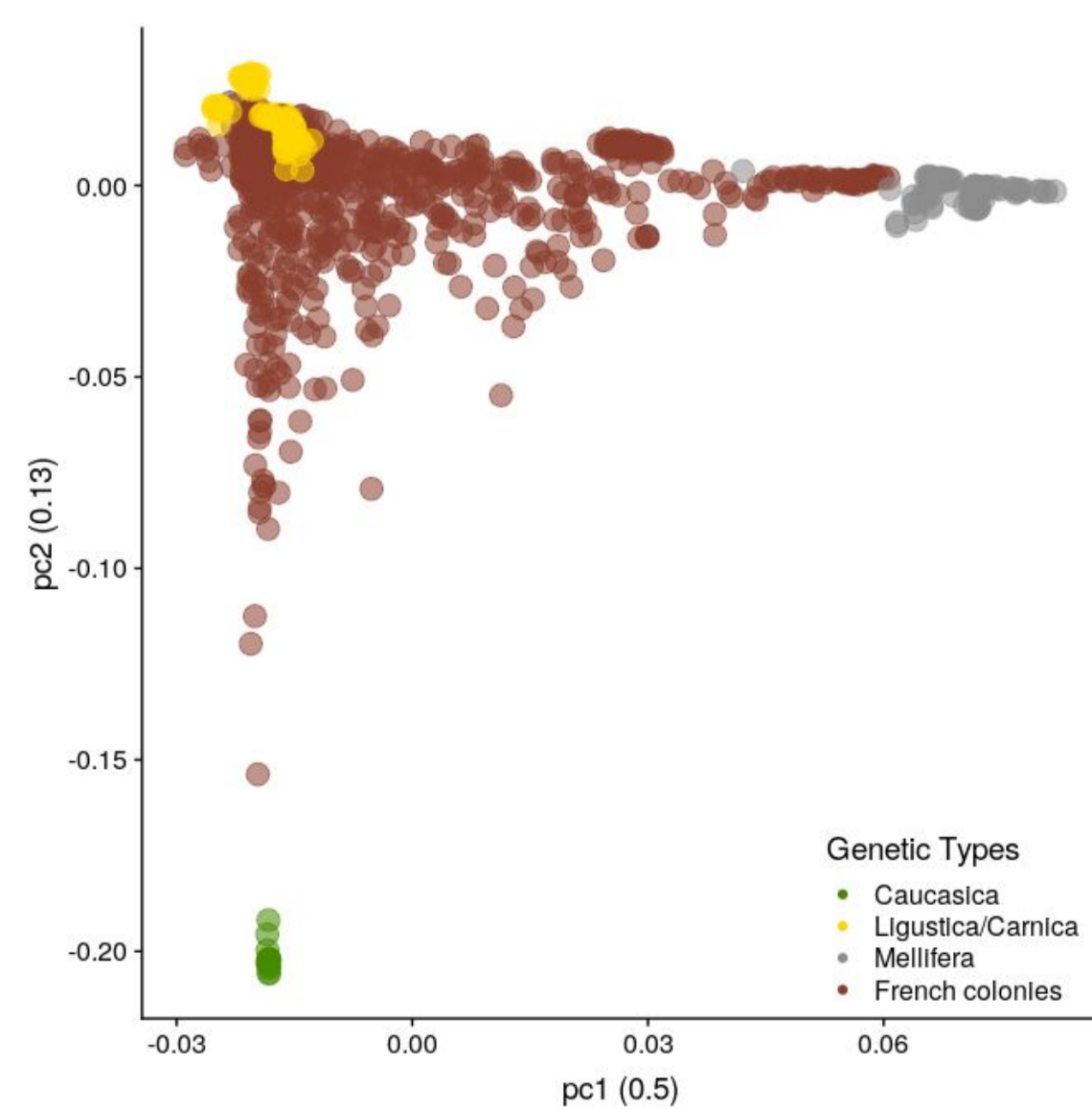
$$\prod_l P(g_l | x_l, d_l, Q)$$

where Q at each marker is f_l

Known genetic types

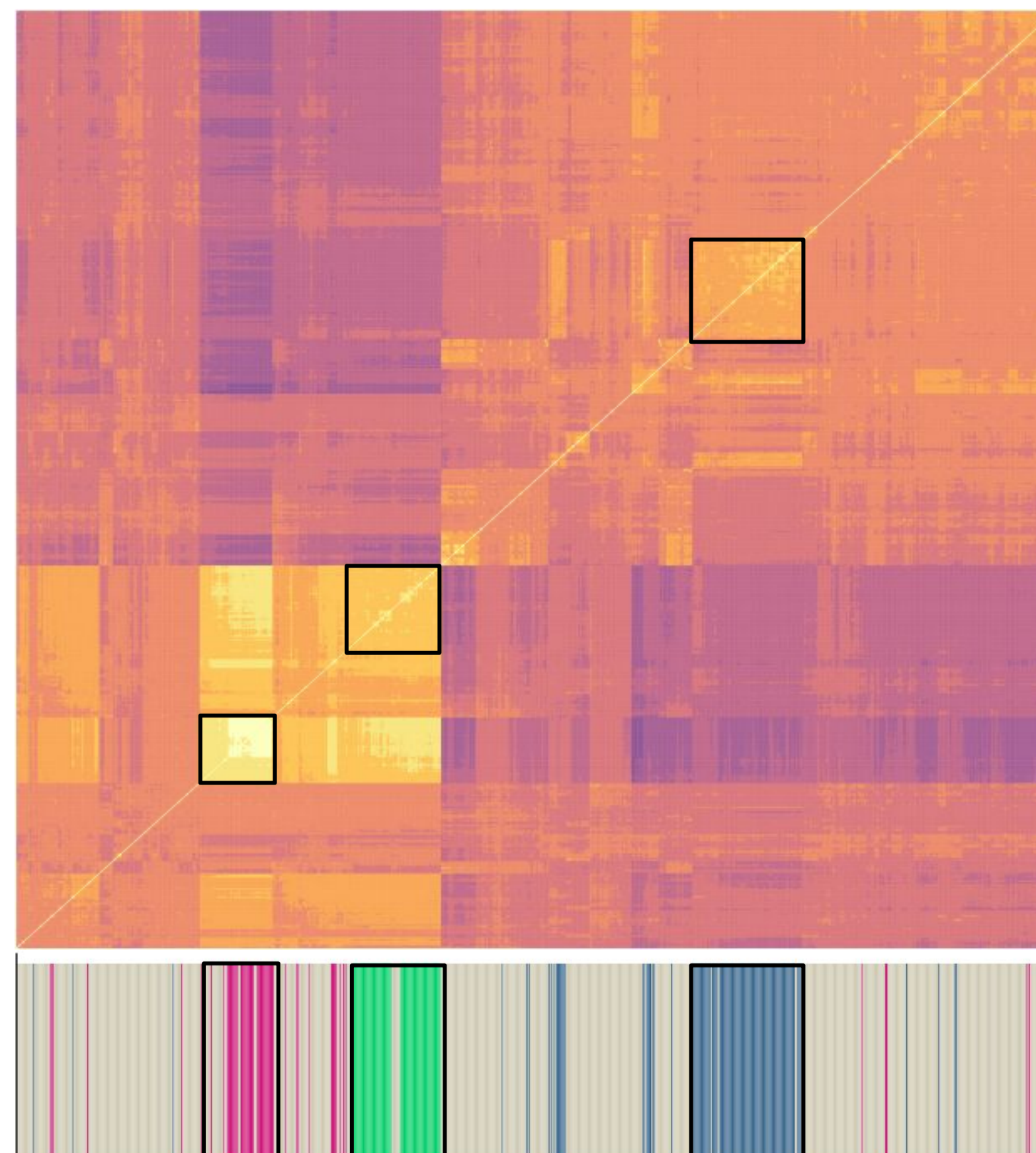


840 French colonies of mixed genetic types



Validate queen genotype with prior knowledge on apicultural practices

Relationships between French colonies (yellow=1, blue=0)



Admixture genetic types

