APIDOLOGIE Supplementary data

Genetic characterization of the honeybee (*Apis mellifera*) population of Rodrigues Island, based on microsatellite and mitochondrial DNA

Authors:

Maéva Angélique Techer[1,2], Johanna Clémencet[1], Patrick Turpin[2], Nicolas Volbert[3], Bernard Reynaud[2], Delatte Hélène[2]

[1]: Université de La Réunion, UMR PVBMT F-97715 Saint Denis cedex 9, La Réunion, France

[2]: CIRAD, UMR PVBMT, 7 chemin de l'Irat, Ligne Paradis, 97410 Saint Pierre, La Réunion, France

[3]: Apicultural Services, Rodrigues Regional Assembly - Agriculture, Citronnelle, Rodrigues

Corresponding authors: Hélène Delatte & Maéva Angélique Techer

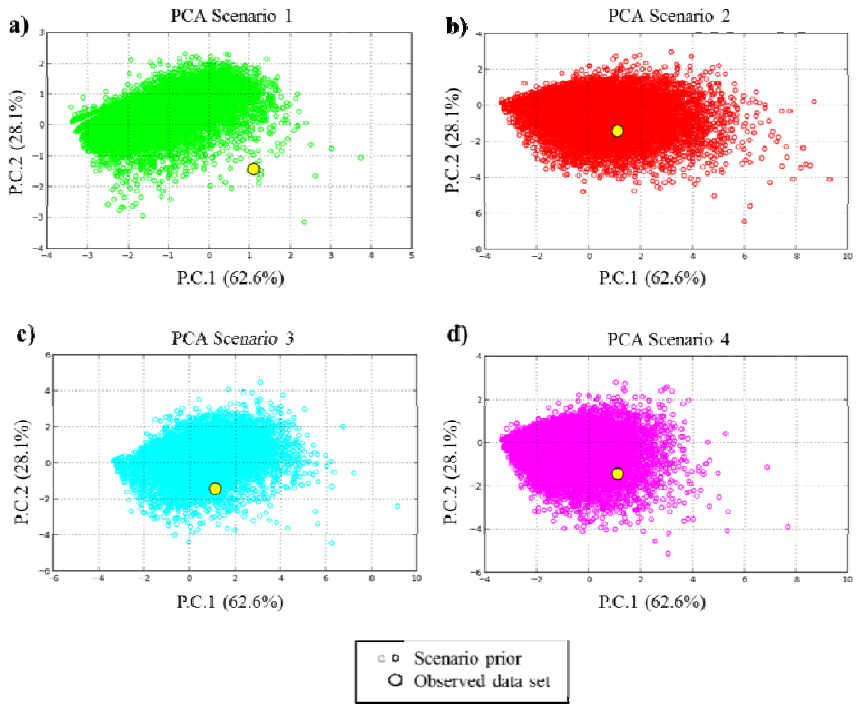
E-Mail: helene.delatte@cirad.fr ; maeva-angelique.techer@cirad.fr

Tel: 00 262 262 49 92 35

Description of the method used and preliminary results for DIYABC analyses:

1) Simulations and prior checking:

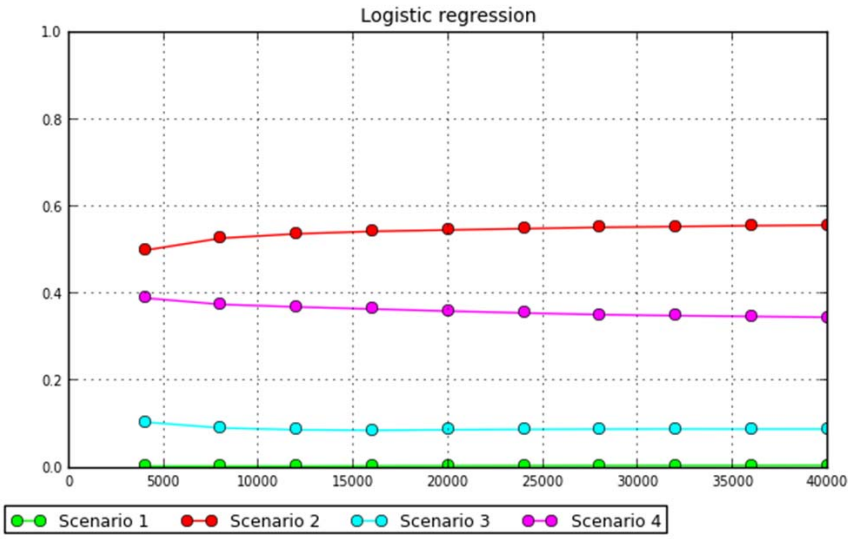
As recommended by Cornuet et al. (2008; 2010), we performed one million simulations of data sets per scenario with DIYABC software. Demographic parameters were sampled into prior distributions (typically uniform distributions, see Tab. 1 in the article for the full prior description). Mutations were assumed to follow a Generalized Stepwise Mutation model with rare insertions/deletions in flanking regions. All genetic parameters prior were the program default values. Then, prior checking was conducted using PCA over 10 000 random simulations. This was done in order to check that the chosen prior distributions allowed simulating data sets to be close to the observed data whatever the scenario envisaged. The results of this preliminary analysis are presented below with the PCA for all four scenarios used (Fig. 2 in the article):



2) Posterior probability for each scenario and confidence choice:

The posterior probability of each scenario was calculated and compared with others by performing a weighted logistic regression on 1% of simulated data sets closest to observed data set (Cornuet et al. 2008; Cornuet et al. 2010). The logistic regression

obtained for posterior probabilities for all four scenarios are presented below (number of simulations in abscissa and posterior probability in ordinate):



Thereafter, the confidence in scenario choice was calculated by evaluating Type I and II error rates, following method described in Cornuet et al. (2010). We first produced 500 simulated data sets (later called, pseudo-observed data sets or PODs) for each scenario and analyzed each of them as the true data, by computing their posterior probability. Type I error was estimated by counting the proportion of PODs simulated under the best scenario X (given scenario) for which X did not have the highest posterior probability. Type II error was estimated by the proportion of PODs that resulted in highest posterior probability of the best scenario, although simulated with other scenarios. Results of the test for confidence are presented in Table S2.

3) Estimations of the scenario demographic parameters:

For the most likely scenario, one million data sets were simulated independently of previous simulations to obtain parameter estimates, using the mode of posterior distribution as a point estimate. Precision of parameter estimation was assessed by computing the relative bias and the relative root mean square error on 500 PODs simulated with the best scenario.

Fig. S1 Box with description of the method used for DIYABC software with preliminary PCA of 10 000 simulated data sets for each scenario and logistic regression of posterior probabilities per scenario.

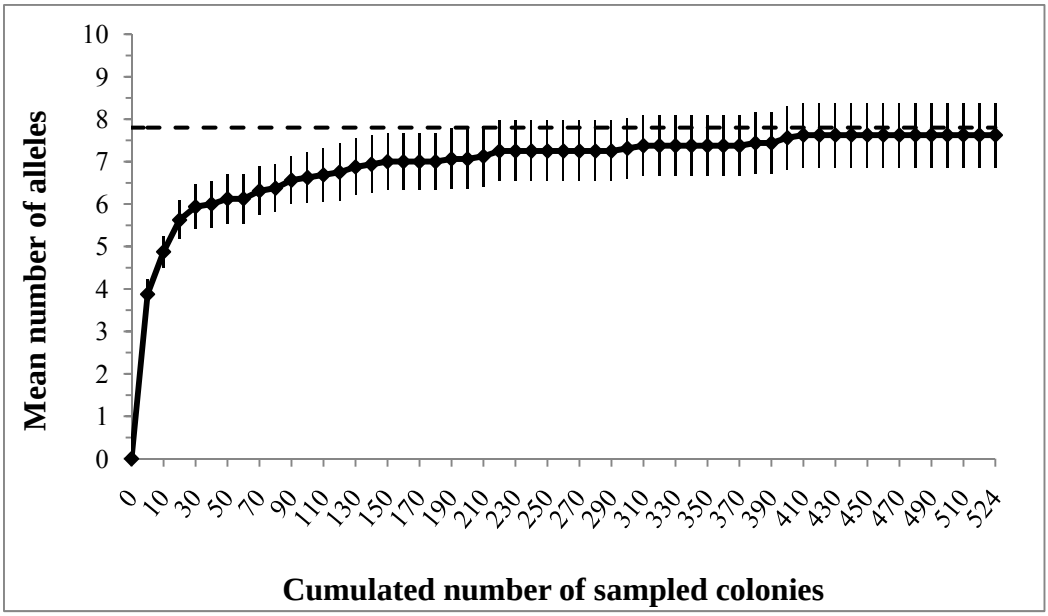


Fig. S2 Sampling effort represented by the cumulated number of colonies sampled (ordered by sampling date) and mean number of alleles per *locus* (n = 16) (± Standard deviations). The dotted line represented the asymptote with x = 7.80.

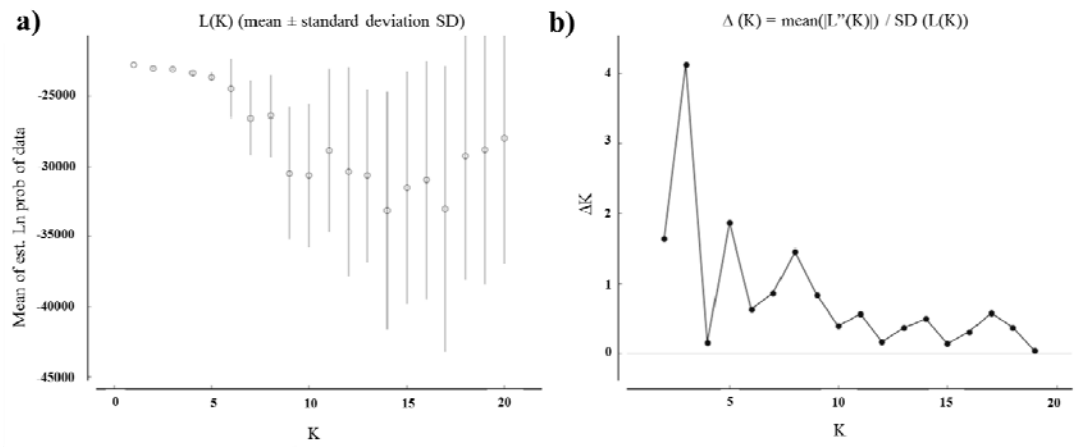


Fig. S3 a) Average likelihood of runs in STRUCTURE (Pritchard et al. 2000) L(K) along with number of K clusters, b) ΔK, estimator of the optimal number of clusters (K) according to Evanno et al. (2005).

Tab. S1 Microsatellite *loci* information with indication of multiplex group (Mix 1 to 4), original marker name by Solignac et al. (2003), name of primers used for the study, primer's nucleotide sequence, microsatellite motive repeated, fluorochrome, range of detected allele size and N_{Allele}, the number of detected alleles for all individuals in Rodrigues (n = 524). Markers removed from the analysis are presented in gray.

Multiplex group	Original marker name	Name of primer	Nucleotide sequence	Motive	Fluorochrome	Range of allele size (bp)	N_{Allele}
Mix 1	A113	A113-F	5'- CTCGAATCGTGGCGTCC -3'	$(TC)_5TT(TC)_8TT(TC)_5$	VIC	202 - 234	8
		A113-R	5'- CCTGTATTTTGCAACCTCGC -3'				
	A024	A24-F	5'- CACAAGTTCCAACAATGC -3'	$(CT)_{11}$	FAM	92 - 106	6
		A24-R	5'- CACATTGAGGATGAGCG -3'				
	AC306	Ac306-a	5'- GAATATGCCGCTGCCACC -3'	$(CT)_{11}$	FAM	165 - 185	6
		Ac306-b	5'- TTTCGTTGCATCCGAGCG -3'				
	AP055	Ap55-1	5'- GATCACTTCGTTTCAACCGT -3'	$(TC)_9..(TC)_{12}$	PET	147 - 207	8
		Ap55-2	5'- CATTCGGTATGGTACGACCT -3'				
	AP081	Ap81-1	5'- GGATCGTCGAGGCGTTGA -3'	$(GT)_8$	NED	124 - 136	4
		Ap81-2	5'- GAAAAGTATTCCGCCGAGCA -3'				
Mix 2	A107	A107-1	5'- CCGTGGGAGGTTTATTGTCG -3'	$(CT)_{23}$	VIC	138 - 184	15
		A107-2	5'- GGTTCGTAACGGATGACACC -3'				
	A029	A29-2	5'- CAACTTCAACTGAAATCCG -3'	$(CA)_{24}$	NED	128 - 175	15
		A29-1	5'- AAACAGTACATTTGTGACCC -3'				
	A088	A88-F	5'- CGAATTAACCGATTTGTCG -3'	$(CT)_{10}..(GGA)_7$	VIC	136 - 149	5
		A88-R	5'- GATCGCAATTATTGAAGGAG -3'				
	AP273	Ap273-a	5'- GATCTTGTGTTAAACAGCCG -3'	$(CT)_8$	PET	106 - 110	3
		Ap273-b	5'- GATCTCTGGCAGACGAAGAG -3'				
	A028	A28-F	5'- GAAGAGCGTTGGTTGCAGG -3'	$(AG)_6(GAG)_6$	FAM	128 - 134	4
		A28-R	5'- GCCGTTCATGGTTACCACG -3'				
	AP289	Ap289-a	5'- AGCTAGGTCTTTCTAAGAGTGTTG -3'	$(GA)_5$	NED	174 - 228	10
		Ap289-b	5'- TTCGACCGCAATAACATTC -3'				
	A124	B124-1	5'- GCAACAGGTCGGGTTAGAG -3'	$(CT)_8..(CT)_{14}..(GGCT)_8$	PET	216 - 232	7
		B124-2	5'- CAGGATAGGGTAGGTAAGCAG -3'				
Mix 3	A035	A35-1	5'- GTACACGGTTGCACGGTTG -3'	$(GT)_{14}$	FAM	94 - 123	10
		A35-2	5'- CTTCGATGGTCGTTGTACCC -3'				
	A008	A8-1	5'- CGAAGGTAAGGTAAATGGAAC -3'	$(GA)_{15}...(GCTCG)_5$	VIC	165 - 181	6
		A8-2	5'- GGCGGTTAAAGTTCTGG -3'				
	AP033	Ap33-1	5'- TTTCTTTTTGTGGACAGCG -3'	$(CT)_{15}$	PET	225 - 247	10
		Ap33-2	5'- AAATATGGCGAAACGTGTG -3'				
Mix 4	AP043	Ap43-1	5'- GGCGTGCACAGCTTATTCC -3'	$(TA)_6GATA(GA)_{10}$	FAM	129 - 183	11
		Ap43-2	5'- CGAAGGTGGTTTCAGGCC -3'				
	AP066	Ap66-1	5'- TTGCATTCGGTCTCCAGC -3'	$(CT)_{11}$	VIC	90 - 102	5
		Ap66-2	5'- ACTTGCCGCGGTATCTGA -3'				
	A043	A43-1	5'- CACCGAAACAAGATGCAAG -3'	$(CT)_{12}$	PET	124 - 154	8
		A43-2	5'- CCGCTCATTAAGATATCCG -3'				

Tab. S2 Number of time each scenario was assigned as the best (highest posterior probability) under 500 pseudo-observed data sets simulated under scenario 1, 2, 3 or 4 with DIYABC (Cornuet et al. 2008).

	Data sets simulated under Scenario 1	Data sets simulated under Scenario 2	Data sets simulated under Scenario 3	Data sets simulated under Scenario 4
Scenario 1 (n_{param} = 1)	439	2	93	68
Scenario 2 (n_{param} = 5)	0	401	100	128
Scenario 3 (n_{param} = 6)	28	85	237	66
Scenario 4 (n_{param} = 8)	33	12	70	238
Total of simulations	**500**	**500**	**500**	**500**