

NOTES AND COMMENTS

Empirical comparison of microsatellite and SNP markers to estimate introgression in *Apis mellifera mellifera*

Melanie Parejo^{a,b*}, Dora Henriques^{c,d}, M Alice Pinto^c, Gabriele Soland-Reckeweg^e and Markus Neuditschko^a

^aAgroscope, Swiss Bee Research Centre, Bern, Switzerland.

^bInstitute of Bee Health, Vetsuisse Faculty, University of Bern, Bern, Switzerland.

^cMountain Research Centre, Polytechnic Institute of Bragança, Bragança, Portugal.

^dCentre of Molecular and Environmental Biology, University of Minho, Braga, Portugal.

^eApigenix, Twann, Switzerland.

(Received 14 October 2017, accepted 16 May 2018)

*Corresponding author: Email: melanie.parejo@alumni.ethz.ch

Online Supplementary Material

Supplementary Materials and methods

Whole-genome sequence data

The first dataset consisted of N=31 drone samples originating from our previous study (Parejo et al., 2016), which were selected to represent the entire range of admixed individuals (e.g. pure *Apis mellifera mellifera*, pure *A. m. carnica*, and a range of hybrids in-between). Detailed information on sequencing and variant calling can be found in the respective publication. Using 3,076,531 SNPs identified from whole-genome sequencing, admixture proportions of the 31 individuals were inferred using ADMIXTURE v1.3.0 (Alexander, Novembre, & Lange, 2009) and a reference sample set of N=334 drone samples (Henriques et al., 2018a). ADMIXTURE was run unsupervised for K=2 with 10,000 iterations and the default termination criterion set to stop when the log-likelihood increases by less than 0.0001 between independent runs.

Microsatellites

The second dataset consisted of N=56 *A. m. mellifera* samples, which were tested in 2016 and selected to include a diverse range of introgression proportions (C-lineage ancestry: [0.002-0.502]). All samples have been genotyped at 12 microsatellite loci (A007, A43, Ap33, Ap226, B24, Ap273, Ac306 Ap289, A28, Ap1, A29, A76) (Soland-Reckeweg, 2006) (www.apigenix.ch). Ancestries of the 56 samples were inferred together with 355 reference

samples from Switzerland, France, Norway, Sweden and Ireland (*A. m. mellifera*), as well as Austria and Slovenia (*A. m. carnica*) using STRUCTURE v2 (Falush, Stephens, & Pritchard, 2003) with the correlated allele frequency model for K=2. This is the approach of the currently commercially available microsatellite-based hybrid test in Switzerland (www.apigenix.com).

Ancestry-informative SNP panel

The samples of the two datasets (N=31 and N=56; NTOT=87) with whole-genome sequence information or microsatellite data, respectively, were both genotyped at 117 SNPs using a recently developed ancestry-informative SNP panel for C-lineage introgression in *A. m. mellifera* (Henriques et al., 2018b).

The SNP panel was designed starting from the 144 top-ranked SNPs selected by Muñoz et al., (2015) to identify and estimate introgression in *A. m. mellifera*. Muñoz and colleagues (2015) had ranked by information content >1183 SNPs that had been 2 / 2 genotyped in numerous samples of *A. m. mellifera*, *A. m. ligustica* and *A. m. carnica* using Illuminas GoldenGate Assay. The top-ranked SNPs had been combined into reduced SNP panels to fit the plexes of the GoldenGate Assay genotyped with the VeraCode technology (48, 96, 144, 192, 384 SNPs) (Muñoz et al., 2015). This technology has now been discontinued.

The new 117-SNP assay designed for high sample throughput genotyping using the iPLEX MassARRAY system is validated and extensively tested in different experiments by Henriques et al. (2018b) resulting in a ready-to-use molecular tool to estimate C-lineage introgression in *A. m. mellifera*.

Here, we apply and test this reduced SNP panel in our samples from Switzerland (NTOT=87). Ancestries of the 87 samples were inferred using ADMIXTURE v1.3.0 (Alexander, Novembre, & Lange, 2009). The software was run unsupervised for K=2 with 10,000 iterations and the default termination criterion set to stop when the log-likelihood increases by less than 0.0001 between independent runs.

References

- Alexander, D. H., Novembre, J., & Lange, K. (2009). Fast model-based estimation of ancestry in unrelated individuals. *Genome Research*, 19(9), 1655-1664.
doi: 10.1101/gr.094052.109

- Henriques, D., Parejo, M., Vignal, A., Wragg, D., Wallberg, A., Webster, M. T. & Pinto, M. A. (2018a). Developing reduced SNP assays from whole-genome sequence data to estimate introgression in an organism with complex genetic patterns, the Iberian honey bee (*Apis mellifera iberiensis*). *Evol Appl.* Accepted Author Manuscript.
doi: 10.1111/eva.12623
- Henriques, D., Browne, K. A., Barnett, M. W., Parejo, M., Kryger, P., Freeman, T., Muñoz, I., Garnery, L., Hight, F., Johnston, J. S., McCormack, G., & Pinto, M. A. (2018b). High sample throughput genotyping for estimating C-lineage introgression in the dark honey bee: an accurate and cost-effective SNP-based tool. Manuscript in preparation.
- Muñoz, I., Henriques, D., Johnston, J. S., Chávez-Galarza, J., Kryger, P., & Pinto, M. A. (2015). Reduced SNP panels for genetic identification and introgression analysis in the dark honey bee (*Apis mellifera mellifera*). *PLoS ONE*, 10(4).
doi: 10.1371/journal.pone.0124365
- Parejo, M., Wragg, D., Gauthier, L., Vignal, A., Neumann, P., & Neuditschko, M. (2016). Using whole-genome sequence information to foster conservation efforts for the European dark honey bee, *Apis mellifera mellifera*. *Frontiers in Ecology and Evolution*, 4(140). doi: 10.3389/fevo.2016.00140
- Pritchard, J. K., Stephens, M., & Donnelly, P. (2000). Inference of population structure using multilocus genotype data. *Genetics*, 155(2), 945-959
- Soland-Reckeweg, G. (2006). *Genetic differentiation and hybridization in the honey bee (*Apis mellifera L.*) in Switzerland*. Universität Bern, Bern, CH

Online Supplementary Material 2.

Admixture proportions as estimated by 3.077 M SNPs (whole-genome), 117 SNPs (ancestry-informative panel) and 12 microsatellites.

	Whole-genome (3.077 M SNPs)		117 ancestry-informative SNPs		12 microsatellites	
Sample_ID	M-lineage	C-lineage	M-lineage	C-lineage	M-lineage	C-lineage
M1	0.833768	0.166232	0.839083	0.160917	-	-
M10	0.721744	0.278256	0.762758	0.237242	-	-
M11	0.696076	0.303924	0.732541	0.267459	-	-
M12	0.652131	0.347869	0.647419	0.352581	-	-
M13	0.569874	0.430126	0.571698	0.428302	-	-
M14	0.477073	0.522927	0.424935	0.575065	-	-
M15	0.412099	0.587901	0.419648	0.580352	-	-
M17	0.251813	0.748187	0.246458	0.753542	-	-
M18	0.201527	0.798473	0.179544	0.820456	-	-
M19	0.172023	0.827977	0.00001	0.99999	-	-
M2	0.832012	0.167988	0.861117	0.138883	-	-
M20	0.171783	0.828217	0.00001	0.99999	-	-
M21	0.053485	0.946515	0.00001	0.99999	-	-
M22	0.042018	0.957982	0.00001	0.99999	-	-
M23	0.893369	0.106631	0.885053	0.114947	-	-
M24	0.896514	0.103486	0.959394	0.040606	-	-
M25	0.893973	0.106027	0.99999	0.00001	-	-
M26	0.897572	0.102428	0.99999	0.00001	-	-
M27	0.00001	0.99999	0.00001	0.99999	-	-
M28	0.042723	0.957277	0.00001	0.99999	-	-
M3	0.820399	0.179601	0.872257	0.127743	-	-
M33	0.127025	0.872975	0.00001	0.99999	-	-
M34	0.199272	0.800728	0.187264	0.812736	-	-
M35	0.185363	0.814637	0.13997	0.86003	-	-
M36	0.147089	0.852911	0.00001	0.99999	-	-
M4	0.814109	0.185891	0.879798	0.120202	-	-
M5	0.806082	0.193918	0.938224	0.061776	-	-
M6	0.803106	0.196894	0.882011	0.117989	-	-
M7	0.788771	0.211229	0.868143	0.131857	-	-
M8	0.783093	0.216907	0.899767	0.100233	-	-
M9	0.755685	0.244315	0.700815	0.299185	-	-

	Whole-genome (3.077 M SNPs)		117 ancestry-informative SNPs		12 microsatellites	
Sample_ID	M-lineage	C-lineage	M-lineage	C-lineage	M-lineage	C-lineage
S1	-	-	0.99999	0.00001	0.989	0.011
S10	-	-	0.927836	0.072164	0.996	0.004
S11	-	-	0.99999	0.00001	0.963	0.037
S12	-	-	0.972177	0.027823	0.991	0.009
S13	-	-	0.99999	0.00001	0.987	0.013
S14	-	-	0.988134	0.011866	0.993	0.007
S15	-	-	0.99999	0.00001	0.924	0.076
S16	-	-	0.875589	0.124411	0.929	0.071
S17	-	-	0.99999	0.00001	0.959	0.041
S18	-	-	0.919314	0.080686	0.763	0.237
S19	-	-	0.954839	0.045161	0.904	0.096
S2	-	-	0.732292	0.267708	0.714	0.286
S20	-	-	0.99999	0.00001	0.882	0.118
S21	-	-	0.902325	0.097675	0.908	0.092
S22	-	-	0.937467	0.062533	0.945	0.055
S23	-	-	0.940062	0.059938	0.96	0.04
S24	-	-	0.916385	0.083615	0.959	0.041
S25	-	-	0.896222	0.103778	0.986	0.014
S26	-	-	0.935692	0.064308	0.982	0.018
S27	-	-	0.910225	0.089775	0.976	0.024
S28	-	-	0.915911	0.084089	0.962	0.038
S29	-	-	0.951805	0.048195	0.968	0.032
S3	-	-	0.921175	0.078825	0.981	0.019
S30	-	-	0.907743	0.092257	0.89	0.11
S31	-	-	0.890753	0.109247	0.962	0.038
S32	-	-	0.934056	0.065944	0.975	0.025
S33	-	-	0.77296	0.22704	0.854	0.146
S34	-	-	0.88209	0.11791	0.954	0.046
S35	-	-	0.966042	0.033958	0.97	0.03
S36	-	-	0.922025	0.077975	0.959	0.041
S37	-	-	0.64762	0.35238	0.759	0.241
S38	-	-	0.923465	0.076535	0.978	0.022
S39	-	-	0.885865	0.114135	0.979	0.021
S4	-	-	0.99999	0.00001	0.931	0.069
S40	-	-	0.369051	0.630949	0.498	0.502
S41	-	-	0.953665	0.046335	0.874	0.126
S42	-	-	0.929599	0.070401	0.944	0.056
S43	-	-	0.907139	0.092861	0.938	0.062
S44	-	-	0.99999	0.00001	0.985	0.015
S45	-	-	0.671705	0.328295	0.702	0.298
S46	-	-	0.894835	0.105165	0.96	0.04
S47	-	-	0.937088	0.062912	0.962	0.038
S48	-	-	0.962081	0.037919	0.989	0.011
S49	-	-	0.954278	0.045722	0.956	0.044
S5	-	-	0.977618	0.022382	0.988	0.012
S50	-	-	0.838448	0.161552	0.886	0.114
S51	-	-	0.949931	0.050069	0.894	0.106
S52	-	-	0.765782	0.234218	0.688	0.312
S53	-	-	0.93953	0.06047	0.97	0.03
S54	-	-	0.99999	0.00001	0.827	0.173
S55	-	-	0.93538	0.06462	0.989	0.011
S56	-	-	0.993187	0.006813	0.964	0.036
S6	-	-	0.993354	0.006646	0.995	0.005
S7	-	-	0.953464	0.046536	0.998	0.002
S8	-	-	0.948025	0.051975	0.994	0.006
S9	-	-	0.964006	0.035994	0.993	0.007

