Data description

Here we store 84 folders in which the supplementary data for the paper "Comparison of infinitesimal and finite locus models for long-term breeding simulations with direct and maternal effects at the example of honeybees" is contained:

- The output of the ten individual simulations with setting $\mathrm{FL200}^{300\,\mathrm{BQs}}_{-0.25}$ is stored in the folders
  Sim_300BQs_200Loci_025_Rep1 to Sim_300BQs_200Loci_025_Rep10.

- The output of the ten individual simulations with setting $\mathrm{FL200}^{300\,\mathrm{BQs}}_{-0.75}$ is stored in the folders
  Sim_300BQs_200Loci_075_Rep1 to Sim_300BQs_200Loci_075_Rep10.

- The output of the ten individual simulations with setting $\mathrm{FL200}^{1000\,\mathrm{BQs}}_{-0.25}$ is stored in the folders
  Sim_1000BQs_200Loci_025_Rep1 to Sim_1000BQs_200Loci_025_Rep10.

- The output of the ten individual simulations with setting $\mathrm{FL200}^{1000\,\mathrm{BQs}}_{-0.75}$ is stored in the folders
  Sim_1000BQs_200Loci_075_Rep1 to Sim_1000BQs_200Loci_075_Rep10.

- The output of the ten individual simulations with setting $\mathrm{FL400}^{300\,\mathrm{BQs}}_{-0.25}$ is stored in the folders
  Sim_300BQs_400Loci_025_Rep1 to Sim_300BQs_400Loci_025_Rep10.

- The output of the ten individual simulations with setting $\mathrm{FL400}^{300\,\mathrm{BQs}}_{-0.75}$ is stored in the folders
  Sim_300BQs_400Loci_075_Rep1 to Sim_300BQs_400Loci_075_Rep10.

- The output of the ten individual simulations with setting $\mathrm{FL400}^{1000\,\mathrm{BQs}}_{-0.25}$ is stored in the folders
  Sim_1000BQs_400Loci_025_Rep1 to Sim_1000BQs_400Loci_025_Rep10.

- The output of the ten individual simulations with setting $\mathrm{FL400}^{1000\,\mathrm{BQs}}_{-0.75}$ is stored in the folders
  Sim_1000BQs_400Loci_075_Rep1 to Sim_1000BQs_400Loci_075_Rep10.

- The output of the simulation with setting $\mathrm{INF}^{300\,\mathrm{BQs}}_{-0.25}$ is stored in the folder
  Sim_300BQs_INF_025.

- The output of the simulation with setting $\mathrm{INF}^{300\,\mathrm{BQs}}_{-0.75}$ is stored in the folder
  Sim_300BQs_INF_075.

- The output of the simulation with setting $\mathrm{INF}^{1000\,\mathrm{BQs}}_{-0.25}$ is stored in the folder
  Sim_1000BQs_INF_025.

- The output of the simulation with setting $\mathrm{INF}^{1000\,\mathrm{BQs}}_{-0.75}$ is stored in the folder
  Sim_1000BQs_INF_075.

We distributed the folders to 17 meta folders which were then compressed using tar and gzip.

Each of the individual 84 folders itself contains five subfolders and one textfile:

- `EBV`

- `Inbreedings`

- `Observed_Values`

- `Pedigrees`

- `TBV`

- `geneset.txt`

The folder `EBV` contains the estimated breeding values of all BQs and their worker groups based on a breeding value estimation that was performed after the last year of the simulations. Direct and maternal breeding values of queens can be found in the subfolders `dir` and `mat`, respectively. The corresponding values for the worker groups (colonies) are stored in `dcol` and `mcol`. Each of the four folders `dir`, `mat`, `dcol`, and `mcol` contains 100 files representing the 100 repetitions that were carried out with each setting. In an individual file, each line represents a simulated year and each column a simulated queen born in that year (resp. her worker group). The entry is the corresponding estimated breeding value. The entries are tab separated.

The folder `TBV` (or rather its subfolder `Breeding_Queens`) contains the true breeding values of all BQs and their worker groups. Direct and maternal breeding values of queens can be found in the subfolders `dir` and `mat`, respectively. The corresponding values for the worker groups (colonies) are stored in `dcol` and `mcol`. Each of the four folders `dir`, `mat`, `dcol`, and `mcol` contains 100 files representing the 100 repetitions that were carried out with each setting. In an individual file, each line represents a simulated year and each column a simulated queen born in that year (resp. her worker group). The entry is the corresponding true breeding value. The entries are tab separated.

The folder `Inbreedings` (or rather its subfolder `Breeding_Queens`) contains the inbreeding coefficients of all BQs. It contains 100 files representing the 100 repetitions that were carried out with each setting. In an individual file, each line represents a simulated year and each column a simulated queen born in that year. The entry is the corresponding inbreeding coefficient. The entries are tab separated.

The folder `Observed_Values` contains the simulated performance test results of all BQs and their colonies. It contains 100 files representing the 100 repetitions that were carried out with each setting. In an individual file, each line represents a simulated year and each column a simulated queen born in that year. The entry is the corresponding performance data. The entries are tab separated.

The folder `Pedigrees` contains the pedigrees corresponding to the simulated data. It contains 100 files representing the 100 repetitions that were carried out with each setting. In an individual file, each line represents a simulated breeding queen. The first column gives the queen ID consisting of two numbers, the second of which is the queen's birthyear. The second column gives the identity of the queens dam, if known. The third column describes the dam of a queen's

2

pseudo sire, whereas the fourth column gives an ID for the pseudo sire itself. The last two columns only contain the numbers 8 and 12, representing that each pseudo sire comprised 8 DPQs and each queen mated with 12 drones.

The file `geneset` contains information on the allele effects and frequencies used for the simulation. The first two lines contain the number of simulated loci. Then separate blocks with information on the respective loci follow. Each block starts with four meaningless lines reading `q`, `0.5`, `2`, and `0`, respectively. The following line indicates the frequency of the first allele (followed by a one). Next come the maternal and direct effect (in this order) of the first allele and finally of the second allele.