Hamilton Rule in the Microbiome

Camille Simonet

01/05/2020

Compiling all results from a given .RData object output which ran all models. We have:

- **Model 1**: *cooperation ~ mean relatedness + log(genome size)*. There are 5 models, one for each cooperative trait. We then use a random effect meta-analysis to estimate an overall effect.

- **Model 2**: *cbind(cooperation, within host relatedness) ~ phylogeny + species*. This bivariate formulation of the model allowing to account for uncertainty in the measures of relatedness. Again there are 5 models, one for each cooperative traits, and a random effect meta-analysis to estimate an overall effect.

- **Model 3**: *within host relatedness ~ 1 + sporulation score + within host relative abundance*. Two ecological factor classically predicted to shape relatedness (infinite island model background). Asking if these predictions hold in the case of microbiome.

- **Model 4**: *cooperation ~ mean relatedness + log(genome size) + sporulation score + mean relative abundance*. Given the effect of those ecological factors on relatedness, are they the ultimate driver of cooperation (i.e. are they sufficient to explain the phylogenetic distribution of relatedness)? This model asks if relatedness retains predictive power after accounting for these two ecological factors.

- **Model 5**: *within host relatedness ~ cooperation + sporulation score + within host relative abundance + secretome size + biofilm + antibiotic degradation + siderophores + quorum sensing*. This models completes the path analysis, asking if cooperative traits and ecological factor predict relatedness.

Additional notes on variables included in the models:

- **cooperation**: 5 cooperative traits. *secretome size* , measured computationally from the number of predicted secreted proteins. Then *biofilm formation, quorum-sensing, secretion systems, cooperative antibiotic degradation and siderophores production* were computed from Gene Ontologies, as the number of genes matching gene ontology terms refering to these biological functions.

- **gram profile**: all models run on secretome also include gram profile as a co-variate, because the algorithm to compute secretome is different for gram negative and gram positive.

- **mean relative abundance**: relative abundance measures taken from midas output,(**NOT** rescaled to the set of species included in the analysis), and averaged over species.

- **genome size**: the number of CDS not involved in the cooperative trait (following reasonning in supplementary equations).
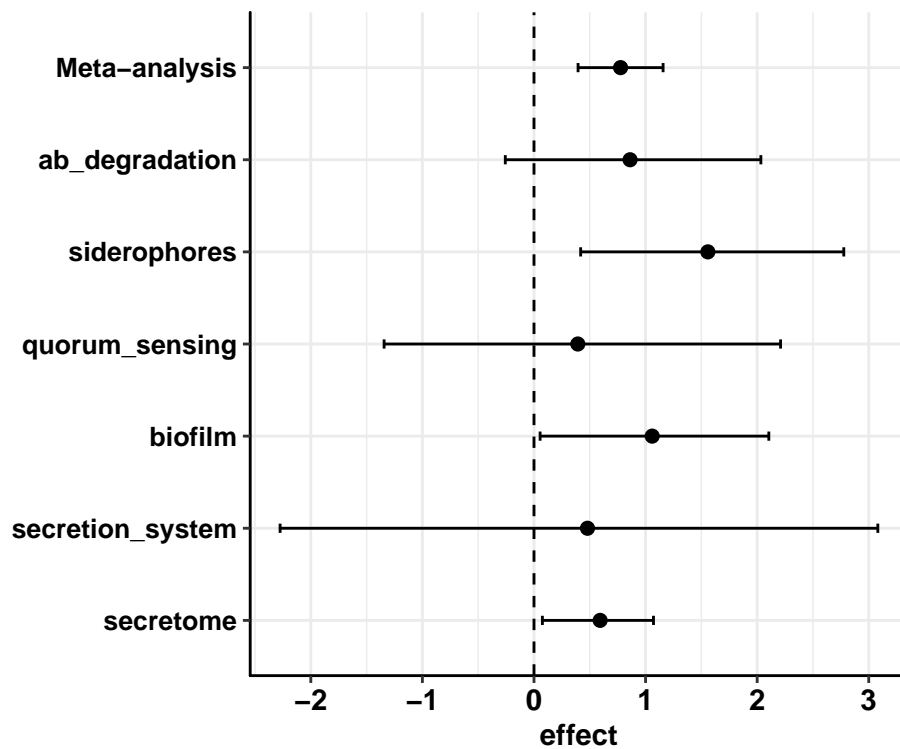
MODEL 1

Figure 1: Mean relatedness model effects coefficients, Models 1

Table 1: Mean relatedness model summary

Predictor	Response	Posterior mean	Posterior sd	HMP lower	HMP higher	pMCMC
mean_relatedness	secretome	0.593	0.255	0.075	1.071	0.019
mean_relatedness	secretion_system	0.479	1.360	-2.275	3.081	0.723
mean_relatedness	biofilm	1.059	0.526	0.055	2.104	0.046
mean_relatedness	quorum_sensing	0.393	0.905	-1.343	2.210	0.669
mean_relatedness	siderophores	1.558	0.599	0.418	2.776	0.010
mean_relatedness	ab_degradation	0.861	0.587	-0.257	2.033	0.142
log(genome size)	secretome	1.289	0.157	0.978	1.592	0.000
log(genome size)	secretion_system	1.156	0.859	-0.467	2.881	0.177
log(genome size)	biofilm	1.245	0.281	0.700	1.799	0.000
log(genome size)	quorum_sensing	0.056	0.597	-1.125	1.238	0.906
log(genome size)	siderophores	1.322	0.323	0.699	1.973	0.000
log(genome size)	ab_degradation	1.201	0.304	0.597	1.800	0.000

Table 2: Mean relatedness model Meta-analysis summary

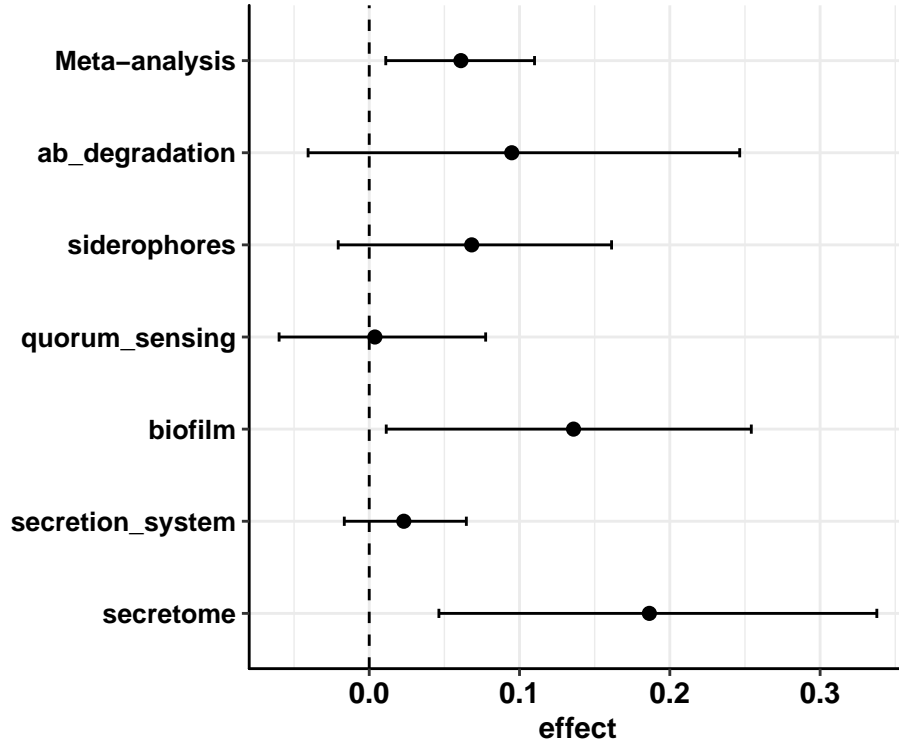| Predictor	Estimate	Std. Error	Z value	P value	CI lower	CI upper
mean_relatedness	0.776	0.194	3.992	6.566e-05	0.395	1.157
log(genome size)	1.227	0.113	10.837	2.304e-27	1.005	1.449

MODEL 2

Figure 2: Within host relatedness model effects coefficients

Table 3: Within host relatedness model summary

Predictor	Response	Posterior mean	Posterior sd	HMP lower	HMP higher	pMCMC
Within host relatedness	secretome	0.186	0.074	0.046	0.338	0.009
Within host relatedness	secretion_system	0.023	0.020	-0.017	0.065	0.236
Within host relatedness	biofilm	0.136	0.063	0.011	0.254	0.026
Within host relatedness	quorum_sensing	0.004	0.034	-0.060	0.077	0.961
Within host relatedness	siderophores	0.068	0.046	-0.021	0.161	0.117
Within host relatedness	ab_degradation	0.095	0.073	-0.041	0.247	0.155
Genome size	secretome	1.332	0.161	1.014	1.644	0.000
Genome size	secretion_system	1.326	0.802	-0.265	2.919	0.098
Genome size	biofilm	1.216	0.271	0.711	1.772	0.000
Genome size	quorum_sensing	-0.026	0.551	-1.112	1.041	1.041
Genome size	siderophores	1.196	0.325	0.573	1.834	0.000
Genome size	ab_degradation	1.123	0.284	0.571	1.676	0.000

Table 4: Within host relatedness model Meta-analysis summary

Predictor	Estimate	Std. Error	Z value	P value	CI lower	CI upper
Within host relatedness	0.061	0.025	2.386	0.017	0.011	0.11
genome size	1.206	0.113	10.701	1.009e-26	0.985	1.427

MODEL 3

Table 5: Abundance and sporulation drivers of relatedness model summary

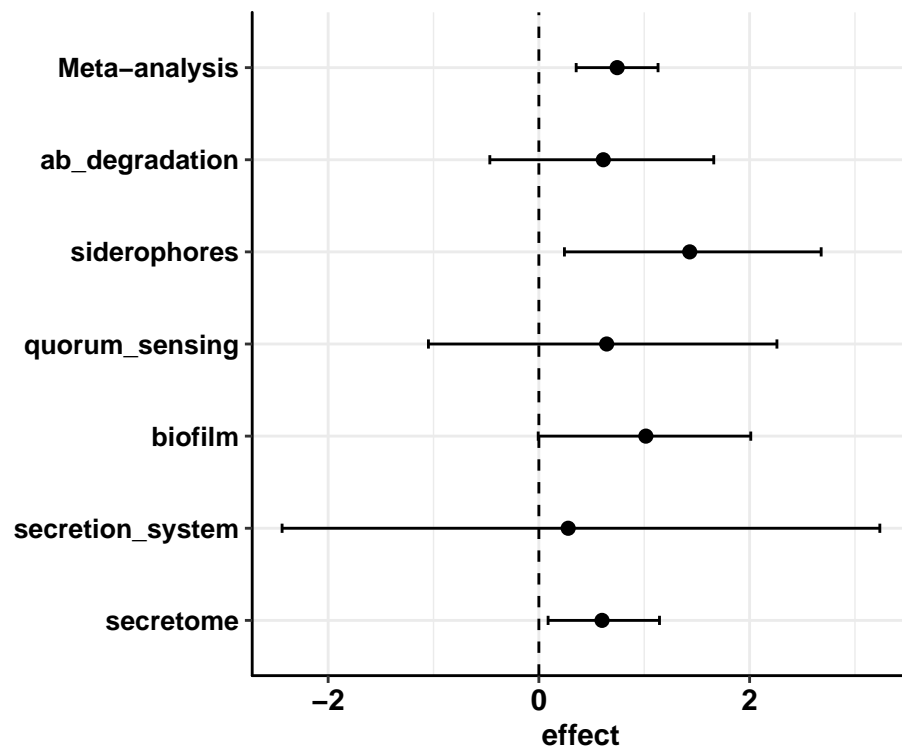| Predictor	Posterior mean	CI lower	CI upper	Effective Sampling	pMCMC
(Intercept)	8.91e-01	6.76e-01	1.11e+00	1.90e+04	5.00e-05
sporulation_score	-6.92e-01	-1.25e+00	-9.97e-02	1.60e+04	1.45e-02
within_host_relative_abundance	9.09e-02	7.95e-03	1.69e-01	2.00e+04	2.48e-02

MODEL 4

Figure 3: Mean relatedness model including relative abundance and sporulation scores in predictors, focus on relatedness effect coefficients
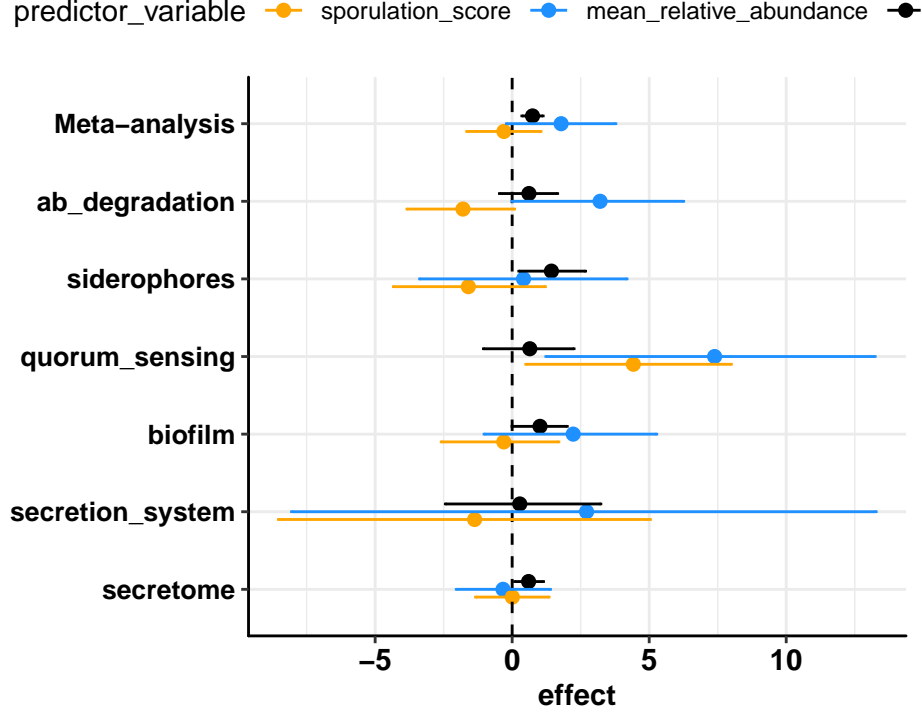
Figure 4: Mean relatedness model including relative abundance and sporulation scores in predictors, all predictors effect coefficients

Table 6: Mean relatedness model including relative abundance and sporulation scores in predictors model summary

Predictor	Response	Posterior mean	Posterior sd	HMP lower	HMP higher	pMCMC
mean_relatedness	secretome	0.599	0.268	0.087	1.145	0.030
mean_relatedness	secretion_system	0.278	1.437	-2.437	3.236	0.842
mean_relatedness	biofilm	1.015	0.520	-0.006	2.011	0.057
mean_relatedness	quorum_sensing	0.644	0.851	-1.048	2.259	0.454
mean_relatedness	siderophores	1.432	0.619	0.243	2.679	0.020
mean_relatedness	ab_degradation	0.612	0.547	-0.465	1.660	0.266
mean_relative_abundance	secretome	-0.333	0.869	-2.038	1.398	0.704
mean_relative_abundance	secretion_system	2.715	5.463	-8.050	13.282	0.600
mean_relative_abundance	biofilm	2.227	1.611	-1.026	5.274	0.172
mean_relative_abundance	quorum_sensing	7.386	3.080	1.229	13.246	0.022
mean_relative_abundance	siderophores	0.415	1.927	-3.378	4.191	0.809
mean_relative_abundance	ab_degradation	3.209	1.600	-0.008	6.256	0.059
sporulation_score	secretome	0.007	0.682	-1.336	1.347	0.988
sporulation_score	secretion_system	-1.370	3.399	-8.529	5.045	0.690
sporulation_score	biofilm	-0.306	1.093	-2.589	1.703	0.767
sporulation_score	quorum_sensing	4.422	1.897	0.498	7.998	0.026
sporulation_score	siderophores	-1.600	1.426	-4.335	1.213	0.261
sporulation_score	ab_degradation	-1.795	0.996	-3.839	0.089	0.067
log(genome size)	secretome	1.280	0.167	0.954	1.605	0.000
log(genome size)	secretion_system	1.297	0.899	-0.516	3.054	0.141
log(genome size)	biofilm	1.289	0.276	0.745	1.827	0.000
log(genome size)	quorum_sensing	-0.048	0.544	-1.114	1.049	0.936
log(genome size)	siderophores	1.401	0.328	0.780	2.071	0.000

Predictor	Response	Posterior mean	Posterior sd	HMP lower	HMP higher	pMCMC
log(genome size)	ab_degradation	1.267	0.274	0.715	1.800	0.000

Table 7: Mean relatedness model including relative abundance and
sporulation scores in predictors Meta-analysis summary

Predictor	Estimate	Std. Error	Z value	P value	CI lower	CI upper
mean_relatedness	0.743	0.198	3.748	1.779e-04	0.354	1.131
mean_relative_abundance	1.786	1.019	1.753	0.08	-0.211	3.783
sporulation_score	-0.302	0.692	-0.437	0.662	-1.659	1.054
log(genome size)	1.235	0.115	10.778	4.389e-27	1.011	1.46

MODEL 5

Table 8: Cooperation, abundance and sporulation drivers of relatedness models summary

Predictor	Posterior mean	CI lower	CI upper	Effective Sampling	pMCMC
(Intercept)	0.875	0.625	1.106	20000.000	0.000
sporulation_score	-0.648	-1.248	-0.068	20000.000	0.032
within_host_relative_abundance	0.094	0.016	0.173	20000.000	0.016
biofilm	0.001	-0.010	0.011	19513.746	0.911
ab_degradation	-0.009	-0.031	0.012	19188.288	0.405
quorum_sensing	-0.002	-0.026	0.021	20000.000	0.852
siderophores	0.006	-0.007	0.020	20000.000	0.357
secretion_system_no4	-0.002	-0.022	0.017	19269.765	0.820
nb_extracellular	-0.000	-0.002	0.002	19516.893	0.876

Table 9: Cooperation traits joint test effect (wald test)

Chi2	df	P.value
1.657	6	0.948

Gelman-Rubin Tests

```
## Loading required package: R.oo
```

```
## Loading required package: R.methodsS3
```

```
## R.methodsS3 v1.8.0 (2020-02-14 07:10:20 UTC) successfully loaded. See ?R.methodsS3 for help.
```

```
## R.oo v1.23.0 successfully loaded. See ?R.oo for help.
```

```
##
## Attaching package: 'R.oo'
```

```
## The following object is masked from 'package:R.methodsS3':
##
##     throw
```

```
## The following objects are masked from 'package:methods':
##
##     getClasses, getMethods
```

```
## The following objects are masked from 'package:base':
##
##     attach, detach, load, save
```

```
## R.utils v2.9.2 successfully loaded. See ?R.utils for help.
```

```
##
## Attaching package: 'R.utils'
```

```
## The following object is masked from 'package:tidyr':
##
##     extract
```

```
## The following object is masked from 'package:utils':
##
##     timestamp
```

```
## The following objects are masked from 'package:base':
##
##     cat, commandArgs, getOption, inherits, isOpen, nullfile, parse,
##     warnings
```

```
## [1] 1.01522
```

```
## [1] 1.029783
```

```
## [1] 1.116164
```

```
## [1] 1.040133
```

```
## [1] 1.016071
```

```
## [1] 1.020529
```

```
## [1] 1.002462
```

```
## [1] 1.000274
```

Reporting here the results of Gelman-Rubin tests. The values are the **maximum observed upper 95% CI** across all fixed and vcv effects for each model.

Table 10: Gelman-Rubin tests of all model.

model	specific.model	fixed.effects.max.upperCI.psrf	vcv.effects.max.upperCI.psrf
model 1	secretome	1.003796	1.001838
model 1	ab_degradation	1.010254	1.003200
model 1	biofilm	1.004855	1.029783
model 1	secretion_system_no4	1.010967	1.001228
model 1	siderophores	1.006954	1.008429
model 1	quorum_sensing	1.015220	1.023403
model 2	secretome	1.006366	1.001174
model 2	ab_degradation	1.009777	1.023040
model 2	biofilm	1.009605	1.005523
model 2	secretion_system_no4	1.025943	1.026677
model 2	siderophores	1.023825	1.005984
model 2	quorum_sensing	1.116164	1.040133
model 3	model_3	1.000885	1.004699
model 4	secretome	1.006870	1.002116
model 4	ab_degradation	1.008485	1.006878
model 4	biofilm	1.010880	1.003122
model 4	secretion_system_no4	1.016071	1.020530
model 4	siderophores	1.012970	1.005877
model 4	quorum_sensing	1.012820	1.005341
model 5	model_5	1.002462	1.000274

Models formulations

The "response of interest" term is one of the cooperative trait. The code replaces whichever cooperative trait the model is ran over by this term to have a general dataframe to be used in a code wrapper function.

Model 1

```
## $fixed
## [1] "response_of_interest ~ mean_relatedness + log(nb_cds_not_involved_in_response) + gram_profile"
##
## $random
## [1] "~ species"
##
## $resid
## [1] "~ units"
```

Model 2

```
## $fixed
## [1] "cbind(response_of_interest, within_host_relatedness) ~ -1 + at.level(trait, 1):at.level(first, \
##
## $random
## [1] "~ us(at.level(first, \"TRUE\"):at.level(trait, 1) + at.level(trait, 2)):species + idh(at.level(
##
## $resid
## [1] "~ idh(at.level(first, \"TRUE\"):at.level(trait, 1)):species.ide + idh(at.level(trait, 2)):units
```

Model 3

```
## $fixed
## [1] "within_host_relatedness ~ 1 + sporulation_score + within_host_relative_abundance"
##
## $random
## [1] "~ species.ide + host + species"
##
## $resid
## [1] "~ units"
```

Model 4

```
## $fixed
## [1] "response_of_interest ~ mean_relatedness + mean_relative_abundance + sporulation_score + log(nb_c
##
## $random
## [1] "~ species"
##
## $resid
## [1] "~ units"
```

Model 5

```
## $fixed
## [1] "within_host_relatedness ~ 1 + sporulation_score + within_host_relative_abundance + biofilm + ab_
##
## $random
## [1] "~ species.ide + host + species"
##
## $resid
## [1] "~ units"
```