

Assessing the influence of the amount of reachable habitat on
genetic structure using landscape and genetic graphs
Supplementary information

Savary, Paul^{*1, 2, 3}, Foltête, Jean-Christophe², van Strien, Maarten J.⁴, Moal, Hervé¹,
Vuidel, Gilles², and Garnier, Stéphane³

¹*ARP-Astrance, 9 Avenue Percier, 75008 Paris, France*

²*ThéMA, UMR 6049 CNRS, Université Bourgogne-Franche-Comté, 32 Rue Mégevand, 25030 Besançon
Cedex, France*

³*Biogéosciences, UMR 6282 CNRS, Université Bourgogne-Franche-Comté, 6 Boulevard Gabriel 21000
Dijon, France*

⁴*Planning of Landscape and Urban Systems, Institute for Spatial and Landscape Planning, ETH Zurich,
Zurich, Switzerland*

Contents

1	Supplementary information 1	2
2	Supplementary information 2	8

*Corresponding author: paul.savary@univ-fcomte.fr

1 Supplementary information 1

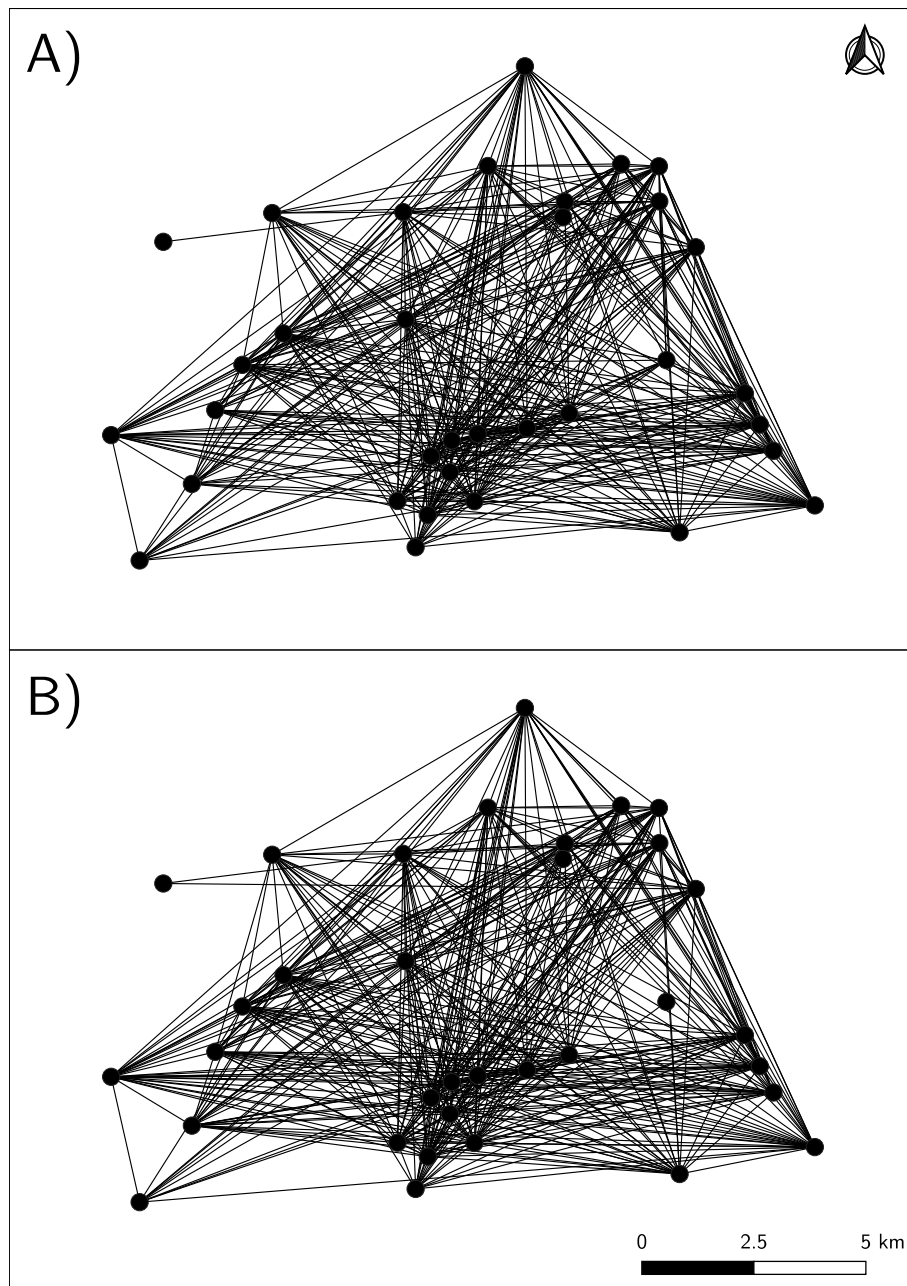


Figure S1: Comparisons of the topology of the genetic graphs pruned using the percolation threshold computed with the D_{PS} (A) or F_{ST} (B). Both graphs include 412 links and 34 nodes.

Table S1: Pearson correlation coefficients between habitat metrics and genetic indices according to the cost scenario used, the MDD considered and the weight given to patch capacities in the metric calculation (β value). The largest correlation coefficient obtained for each genetic index, habitat metric and β value are displayed. The 'Signif.' column indicates whether the correlation is still significant after p -value adjustment (*: $p < 0.05$, **: $p < 0.01$, ***: $p < 0.001$). For the cost scenarios, refer to table 1 in the main document. 'DistNN' means 'Distance to the Nearest Neighbour'.

Genetic index	Habitat metric		MDD	β	Correlation	
	Metric	Cost sc.			$r_{Pearson}$	Signif.
AR	Capacity	w50-R1000	1500	\emptyset	0.426	
AR	Local.Buffer	Euc.	200	\emptyset	0.339	
AR	DistNN	w1000-R1000	\emptyset	\emptyset	0.125	
AR	F	w50-R1000	2000	0	0.480	
AR	F	w50-R1000	2000	1	0.400	
AR	Large.Buffer	w50-R1000	1500	\emptyset	0.440	
AR	BC	w1000-R1000	7500	0	0.324	
AR	BC	w1000-R1000	7500	1	0.325	
Priv. AR	Capacity	w1000-R1000	1500	\emptyset	0.263	
Priv. AR	Local.Buffer	w1000-R50	500	\emptyset	0.284	
Priv. AR	DistNN	Euc.	\emptyset	\emptyset	-0.223	
Priv. AR	F	Euc.	1500	0	-0.507	
Priv. AR	F	Euc.	1500	1	-0.484	
Priv. AR	Large.Buffer	w50-R1000	2500	\emptyset	0.448	
Priv. AR	BC	Euc.	4500	0	-0.380	
Priv. AR	BC	w1000-R50	1500	1	0.397	
MIW _{comp.DPS}	Capacity	w50-R1000	1500	\emptyset	0.196	
MIW _{comp.DPS}	Local.Buffer	Euc.	200	\emptyset	0.211	
MIW _{comp.DPS}	DistNN	w1000-R1000	\emptyset	\emptyset	0.167	
MIW _{comp.DPS}	F	Euc.	3500	0	0.466	
MIW _{comp.DPS}	F	Euc.	3500	1	0.383	
MIW _{comp.DPS}	Large.Buffer	w50-R1000	3500	\emptyset	-0.347	
MIW _{comp.DPS}	BC	Euc.	3500	0	0.351	
MIW _{comp.DPS}	BC	Euc.	1500	1	0.291	
MIW _{prun.DPS}	Capacity	w50-R1000	1500	\emptyset	0.144	
MIW _{prun.DPS}	Local.Buffer	Euc.	300	\emptyset	0.190	
MIW _{prun.DPS}	DistNN	w1000-R1000	\emptyset	\emptyset	0.156	
MIW _{prun.DPS}	F	Euc.	3500	0	0.614	*
MIW _{prun.DPS}	F	Euc.	3000	1	0.544	
MIW _{prun.DPS}	Large.Buffer	w50-R50	4500	\emptyset	-0.393	
MIW _{prun.DPS}	BC	Euc.	3500	0	0.467	
MIW _{prun.DPS}	BC	Euc.	1500	1	0.408	
MIW _{comp.FST}	Capacity	w50-R50	1500	\emptyset	0.224	
MIW _{comp.FST}	Local.Buffer	w50-R1000	500	\emptyset	0.189	
MIW _{comp.FST}	DistNN	Euc.	\emptyset	\emptyset	-0.198	
MIW _{comp.FST}	F	w50-R50	2500	0	0.508	
MIW _{comp.FST}	F	w50-R50	2000	1	0.527	
MIW _{comp.FST}	Large.Buffer	w50-R50	1500	\emptyset	0.264	
MIW _{comp.FST}	BC	w1000-R1000	1500	0	0.475	
MIW _{comp.FST}	BC	w1000-R1000	1500	1	0.622	*
MIW _{prun.FST}	Capacity	w50-R50	1500	\emptyset	0.224	
MIW _{prun.FST}	Local.Buffer	w50-R1000	500	\emptyset	0.187	
MIW _{prun.FST}	DistNN	Euc.	\emptyset	\emptyset	-0.199	
MIW _{prun.FST}	F	w50-R50	2500	0	0.509	
MIW _{prun.FST}	F	w50-R50	2000	1	0.529	
MIW _{prun.FST}	Large.Buffer	w50-R50	1500	\emptyset	0.264	
MIW _{prun.FST}	BC	w1000-R1000	1500	0	0.475	
MIW _{prun.FST}	BC	w1000-R1000	1500	1	0.623	*

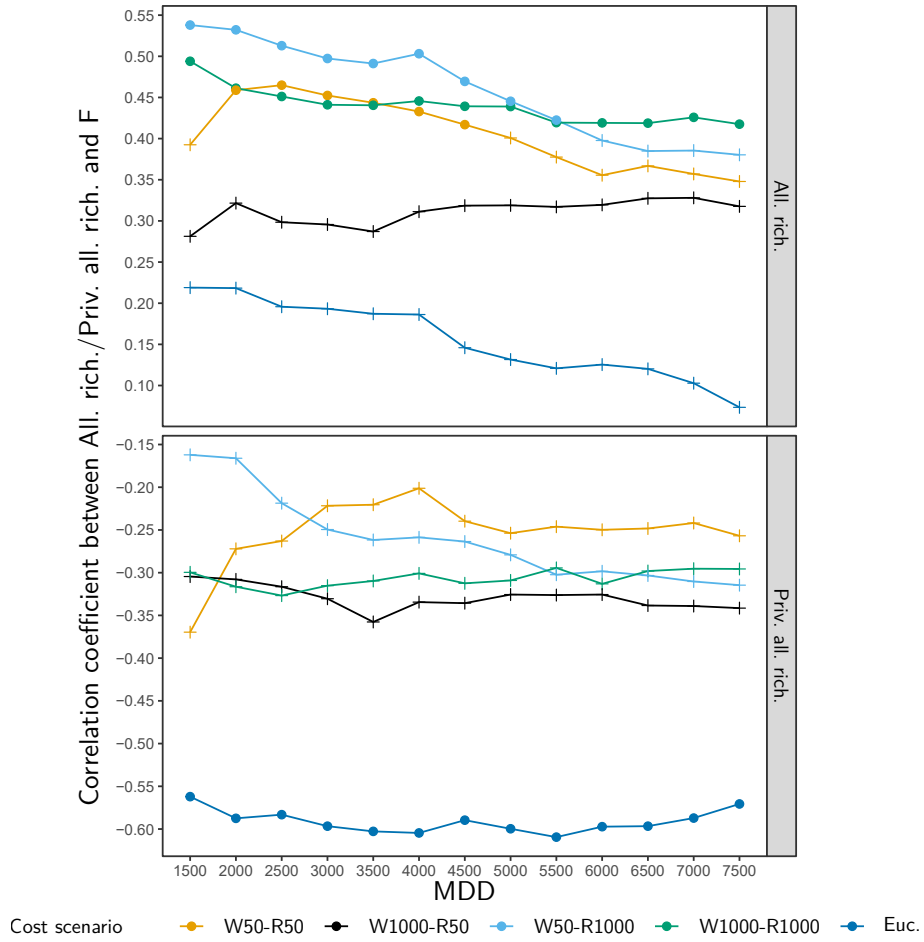


Figure S2: Variation of the Spearman correlation coefficients between total allelic richness (top) or private allelic richness (bottom) and the F metric according to the cost scenarios and MDD used to compute these indices. x axis indicates the dispersal kernels used to compute the metrics and corresponds to the MDD (maximum dispersal distance). In this figure, the F metric was computed without weighting patch capacities ($\beta = 0$). Point colours refer to the cost scenario used to compute cost-distances (see Table 1). Crosses indicate that the correlation is not significant after p -value adjustment.

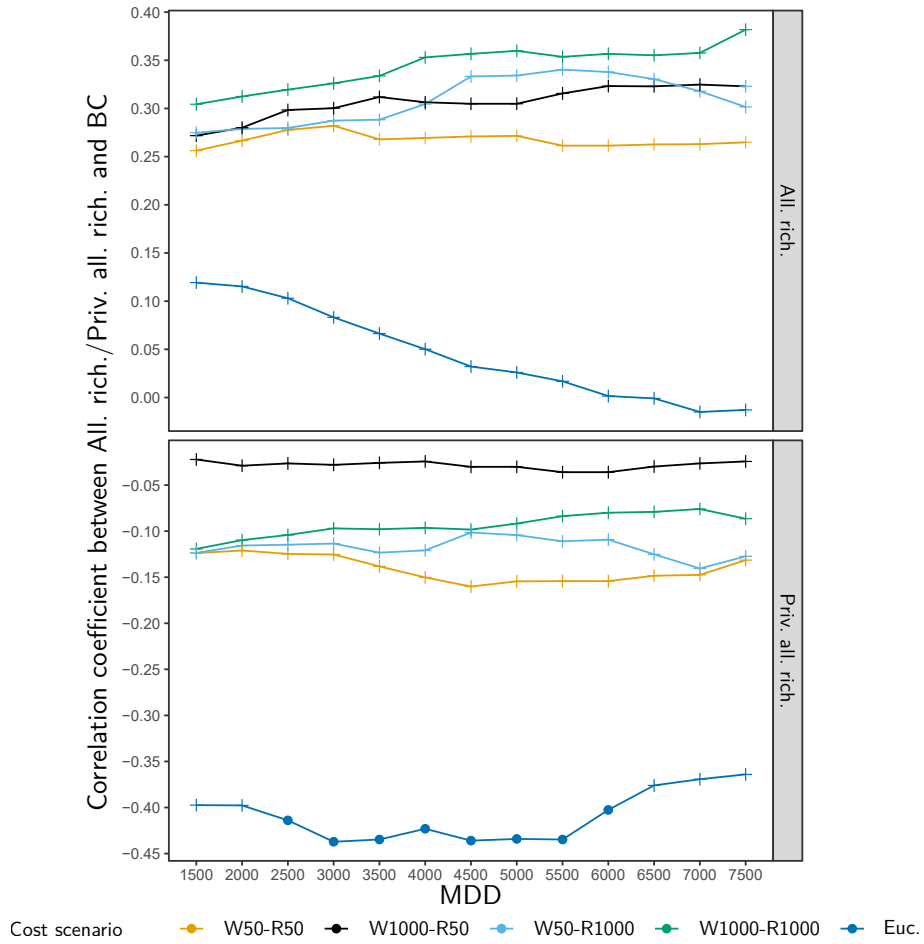


Figure S3: Variation of the Spearman correlation coefficients between total allelic richness (top) or private allelic richness (bottom) and the BC metric according to the cost scenarios and MDD used to compute these indices. x axis indicates the dispersal kernels used to compute the metrics and corresponds to the MDD (maximum dispersal distance). In this figure, the BC metric was computed without weighting patch capacities ($\beta = 0$). Point colours refer to the cost scenario used to compute cost-distances (see Table 1 in the main document). Crosses indicate that the correlation is not significant after p -value adjustment.

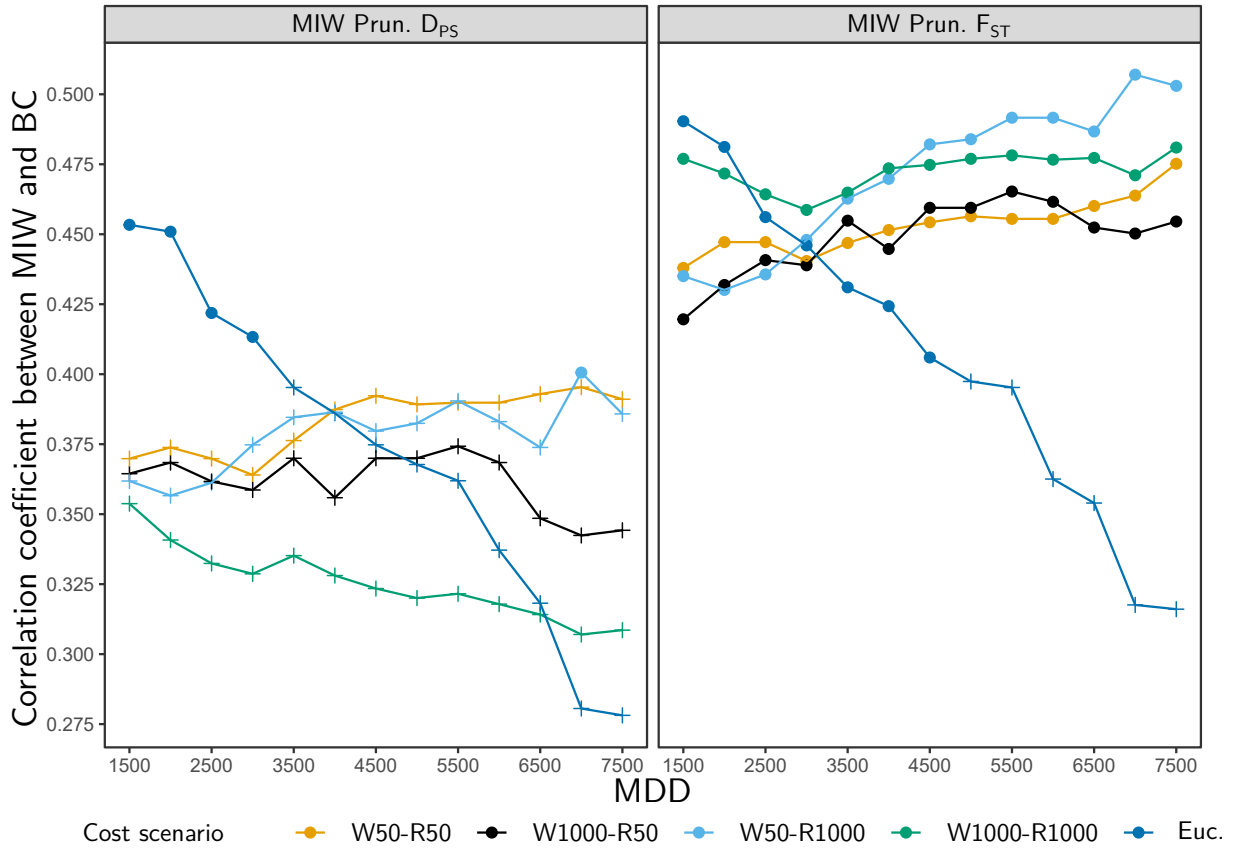


Figure S4: Variation of the Spearman correlation coefficients between the MIW index computed from the pruned genetic graphs and the BC metric according to the genetic distance, cost scenarios and MDD used to compute these indices. x axis indicates the dispersal kernels used to compute the metrics and corresponds to the MDD (maximum dispersal distance). In this figure, the BC metric was computed without weighting patch capacities ($\beta = 0$). Point colours refer to the cost scenario used to compute cost-distances (see Table 1 in the main document). The left and right panels display the variations observed when computing MIW from a genetic graph weighted with D_{PS} and F_{ST} values respectively. Crosses indicate that the correlation is not significant after p -value adjustment.

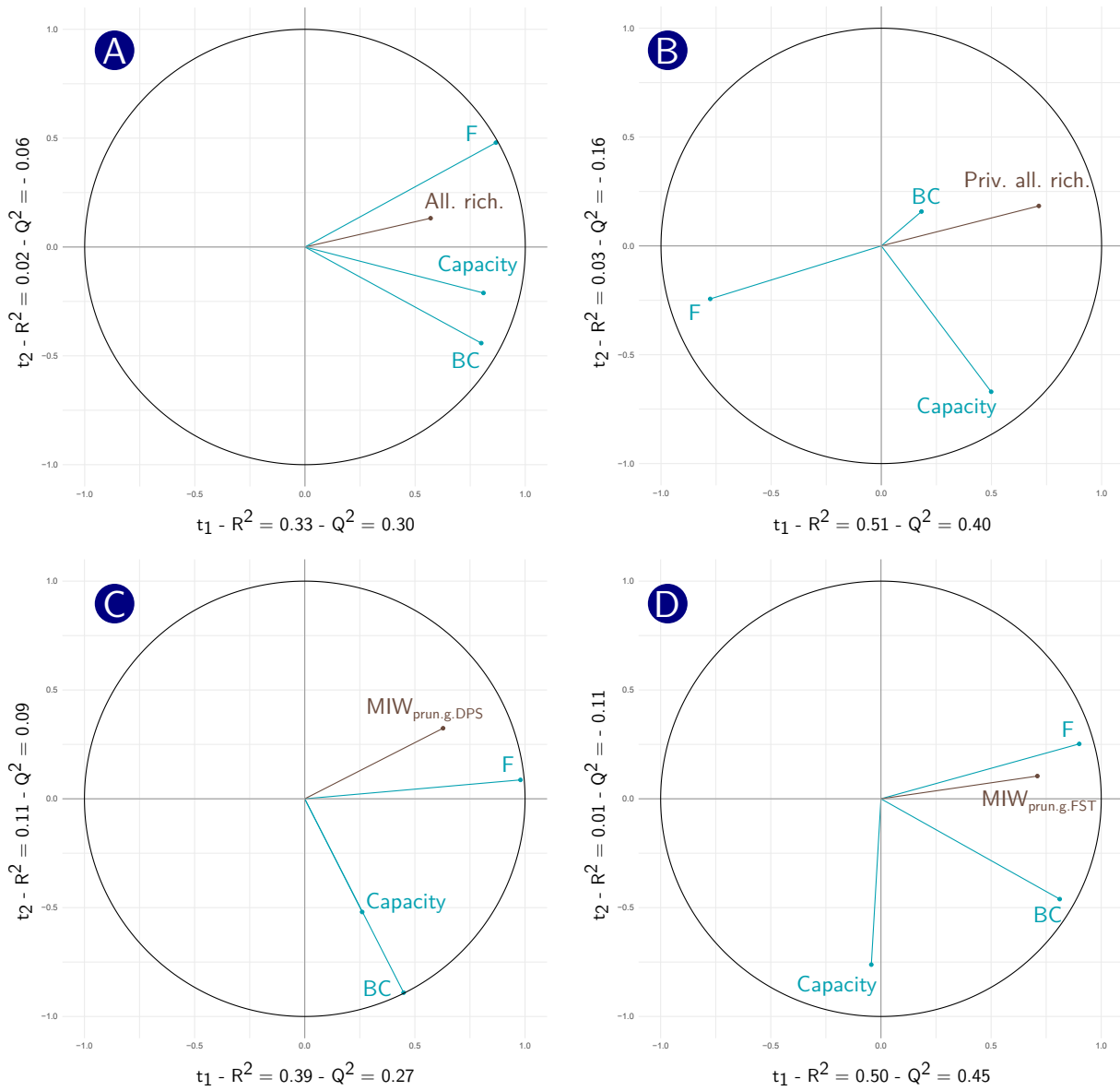


Figure S5: Projection of both the response variable (genetic indices) and the predictor variables (habitat metrics: capacity, F, BC) of the best PLS-R1 regression for each genetic index (according to the Q^2 value) on the obtained factorial space. (A) Response variable: allelic richness. Predictor variables: Capacity computed under the cost scenario w50-R1000, F computed under the cost scenario w50-R1000 with a MDD of 2000 m and $\beta = 0$, BC computed under the cost scenario w1000-R1000 with a MDD of 7500 m and $\beta = 1$ (B) Response variable: private allelic richness. Predictor variables: Capacity computed under the cost scenario w50-R1000, F computed under the Euclidean cost scenario with a MDD of 2500 m and $\beta = 0$, BC computed under the cost scenario w1000-R1000 with a MDD of 6000 m and $\beta = 1$ (C) Response variable: $MIW_{prun.g.DPS}$. Predictor variables: Capacity computed under the cost scenario w50-R1000, F computed under the Euclidean cost scenario with a MDD of 3000 m and $\beta = 0$, BC computed under the cost scenario w1000-R1000 with a MDD of 7500 m and $\beta = 1$ (D) Response variable: $MIW_{prun.g.FST}$. Predictor variables: Capacity computed under the cost scenario w1000-R50, F computed under the cost scenario w50-R1000 with a MDD of 2500 m and $\beta = 0$, BC computed under the Euclidean cost scenario with a MDD of 3000 m and $\beta = 1$

2 Supplementary information 2

Rationale behind the use of the Q2 to analyse PLS regression results

When performing a PLS-R1 regression, we express the response variable y as a linear combination of H components t_1, t_2, \dots, t_H such that:

$$y = c_1 t_1 + c_2 t_2 + \dots + c_H t_H$$

where c_1, c_2, \dots, c_H are regression coefficients and t_1, t_2, \dots, t_H are components obtained such that:

$$t_1 = w_{11}x_1 + w_{12}x_2 + \dots + w_{1p}x_p$$

and

$$w_{1j} = \frac{\text{cov}(x_j, y)}{\sqrt{\sum_{j=1}^p \text{cov}^2(x_j, y)}}$$

Therefore, t_1, t_2, \dots, t_H components are also linear combinations of the predictor variables x_1, x_2, \dots, x_p . From that, we can express y as a function of x_1, x_2, \dots, x_p :

$$\begin{aligned} y = & c_1 w_{11}x_1 + c_1 w_{12}x_2 + \dots + c_1 w_{1p}x_p + \\ & c_2 w_{21}x_1 + c_2 w_{22}x_2 + \dots + c_2 w_{2p}x_p + \\ & \dots \\ & c_H w_{H1}x_1 + c_H w_{H2}x_2 + \dots + c_H w_{Hp}x_p \end{aligned} \tag{1}$$

The number H of components to compute is determined through a cross-validation. For each value of h , a model with h components is computed, either from all the observations or leaving one (Leave One Out cross Validation, LOOV) or a block of observations (k -fold cross validation) out. From these models, predicted values of y are computed, either \hat{y}_{hi} , the prediction of y_i from the model with h components calibrated from all the observations, or $y_{h(\hat{-i})}$, the prediction of y_i from the model with h components calibrated from a subset of the observations in which observation i is absent. Two criteria are then computed to assess the goodness of fit of these models:

$$RSS_h = \sum (y_i - \hat{y}_{hi})^2$$

and

$$PRESS_h = \sum (y_i - y_{h(\hat{-i})})^2$$

which are respectively referred to as the Residual Sum of Squares (RSS) and Prediction Error Sum of Squares (PRESS). Adding a component is relevant if:

$$\sqrt{PRESS_h} \leq 0.95 \sqrt{RSS_{h-1}}$$

which means that when adding another component the prediction error is lower than 90.25 % of the residual sum of squares without adding this component:

$$PRESS_h \leq 0.9025 RSS_{h-1}$$

Then,

$$\frac{PRESS_h}{RSS_{h-1}} \leq 0.9025$$

and

$$1 - \frac{PRESS_h}{RSS_{h-1}} \geq 0.0975$$

Accordingly, the criterion Q^2 is equal to:

$$Q^2 = 1 - \frac{PRESS_h}{RSS_{h-1}}$$

The value of Q^2 is computed for every component h of the models. A component is considered as having a significant effect in the model if it improves the prediction of y , and therefore if $Q^2 > 0.0975$ ([Tenenhaus 1998](#)).

References

Tenenhaus, M. (1998). *La régression PLS: théorie et pratique*. Editions TECHNIP.