Supplement of

Geodiversity primarily shapes large-scale limnology and aquatic species distribution in the northern Neotropics

Laura Macario-González et al.

Correspondence to: Sergio Cohuo (sergio.cd@chetumal.tecnm.mx

S1. Structural Equation Modelling, models performance and results.

The Structural Equation Modelling technique was used to evaluate the direct and indirect influence of geodiversity and related environmental and limnological variables on ostracode species composition. Given regional factors such as the highly variable topography in Central America, diverse geological history and environmental variability within regions (latitude), mainly related to precipitation and temperature, we assume that elevation gradients, bedrock and latitude were primary factors determining biological composition in aquatic systems. These three factors were then used as exogenous variables and we also test the performance of geodiversity (considered as the sum of topography and geological history) as exogenous variable. Limnology was considered as an endogenous variable as water chemistry and physical parameters of lakes, are strongly influenced by geodiversity. The effect of vulcanism, evaporation/precipitation and marine-freshwater interactions was also considered as explanatory variables, by using the presence of conservative and specific major anion and cations.

To evaluate the direct and indirect contribution of each variable to biological composition, we analyze five different SEM models. Models were fitted with a covariance data matrix with a set of uncorrelated variables. Model performances were evaluated with the Chi-square test, the Root Mean Square Error (RMSE), comparative fit index (CFI) and standardized root mean squared residuals (srmr).

Model 1, 2 and 3: The first three models tested the simplified relationships between geodiversity, limnology, and species composition. On all these models, geodiversity and related variables were considered as exogenous variables, but all models differ in the predictors used for both exogenous and endogenous variables. Model 1, uses bedrock, elevation, and TOC as predictors of geodiversity as they received high scores in the PCA. Limnology (as endogenous variable) uses as indicators conductivity, pH and selected major ions such as HCO_{3-}. This model did not converge and therefore did not find a statistical solution. Model 2 was then constructed with a reduced number of predictors. Geodiversity was constructed only with elevation and latitude as predictors, whereas limnology only with conductivity. The selection of these variables resulted from the fact that elevation was directly related with water temperature in lakes and latitude with presence of carbonates given reduction in precipitation and increase evaporation which is typical in the highlands of Central America. This model received acceptable scores but was considered too simplified. Model 3 was a variant of the model 2 but incorporates again variables identified in PCA as most influential for limnological regions. For instance, geodiversity was constructed with elevation and carbonates as in Yucatán Peninsula carbonates dominates mineral composition on lake sediments. This model shows poor fit of data, Chi-square test reject the model (highly significant Chi-square) and some paths were not significant.

Model 4. This model was constructed on the basis of the model 2 and 3 with respect of predictors of geodiversity and limnology. On this model, we tested the direct effect of geodiversity on species composition. Furthermore, we evaluate the individual influence of altitude and conductivity for species composition. These two variables were selected considering that PCA suggests major influence on limnological regionalization in the study area. This model had a better fit than model 3, but as Chi square was highly significant, the model was rejected.

Model 5. This model shows the best performance and was the reported in the main text of this study. This model was constructed using geodiversity as unique exogenous and latent variable, constructed out from mineralogical components of sediments such as carbonates and TOC. Limnology and altitude were considered as endogenous variables. Furthermore, the individual influence of geodiversity, altitude and conductivity over biological composition were tested. The global fit resulted in CFI = 0.961, RMSEA = 0.01 and SRMR=0.036, Chi-square = 2.6, d.f.= 5. P-value = 0.757.

Model 5, optimal model full output:

```
lavaan 0.6-9 ended normally after 94 iterations

  Estimator                                         ML
  Optimization method                           NLMINB
  Number of model parameters                        16

  Number of observations                            66

Model Test User Model:

  Test statistic                                 2.628
  Degrees of freedom                                 5
  P-value (Chi-square)                           0.757

Model Test Baseline Model:

  Test statistic                                77.377
  Degrees of freedom                                15
  P-value                                        0.000

User Model versus Baseline Model:

  Comparative Fit Index (CFI)                    0.961
  Tucker-Lewis Index (TLI)                       0.978

Loglikelihood and Information Criteria:

  Loglikelihood user model (H0)               -325.353
  Loglikelihood unrestricted model (H1)       -324.040

  Akaike (AIC)                                 682.707
  Bayesian (BIC)                               717.741
  Sample-size adjusted Bayesian (BIC)          667.370

Root Mean Square Error of Approximation:

  RMSEA                                          0.010
  90 Percent confidence interval - lower         0.000
  90 Percent confidence interval - upper         0.119
  P-value RMSEA <= 0.05                          0.815

Standardized Root Mean Square Residual:
```

```
  SRMR                                           0.036

Parameter Estimates:

  Standard errors                             Standard
  Information                                 Expected
  Information saturated (h1) model          Structured

Latent Variables:
                        Estimate  Std.Err  z-value  P(>|z|)   Std.lv
Std.all
  geology =~
    Carbonates             1.000                               1.372
0.642
    TOC                    0.027    0.049    0.549    0.583    0.037
0.074
  limnology =~
    cond                   1.000                               0.399
0.824
    HCO3                   0.448    0.109    4.099    0.000    0.179
0.586
  speciescomposition =~
    diversity              1.000                               0.356
1.000
  Altitud =~
    Altitude               1.000                               0.871
1.000

Regressions:
                       Estimate  Std.Err  z-value  P(>|z|)   Std.lv  Std.all
  limnology ~
    geology               0.306    0.085    3.590    0.000    1.053    1.053
  Altitud ~
    geology              -0.366    0.093   -3.934    0.000   -0.577   -0.577
  speciescomposition ~
    limnology            -1.575    4.499   -0.350    0.726   -1.765   -1.765

Covariances:
                        Estimate  Std.Err  z-value  P(>|z|)   Std.lv
Std.all
  geology ~~
   .speciescompstn         0.970    2.596    0.374    0.709    0.865
0.865
 .speciescomposition ~~
   .Altitud               -0.080    0.041   -1.945    0.052   -0.138   -
0.138
 .cond ~~
   .diversity             -0.032    0.026   -1.205    0.228   -0.032     -
Inf

Variances:
                   Estimate  Std.Err  z-value  P(>|z|)   Std.lv  Std.all
   .Carbonates        2.685    0.644    4.172    0.000    2.685    0.588
   .TOC               0.249    0.043    5.739    0.000    0.249    0.995
   .cond              0.075    0.033    2.261    0.024    0.075    0.321
   .HCO3              0.061    0.012    4.969    0.000    0.061    0.656
   .diversity         0.000                               0.000    0.000
   .Altitude          0.000                               0.000    0.000
    geology           1.882    0.781    2.411    0.016    1.000    1.000
   .limnology        -0.017    0.047   -0.372    0.710   -0.110   -0.110
   .speciescompstn    0.668    2.735    0.244    0.807    5.265    5.265
   .Altitud           0.506    0.106    4.762    0.000    0.667    0.667

R-Square:
                   Estimate
    Carbonates        0.412
    TOC               0.005
    cond              0.679
    HCO3              0.344
    diversity         1.000
    Altitude          1.000
    limnology            NA
    speciescompstn   -4.265
    Altitud           0.333
```

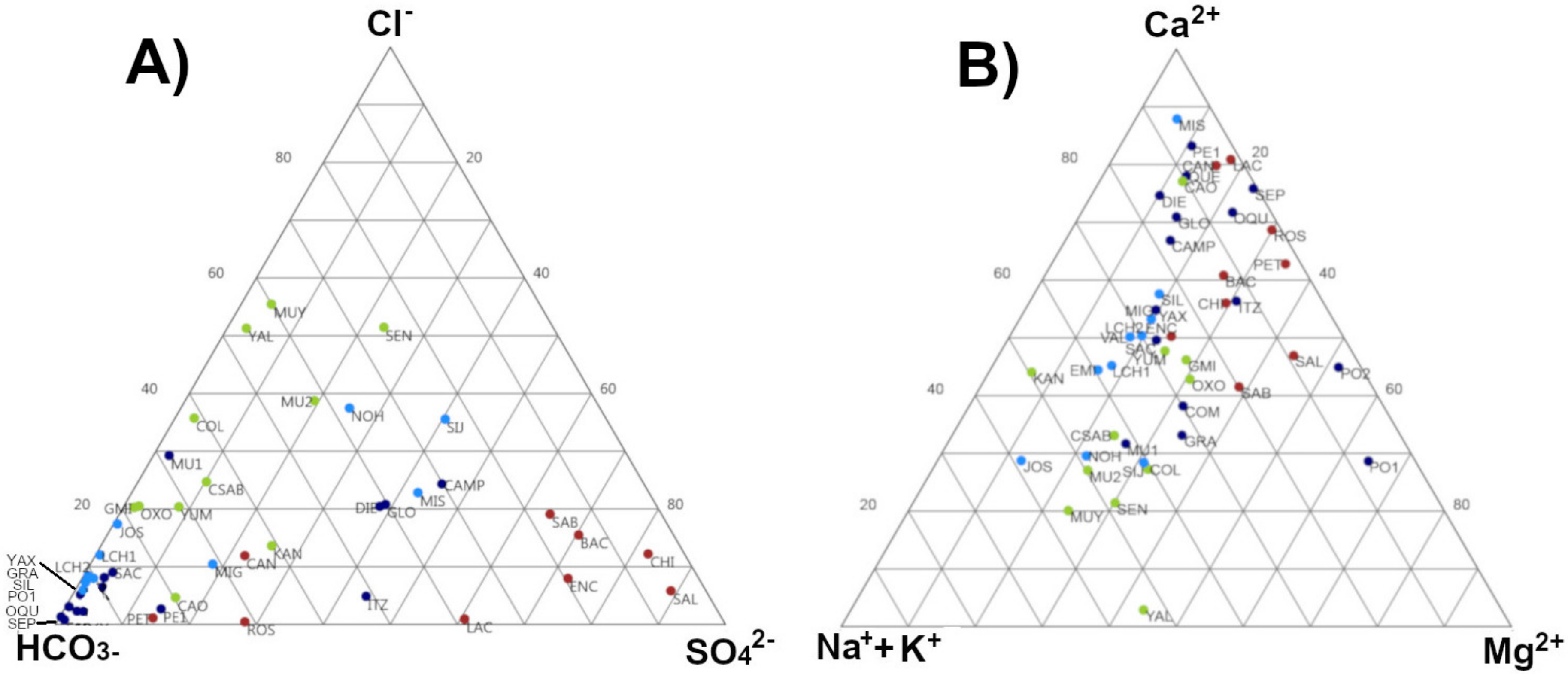

Figure S1. Ternary plots showing major cations and anion proportions [%] of 76 aquatic systems of the northern Neotropical region. Major anions (A) and major cations (B) from YG limnological group; Abbreviations correspond to those in Table S1, colors representing limnological subregions, correspond to the cluster analysis dendrogram (Fig. 2).

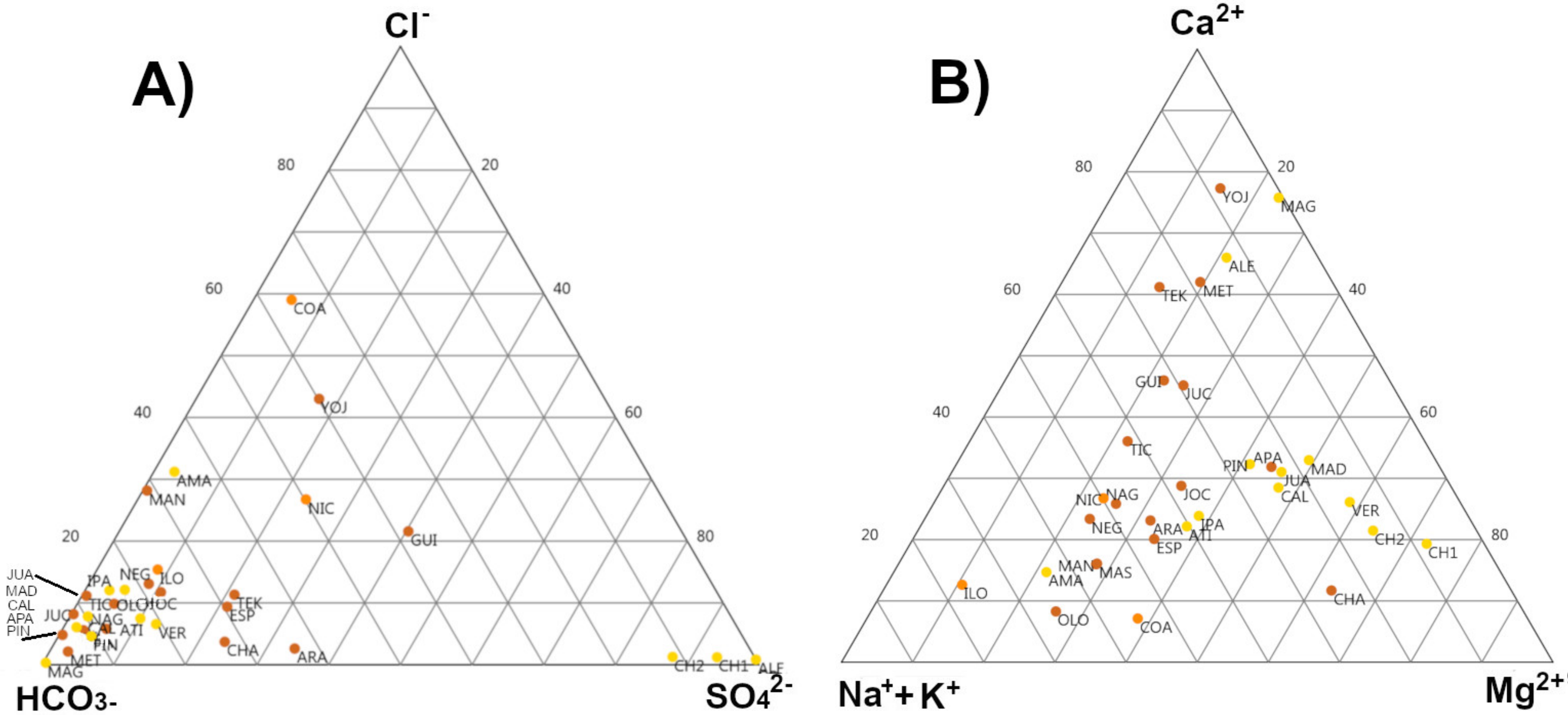

Figure S2. Ternary plots showing major cations and anion proportions [%] of 76 aquatic systems of the northern Neotropical region. A) major anion from GSHN limnological group; d) major cation from GSHN limnological group. Abbreviations correspond to those in Table S1 and colors, representing limnological subregions, correspond to the cluster analysis dendrogram (Fig.

Table S2.1. Loading values of the 13 variables used for principal component analysis of the YG region ($P < 0.05$). Significant scores for components 1 and 2 in bold.

Variable	PC1 (23.47 %)	PC2 (14.63 %)
Temperature	0.02	0.35
DO	0.06	0.45
pH	0.01	**0.73**
Conductivity	**0.77**	0.04
HCO_3^-	0.13	0.05
Cl^-	**0.64**	0.06
Na+	**0.72**	0.04
Ca	0.51	0.11
Mg2+	**0.65**	0.04
TOC	0.10	0.13
Altitude	0.40	0.12
SO_4^{2-}	0.47	0.12
Age	0.43	0.47

Table S2.2. Loading values of the 13 variables used for principal component analysis of the GSHN limnological region ($P < 0.05$). Significant scores for components 1 and 2 in bold.

Variable	PC1 (26.83%)	PC2 (22.89%)
Conductivity	0.57	0
HCO_3^-	**0.70**	0.11
Cl^-	**0.62**	0.08
Na+	**0.72**	0.07
Ca	0.33	0.15
Mg2+	0.48	0.05
TN	0.01	0.62
TOC	0.01	**0.79**
Phyllosilicates	0.11	0.20
Altitude	0.11	0.09
CO_3^-	**0.56**	0.17
Bedrock	0.12	**0.75**
Age	0.23	**0.80**

Table S3. List of ostracode species found in our study.

#	Species name	Species code	#	Species name	Species code
1	*Alicenula serricaudata* (Klie, 1935a)	ASE	36	*Hemicypris* sp.1	*Hemicypris* sp.
2	*Alicenula yucatanensis* sp. nov.	AYU	37	*Heterocypris nicaraguensis* Hartmann 1959	HNI
3	*Candona* sp. 1	*Candona* sp.	38	*Heterocypris punctata* (Keyser, 1975)	HPU
4	*Cypretta campechensis* Cohuo-Durán *et al*., 2013	CCA	39	*Keysercypria* sp. 1	*Keysercypria* sp. 1
5	*Cypretta* cf. *campechensis* Cohuo-Durán et al., 2013	CCAf	40	*Keysercypria* sp. 2	*Keysercypria* sp. 2
6	*Chlamydotheca* cf. *colombiensis* Roessler 1985	CCOf	41	*Keysercypria* sp. 3	*Keysercypria* sp. 3
7	*Cypretta elongata* sp. nov.	CEL	42	*Keysercypria* sp. 4	*Keysercypria* sp. 4
8	*Cypretta* cf. *elongata*	CELf	43	*Keysercypria* sp. 5	*Keysercypria* sp. 5
9	*Cypria gibbera* Furtos, 1936a	CGI	44	*Keysercypria granadae* (Hartmann 1959)	KGR
10	*Chlamydotheca* sp. 1	*Chlamydotheca* sp.	45	*Limnocythere floridensis* Keyser, 1975	LFL
11	*Cytheridella ilosvayi* Daday, 1905	CIL	46	*Limnocythere* sp. 1	*Limnocythere* sp.
12	*Cypretta maya* Cohuo-Durán *et al*., 2013	CMA	47	*Limnocytherina royi* Hartmann 1959	LRO
13	*Cypria petenensis* Ferguson et al., 1964	CPE	48	*Limnocythere* cf. *stationis* Vávra, 1891	LST
14	*Cypretta spinosa* Cohuo-Durán *et al*., 2013	CSP	49	*Neocypridopsi*s sp. 1	*Neocypridopsis* sp.
15	*Cyprinotus unispinifera* Furtos, 1936b	CUN	50	*Pseudocandona antilliana* Broodbakker, 1983c	PAN
16	*Chlamydotheca unispinosa* (Baird, 1862)	CUNI	51	*Peryssocytheridea* cf. *cribrosa* Klie, 1933a	PCRf
17	*Cypridopsis* sp. [Ca 1]	*CVI sp 1*	52	*Penthasilenula* sp. 1	*Penthasilenula* sp.
18	*Cypridopsis* sp. [Ca 2]	*CVI sp 2*	53	*Potamocypris islagrandensis* Hoff 1943b	PIS
19	*Cypridopsis* sp. [Ca 3]	*CVI sp 3*	54	*Pericythere marginata* Hartmann, 1959	PMA
20	*Cypridopsis* sp. 4	*CVI sp 4*	55	*Paracythereis opesta* (Brehm, 1939)	POP
21	*Cypridopsis* sp. 5	*CVI sp 5*	56	*Potamocypris* sp.1	*Potamocypris* sp. 1
22	*Cypridopsis* sp. 6	*CVI sp 6*	57	*Potamocypris* sp. 2	*Potamocypris* sp. 2
23	*Cypridopsis* sp. 7	*CVI sp 7*	58	*Potamocypris* sp. 3	*Potamocypris* sp. 3
24	Cypricercinae sp. 1	CYP 1	59	*Pseudocandona* sp. 1	*Pseudocandona* sp. 1
25	Cypricercinae sp. 2	CYP 2	60	*Pseudocandona* sp. 2	*Pseudocandona* sp. 2
26	Cypricercinae sp. 3	CYP 3	61	*Pseudostrandesia* sp. 1	*Pseudostrandesia* sp. 1
27	*Cyprinotinae* sp. 1	CYP 4	62	*Pseudostrandesia* sp. 2	*Pseudostrandesia* sp. 2
28	*Cypria* sp. 1	*Cypria* sp. 1	63	*Strandesia bicuspis* (Claus 1892)	SBI
29	*Cypria* sp. 4	*Cypria* sp. 4	64	*Stenocypris cylindrical major* (Baird, 1859b)	SCY
30	*Cypria* sp. 5	*Cypria* sp. 5	65	*Strandesia intrepida* Furtos, 1936b	SIN
31	*Cyprideis* cf. *salebrosa*	*Cyprideis* sp.	66	*Strandesia* sp. 1	*Strandesia* sp.
32	Cyprididae sp. [Ca 1]	Cyprididae sp. 1	67	*Tanycypris* sp. 1	*Tanycypris* sp.
33	*Desc. 1*	Cyprididae sp.2	68	*Thalassocypria* sp.1	*Thalassocypria* sp.
34	*Diaphanocypris meridana* (Furtos, 1936b)	DME	69	*Vestalenula* sp.1	*Vestalenula* sp.
35	*Darwinula stevensoni* (Brady & Robertson, 1885)	DST	70	*Vestalenula pagliolii* (Pinto and Kotzian 1961)	VPA

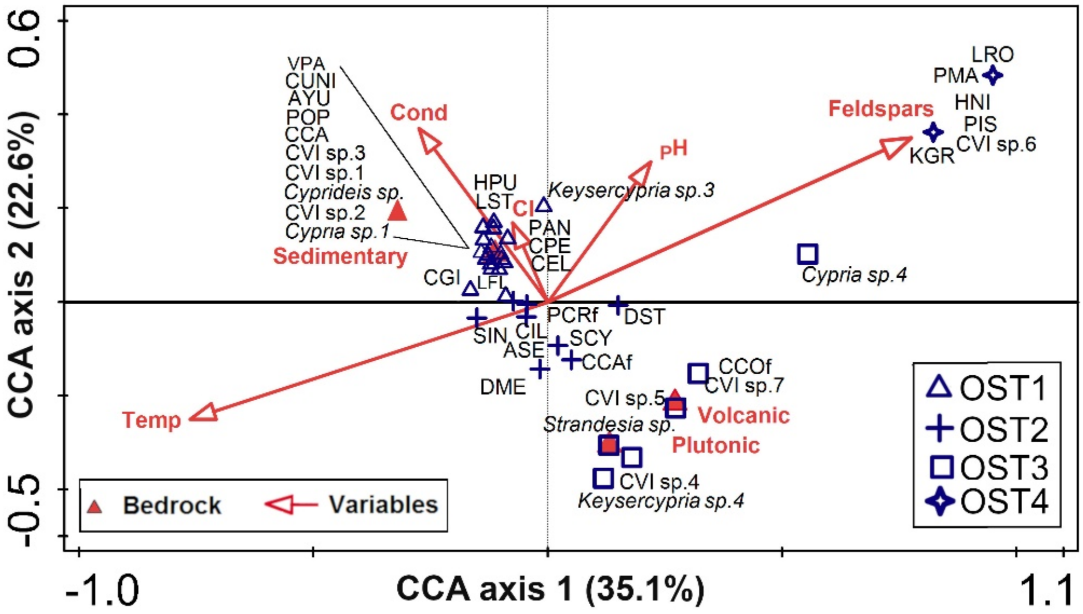

Figure S3. Canonical Correspondence Analysis (CCA) biplot for axis 1 and 2 showing the relationship between 39 ostracode species and eight forward selected explanatory variables (Monte Carlo p=0.001). Blue symbols represent species belonging to ostracode groups identified by NMDS analysis. Filled triangles represent categorical bedrock type (Sedimentary, Volcanic, Plutonic) and red arrows are showing numerical variables used on the analysis. The first two axes explain 57.7% of the total variance. For species name abbreviations see Table S3. Abbreviations for variables are as follows: Temperature (Temp), conductivity (cond), chloride (Cl^-).

Table S4.1 Forward selection results of explanatory variables and percentage of variance in the species data explained by each variable (P <0.01).

Variable	% variance in species data	*P*-value (999 permutations)
Sedimentary*	7.9	0.001
Feldspars	6.7	0.001
Cl^-	3.7	0.001
Temperature	3.8	0.001
Volcanic*	2.8	0.008
Plutonic*	2.8	0.008
pH	1.9	0.044
Conductivity	1.4	0.284

*Bedrock type

Table S4.2 Summary of results from canonical correspondence analysis based on 39 ostracode species and eight forward-selected explanatory variables.

Axes	1	2	3	4	Total inertia
Eigenvalues	0.68	0.44	0.34	0.22	6.91
Species-environment correlations	0.93	0.83	0.8	0.72	
Cumulative percentage variance of species data	9.91	16.28	21.21	24.45	
Cumulative percentage of species-environment relation	35.17	57.79	75.28	86.79	
Sum all unconstrained eigenvalues					6.91
Sum all canonical eigenvalues					1.94

DCA values = 6.8 SD (First axis) and 4.9 SD (Second axis).