Principal Component Analysis (PCA)

Discovering the Multiverse

Dr. Erola Fenollosa

erola.fenollosa@gmail.com

@Erola_Fenollosa

www.erolafenollosa.weebly.com

YouTube

Environmental Data Scientist

Aims of the session

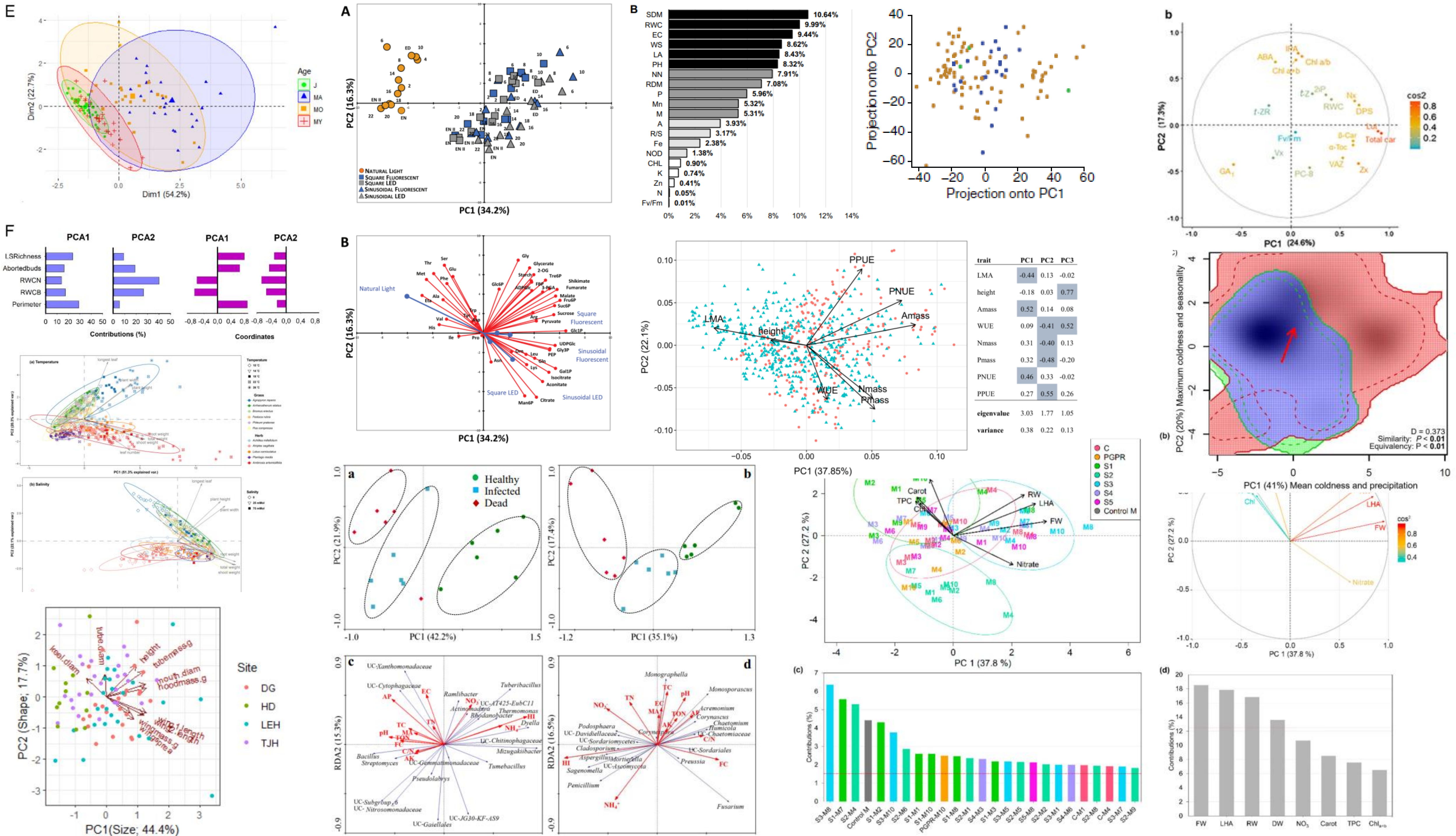
1) Understand PCA in scientific **articles**

2) Recognise different **applications** of PCA

3) Identify **what is needed** to build and report a PCA and when is not appropriated to use it

4) Be conscient of the **limitations** of PCA through its mechanics

EXTRA: **Build** your own PCA



E
Age
J
MA
MO
MY
Dim1 (54.2%)
Dim2 (22.7%)
A
PC1 (34.2%)
PC2 (16.3%)
Natural Light
Square Fluorescent
Square LED
Sinusoidal Fluorescent
Sinusoidal LED
B
SDM 10.64%
RWC 9.99%
EC 9.44%
WS 8.62%
LA 8.43%
PH 8.32%
NN 7.91%
RDM 7.08%
P 5.96%
Mn 5.32%
M 5.31%
A 3.93%
R/S 3.17%
Fe 2.38%
NOD 1.38%
CHL 0.90%
K 0.74%
Zn 0.41%
N 0.05%
Fv/Fm 0.01%
Projection onto PC1
Projection onto PC2
b
PC1 (24.6%)
PC2 (17.3%)
cos2
ABA
IAA
Chl a/b
Chl a+b
t-ZR
t-Z
2iP
RWC
Nx
DPS
Lut
Total car
Fv/Fm
β-Car
α-Toc
Vx
VAZ
Zx
GA1
PC-8
F
PCA1
PCA2
LSRichness
Abortedbuds
RWCN
RWCB
Perimeter
Contributions (%)
Coordinates
B
Natural Light
Square Fluorescent
Sinusoidal Fluorescent
Square LED
Sinusoidal LED
PC1 (34.2%)
PC2 (16.3%)
PC1 (37.85%)
PC2 (22.1%)
PPUE
PNUE
Amass
LMA
height
WUE
Nmass
Pmass
trait PC1 PC2 PC3
LMA -0.44 0.13 -0.02
height -0.18 0.03 0.77
Amass 0.52 0.14 0.08
WUE 0.09 -0.41 0.52
Nmass 0.31 -0.40 0.13
Pmass 0.32 -0.48 -0.20
PNUE 0.46 0.33 -0.02
PPUE 0.27 0.55 0.26
eigenvalue 3.03 1.77 1.05
variance 0.38 0.22 0.13
PC1 (41%) Mean coldness and precipitation
PC2 (20%) Maximum coldness and seasonality
D = 0.373
Similarity: P < 0.01
Equivalency: P < 0.01
a
Healthy
Infected
Dead
PC1 (42.2%)
PC2 (21.9%)
b
PC1 (35.1%)
PC2 (17.4%)
c
RDA2 (15.3%)
d
RDA2 (16.5%)
C
PGPR
S1
S2
S3
S4
S5
Control M
PC 1 (37.8 %)
PC 2 (27.2 %)
RW
LHA
FW
Nitrate
Carot
TPC
Chl
(c)
Contributions (%)
(d)
FW
LHA
RW
DW
NO3
Carot
TPC
Site
DG
HD
LEH
TJH
PC1(Size; 44.4%)
PC2 (Shape; 17.7%)
keel.diam
tube.diam
height
tubemass.g
mouth.diam
hoodmass.g
wing1.length
wing2.length
wingmass.g
wingspread

Principal component Analysis (PCA) is a linear dimensionality reduction technique with applications in exploratory data analysis, visualization and data preprocessing.

When is it useful to use a PCA?

- Reduce dimensionality → You have multiple variables with trade-offs
- Contrast sites, genotypes, conditions multidimensionaly: understand what makes groups diferent
- Detect outliers within groups
- Contrast groups variance (*i.e.* Plasticity)
- Estimate level of similarity by overlap

The data

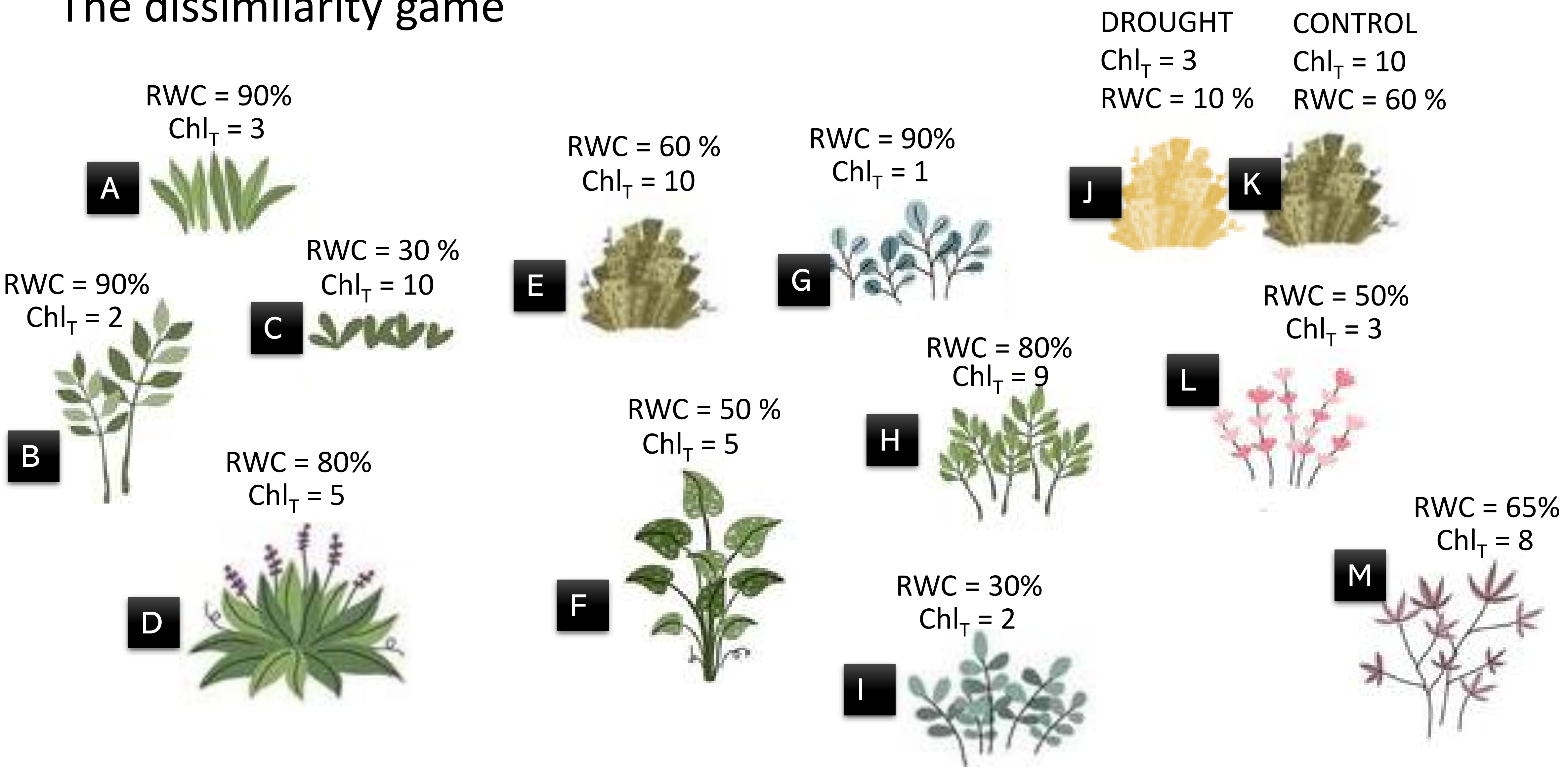
This a whole topic itself, but ideally your data should:

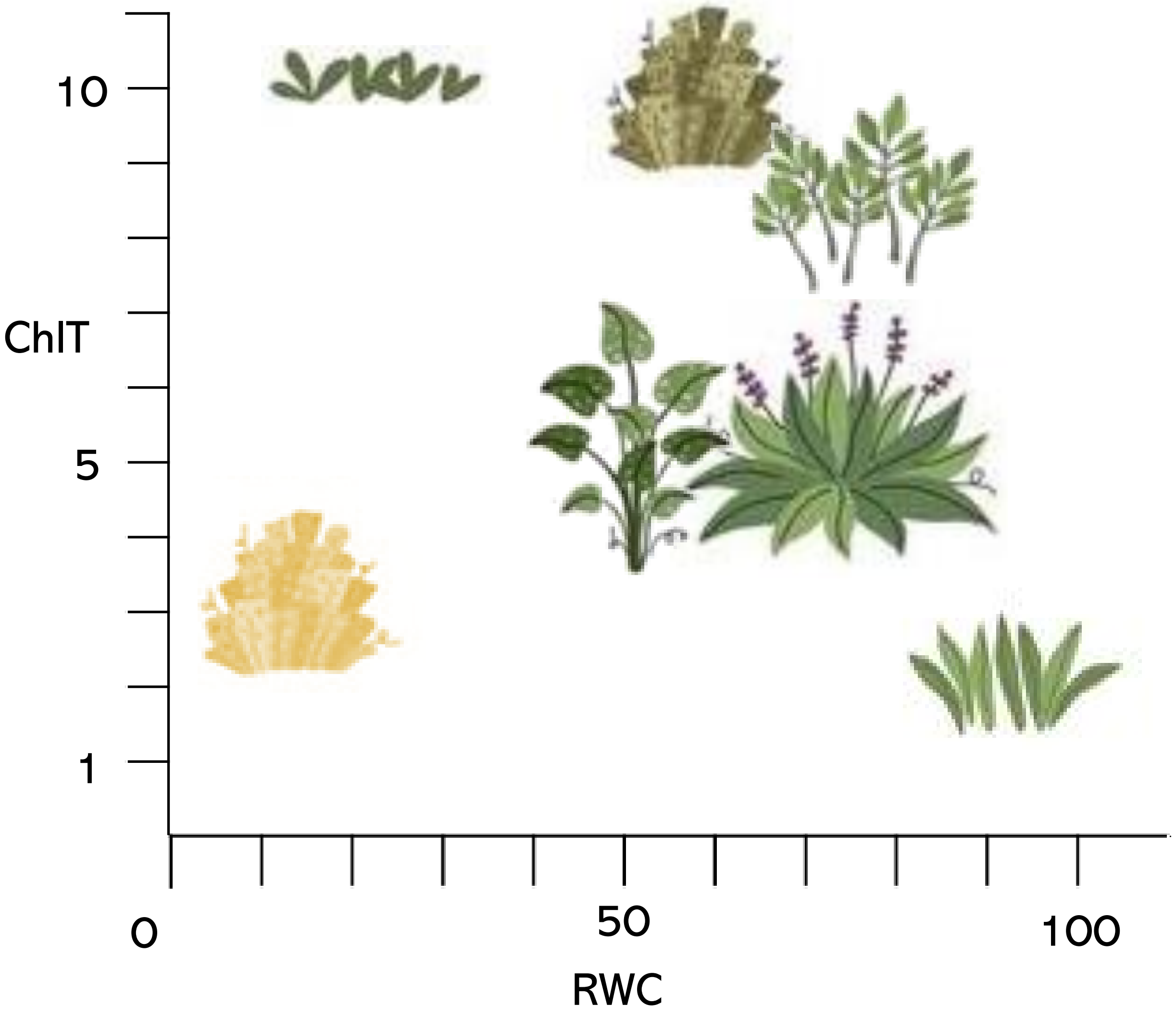
- Not contain missing values
- Contain more than two numerical variables
- Can contain one or more factors or grouping variables
- Data does not contain predictor categorical variables (if so, different analysis should be used)
- Be paired: each row represents the same biological replicate

	A	B	C	D	E	F	G	H	I
1	season	Species	leafangle	leafTemp	fvfm	h	area	lma	looh
2	estiu	AC	80	25.4	0.758	11.3333	2.201	66.7878	20.6107
3	estiu	AC	90	26.2	0.761	11.8317	3.066	67.8408	20.5299
4	estiu	AC	130	26.6	0.733	9.58427	1.193	74.6018	29.1841
5	estiu	AC	110	27	0.677	11.9268	1.333	92.2731	40.1276
6	estiu	AC	130	25.3	0.779	15.2143	4.102	75.0853	4.63958
7	estiu	AC	90	24.9	0.738	11.9167	1.321	81.7562	19.0654
8	estiu	AC	60	26.8	0.747	9.5	1.779	80.9444	40.9843
9	estiu	AC	120	28.9	0.756	7.32792	4.259	72.3174	39.7237
10	estiu	ACs	150	29.2	0.742	8.86486	1.764	104.875	28.6699
11	estiu	ACs	130	52.1	0.766	16.1442	0.942	110.403	18.4774
12	estiu	ACs	130	35.5	0.786	16.8007	2.823	104.853	16.1958
13	estiu	ACs	90	46.2	0.655	14.4479	3.9994	105.516	7.76234
14	estiu	ACs	160	27	0.69	14.4771	2.353	111.347	11.813
15	estiu	ACs	130	45	0.781	10.7487	1.549	123.305	42.9233
16	estiu	ACs	100	46.9	0.768	13.3691	2.852	104.488	12.0324
17	estiu	ACs	170	42.2	0.647	8.24082	4.242	115.512	15.8204
18	estiu	CM	170	34.6	0.799	6.64053	3.259	207.426	27.946
19	estiu	CM	150	38.4	0.749	5.98673	2.682	252.796	21.5128
20	estiu	CM	180	24.2	0.754	7.22546	2.6	250.769	28.0027
21	estiu	CM	160	23.6	0.8	6.53992	3.014	261.778	26.7143
22	estiu	CM	170	39.4	0.85	6.32127	2.096	210.878	24.1357
[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]

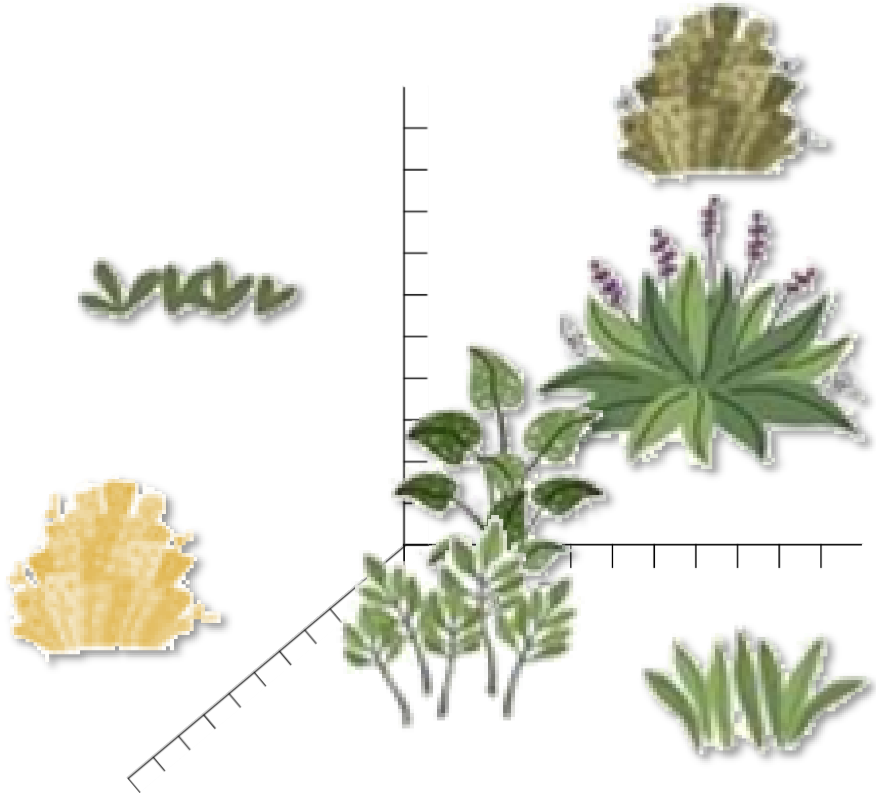
The dissimilarity concept. Everything is relative!

The dissimilarity game





And now… α-Toc

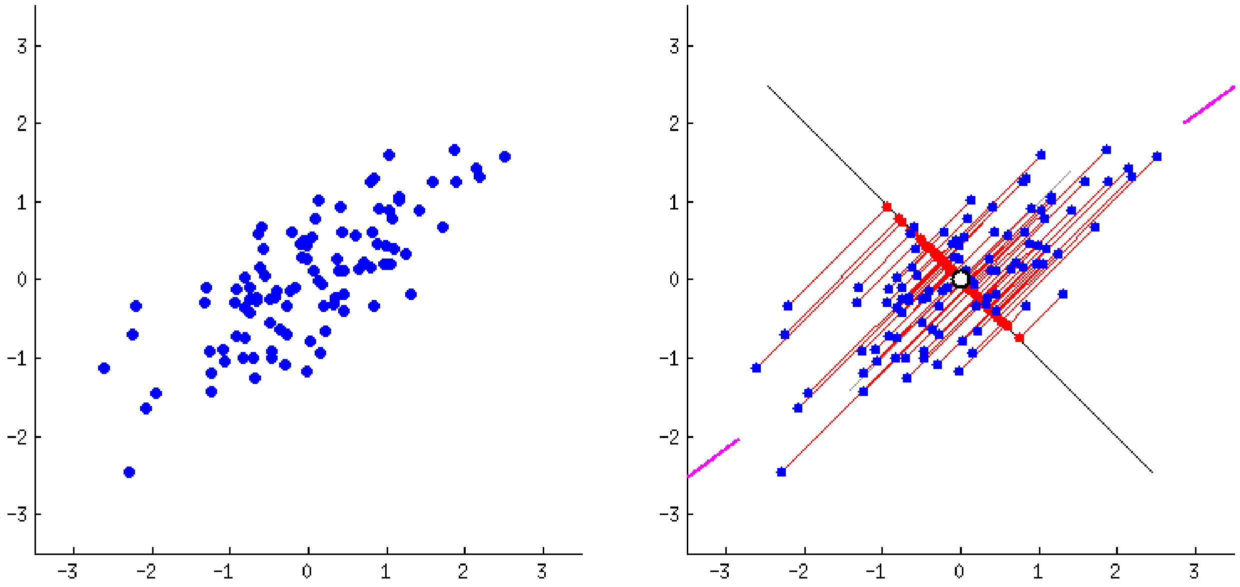


Take an object. If you had to describe the object in just two dimensions how would you do it? How would you take a picure of it so the observer can identify it?

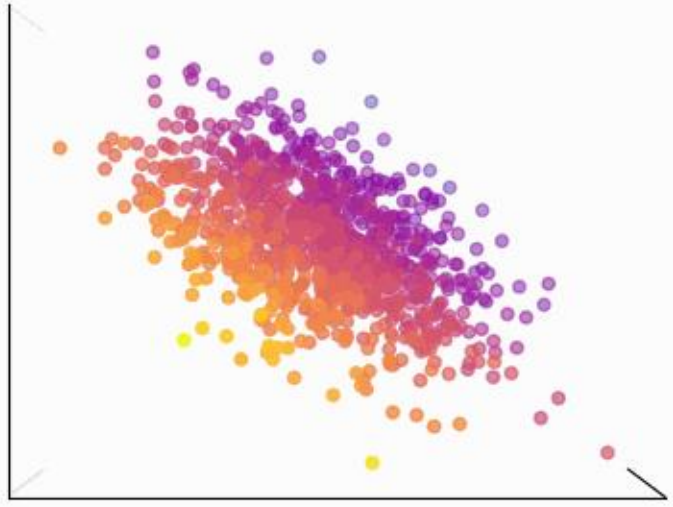
Take a picure and discuss with the person on your left

Nº of components = Nº of variables

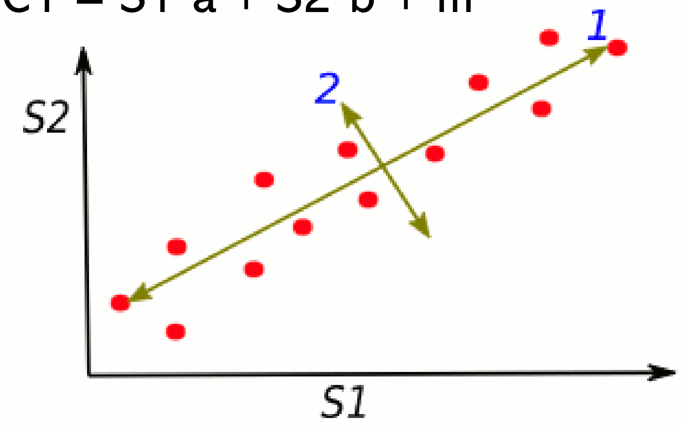
Dimensions reduction



Multiple axis of variation



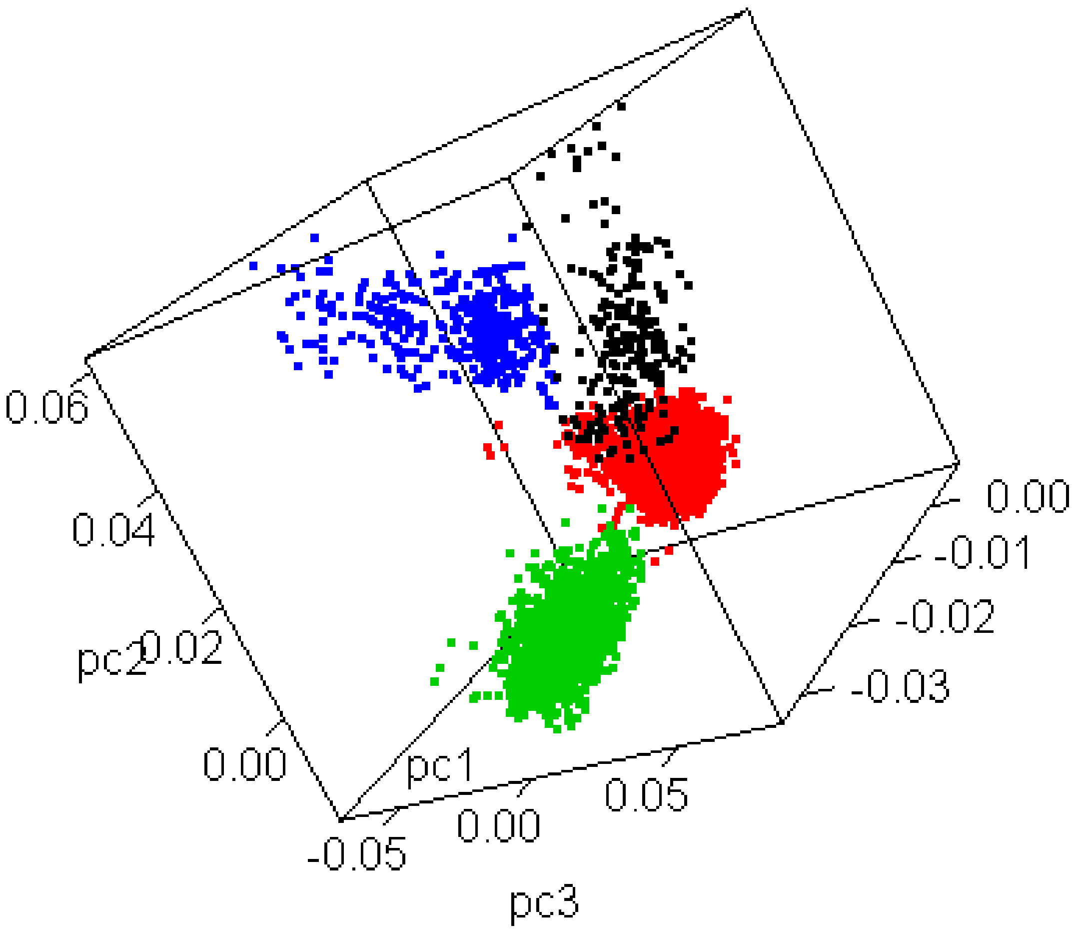
PC1 = S1*a + S2*b + m



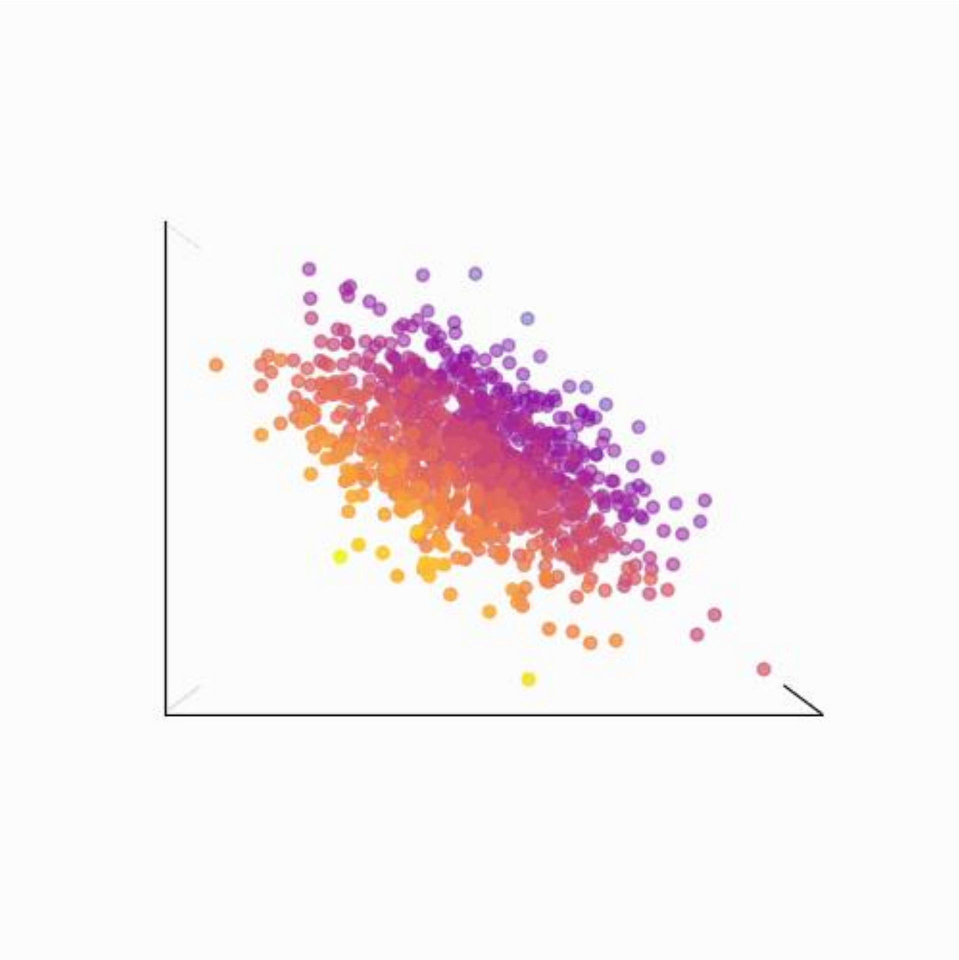
The data is linearly transformed onto a new coordinate system such that the directions (principal components) capturing the largest variation in the data can be easily identified.

Which PC captures higher variation?

You can find grups considering multiple variables



But... What is the deal? The hidden dimensions

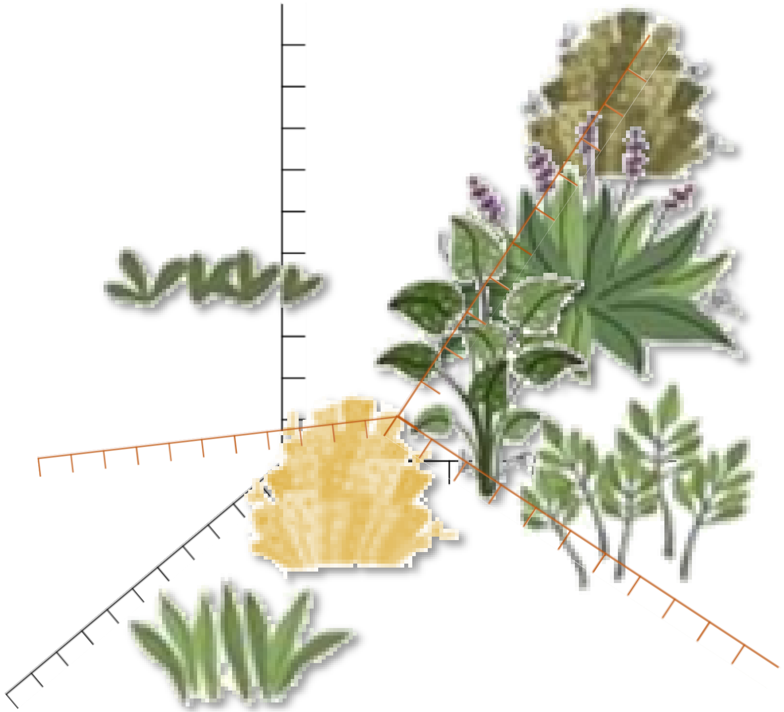


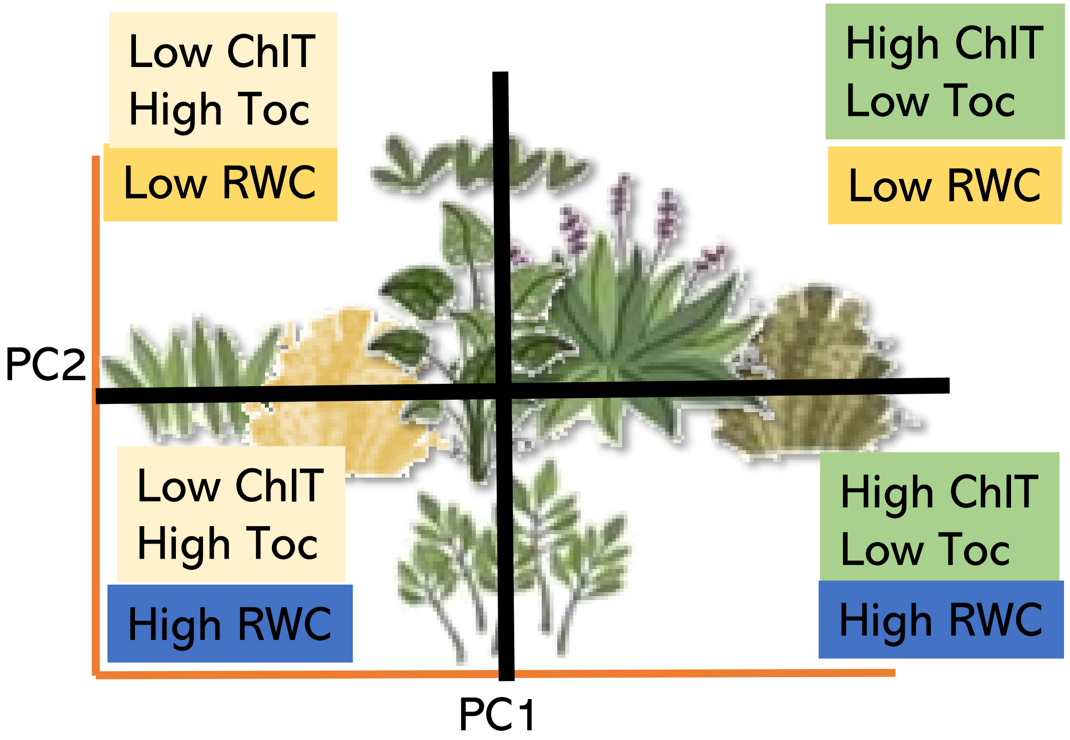
Always report the percentatge of variance explain by each component

What does PC1 and PC2 mean?

They are just new variables constituted from old variables, we could say they are just formulas. For exemple, in the ChlT, RWC and a-Toc exemple, PC1 could be something like:

Which species are drought tolerant?



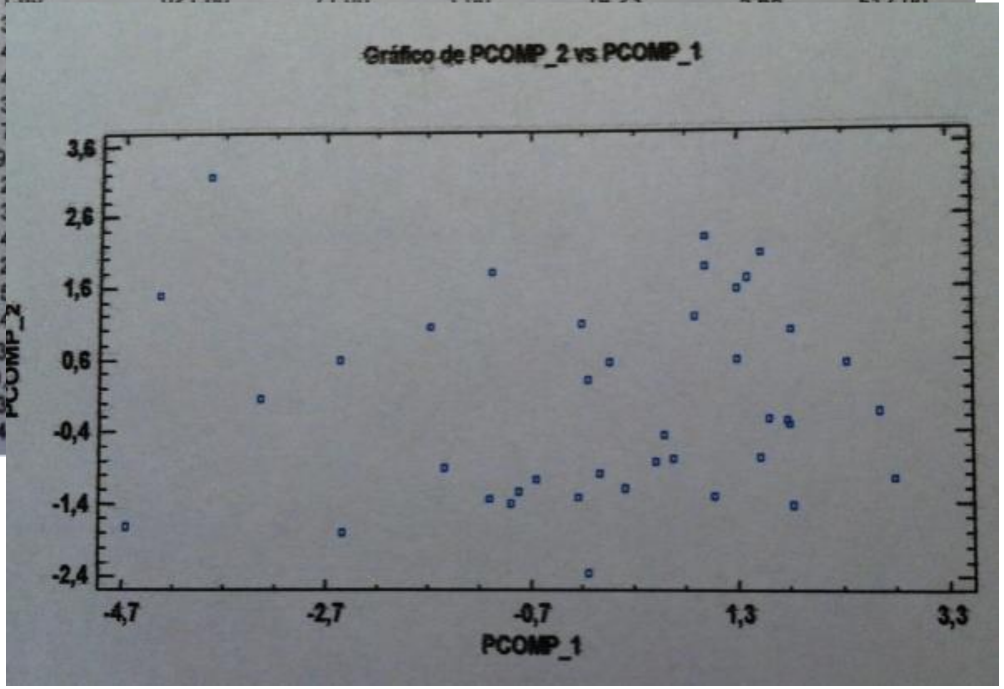


PC1 = 0.5 ChlT + 0.1 RWC - 0.4 aTOC

PC2 = - 0.1 ChlT – 0.8 RWC + 0.1 aTOC

Let's see some examples. Find a partner and try to understand PC1 and PC2 of two examples.

Comarca	Temp media anual	Precipitaciones	Humedad	Vel. Viento	Media Temp máx	Media Temp mín	Altitud
Alt Camp	14,70	450,20	67,00	2,50	20,17	9,93	290,00
Alt Empordà	15,90	762,80	66,00	3,50	20,63	10,92	24,00
Alt Penedès	14,50	522,70	77,00	1,90	21,34	8,88	238,00
Alt Urgell	11,50	402,50	59,00	1,70	18,77	4,99	849,00
Alta Ribagorça	10,00	615,40	65,00	1,20	18,92	2,55	824,00
Anoia	14,20	487,00	65,00	2,30	20,38	9,86	312,00
Bages	13,50	420,20	70,00	1,10	21,09	6,80	349,00
Baix Camp	15,50	505,60	70,00	3,70	20,83	10,88	231,00
Baix Ebre	15,60	513,00	69,00	4,80	20,63	11,59	179,00
Baix Empordà	14,90	818,60	73,00	2,10	22,14	8,43	29,00
Baix Llobregat	15,60	511,70	66,00	1,70	21,77	11,18	220,00
Baix Penedès	16,60	601,40	71,00	2,20	21,94	11,59	60,00
Barcelonès	15,30	540,20	68,00	4,20	19,94	11,71	411,00
Berguedà	11,40	589,40	71,00	1,20	18,10	6,25	860,00
Cerdanya	8,50	439,60	66,00	3,50	17,29	0,57	1096,00
Conca de Barberà	13,50	366,00	69,00	3,40	19,43	8,18	441,00
Garraf	15,50	661,60	75,00	0,60	22,00	10,60	171,00
Garrigues	13,20	359,70	66,00	2,60	18,57	7,63	490,00
Garrotxa	12,50	740,90	75,00	1,40	19,44	6,45	422,00
Gironès	15,80	599,90	71,00	1,40	22,38	9,97	100,00
Maresme	17,10	521,40	70,00	2,60	21,23	13,32	45,00
Montsià	15,80	421,80	71,00	2,50	19,97	11,96	7,00
Noguera	13,60	332,80	74,00	1,10	20,55	7,30	245,00
Osona	11,80	621,80	71,00	1,00	19,23	5,85	517,00
Pallars Jussà	13						
Pla d'Urgell	14						
Pla de l'Estany	14						
Priorat	13						
Ribera d'Ebre	17						
Ripollès	9						
Segarra	12						
Segrià	13						
Selva	14						
Solsonès	12						
Tarragonès	15						
Terra Alta	14						
Urgell	13						
Val d'Aran	9						
Vallès Occidental	13						
Vallès Oriental	14						



Análisis de Componentes Principales

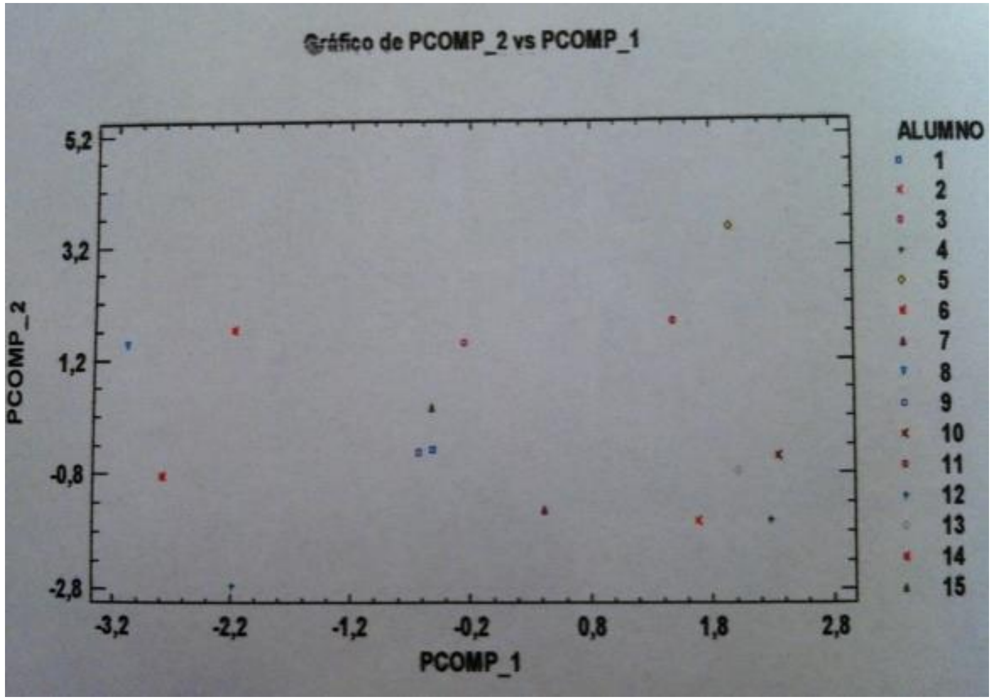
Componente		*Porcentaje de*	*Porcentaje*
Número	*Valor propio*	*Varianza*	*Acumulado*
1	3,61041	51,577	51,577
2	1,8083	25,833	77,410
3	0,834195	11,917	89,327
4	0,461938	6,599	95,926
5	0,207834	2,969	98,895
6	0,0714179	1,020	99,916
7	0,00590614	0,084	100,000

Tabla de Pesos de los Componentes

	Componente	*Componente*
	1	2
Temperatura media anual	0,521913	-0,0282158
Precipitaciones	-0,0551489	0,590282
Humedad	0,0470988	0,621072
Velocidad Viento	0,1257	-0,486115
Media temperaturas máximas	0,461348	0,0936507
Media temperaturas mínimas	0,49067	-0,0585732
Altitud	-0,502939	-0,128579

https://jllopisperez.com/2012/12/29/tema-17-analisis-de-componentes-principales/

ALUMNO	LENGUA	MATEMÁTICAS	FÍSICA	INGLÉS	FILOSOFÍA	HISTORIA	QUÍMICA	EDUCACIÓN FÍSICA
1	5	5	5	5	5	5	5	5
2	7	4	3	8	4	7	3	8
3	5	8	7	6	5	6	7	5
4	7	2	4	8	7	7	3	6
5	8	9	10	8	8	7	9	4
6	4	9	8	4	3	4	7	5
7	6	4	4	6	5	5	3	7
8	4	7	8	3	3	2	8	3
9	5	5	4	5	6	5	5	1
10	7	4	5	7	8	8	4	6
11	7	8	8	7	7	6	7	9
12	4	3	3	4	3	2	1	4
13	7	4	4	7	8	7	4	5
14	3	5	5	2	3	3	5	7
15	5	6	6	5	5	5	6	6



Análisis de Componentes Principales

Componente		*Porcentaje de*	*Porcentaje*
Número	*Valor propio*	*Varianza*	*Acumulado*
1	3,71043	46,380	46,380
2	2,86078	35,760	82,140
3	0,953481	11,919	94,059
4	0,215574	2,695	96,753
5	0,151316	1,891	98,645
6	0,0628091	0,785	99,430
7	0,0317443	0,397	99,827
8	0,0138659	0,173	100,000

Tabla de Pesos de los Componentes

	Componente	*Componente*
	1	2
LENGUA	0,500113	0,0853043
MATEMATICAS	-0,112909	0,555049
FÍSICA	-0,0517681	0,574789
INGLÉS	0,498752	0,036556
FILOSOFÍA	0,450292	0,121881
HISTORIA	0,49264	0,0635768
QUIMÍCA	-0,0726488	0,573763
EDUCACIÓN FÍSICA	0,187002	-0,0694516

https://jllopisperez.com/2012/12/29/tema-17-analisis-de-componentes-principales/

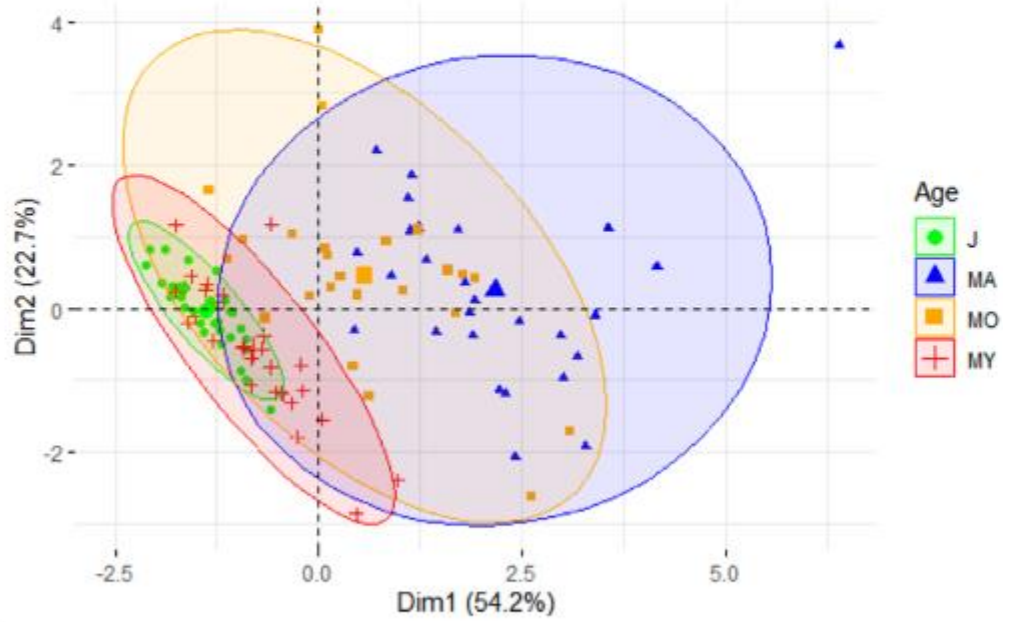
PNAS RESEARCH ARTICLE | PLANT BIOLOGY

Ancient trees are essential elements for high-mountain forest conservation: Linking the longevity of trees to their ecological function

Ot Pasques[a,b] and Sergi Munné-Bosch[a,b,1]

Edited by Richard Dixon, University of North Texas, Denton, TX; received October 13, 2023; accepted December 24, 2023

E



F

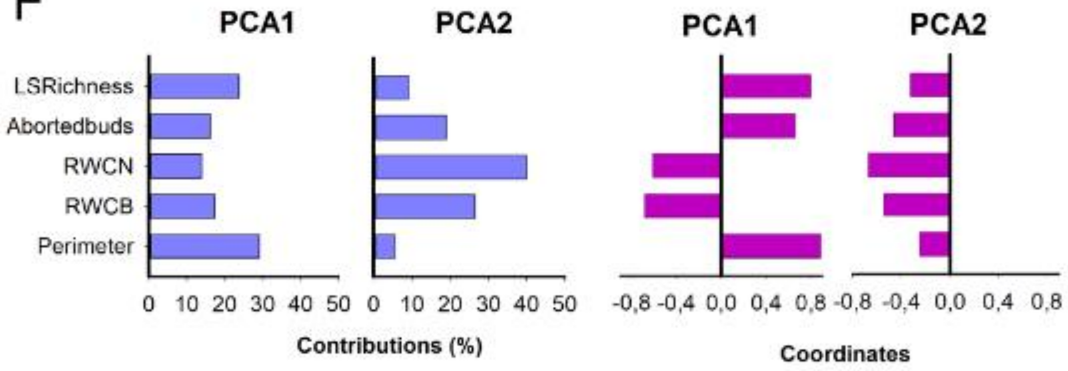


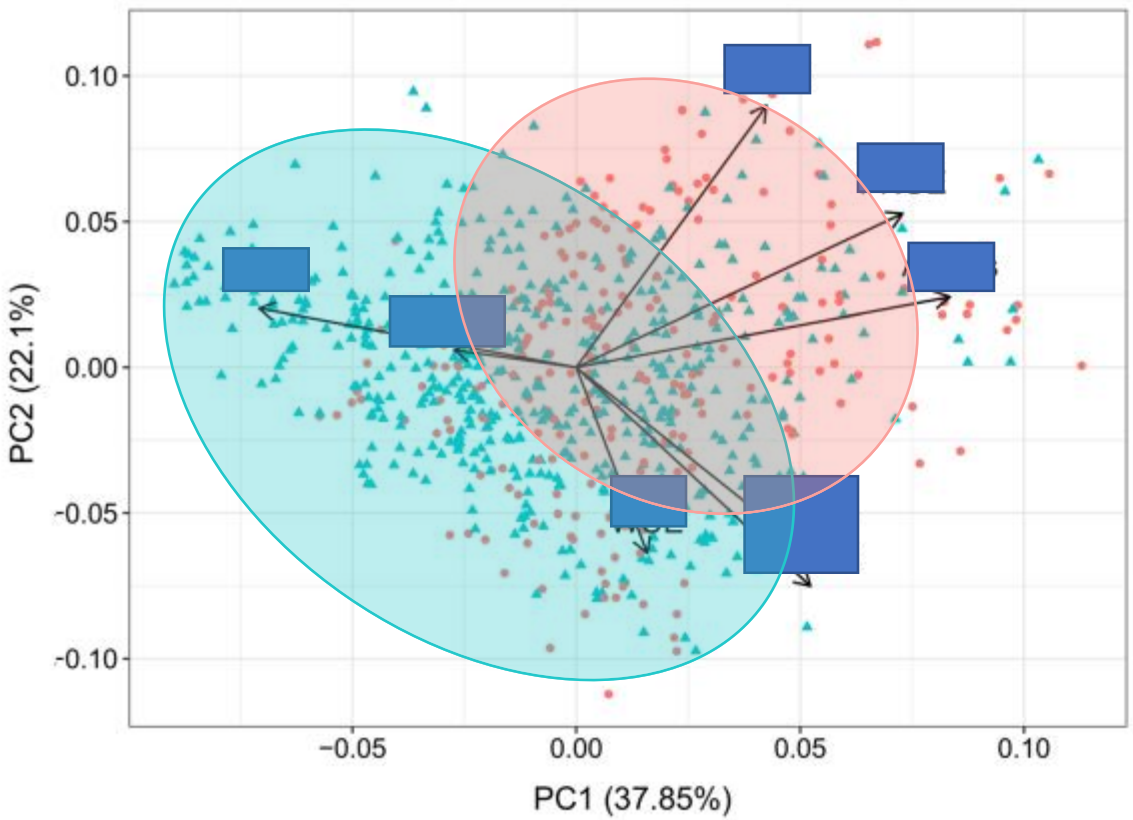
Fig. 3. Physiological and ecological consequences of attaining extraordinarily advanced ages. (*A*) Accumulated occurrence of longevity-related physiological traits with the increase in tree DBH. The presence of an apical dominance break, modular senescence, fissured or stripped bark, and exposed roots were longevity traits found in most of the mature old and ancient trees. (*B*) Accumulated occurrence of ancient human footprints and coexisting organisms with an increasing DBH. Red= forest S; yellow = forest P; blue = forest A; pink = forest C. (*C*) Linear correlation between tree DBH and lichen species richness (α-diversity), including long-living dead trees. Lichen species richness positively correlated with the DBH of living trees (*C*, *N* = 121 studied trees). (*D*) α-Diversity present in each developmental stage in living and dead mature old and ancient trees. The highest diversity values were recorded at mature ancient living trees of forests *A* and *C*, as well as in dead mature ancient trees of forest S. Differences in letters indicate significant differences in size based on one-way ANOVA ($P = 0.05$) and Tukey's test comparing the effect of size and the DBH on lichen species richness between the studied forests. Bars sharing at least one letter show no significant differences between them. (*E*) Principal component analysis (PCA) of the variables linking tree physiological traits to the ecosystem functions of extremely old trees. Biplot representing individual trees resulting from the PCA. Colors represent the different developmental stage, while the colored ellipses represent the 95% CIs (green = juvenile trees; red = mature young trees; orange = mature old trees; and blue = mature ancient trees). (*F*) Bar plots for the contributions and coordinates for each variable in the PCA. LSPRichness, lichen species richness; Abortedbuds, number of aborted buds/m^2; RWCN, somatic tissue RWC; RWCB, meristematic unit RWC; perimeter, DBH. (*G*) Photographs illustrating the irreplaceable functions of the oldest trees in the ecosystem. *Upper left*: a vulnerable rare lichen species, *Letharia vulpina*, which was described in the mature forests studied, growing only on some of the oldest ancient trees; *Upper right*: extremely old trees harboring vascular plants, such as *Sempervivum montanum*; *Bottom left*: complete exposure of the main roots of the oldest roots provide microhabitats and wet substrates for bryophytes and certain lichen species; *Bottom right*: the oldest trees of the high-mountain mature forests have faced harsh environmental conditions for several centuries, which have directly affected their morphological trunk structures, producing scars and bark crevices that some ant colonies take advantage of to create their own habitat. Linear correlations were set as significant with $P < 0.05$ and very significant with $P < 0.01$.

Biol Invasions (2021) 23:253–266
https://doi.org/10.1007/s10530-020-02368-5

ORIGINAL PAPER

Functional segregation of resource-use strategies of native and invasive plants across Mediterranean biome communities

Javier Galán Díaz · Enrique G. de la Riva · Jennifer L. Funk · Montserrat Vilà



trait	PC1	PC2	PC3
LMA	-0.44	0.13	-0.02
height	-0.18	0.03	0.77
Amass	0.52	0.14	0.08
WUE	0.09	-0.41	0.52
Nmass	0.31	-0.40	0.13
Pmass	0.32	-0.48	-0.20
PNUE	0.46	0.33	-0.02
PPUE	0.27	0.55	0.26
eigenvalue	3.03	1.77	1.05
variance	0.38	0.22	0.13

Fig. 1 Principal Component Analysis (PCA) of eight plant traits from 137 natives (blue triangles) and invasive (red dots) plant species in Mediterranean communities (4–5 replicates per species). The table shows the loadings and variance associated with each principal component with eigenvalues over 1. The most relevant traits of each principal component have been shaded. Traits: LMA: leaf mass per area, Amass: mass-based photosynthetic rate, WUE: instantaneous water use efficiency, Nmass: mass-based leaf nitrogen concentration, Pmass: mass-based leaf phosphorus concentration, PNUE: photosynthetic nitrogen-use efficiency, PPUE: photosynthetic phosphorus-use efficiency, and Height: vegetative plant height

doi:10.1038/nature16489

The global spectrum of plant form and function

Sandra Díaz[1], Jens Kattge[2,3], Johannes H. C. Cornelissen[4], Ian J. Wright[5], Sandra Lavorel[6], Stéphane Dray[7], Björn Reu[8,9], Michael Kleyer[10], Christian Wirth[2,3,11], I. Colin Prentice[5,12], Eric Garnier[13], Gerhard Bönisch[2], Mark Westoby[5], Hendrik Poorter[14], Peter B. Reich[15,16], Angela T. Moles[17], John Dickie[18], Andrew N. Gillison[19], Amy E. Zanne[20,21], Jérôme Chave[22], S. Joseph Wright[23], Serge N. Sheremet'ev[24], Hervé Jactel[25,26], Christopher Baraloto[27,28], Bruno Cerabolini[29], Simon Pierce[30], Bill Shipley[31], Donald Kirkup[32], Fernando Casanoves[33], Julia S. Joswig[2], Angela Günther[2], Valeria Falczuk[1], Nadja Rüger[3,23], Miguel D. Mahecha[2,3] & Lucas D. Gorné[1]

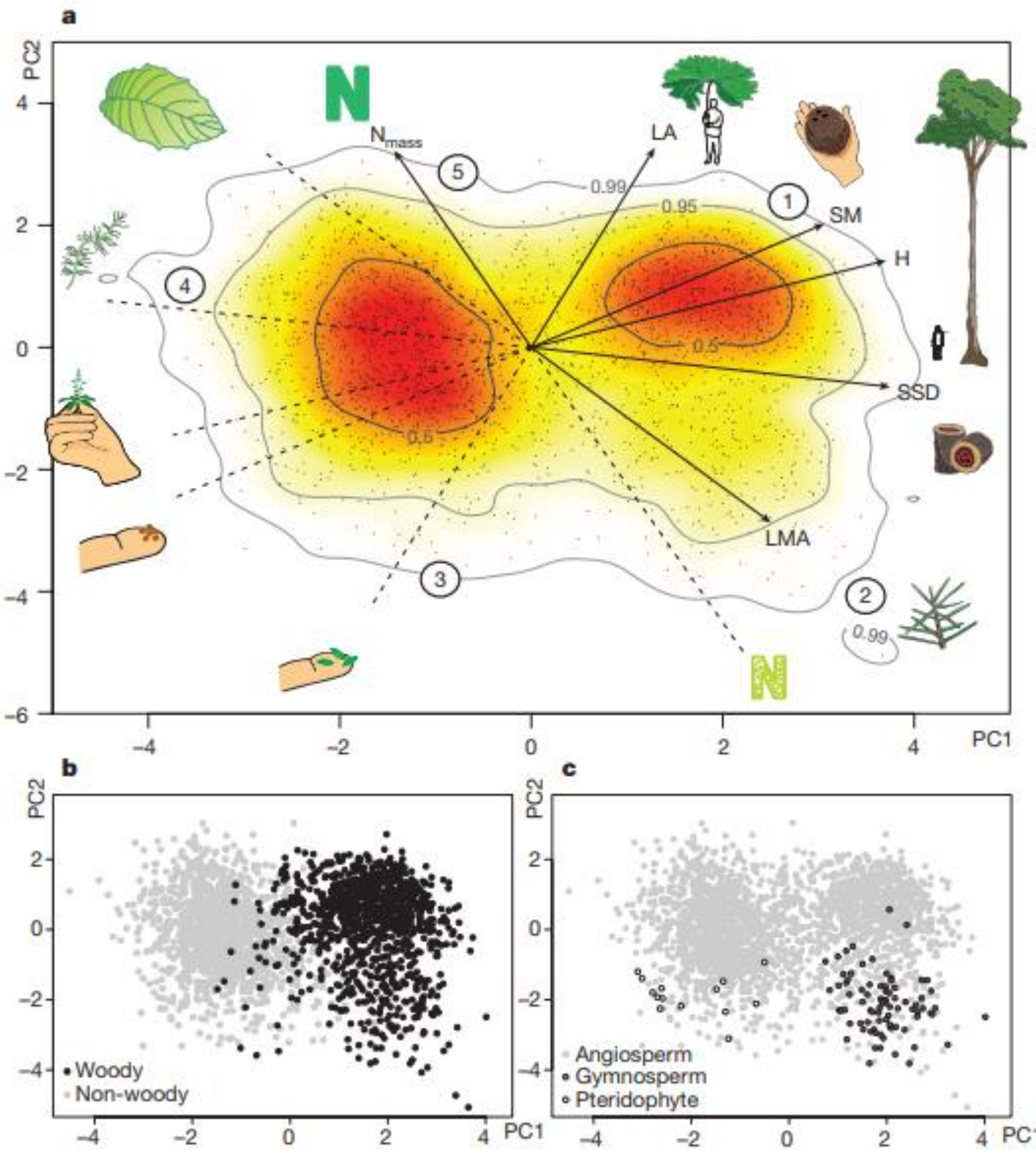


Figure 2 | The global spectrum of plant form and function. **a**, Projection of global vascular plant species (dots) on the plane defined by principal component axes (PC) 1 and 2 (details in Extended Data Table 1 and Extended Data Fig. 2). Solid arrows indicate direction and weighing of vectors representing the six traits considered; icons illustrate low and high extremes of each trait vector. Circled numbers indicate approximate position of extreme poles of whole-plant specialization, illustrated by typical species (Extended Data Table 2). The colour gradient indicates regions of highest (red) to lowest (white) occurrence probability of species in the trait space defined by PC1 and PC2, with contour lines indicating 0.5, 0.95 and 0.99 quantiles (see Methods, kernel density estimation). Red regions falling within the limits of the 0.50 occurrence probability correspond to the functional hotspots referred to in main text. **b**, **c**, location of different growth-forms (**b**) and major taxa (**c**) in the global spectrum.

A lot of exemples to practice

Science of the Total Environment 749 (2020) 142356

Contents lists available at ScienceDirect

Science of the Total Environment

journal homepage: www.elsevier.com/locate/scitotenv

C3 and C4 plant systems respond differently to the concurrent challenges of mercuric oxide nanoparticles and future climate CO_2

Hamada AbdElgawad [a], Yasser M. Hassan [a], Modhi O. Alotaibi [b,*], Afrah E. Mohammed [b], Ahmed M. Saleh [c,d,**]

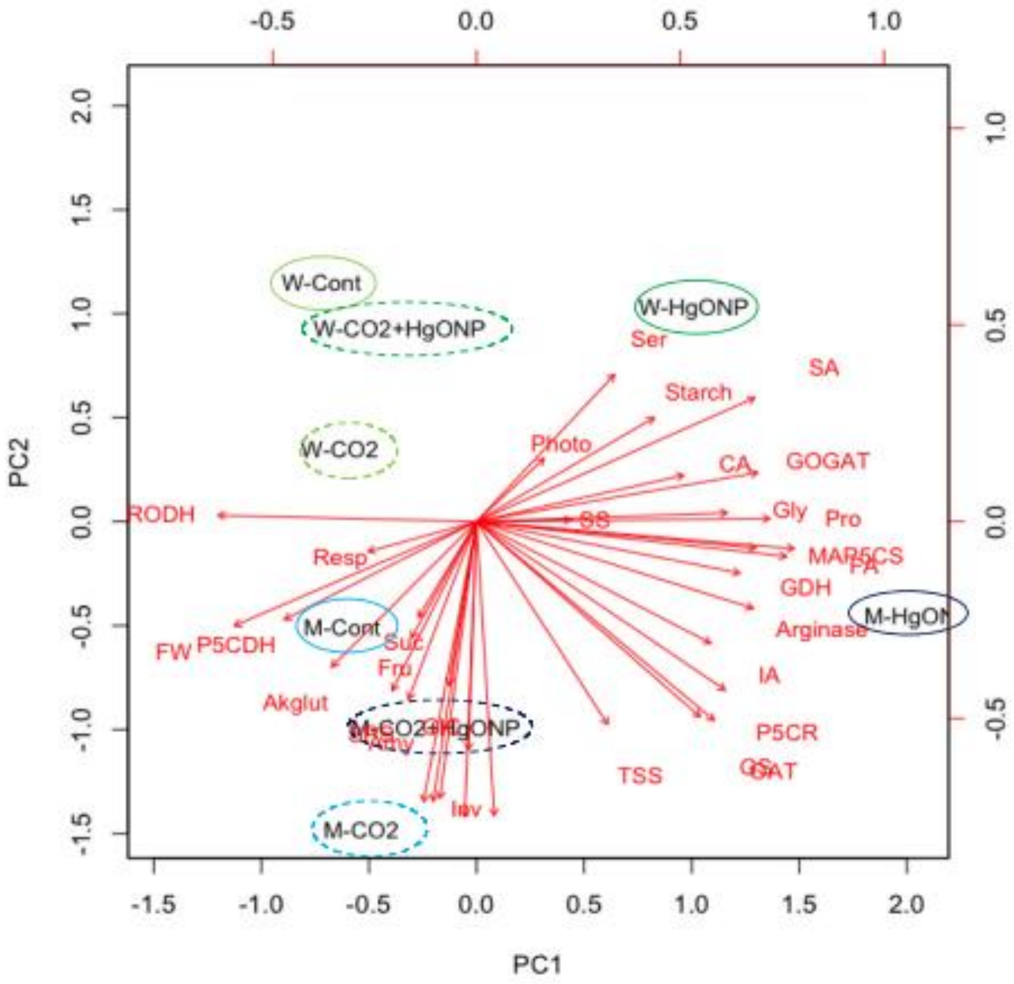


Fig. 5. Principal component analysis (PCA) of growth, photosynthesis, respiration and metabolites and enzymes involved in sugars and proline biosynthesis in wheat (W) and maize (M) grown under control conditions (aCO_2), HgO-NPs, eCO_2 or coexistence of HgO-NPs and eCO_2. Variances explained by the first two components (PC1 and PC2) appear in parentheses.

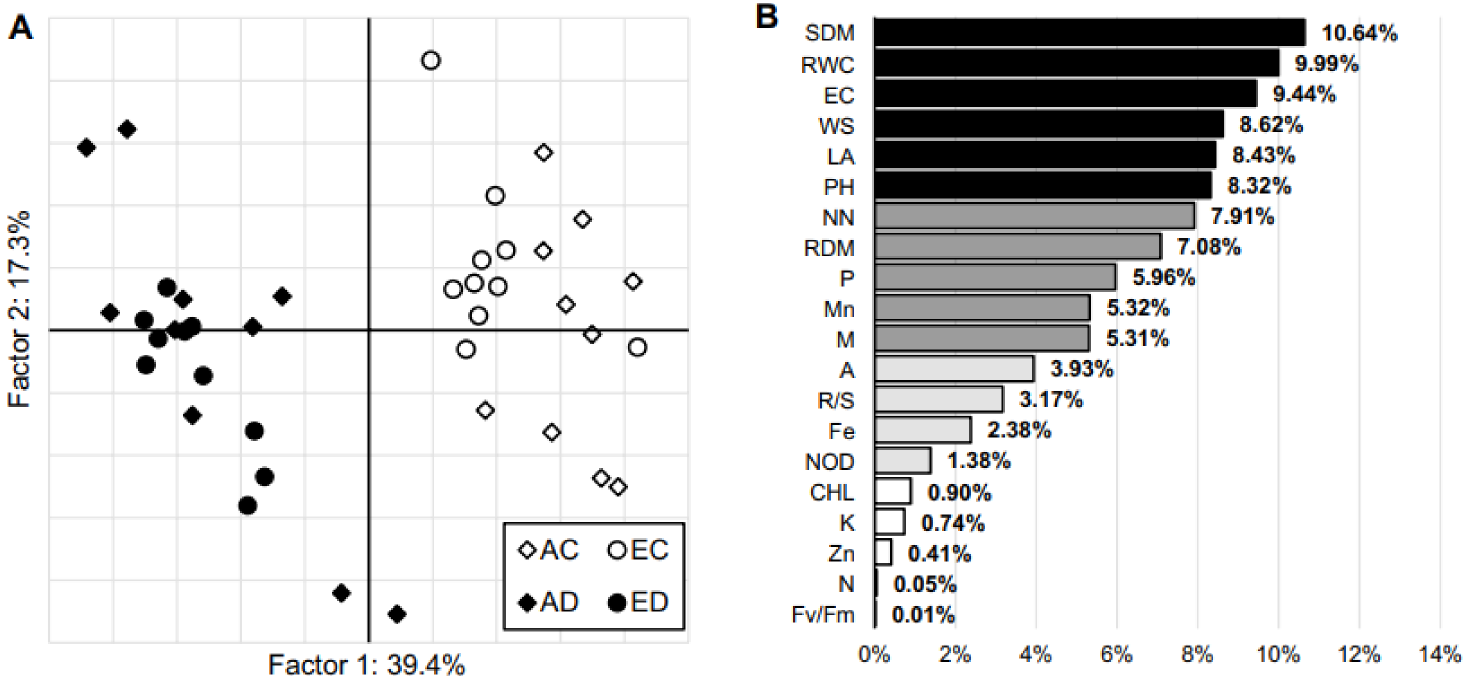


Fig. 2 The result of PCA in experiment II. **a** Projection of the cases on factor-plane (A/E: soybean varieties—Alíz/Emese, C/D: control/drought stressed plants), **b** the PC1 loading values of the 20 measured parameters (abbreviations of the traits are in Table 1)

Acta Physiologiae Plantarum (2019) 41:56
https://doi.org/10.1007/s11738-019-2842-9

ORIGINAL ARTICLE

Selection of plant physiological parameters to detect stress effects in pot experiments using principal component analysis

Anna Füzy[1] · Ramóna Kovács[1] · Imre Cseresnyés[1] · István Parádi[1,2] · Tibor Szili-Kovács[1] · Bettina Kelemen[1] · Kálmán Rajkai[1] · Tünde Takács[1]

What is principal component analysis?

Markus Ringnér

Nature Biotechnology **26**, 303–304 (2008) | Cite this article

Principal component analysis is often incorporated into genome-wide expression studies, but what is it and how can it be used to explore high-dimensional data?

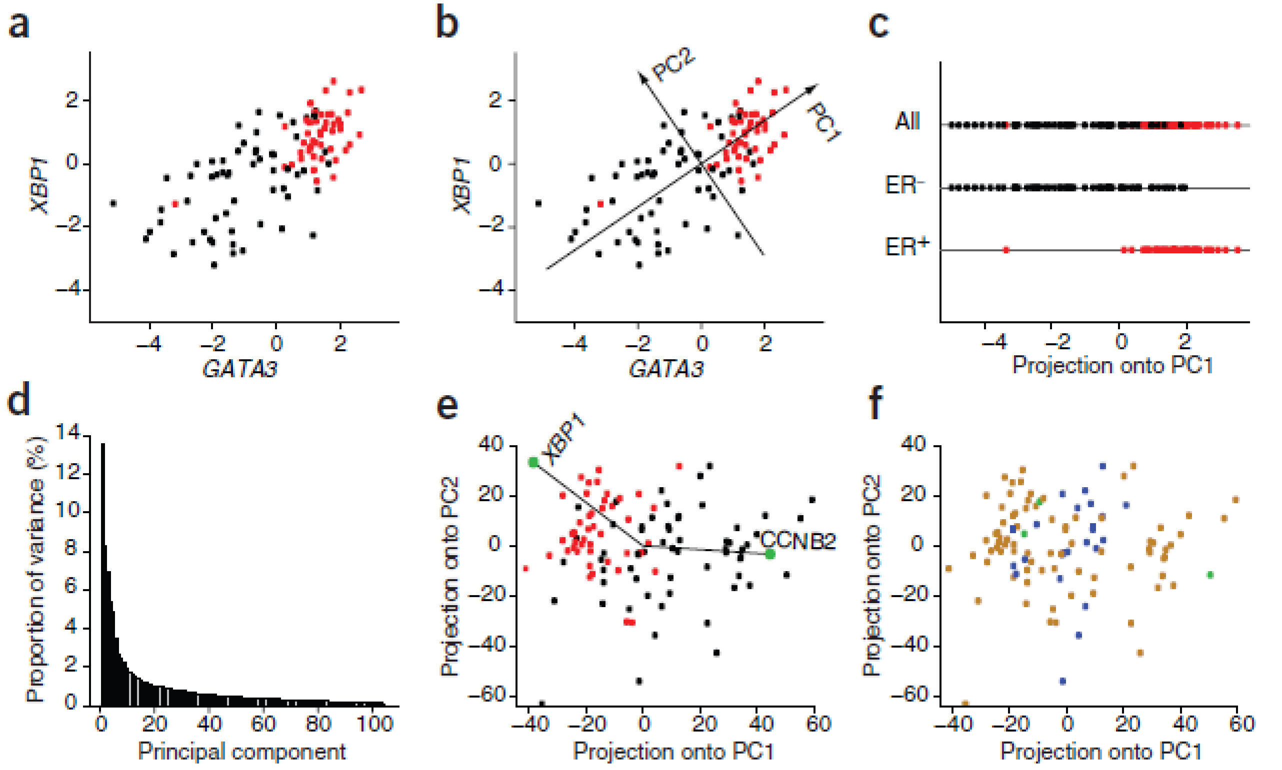


Figure 1 Principal component analysis (PCA) of a gene expression data set. (**a**) Each dot represents a breast cancer sample plotted against its expression levels for two genes. (In **a–c**, **e**, samples are colored according to estrogen receptor (ER) status: ER^+, red; ER^-, black). (**b**) PCA identifies the two directions (PC1 and PC2) along which the data have the largest spread. (**c**) Samples plotted in one dimension using their projections onto the first principal component (PC1) for ER^+, ER^- and all samples separately. (**d**) The variance of the principal components when PCA is applied to all 8,534 genes with expression levels for all samples. (**e**) PCA biplot with samples plotted in two dimensions using their projections onto the first two principal components, and two genes plotted using their weights for the components (green points). The scale shown is for the samples; for the genes, the scale should be divided by 950. (**f**) Samples plotted as in **e** but colored according to *ERBB2* status (blue, *ERBB2*$^+$; brown, *ERBB2*$^-$; green, unknown).

Enhanced Secondary- and Hormone Metabolism in Leaves of Arbuscular Mycorrhizal *Medicago truncatula*[1[OPEN]]

Lisa Adolfsson,[a,2] Hugues Nziengui,[a,2] Ilka N Abreu,[b,3] Jan Šimura,[c,3] Azeez Beebo,[a,3] Andrei Herdean,[a] Jila Aboalizadeh,[a] Jitka Široká,[c] Thomas Moritz,[b] Ondřej Novák,[c] Karin Ljung,[b] Benoît Schoefs,[d] and Cornelia Spetea[a,4]

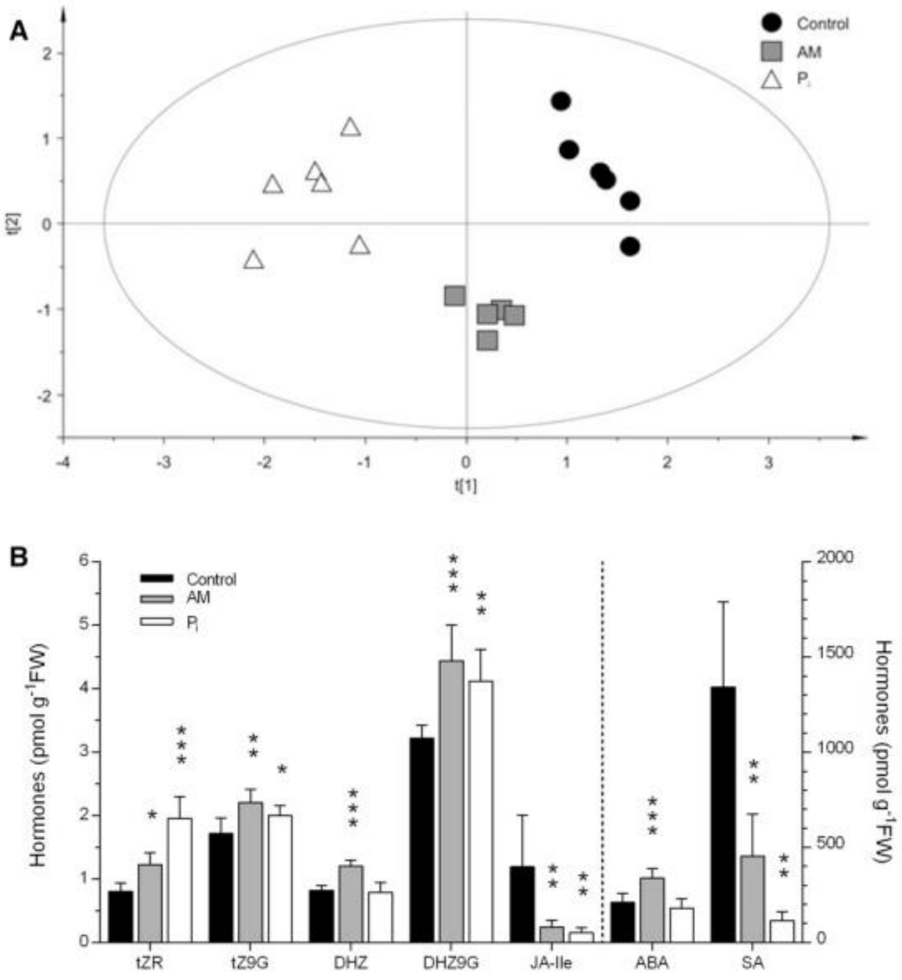


Figure 3. Quantification of hormones in *M. truncatula* leaves. A, PCA score plot (explained variance $R^2 = 0.722$ and predicted variance $Q^2 = 0.0801$; ellipse, Hotelling's T2 [95%]). B, Content of cytokinin species (tZR, trans-zeatin riboside; tZ9G, trans-zeatin 9-glucoside; DHZ, dihydrozeatin; DHZ9G, dihydrozeatin 9-glucoside) and the stress-related hormones JA-Ile, ABA, and SA. Bars represent means ± SD from six plants. Asterisks indicate significant differences between treatments and the control (one-way ANOVA, $P < 0.05$ [*], $P < 0.01$ [**], and $P < 0.001$ [***]; GraphPad Prism). FW, Fresh weight.

Research Article

Performance of *Ambrosia artemisiifolia* and its potential competitors in an experimental temperature and salinity gradient and implications for management

Hana Skálová[1,*], Wen-Yong Guo[1], Lenka Moravcová[1] and Petr Pyšek[1,2]

[1]*Institute of Botany, The Czech Academy of Sciences, Zámek 1, CZ-252 43 Průhonice, Czech Republic*

[2]*Department of Ecology, Faculty of Science, Charles University, Viničná 7, CZ-128 44 Prague, Czech Republic*

Author e-mails: *hana.skalova@ibot.cas.cz (HS), guowyhgy@gmail.com (WYG), lenka.moravcova@ibot.cas.cz* (LM), *pysek@ibot.cas.cz (PP)*

**Corresponding author*

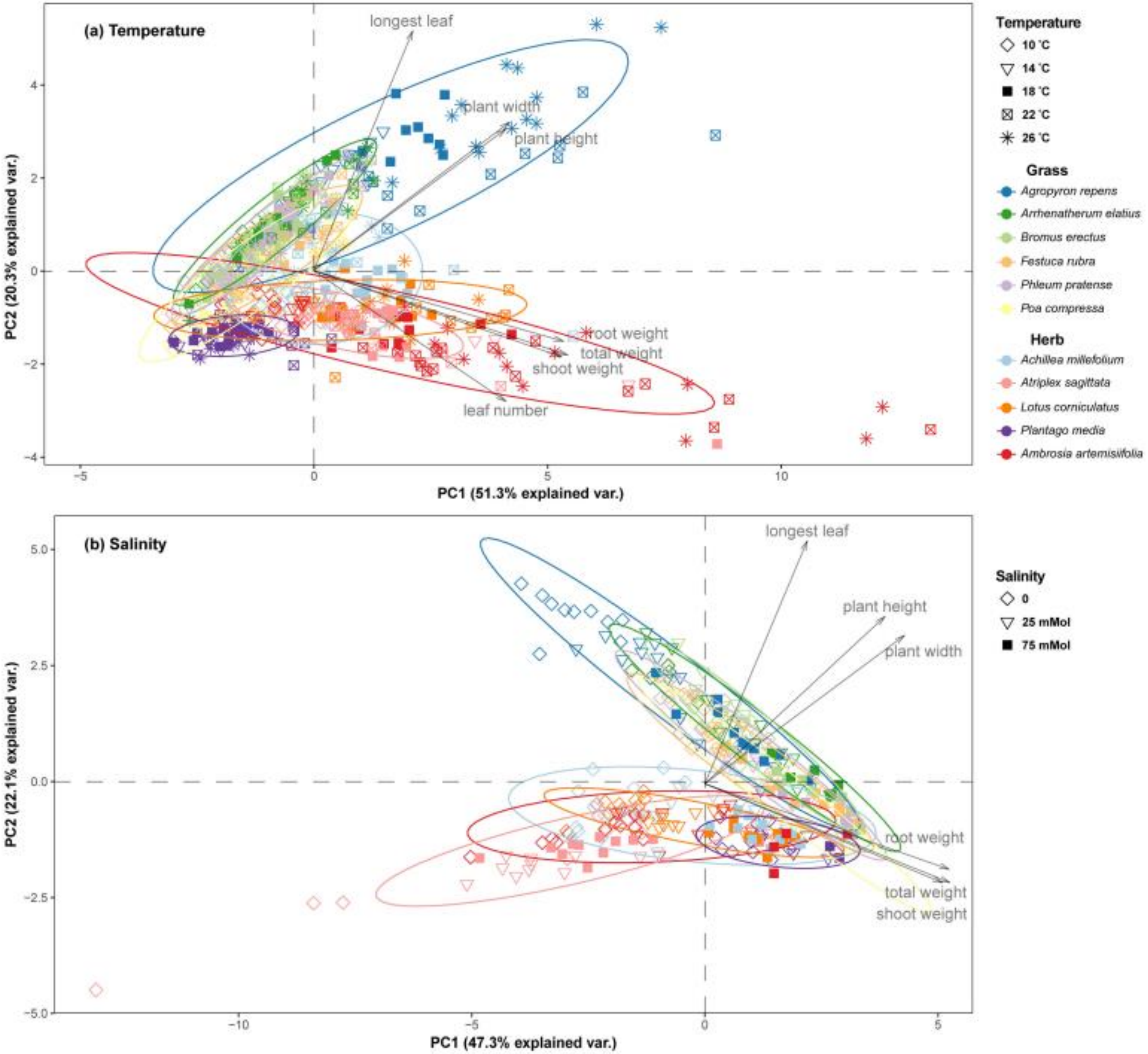


Figure 1. Principal component analysis (PCA) of the plant characteristics measured at different temperatures and salinities. Different colours indicate different species and shapes indicate treatment levels. The ellipses define the 95% confidence intervals of the species. Factor loadings from the principal components analyses of (a) temperature and (b) salinity are shown. The arrows

Journal of Experimental Botany, Vol. 68, No. 16 pp. 4463–4477, 2017
doi:10.1093/jxb/erx220 Advance Access publication 29 June 2017
This paper is available online free of all access charges (see http://jxb.oxfordjournals.org/open_access.html for further details)

RESEARCH PAPER

Getting back to nature: a reality check for experiments in controlled environments

Maria Grazia Annunziata[1], Federico Apelt[1], Petronia Carillo[2], Ursula Krause[1], Regina Feil[1], Virginie Mengin[1], Martin A. Lauxmann[1], Karin Köhl[1], Zoran Nikoloski[1,3], Mark Stitt[1] and John E. Lunn[1,*]

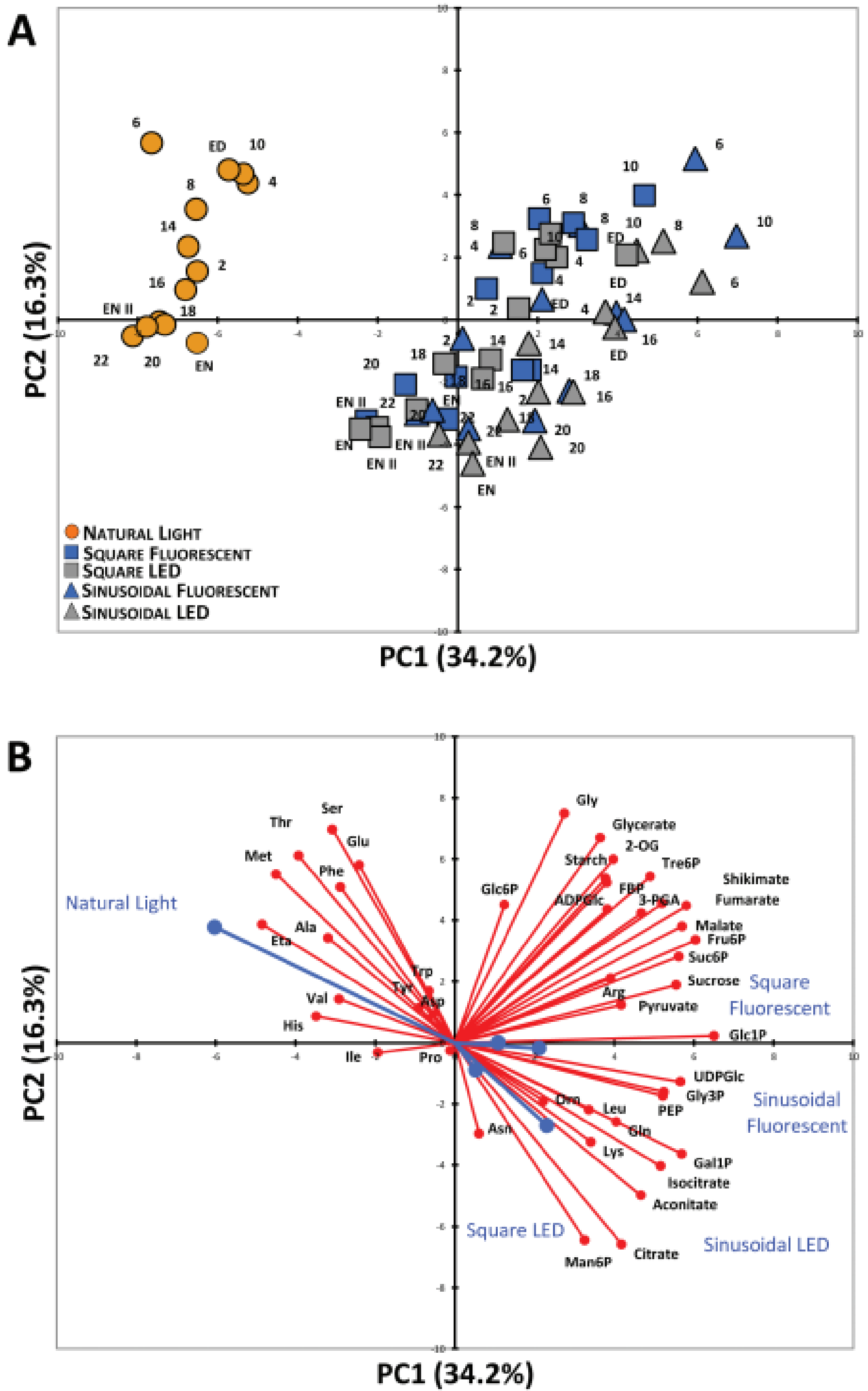


Fig. 1. Principal component analysis (PCA) of metabolite data from Arabidopsis plants. (A) PCA of metabolite data from plants grown in a naturally illuminated greenhouse (orange circles) or in controlled environment chambers with a 12-h photoperiod and daily light integral (DLI) of 7 mol m^{-2} d^{-1}. The artificial illumination was provided by white fluorescent tubes (blue symbols) or LED lights (grey symbols), with either a constant (squares) or sinusoidal (triangles) light profile during the day. Numbers indicate the time of harvest in hours after dawn (zeitgeber time, ZT); ED, end of day (ZT12); EN I, end of preceding night (ZT0); EN II, end of night (ZT24). The percentages of total variance represented by principal component 1 (PC1) and principal component 2 (PC2) are shown in parentheses. (B) The loadings of individual metabolites (red) on the principal components shown in (A) and the (average) loadings of the individual experimental conditions (blue). Glucose and fructose were not included in the PCA due to the very high variability in the data.

agronomy

MDPI

Article

Microbe-Plant Growing Media Interactions Modulate the Effectiveness of Bacterial Amendments on Lettuce Performance Inside a Plant Factory with Artificial Lighting

Thijs Van Gerrewey [1,2,3,4], Maarten Vandecruys [3], Nele Ameloot [4], Maaike Perneel [5], Marie-Christine Van Labeke [6], Nico Boon [2] and Danny Geelen [1,*]

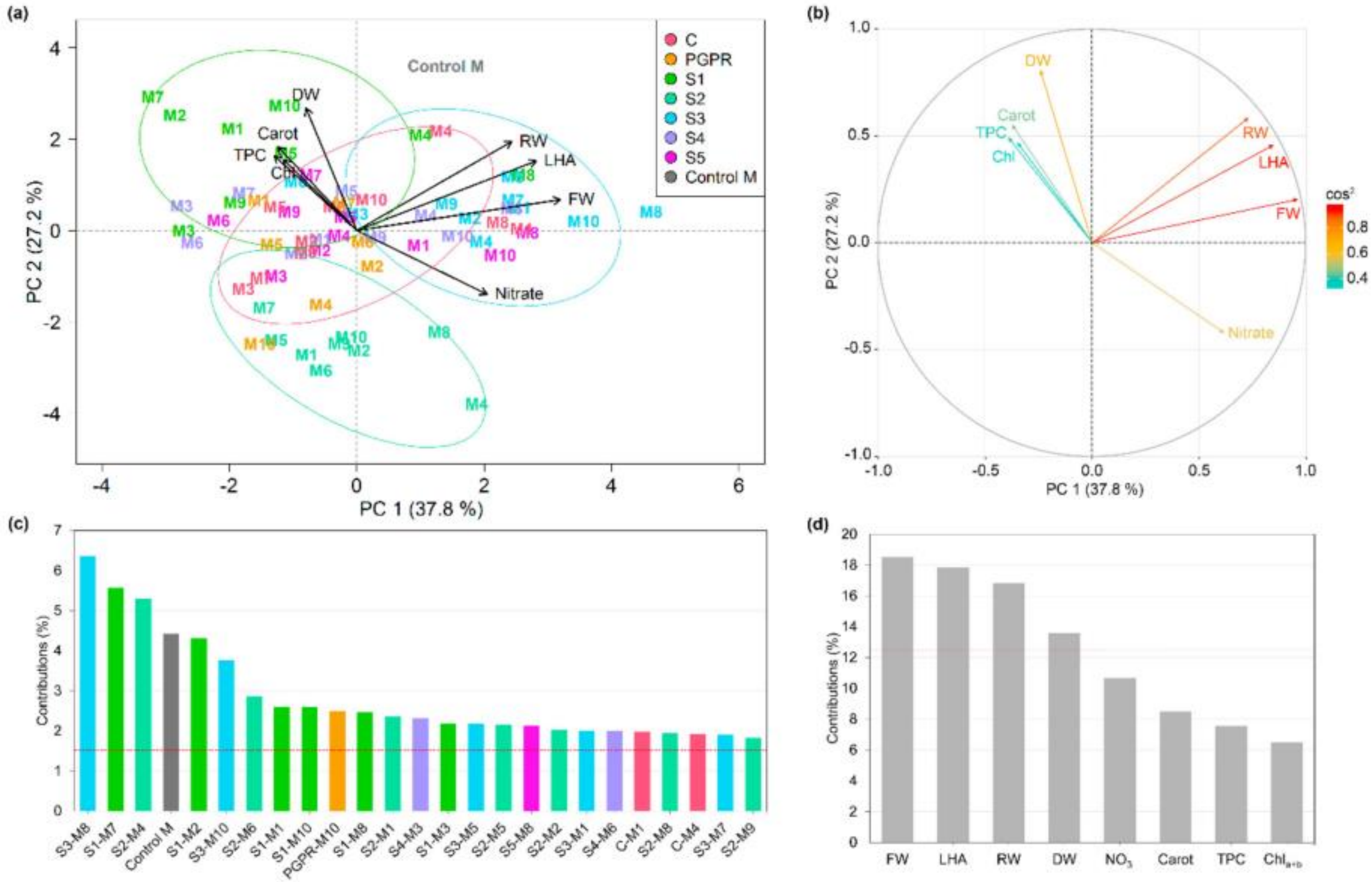


Figure 4. Principal component analysis (PCA) of the lettuce yield and quality variables under different BCI-plant growing medium treatments. (**a**) PCA biplot of individual samples to PC 1 and PC 2. Symbols indicate the type of plant growing medium (M1–10 and control M, the commercial plant growing medium) and colors indicate BCI treatment (S1–5, negative control C, and positive control PGPR). Ellipses denote 95% confidence interval of C, S1, S2, and S3. The plant performance parameters are shoot fresh weight (FW), lettuce head area (LHA), root fresh weight (RW), shoot dry weight (DW), total phenolic content (TPC), Nitrate content, chlorophyll a+b (Chl), and carotenoids (Carot); (**b**) Quality of representation (cos^2) correlation circle of variables to PC 1 and PC 2. The color gradient indicates the quality of representation of the variables; (**c**) Contribution plot of the top 25 samples to PC 1 and PC 2. Colors are the same as in a. The dashed line indicates the expected average contribution if the contribution of the samples were uniform; (**d**) Contribution plot of variables to PC 1 and PC 2. The dashed line indicates the expected average contribution if the contribution of the variables were uniform.

Applied Microbiology and Biotechnology (2019) 103:1497–1509
https://doi.org/10.1007/s00253-018-9564-6

ENVIRONMENTAL BIOTECHNOLOGY

Deciphering differences in the chemical and microbial characteristics of healthy and *Fusarium* wilt-infected watermelon rhizosphere soils

Tianzhu Meng[1] · Qiujun Wang[1] · Pervaiz Abbasi[2] · Yan Ma[1]

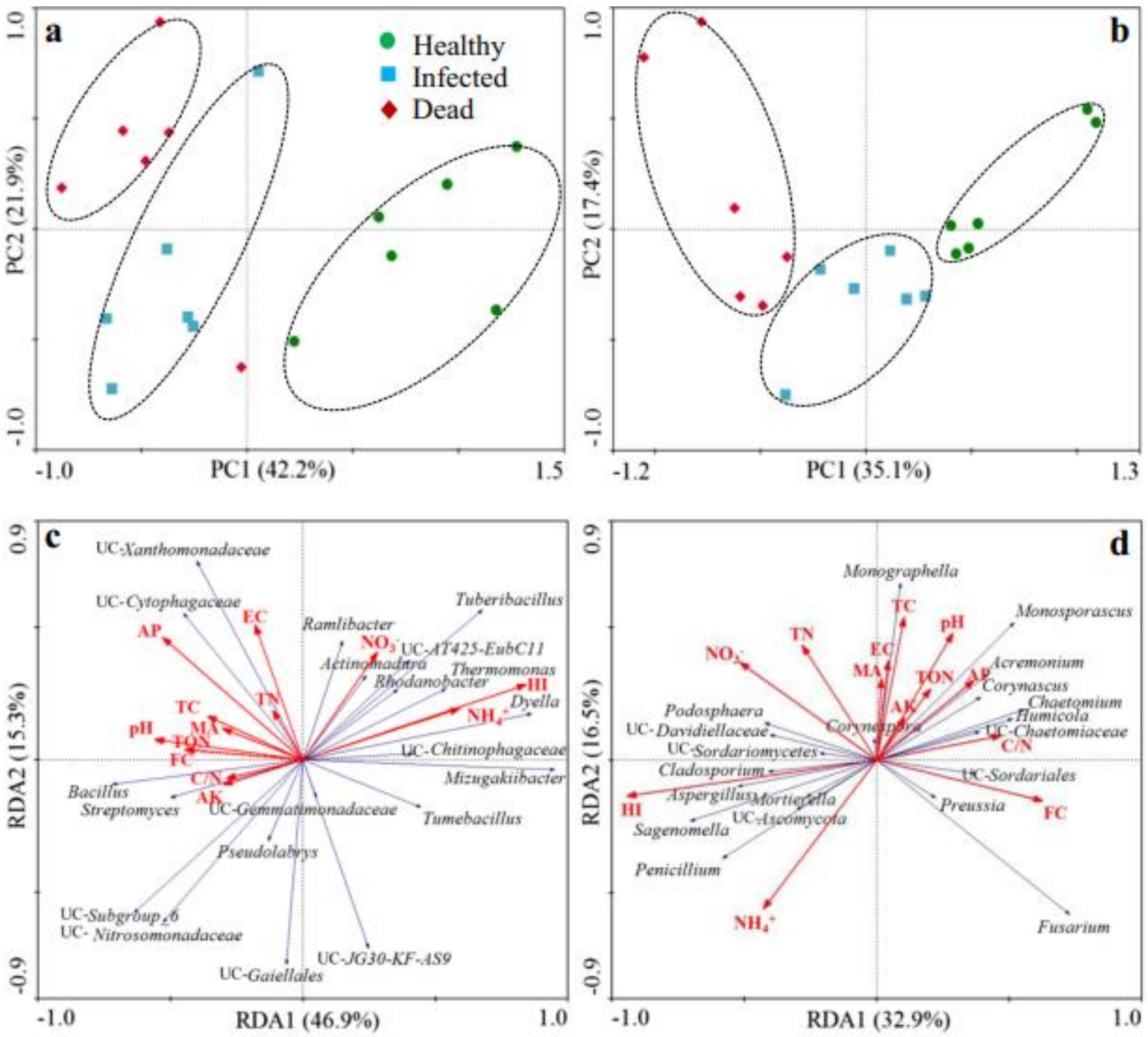
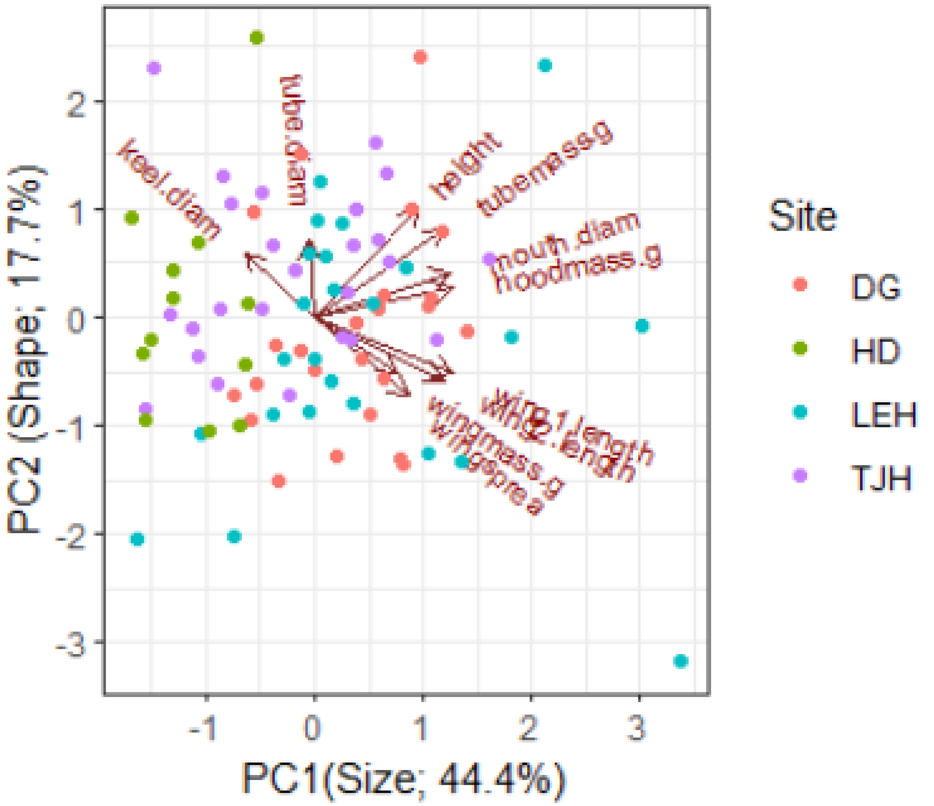
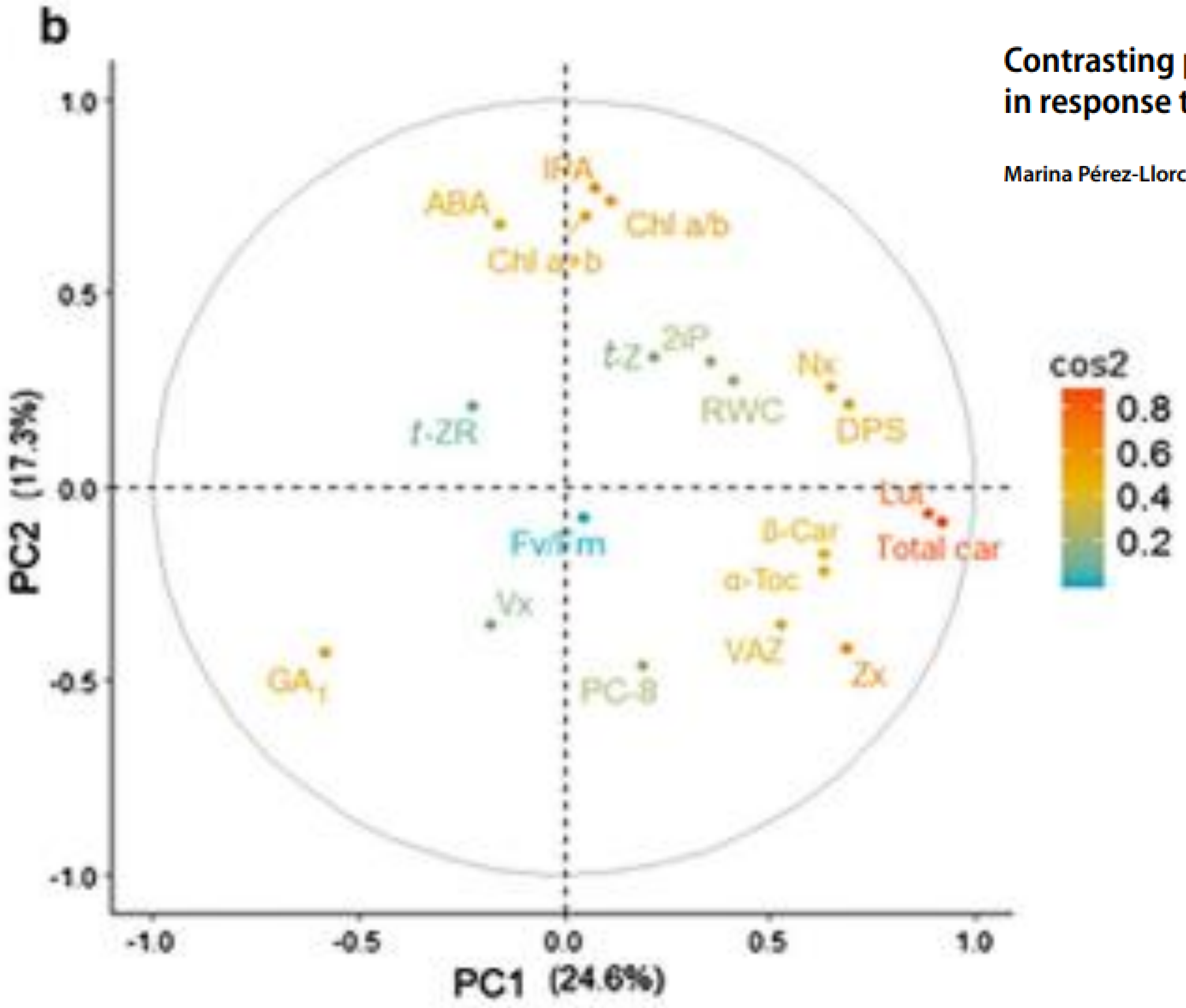


Fig. 5 Principal component analysis (PCA) and redundancy analysis (RDA) of the microbial communities based on genera distributions in the rhizosphere soils of healthy, *Fusarium oxysporum*-infected, and dead watermelon plants. Healthy, the watermelon plants were healthy and not infected by *F. oxysporum*. Infected, the watermelon plants were infected by *F. oxysporum* and showed typical *Fusarium* wilt symptoms. Dead, the watermelon plants were infected by *F. oxysporum* and died. PCA of bacterial (**a**) and fungal (**b**) communities at the genus level. RDA ordination plots show the relationships between the top 20 bacterial (**c**) and fungal (**d**) genera and soil environmental factors. All of the environmental variables (red lines with arrows) shown were tested by partial Monte-Carlo permutations at the $P < 0.05$ level and selected according to their marginal effects in descending order. C/N, ratio of TC to TN; MA, soil total microbial activity. The health index (HI) denotes a healthy plant as "2," the plant infected by *F. oxysporum* as "1," and the dead plant infected by *F. oxysporum* as "0"

https://uw.pressbooks.pub/appliedmultivariatestatistics/chapter/pca/



Biplot of the first two principal components from a PCA of the Darlingtonia plant data. Each plant is represented by a symbol, with colors corresponding to the four sites. The red vectors point in the directions in which variables increase most strongly



Planta (2019) 250:1409–1422
https://doi.org/10.1007/s00425-019-03234-y

ORIGINAL ARTICLE

Contrasting patterns of hormonal and photoprotective isoprenoids in response to stress in *Cistus albidus* during a Mediterranean winter

Marina Pérez-Llorca[1,2] · Andrea Casadesús[1] · Sergi Munné-Bosch[1,2] · Maren Müller[1]

Conservation Physiology

Volume 7 • 2019 10.1093/conphys/coz075

Research article

Increased chilling tolerance of the invasive species *Carpobrotus edulis* may explain its expansion across new territories

Erola Fenollosa [1,2,*] **and Sergi Munné-Bosch** [1,2,†]

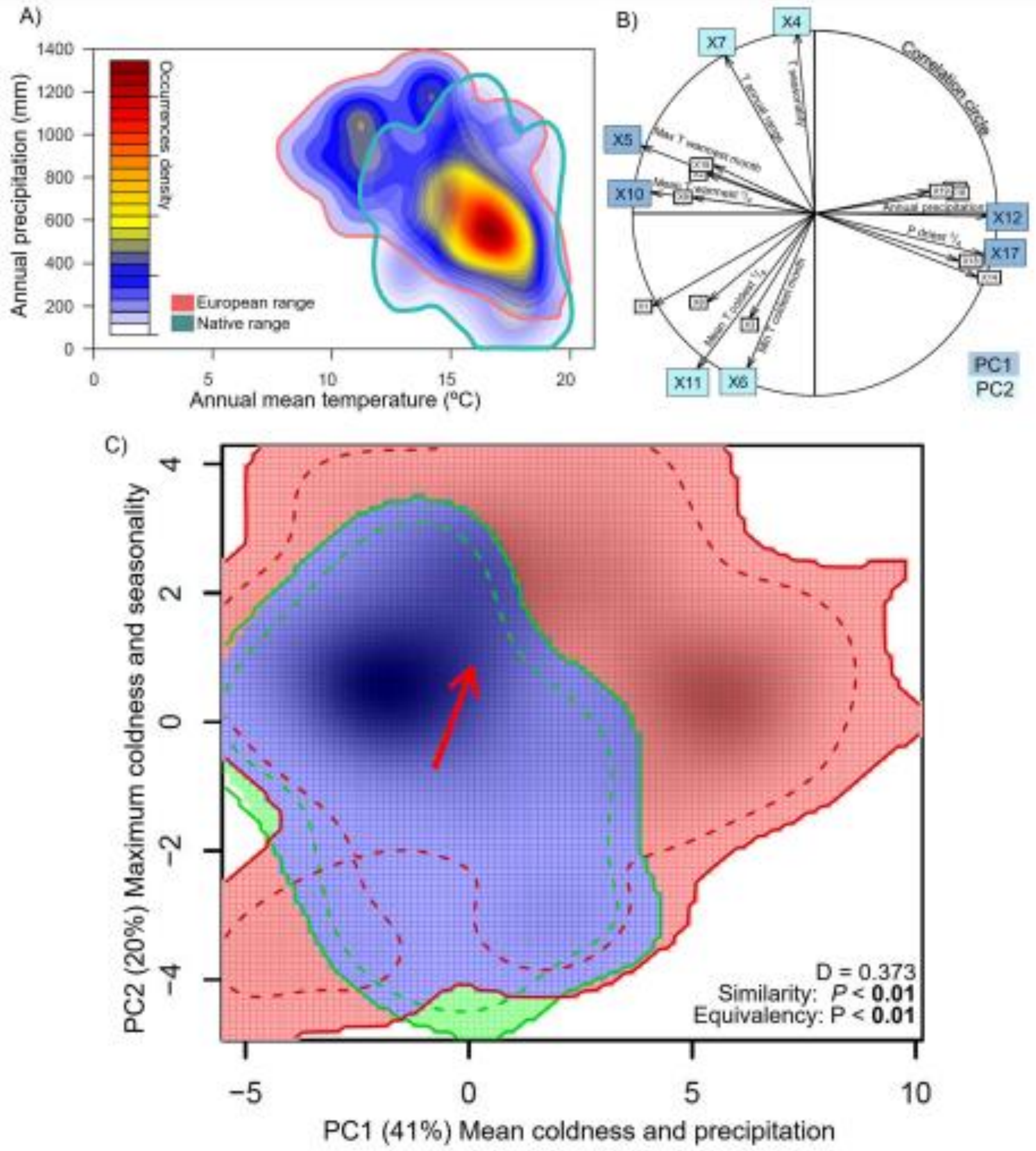


Figure 1: (A) Kernel density estimation for *C. edulis* occurrences in response to annual mean temperature and precipitation. (B) Correlation circle for the PCA-env analysis, with the 19 bioclimatic WorldClim variables (X1-19). Bioclimatic variables full names can be found at: http://worldclim.org/bioclim. (C) Niche dynamics: stability, expansion and unfilling (in blue, red and green respectively) in the multivariate climatic space for native compared to the European niche of *C. edulis* considering the two first components from the PCA-env. D Stands for Schoener's D overlap value. Solid and dashed lines delineate 100 and 75% of the available background environment, respectively.

Why wouldn’t work a PCA?

- No lineality
- Too much variables
- Data is not paired

Which percentatge of variable explained is acceptable for you?

How to report a PCA

- Show % variance explained
- Explore Components weights and find a biological explanation, report variable vectors or components weights.

Oecologia (2021) 196:747–761
https://doi.org/10.1007/s00442-021-04971-2

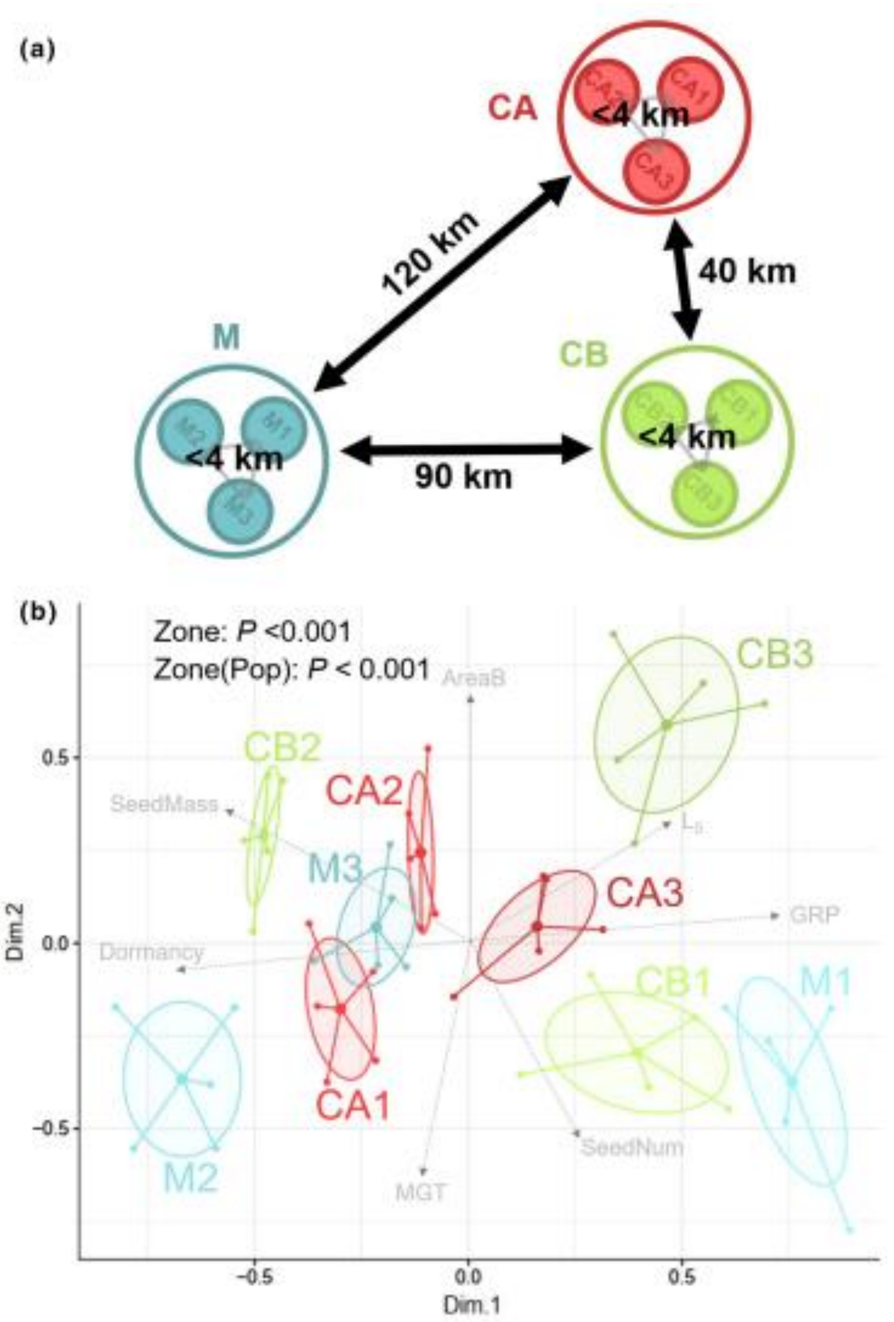
POPULATION ECOLOGY – ORIGINAL RESEARCH

Check for updates

Geographic patterns of seed trait variation in an invasive species: how much can close populations differ?

Erola Fenollosa[1,2] · Laia Jené[1] · Sergi Munné-Bosch[1,2]

Fig. 1 **a** Relative location of the nine studied populations (filled circles) of *C. edulis* distributed in three differentiated zones: Maresme (M), Costa Brava (CB) and Cap de Creus (CA). **b** Results of multidimensional scaling analysis (MDS) evaluating differences in nine seed traits among studied populations. Traits indicated in grey have significant ($P < 0.01$) contribution population variability. Ellipses represent 95% of confidence intervals. *P*-values correspond to PERMANOVA results for Zone and Population (nested in Zone) factors



Other multidimensional techniques

- MDS, NDMS
- CCA
- Discriminant analysis

Difficult question: What is the difference between PCA and MDS?

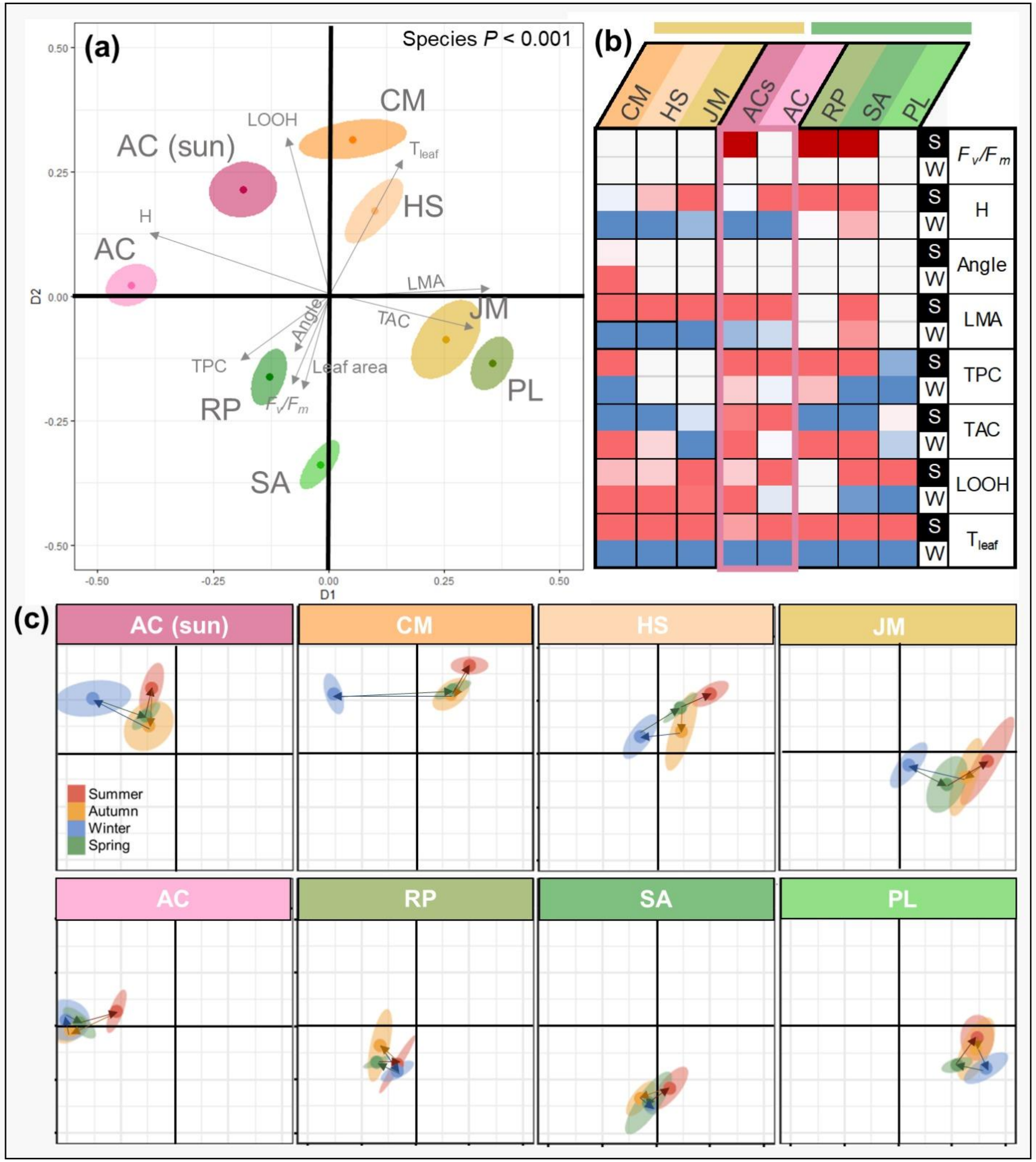
Submitted to PPL 2024

Invasion amidst the shadows: A higher water use and improved physiological performance relative to natives underlies a potentially invader's success

Fenollosa, E.[1,2*], Munné-Bosch, S.[1,2], Pintó-Marijuan, M.[1,2]

1. Department of Evolutionary Biology, Ecology and Environmental Sciences, University of Barcelona, Avinguda Diagonal 643, 08028, Barcelona, Spain
2. Institute of Research in Biodiversity (IRBio-UB), Avinguda Diagonal 643, 08028, Barcelona, Spain

*Correspondence: Erola Fenollosa (erola.fenollosa@gmail.com)



The basic steps to build a PCA

- Standarize
- Compute (Check variance %)
- Understand the components
- Plot

http://www.sthda.com/english/articles/31-principal-component-methods-in-r-practical-guide/112-pca-principal-component-analysis-essentials/

Test and validate programming in R with: ChatGPT

Aims of the session

1) Understand PCA in scientific **articles**

2) Recognise different **applications** of PCA

3) Identify **what is needed** to build and report a PCA and when is not appropriated to use it

4) Be conscient of the **limitations** of PCA through its mechanics

EXTRA: **Build** your own PCA

Principal Component Analysis (PCA)

Discovering the Multiverse

Dr. Erola Fenollosa

erola.fenollosa@gmail.com

@Erola_Fenollosa

www.erolafenollosa.weebly.com

Environmental Data Scientist

WINEQUALITY:
https://archive.ics.uci.edu/dataset/186/wine+quality

IRIS
https://archive.ics.uci.edu/dataset/53/iris

PIZZA, DIABETES, USArrests
https://github.com/f-imp/Principal-Component-Analysis-PCA-over-3-datasets/tree/master