Do Entropic Biodiversity Methods Outcompete Alternatives ?

Evaluation of entropy/information approaches is sporadic:

- simulation
- empirical

(Sherwin et al. 2017)

E&ERC
UNSW
Evolution & Ecology Research Centre

UNSW SYDNEY
Australia's Global University

Bill Sherwin
W.Sherwin@unsw.edu.au

Evaluation Criteria
Biodiversity measures must:

Forecast under Natural / Artificial processes

Measure to know if hypothesis / policy correct?

EG: Evaluation of 2 Entropic methods

Assess gene-frequency differentiation between groups, times or locations

Incorporating functional differences between variants.

Differentiation:
Bray-Curtis (Bray & Curtis 1957)

USE:

- Widespread: Ecology, Metagenomics (Peng et al. 2020).
- Molecular Ecology (AFD, Berner 2019a,b)

NO Forecasting

- Unlike Hill-measure entropies

A function of a Hill-entropy

Forecasting Differentiation with Bray-Curtis

Bray-Curtis for 2 locations, 2 variants (EG SNP genes)

- $B = |p_1 - p_2|$, averaged over many genes

  where p_1, p_2 are proportions of one of the two variants, in locations 1,2

OR for multiple variants

- $B = \frac{\sum_{j=1}^{S} |a_{1j} - a_{2j}|}{\sum_{j=1}^{S} (a_{1j} + a_{2j})}$,

  a_{1j} and a_{2j} are abundances in each location, for variant j $(1 \leq j \leq S)$

(Chao & Chiu 2016, Ricotta & Podani 2017, Ricotta et al. 2021)

Differentiation with Bray-Curtis & Competitors

Bray-Curtis for 2 locations, 2 variants (EG SNP alleles)

- $B = |p_1 - p_2|$
  where p_1, p_2 are proportions of one of the two variants in locations 1,2

Competitors of Bray-Curtis

$$G_{ST} = [H_T - \overline{H_1, H_2}]/H_T \approx F_{ST} = \sigma_p^2/(\bar{p} * \overline{1-p})$$ (Halliburton '04)

where

- H_T is the Hardy-Weinberg (Binomial) expected heterozygosity – ~~ entropy
  eg $H_T = 1 - \bar{p}^2 - (1-\bar{p})^2$; $H_1 = 1 - \bar{p}_1^2 - \bar{q}_1^2$
- $\bar{p}$ is the average *p* over the two locations (1,2).
- σ_p^2 is the variance of *p* values between locations,

Forecasting Differentiation with Bray-Curtis

G_{ST} has an expected equilibrium forecast

- $G_{ST} = 1/(1 + 8N(2m + \mu))$ (Halliburton '04)

N – effective population size at each location
m – dispersal per generation between locations (0≤*m*≤1)
μ – mutation rate per generation ($0 \leq \mu \leq 1$)

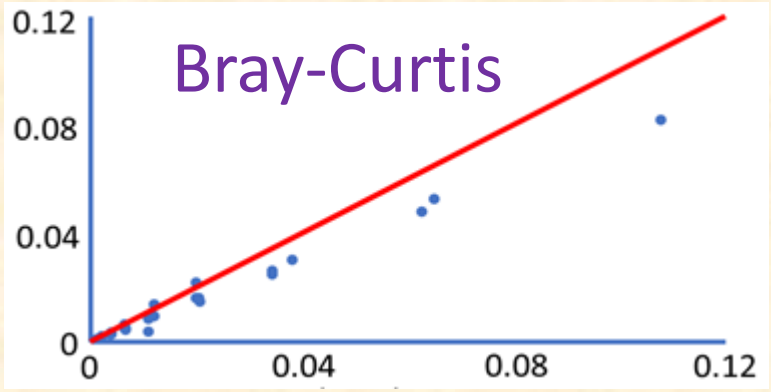
New Equilibrium forecast for Bray-Curtis

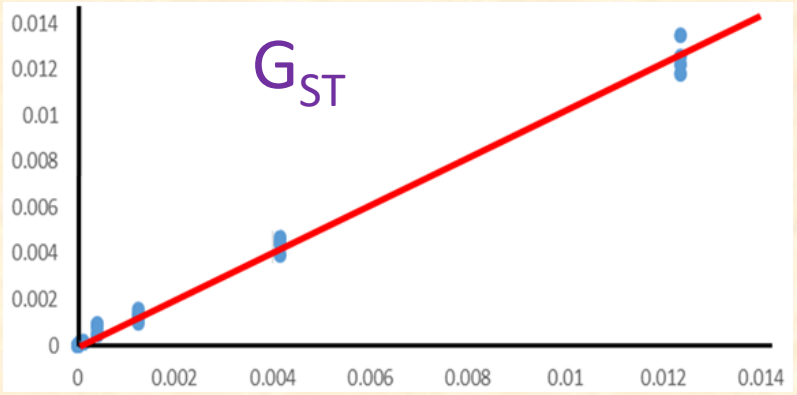
- $B = \sqrt{\frac{2 * {}^2D - 2}{{}^2D(1 + 4N(2m + \mu))}}$ (Sherwin '21)

${}^2D$ – Second order Hill Entropy within-location
or ${}^2D = 1/(1 - H)$ or $H = 1 - 1/{}^2D$

Checks of Forecasting by Simulation

Simulated





Predicted

SO...Corrected Equilibrium forecast For Bray-Curtis

- $$B = 0.8\sqrt{\frac{2\,{}^{2}D-2}{{}^{2}D(1+4N(2m+\mu))}}$$

Sherwin '21

Evaluating Bray Curtis

Forecasting Currently for

- Equilibrium, 2 variants
- Non-equilibrium (time "*t*" after dispersal *m* = 0)

NB BAD dependence on variation within-location ^{2}D

- *also* a problem for G_{ST}
- *not* for other entropic differentations:

Mutual Information, Morisita-Horn/Jost

EG: Evaluation of 2 Entropic methods

Assess gene-frequency differentiation between groups, times or locations

Incorporating functional differences between variants.

Functional Differentiation

Differentiation depends not just upon differences of

- Number of different types
- Frequency of variant types

BUT Also upon how variants differ from one another, eg:

- Colour
- Physiology
- DNA Sequence
- Etc …….

Functional Differentiation – **Past Problems**

- Insensitive to function !!
- Negative diversity !
- Between-location differentiation dependent on variability within-locality
- Etc …..

Functional Diversity - A Novel Approach

- Avoids counterintuitive problems
- Based on 3 Hill entropies (proportion only)

$$ {}^{q}D = \left(\sum_{i=1}^{S} p_i^q\right)^{1/(1-q)} $$

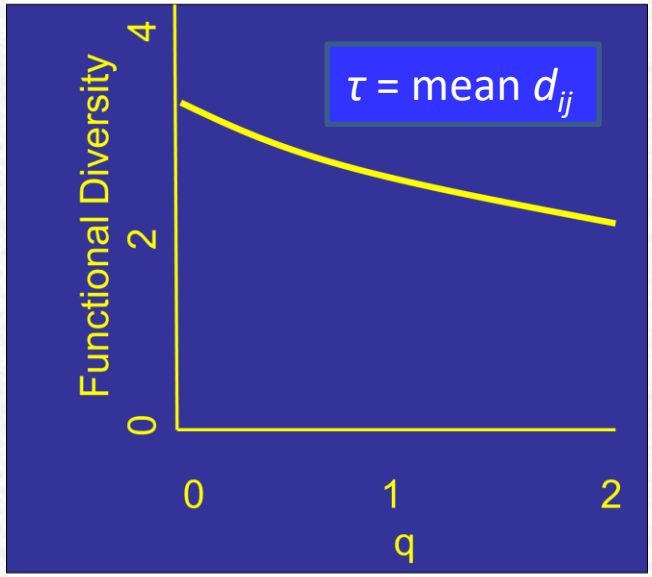
- *S=4*= number of types; p_i=proportions; q=0,1,2 is "order"

- Add in d_{ij} functional distance (Chao et al 2020)
- eg difference of gene's DNA sequence at 1 location

d_{ij}	Seq 1	Seq 2	Seq 3	Seq 4
Seq 1	0	2	8	8
Seq 2	-	0	8	8
Seq 3		-	0	6
Seq 4			-	0

Functional Diversity - A Novel Approach

d_{ij}	Seq 1	Seq 2	Seq 3	Seq 4
Seq 1	0	2	8	8
Seq 2	-	0	8	8
Seq 3		-	0	6
Seq 4			-	0



Tau cutoff applied to d_{ij}

- if $d_{ij} < \tau$, treat as same type
- Low τ emphasizes finer distinctions
- Can use all possible τ-cutoffs, to give τ-profile

(Chao et al '20).

Evaluation of 2 Entropic methods

Bray-Curtis differentiation B

- can be forecast
- depends on within-location diversity
  (some competitors do not)

Functional diversity τ method

- Avoids pitfalls of competitors

‘Entropy’ Topical Collection

- Do Entropic Approaches Improve Understanding of Biology?
- https://www.mdpi.com/journal/entropy/special_issues/entropy_biology
- Editors Sherwin & Niven

Equations Etc

- Sherwin ‘21 Is Bray-Curtis differentiation meaningful in Molecular Ecology? **Molec. Ecol. Res.** (submitted) https://doi.org/10.22541/au.161839260.09775220/v1

- Chao et ‘20 ***Ecol. Monogr.*** 89:e01343

- chao.shinyapps.io/**SpadeR**
- georges.biomatix.org/**dartR**
- biology-assets.anu.edu.au/**GenAlEx**

THANKS to Collaborators & Students
$$: Academia Sinica Taiwan, EU

W.Sherwin@
unsw.edu.au

Functional diversity of order *q*

Chao et al ‘2020

$$= \left(\sum_{i=1}^{S} p_i \left(\sum_{j=1}^{S} [1 - f(d_{ij}(\tau))] p_j \right)^{q-1} \right)^{1/(1-q)}$$