

Il calcolo della somiglianza con dati binari nelle analisi biogeografiche

MAURIZIO BIONDI

*Dipartimento di Scienze Ambientali, Università degli Studi dell'Aquila,
67010 Coppito, L'Aquila;
e-mail: biondi@univaq.it*

Key words: biogeography, Cluster Analysis, binary data, similarity indices

SUMMARY

In this work some similarity indices used in Q-mode and R-mode in biogeographical analyses are compared. The indices were selected by some criterions useful to establish their congruence. The comparison was finalized to put in evidence eventual analogies in the behaviour of different coefficients when they are used in the cluster analysis. Finally, two indices recently proposed, "GECO Coefficient" and "Tripartite Similarity Index", were also discussed.

INTRODUZIONE

Calcolare il livello di somiglianza faunistica e floristica tra differenti aree geografiche è stata da sempre un'esigenza fortemente sentita da molti biogeografi. I primi indici di associazione (cfr. Shi, 1993) proposti per elaborare matrici binarie con dati di presenza/assenza di taxa all'interno di unità geografiche differenziate, risalgono alla seconda metà dell'800, ma è solo nella seconda metà del secolo scorso che, anche grazie alla diffusione di più potenti strumenti di calcolo, hanno avuto il loro massimo sviluppo. Un indice di associazione misura la somiglianza sia tra oggetti (ad esempio, aree geografiche) sia tra descrittori (ad esempio, specie) e rappresenta una procedura di analisi che può essere fine a se stessa oppure un "passaggio obbligato" per successive analisi statistiche di classificazione (ad esempio, Cluster Analysis) o di ordinamento (ad esempio, Principal Coordinates Analysis, Multidimensional Scaling, ecc.). Per evitare qualsiasi tipo di confusione va quindi tenuto presente che negli studi di biogeografia: a) ogni analisi che computa una matrice di associazione tra "aree geografiche" va definita come "Q-mode analysis"; b) ogni analisi che computa una matrice di associazione tra "distribuzioni di taxa" va definita come "R-mode analysis".

Nel calcolo degli indici di associazione per matrici binarie, per ogni confronto a coppie i dati vengono generalmente ordinati sulla base di una semplice

tabella 2x2 (Tab. I). La simbologia adottata comprende: a = numero di taxa (attributi) condivisi dalle due aree geografiche considerate (OGU - Operational Geographic Unit); b = numero di taxa presenti nella prima OGU ma assenti nella seconda; c = numero di taxa presenti nella seconda OGU ma assenti nella prima, d = numero di taxa assenti sia nella prima sia nella seconda OGU ma presenti in altre OGU della regione geografica considerata.

MATERIALI E METODI

A partire da circa 40 diversi indici di associazione sinora proposti in letteratura (cfr. Shi, 1993), è stata operata una selezione sulla base dei seguenti quattro criteri:

		Area x_2		
		1	0	
Area x_1	1	a	b	$a+b$
	0	c	d	
		$a+c$	$b+d$	

dove:
 a = taxa comuni alle due aree x_1 e x_2
 b = taxa esclusivi di x_1
 c = taxa esclusivi di x_2
 d = taxa assenti sia da x_1 che x_2 ma presenti in almeno una delle aree altre considerate

Tab. I - Spiegazione dei parametri utilizzati nelle formule degli indici di associazione considerati.

- criterio 1: il valore dell'indice (S) deve essere compreso tra un limite inferiore "0" (massima dissimiglianza) e un limite superiore "1" (massima somiglianza); sono stati quindi esclusi indici con limiti non definiti in quanto di difficile interpretazione e indici con intervalli definiti ma differenti da "0" e "1", in quanto non direttamente confrontabili;
- criterio 2: quando le OGU, quindi le aree, non condividono alcuna specie, ovvero quando il parametro $a = 0$, allora l'indice $S = 0$;
- criterio 3: quando le OGU condividono tutti i taxa ovvero quando i parametri $b = c = 0$, allora l'indice $S = 1$;
- criterio 4: l'indice S non deve dipendere di norma dal numero di "coassenze" (parametro d), in quanto rappresenta una misura di difficile valutazione nel calcolo della somiglianza; tuttavia sono stati considerati due indici, di Russel & Rao (RR) e di Baroni Urbani & Buser (BUB), che anche se includono il parametro d nella loro formula, considerano questa misura subordinatamente ad a (numero copresenze) e rispettano le seguenti due condizioni: 1) se $a = 0$ allora $S = 0$ per qualsiasi valore di d ; 2) se $d = 0$ allora $S = 0$ solo se $a = 0$.

Sulla base di queste indicazioni, sono stati quindi selezionati per la discussione critica i seguenti indici (cfr. Tab. II): Coefficient of Community (CC), Coincidence

index (CI), Legendre & Legendre's index (LL), Sokal & Sneath's 2 index (SS2), Lance & Williams' index (LW), Kulczynski's 2 index (K2), Coefficient of closeness (CCL), Correlation ratio index (CR), Russel & Rao's index (RR), Baroni Urbani & Buser index (BUB), Simpson's index (SI) e Braun-Blanquet's index (BB).

Tab. II - Formule degli indici di associazione confrontati.

Indice di associazione	Formula
CC - Coefficient of community (= Jaccard's index)	$a/(a+b+c)$
CI - Coincidence index (= Czekanowski's index, Dice's index, Sorensen's index, Pirlor's index, Burt's index)	$2a/(2a+b+c)$
LL - Legendre & Legendre's index	$3a/(3a+b+c)$
SS2 - Sokal & Sneath's 2 index	$(a/2)/[(a/2)+b+c]$
LW - Lance & Williams' index (= Bray & Curtis nonmetric coefficient)	$1-[(b+c)/(2a+b+c)]$
K2 - Kulczynski's 2 index	$\{[a/(a+b)]+[a/(a+c)]\}/2$
CCL - Coefficient of closeness (= Ochiai's index; Otsuka's index)	$a/[(a+b)(a+c)]^{1/2}$
CR - Correlation ratio index (= Sorgenfrei's index)	$a^2/[(a+b)(a+c)]$
RR - Russel & Rao's index	$a/(a+b+c+d)$
BUB - Baroni Urbani & Buser's index	$[(\sqrt{ad})+a]/[(\sqrt{ad})+a+b+c]$
SI - Simpson's index	$a/(a+b)$ con $(a+b)\leq(a+c)$
BB - Braun-Blanquet's index	$a/(a+c)$ con $(a+b)\leq(a+c)$

Va tuttavia sottolineato che anche altre proprietà vengono ritenute “utili” per definire un “buon” indice di associazione (cfr. Shi, 1993), come ad esempio:

- che l'indice S sia preferibilmente “metrico”; questa proprietà è particolarmente importante quando l'indice viene utilizzato in metodi di ordinamento (ad esempio, Principal Coordinates Analysis, Multidimensional Scaling, ecc.) perché in questo caso la distorsione nel processo di conversione della matrice di somiglianza in matrice delle distanze sarà minima; quando invece S viene utilizzato in metodi di classificazione (ad esempio, Cluster Analysis) non è necessario che sia metrico. Tra gli indici considerati non risultano metrici: CI, SI, K2 e CCL;
- che l'indice S sia preferibilmente “simmetrico”, ovvero che entrambi i parametri b e c siano inclusi nella formula; nel caso di elevata sproporzione nella dimensione delle OGU considerate, tuttavia, l'utilizzo di indici asimmetrici, come SI e BB, può risultare particolarmente indicato.

Nella presente analisi vengono inoltre discussi anche due indici recentemente proposti: il “Tripartite Similarity Index” (TSI) (Tulloss, 1997), basato su tre differenti “funzioni dei costi” (U, S, R) che possono essere valutate separatamente

o unitamente per fornire dettagliate informazioni sui fattori che determinano i valori di somiglianza calcolati, e il “GECO Coefficient” (Hennig e Hausdorf, 2006) indicato per calcolare la somiglianza tra le distribuzioni delle specie anche sulla base delle distanze delle aree geografiche considerate.

RISULTATI

Gli indici di associazione “tradizionali”

I 12 indici di associazione, selezionati sulla base dei criteri sopra descritti (Tab. II), sono stati confrontati utilizzando come dati di riferimento la presenza/assenza nelle isole di Capo Verde dei 26 taxa (specie e sottospecie) che compongono la fauna erpetologica di questo arcipelago (Arechavaleta et al., 2005) (Tab. III), le cui conoscenze risultano sufficientemente complete e caratterizzate da elevata stabilità.

Tab. III - Matrice dei dati utilizzati per i confronti (cfr. testo).

	S_Antao	S_Vicente	S_Luzia	Branco	Raso	S_Nicolau	Sal	Boavista	Maio	Santiago	Fogo	Brava
tax_1	0	0	0	0	0	0	0	1	0	0	0	0
tax_2	1	1	1	0	0	1	0	0	0	1	1	1
tax_3	0	0	0	0	0	0	1	0	0	0	0	0
tax_4	0	0	0	0	1	0	0	0	0	0	0	0
tax_5	1	1	0	0	0	0	1	1	0	1	1	0
tax_6	0	1	0	0	0	0	0	0	0	0	0	0
tax_7	1	0	0	0	0	0	0	0	0	0	0	0
tax_8	0	0	0	0	0	1	0	0	0	0	0	0
tax_9	0	0	0	0	1	0	0	0	0	0	0	0
tax_10	0	1	1	1	0	0	0	0	0	0	0	0
tax_11	0	0	0	0	0	1	0	0	0	1	1	0
tax_12	0	0	0	1	0	0	0	0	0	0	0	0
tax_13	0	0	0	0	1	0	0	0	0	0	0	0
tax_14	0	0	0	0	0	0	0	1	0	0	0	0
tax_15	0	0	0	0	0	0	0	0	1	0	0	0
tax_16	0	0	0	0	0	0	0	0	0	0	0	1
tax_17	0	0	0	0	0	0	0	0	0	0	1	1
tax_18	0	0	0	0	0	0	0	0	0	1	0	0
tax_19	1	0	0	0	0	0	0	0	0	0	0	0
tax_20	0	0	0	0	0	0	0	1	1	1	1	1
tax_21	0	0	0	0	0	1	0	0	0	0	1	0
tax_22	0	0	0	0	0	0	0	0	1	0	0	0
tax_23	0	0	0	0	0	0	1	1	0	0	0	0
tax_24	0	0	0	0	0	0	0	0	0	1	1	0
tax_25	0	1	1	1	1	0	0	0	0	0	0	0
tax_26	0	0	0	0	0	0	0	0	0	1	1	1

Le matrici di somiglianza sono state elaborate attraverso la tecnica della Cluster Analysis, utilizzando il metodo WPGMA (Weighted Pair Group Method using the Arithmetic averaging) (Sneath e Sokal, 1973), per costruire i dendrogrammi riportati nelle Figg. 1-5, 7-8, 10-14.

I valori di somiglianza calcolati con i 12 indici selezionati (Tab. IV) sono stati confrontati utilizzando l'indice di cograduazione di Kendall (cfr. Palumbo, 1977), il quale risulta particolarmente indicato per valutare la concordanza tra due serie di valori allo scopo di mettere in evidenza eventuali analogie nel loro comportamento. Dall'analisi della tabella di confronto dell'indice di cograduazione (Tab. V) si può osservare come alcuni dei valori ottenuti equivalgano a 1, evidenziando il comportamento totalmente "monotono" di parte degli indici di associazione confrontati.

A tale proposito va osservato che il gruppo di coefficienti monotoni più consistente comprende CC, CI, LL, SS2 e LW (Fig. 6) il cui calcolo restituisce, al di là del metodo di clustering utilizzato, esattamente la stessa struttura di dendrogramma. I dendrogrammi ottenuti si differenziano, infatti, soltanto nei livelli di somiglianza ai quali si formano i vari raggruppamenti (clusters) (Figg. 1-5). Questo risultato può essere facilmente spiegato osservando le relative formule degli indici attribuiti a questo gruppo (Tab. II), le quali si differenziano esclusivamente sulla base del diverso contributo che il parametro a (attributi in comune) fornisce al calcolo della somiglianza. Il contributo più basso viene infatti fornito nell'indice SS2, che considera il parametro a pari ad $a/2$, mentre in CC equivale ad a , in CI e LW equivale a $2a$ e, infine, in LL è pari a $3a$. Nei rispettivi dendrogrammi (Figg. 1-5), quindi, si osserva un graduale spostamento verso valori più elevati nei livelli di aggregazione, partendo da SS2 e arrivando a LL, passando per CC, CI e LW; va evidenziata inoltre la perfetta coincidenza, nella struttura e nei valori di aggregazione dei clusters, tra i dendrogrammi ottenuti con l'applicazione degli indici CI e LW (Figg. 3-4).

Un secondo gruppo comprendente indici con comportamento monotono include CR e CCL (Tabb. II, V; Figg. 7-9). Anche in questo caso, si può osservare dall'analisi delle rispettive formule come questi due indici forniscano la stessa struttura di dendrogramma ma con valori diversi nei livelli di aggregazione, i quali risultano più elevati in CCL, dove il denominatore " $[(a+b)(a+c)]$ " è posto sotto radice quadrata, e meno elevati in CR, dove lo stesso denominatore non è posto sotto radice e il numeratore (parametro a : attributi in comune) è elevato al quadrato. Per identici valori dei parametri a , b e c , quindi, l'indice CR fornirà sempre valori proporzionalmente più bassi rispetto a quelli ottenuti con CCL, ovviamente a eccezione dei valori limite 0 e 1 (Fig. 9) che coincidono nei due indici.

Nell'ambito degli indici simmetrici che nel contempo non includono nel loro calcolo le coassenze (parametro d), l'indice K2 non è risultato monotono con

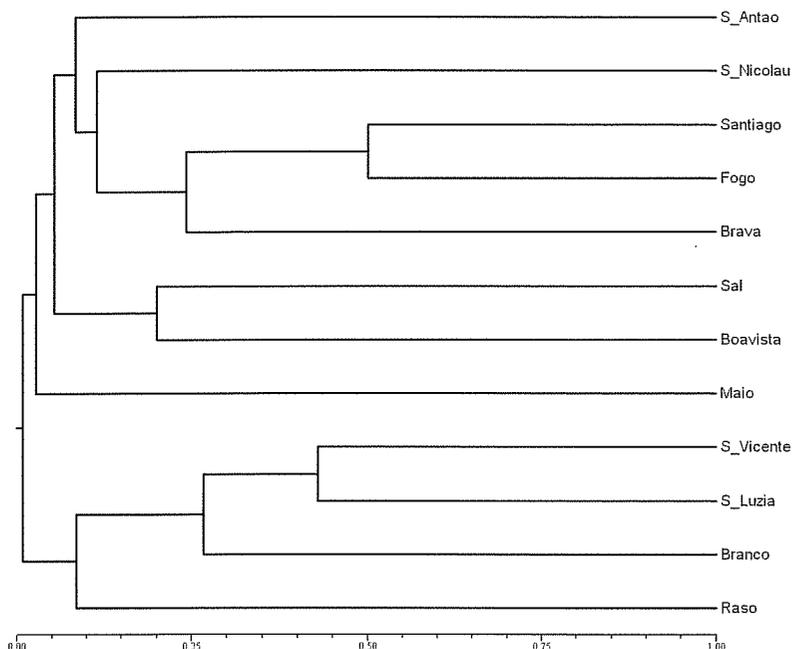


Fig. 1 - Dendrogramma ottenuto con SS2 e WPGMA (cfr. testo).

nessuno degli altri indici considerati, pur presentando un'elevata somiglianza nei valori con altri coefficienti, in particolare con CCL, rispetto al quale risulta meno sensibile alle diverse dimensioni delle OGU confrontate in termini di numero di attributi presenti (Tab. IV).

Per quanto riguarda gli indici RR e BUB, essi includono nella loro formula anche il parametro d (numero di coassenze), il quale però contribuisce al calcolo della somiglianza in modo subordinato alla presenza di almeno un attributo positivo condiviso ($a \neq 0$). Dal calcolo dell'indice di cograduazione di Kendall (Tab. V), si può osservare come il valore più basso (0,793) sia proprio riferito al confronto tra BUB e RR. Ancora una volta l'analisi attenta delle due rispettive formule ci fornisce la spiegazione di questa significativa differenza nei relativi valori di somiglianza. Va innanzitutto evidenziato come sia in BUB sia in RR, nel caso il numero di coassenze sia uguale a "0", entrambi gli indici si riducano alla formula $a/(a+b+c)$, ovvero risultano identici a CC. Tuttavia, proprio il diverso peso che BUB e RR attribuiscono al parametro d spiega le notevoli differenze tra i valori ottenuti. Nel primo caso le coassenze compaiono sotto radice quadrata come prodotto $a \cdot d$, sia al numeratore che al denominatore, mentre in RR il parametro d compare come valore assoluto esclusivamente al denominatore, abbassando, all'aumentare del numero di coassenze, in

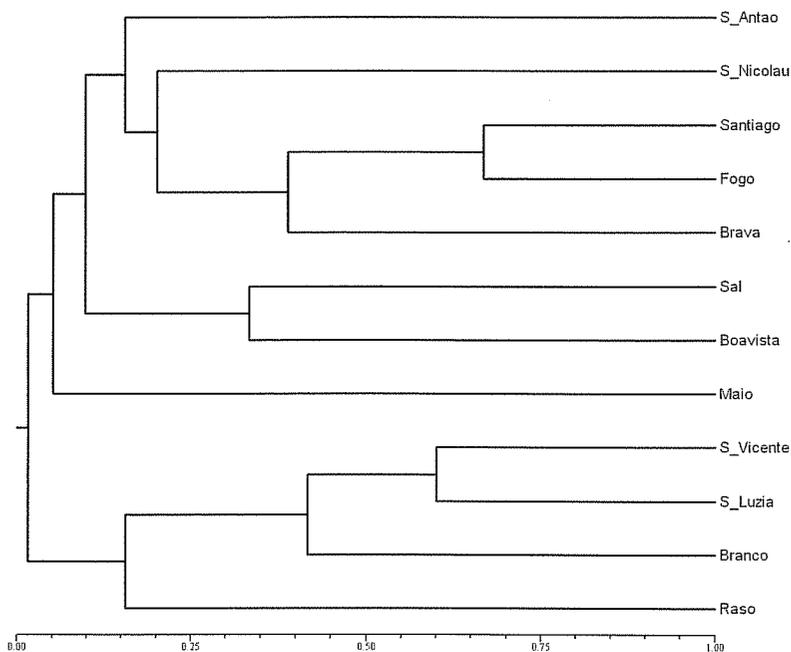


Fig. 2 - Dendrogramma ottenuto con CC e WPGMA (cfr. testo).

modo molto sensibile il valore dell'indice (cfr. Tab. IV). Ad esempio, nel set di dati qui utilizzati per confrontare i diversi indici considerati (Tab. III), il numero di coassenze tra le coppie di OGUs risulta notevolmente elevato e quindi le differenze nei valori di somiglianza e nei dendrogrammi ottenuti dall'applicazione di BUB e RR sono molto differenti (Tab. IV; Figg. 11-12). È importante inoltre sottolineare come le coassenze possano risultare "utili" nel calcolo della somiglianza solo nel caso in cui la conoscenza della distribuzione degli attributi (taxa) nelle OGUs (aree) considerate sia sufficientemente elevata e stimabile almeno al 70% della situazione reale (Biondi, 1988). Inoltre, l'indice RR andrebbe evitato quando le coassenze sono in numero molto più elevato rispetto al numero delle copresenze, in quanto in questo caso esprimerebbe valori decisamente bassi, come risulta dall'esempio qui riportato, riferito ai Rettili dell'arcipelago di Capo Verde (Tab. IV; Fig. 11).

Nel confronto tra gli indici di associazione proposti, sono stati infine considerati SI e BB che hanno la caratteristica di essere indici "asimmetrici", in quanto includono nella loro formula (cfr. Tab. II) o il solo parametro b (attributi esclusivi della OGU con il minor numero di taxa) come in SI, oppure il solo parametro c (attributi esclusivi della OGU con il maggior numero di taxa) come in BB. Questi indici trovano un'applicazione utile nel caso in cui le OGUs

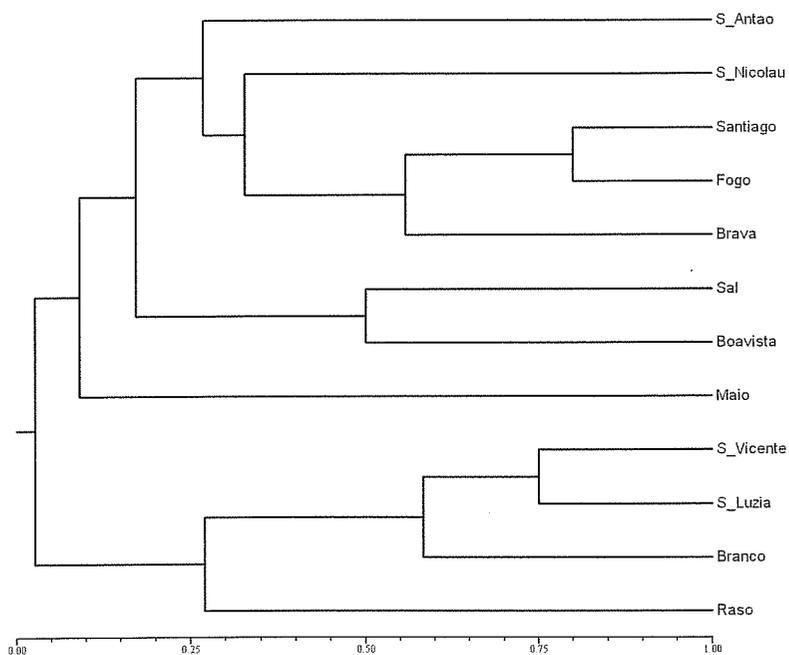


Fig. 3 - Dendrogramma ottenuto con CI e WPGMA (cfr. testo).

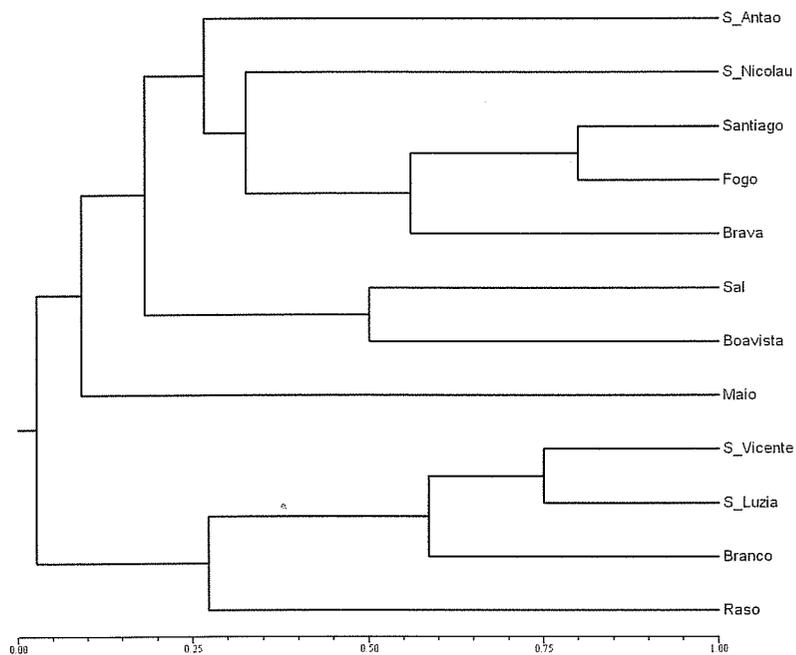


Fig. 4 - Dendrogramma ottenuto con LW e WPGMA (cfr. testo).

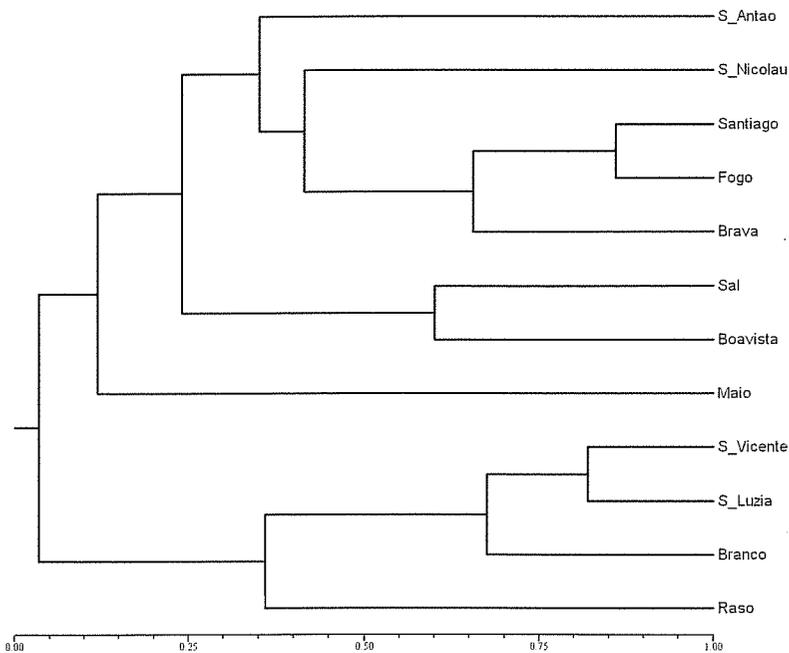


Fig. 5 - Dendrogramma ottenuto con LL e WPGMA (cfr. testo).

confrontate siano tra loro particolarmente sbilanciate in termini di numero di attributi presenti. Tale sbilanciamento potrebbe essere dovuto, ad esempio, a forti differenze nella superficie delle aree considerate (piccole isole confrontate con grandi isole). I valori ottenuti con il calcolo dei due indici SI e BB mostrano significative differenze (cfr. Tab. IV) con un valore dell'indice di cograduazione di Kendall pari a 0,844 (cfr. Tab. V). I due dendrogrammi ottenuti (Figg. 13-14) mostrano tra loro una struttura nei raggruppamenti fortemente differenziata, dovuta alla diversa "filosofia" dei due indici: SI considera infatti il numero di copresenze in rapporto al numero degli attributi totali presenti nella OGU che ne presenta di meno, sovrastimando la loro somiglianza; mentre BB considera il numero di copresenze in rapporto al numero di attributi totali presenti nella OGU che ne presenta di più, sottostimando la loro somiglianza.

Per il calcolo degli indici di associazione sopra discussi, diversi algoritmi sono disponibili gratuitamente in rete, in particolare come applicazioni su piattaforma Java. Tuttavia, è il software commerciale che risulta più completo e più "friendly" nell'utilizzo. Tra i programmi più diffusi si ricordano i prodotti della Exeter Software, quali Syn-Tax 2000 e NtSyspc; MVSP della Kovach Computing Services e gli applicativi operanti in ambiente Microsoft Excel, quali XLStat e StatistiXL.

Tab. IV - Matrice dei valori di somiglianza ottenuti con il calcolo degli indici di associazione (cfr. resto) (segue).

	a	b	c	d	CC	CI	LL	SS2	LW	K2	CCL	CR	RR	BUB	SI	BB
sa-sv	2	2	3	19	0,29	0,44	0,55	0,17	0,44	0,45	0,45	0,20	0,08	0,62	0,50	0,40
sa-sl	1	3	2	20	0,17	0,29	0,38	0,09	0,29	0,29	0,29	0,08	0,04	0,52	0,33	0,25
sa-br	0	4	3	19	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00
sa-ra	0	4	4	18	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00
sa-sn	1	3	3	19	0,14	0,25	0,33	0,08	0,25	0,25	0,25	0,06	0,04	0,47	0,25	0,25
sa-sal	1	3	2	20	0,17	0,29	0,38	0,09	0,29	0,29	0,29	0,08	0,04	0,52	0,33	0,25
sa-bo	1	3	4	18	0,13	0,22	0,30	0,07	0,22	0,23	0,22	0,05	0,04	0,43	0,25	0,20
sa-ma	0	4	3	19	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00
sa-st	2	2	5	17	0,22	0,36	0,46	0,13	0,36	0,39	0,38	0,14	0,08	0,53	0,50	0,29
sa-fo	2	2	6	16	0,20	0,33	0,43	0,11	0,33	0,38	0,35	0,13	0,08	0,49	0,50	0,25
sa-bv	1	3	4	18	0,13	0,22	0,30	0,07	0,22	0,23	0,22	0,05	0,04	0,43	0,25	0,20
sv-sl	3	2	0	21	0,60	0,75	0,82	0,43	0,75	0,80	0,77	0,60	0,12	0,85	1,00	0,60
sv-br	2	3	1	20	0,33	0,50	0,60	0,20	0,50	0,53	0,52	0,27	0,08	0,68	0,67	0,40
sv-ra	1	4	3	18	0,13	0,22	0,30	0,07	0,22	0,23	0,22	0,05	0,04	0,43	0,25	0,20
sv-sn	1	4	3	18	0,13	0,22	0,30	0,07	0,22	0,23	0,22	0,05	0,04	0,43	0,25	0,20
sv-sal	1	4	2	19	0,14	0,25	0,33	0,08	0,25	0,27	0,26	0,07	0,04	0,47	0,33	0,20
sv-bo	1	4	4	17	0,11	0,20	0,27	0,06	0,20	0,20	0,20	0,04	0,04	0,39	0,20	0,20
sv-ma	0	5	3	18	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00
sv-st	2	3	5	16	0,20	0,33	0,43	0,11	0,33	0,34	0,34	0,11	0,08	0,49	0,40	0,29
sv-fo	2	3	6	15	0,18	0,31	0,40	0,10	0,31	0,33	0,32	0,10	0,08	0,45	0,40	0,25
sv-bv	1	4	4	17	0,11	0,20	0,27	0,06	0,20	0,20	0,20	0,04	0,04	0,39	0,20	0,20
sl-br	2	1	1	22	0,50	0,67	0,75	0,33	0,67	0,67	0,67	0,44	0,08	0,81	0,67	0,67
sl-ra	1	2	3	20	0,17	0,29	0,38	0,09	0,29	0,29	0,29	0,08	0,04	0,52	0,33	0,25
sl-sn	1	2	3	20	0,17	0,29	0,38	0,09	0,29	0,29	0,29	0,08	0,04	0,52	0,33	0,25
sl-sal	0	3	3	20	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00
sl-bo	0	3	5	18	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00
sl-ma	0	3	3	20	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00
sl-st	1	2	6	17	0,11	0,20	0,27	0,06	0,20	0,24	0,22	0,05	0,04	0,39	0,33	0,14
sl-fo	1	2	7	16	0,10	0,18	0,25	0,05	0,18	0,23	0,20	0,04	0,04	0,36	0,33	0,13
sl-bv	1	2	4	19	0,14	0,25	0,33	0,08	0,25	0,27	0,26	0,07	0,04	0,47	0,33	0,20
br-ra	1	2	3	19	0,17	0,29	0,38	0,09	0,29	0,29	0,29	0,08	0,04	0,52	0,33	0,25
br-sn	0	3	4	19	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00
br-sal	0	3	3	20	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00
br-bo	0	3	5	18	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00
br-ma	0	3	3	20	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00
br-st	0	3	7	16	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00
br-fo	0	3	8	15	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00
br-bv	0	3	5	18	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00
ra-sn	0	4	4	18	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00
ra-sal	0	4	3	19	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00
ra-bo	0	4	5	17	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00
ra-ma	0	4	3	19	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00
ra-st	0	4	7	15	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00
ra-fo	0	4	8	14	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00
ra-bv	0	4	5	17	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00

Tab. IV - Matrice dei valori di somiglianza ottenuti con il calcolo degli indici di associazione (cfr. testo).

	a	b	c	d	CC	CI	LL	SS2	LW	K2	CCL	CR	RR	BUB	SI	BB
sn-sal	0	4	3	19	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00
sn-bo	0	4	5	17	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00
sn-ma	0	4	3	19	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00
sn-st	2	2	5	17	0,22	0,36	0,46	0,13	0,36	0,39	0,38	0,14	0,08	0,53	0,50	0,29
sn-fo	3	1	5	17	0,33	0,50	0,60	0,20	0,50	0,56	0,53	0,28	0,12	0,63	0,75	0,38
sn-bv	1	3	4	18	0,13	0,22	0,30	0,07	0,22	0,23	0,22	0,05	0,04	0,43	0,25	0,20
sal-bo	2	1	3	20	0,33	0,50	0,60	0,20	0,50	0,53	0,52	0,27	0,08	0,68	0,67	0,40
sal-ma	0	3	3	20	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00
sal-st	1	2	6	17	0,11	0,20	0,27	0,06	0,20	0,24	0,22	0,05	0,04	0,39	0,33	0,14
sal-fo	1	2	7	16	0,10	0,18	0,25	0,05	0,18	0,23	0,20	0,04	0,04	0,36	0,33	0,13
sal-bv	0	3	5	18	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00
bo-ma	1	4	2	19	0,14	0,25	0,33	0,08	0,25	0,27	0,26	0,07	0,04	0,47	0,33	0,20
bo-st	2	3	5	16	0,20	0,33	0,43	0,11	0,33	0,34	0,34	0,11	0,08	0,49	0,40	0,29
bo-fo	2	3	6	15	0,18	0,31	0,40	0,10	0,31	0,33	0,32	0,10	0,08	0,45	0,40	0,25
bo-bv	2	3	4	17	0,22	0,36	0,46	0,13	0,36	0,37	0,37	0,13	0,08	0,53	0,40	0,33
ma-st	1	2	6	17	0,11	0,20	0,27	0,06	0,20	0,24	0,22	0,05	0,04	0,39	0,33	0,14
ma-fo	1	2	7	16	0,10	0,18	0,25	0,05	0,18	0,23	0,20	0,04	0,04	0,36	0,33	0,13
ma-bv	1	2	4	19	0,14	0,25	0,33	0,08	0,25	0,27	0,26	0,07	0,04	0,47	0,33	0,20
st-fo	6	1	2	17	0,67	0,80	0,86	0,50	0,80	0,80	0,80	0,64	0,23	0,84	0,86	0,75
st-bv	3	4	2	17	0,33	0,50	0,60	0,20	0,50	0,51	0,51	0,26	0,12	0,63	0,60	0,43
fo-bv	4	4	1	17	0,44	0,62	0,71	0,29	0,62	0,65	0,63	0,40	0,15	0,71	0,80	0,50

Tab. V - Matrice dei valori del coefficiente di cograduazione di Kendall calcolato tra i valori di somiglianza ottenuti con il calcolo degli indici di associazione (cfr. testo).

	CC	CI	LL	SS2	LW	CR	K2	CCL	RR	BUB	SI	BB
CC	1,000	1,000	1,000	1,000	1,000	0,981	0,950	0,981	0,914	0,956	0,911	0,965
CI		1,000	1,000	1,000	1,000	0,981	0,950	0,981	0,914	0,956	0,911	0,965
LL			1,000	1,000	1,000	0,981	0,950	0,981	0,914	0,956	0,911	0,965
SS2				1,000	1,000	0,981	0,950	0,981	0,914	0,956	0,911	0,965
LW					1,000	0,981	0,950	0,981	0,914	0,956	0,911	0,965
CR						1,000	0,965	1,000	0,911	0,944	0,924	0,950
K2							1,000	0,965	0,911	0,911	0,960	0,914
CCL								1,000	0,911	0,944	0,924	0,950
RR									1,000	0,793	0,866	0,883
BUB										1,000	0,868	0,943
SI											1,000	0,844
BB												1,000

Tab. VI - Matrice delle distanze medie tra le isole dell'arcipelago di Capo Verde (in km).

	SA	SV	SL	Br	Ra	SN	Sal	Bo	Ma	St	Fo	Bv
S_Antao	0	32	56	69	77	109	244	274	296	277	251	250
S_Vicente		0	26	38	48	83	219	247	266	245	223	223
S_Luzia			0	15	23	56	195	221	241	224	207	212
Branco				0	9	48	188	210	227	209	193	200
Raso					0	37	178	201	219	202	190	197
S_Nicolau						0	142	165	195	186	188	202
Sal							0	69	170	199	254	282
Boavista								0	107	145	214	248
Maio									0	51	134	170
Santiago										0	83	119
Fogo											0	36
Brava												0

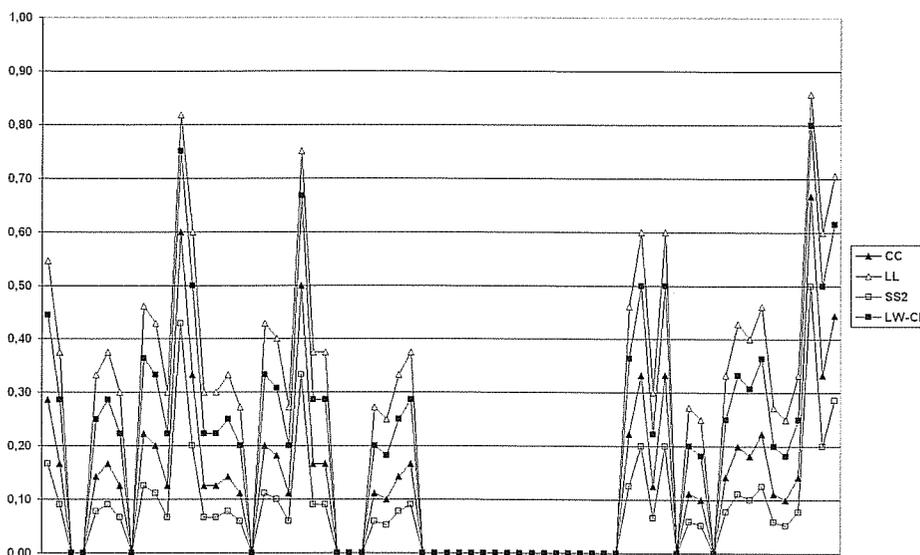


Fig. 6 - Andamento monotono degli indici CC, LL, SS2, LW, CI (cfr. testo).

Il “Tripartite Similarity Index”

Il “Tripartite Similarity Index” (TSI) è stato sviluppato da Tulloss (1997) nel tentativo di evitare alcune discrepanze e inconsistenze evidenziate con l’utilizzo

dei tradizionali indici di associazione. Quest'ultimi, infatti, risultano non o scarsamente sensibili a: 1) le dimensioni delle due liste di taxa confrontate: grandi differenze nelle dimensioni dovrebbero invece portare a una riduzione nel valore dell'indice di somiglianza; 2) la dimensione del contingente rappresentato dai taxa condivisi dalle due liste: un aumento nella differenza tra il numero di taxa della lista più piccola e il numero di taxa in comune dovrebbe portare invece a una riduzione del valore dell'indice di somiglianza; 3) la percentuale di taxa che ciascuna delle due liste presenta in comune con l'altra: un aumento di questa percentuale dovrebbe invece portare a un conseguente aumento del valore dell'indice di somiglianza (Tulloss, 1997).

Lo scopo del TSI è quello, quindi, di ovviare alla scarsa "sensibilità" degli indici di associazione tradizionali. Esso viene espresso come prodotto di tre "funzioni dei costi" (U, S, R), sulla base delle seguenti formule:

$$1) \quad TSI = \sqrt{U \cdot S \cdot R} \text{ con } 0 \leq TSI \leq 1$$

$$2) \quad U = \frac{\log\left(1 + \frac{\min(b,c) + a}{\max(b,c) + a}\right)}{\log 2} \text{ con } 0 < U \leq 1$$

Questa prima funzione riduce il valore della somiglianza nel caso in cui le coppie di liste di taxa presentino una elevata differenza tra il numero di attributi esclusivi; ovvero, più è elevata la differenza $|b-c|$ meno è elevato il valore della funzione U e quindi di TSI;

$$3) \quad S = \sqrt{\frac{\log 2}{\log\left(2 + \frac{\min(b,c)}{a+1}\right)}} \text{ con } 0 < S \leq 1$$

Questa seconda funzione considera la differenza tra il numero di attributi comuni alla coppia di liste di taxa in rapporto alla lista che comprende il minor numero di taxa; quando questa differenza aumenta, diminuisce di conseguenza il valore di TSI e viceversa;

$$4) \quad R = \frac{\log\left(1 + \frac{a}{a+b}\right) \cdot \log\left(1 + \frac{a}{a+c}\right)}{(\log 2)^2} \text{ con } 0 \leq R \leq 1$$

Questa terza funzione, infine, considera il numero di taxa comuni alle due liste rispetto al numero di taxa totali inclusi sia nella lista più grande ($a+b$) e sia in quella più piccola ($a+c$); il valore della funzione R, e quindi dell'indice TSI,

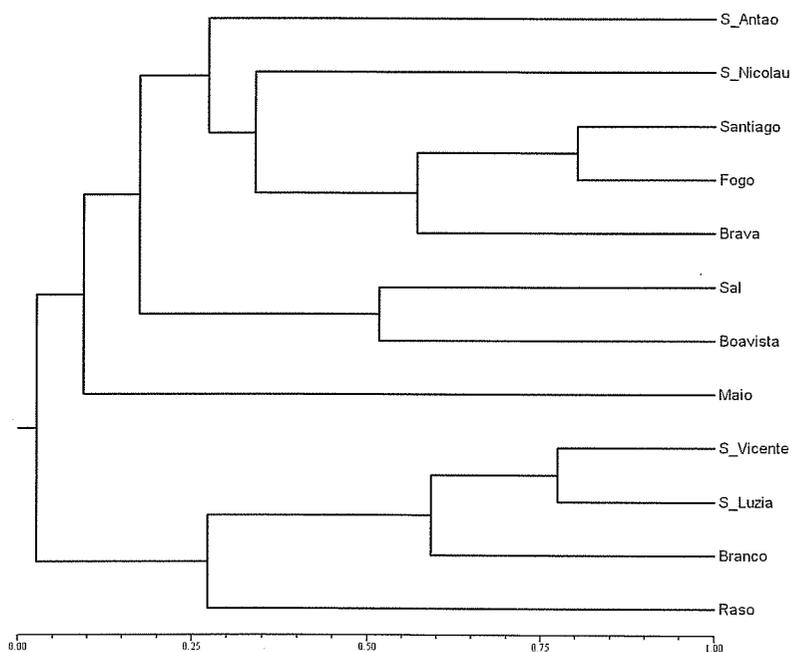


Fig. 7 - Dendrogramma ottenuto con CCL e WPGMA (cfr. testo).

aumenta quando la percentuale dei taxa in comune aumenta per entrambe le liste.

Nelle tre funzioni sopra riportate i parametri *a*, *b* e *c* assumono lo stesso significato di quanto riportato in Tab. I.

Anche in questo caso, l'indice TSI è stato calcolato sulla base dei dati in Tab. III. Ciò è stato possibile grazie a un calcolatore disponibile on-line e accessibile al sito <http://www.amanitabear.com> che permette, dopo una breve procedura di registrazione, di calcolare la matrice dei valori del TSI su dati inseriti dell'utente. Il dendrogramma costruito utilizzando il TSI e il metodo di clustering WPGMA (Sneath e Sokal, 1973) (Fig. 15) si discosta da quelli ottenuti con il calcolo dei tradizionali indici di associazione principalmente per i seguenti punti:

- rispetto alla gran parte degli indici di associazione tradizionali considerati (CC, CI, LL, SS2, LW, K2, CCL, CR, RR, BUB e SI) (Figg. 1-5; 7-8; 10-13) la struttura del dendrogramma si differenzia principalmente per la posizione di S_Antao, che viene collocato dal TSI nel cluster (S_Antao(S_Vicente(S_Luzia-Branco))), mentre gli indici tradizionali sopra elencati lo inseriscono nel cluster (S_Antao(S_Nicolau(Brava(Santiago-Fogo))));
- gli indici tradizionali, a eccezione di BB, collocano il cluster (Branco(S_

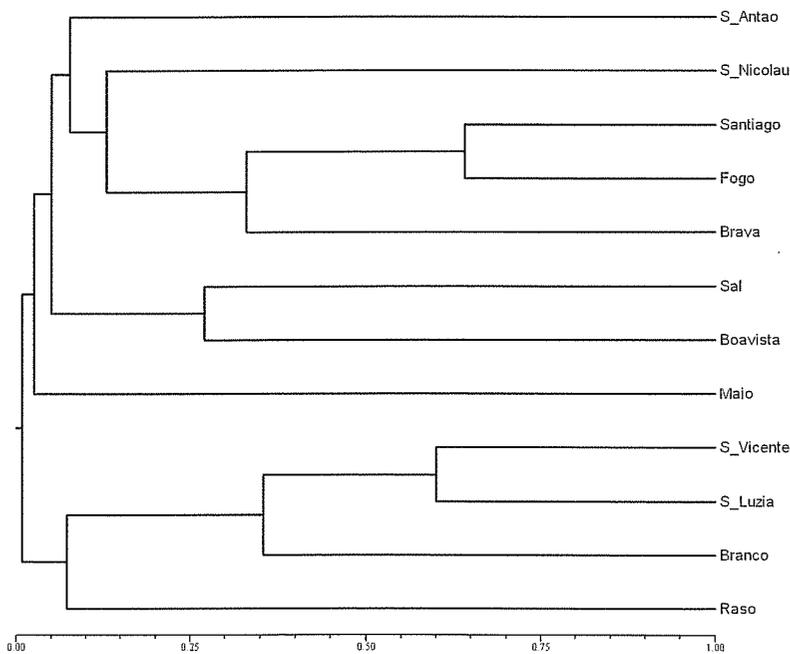


Fig. 8 - Dendrogramma ottenuto con CR e WPGMA (cfr. testo).

- Vicente-S_Luzia)) insieme a Raso in una posizione nettamente distinta rispetto alle altre OGUs, mentre TSI situa questo gruppo a ridosso del cluster (S_Nicolau(Brava(Santiago-Fogo)));
- inoltre va rilevata la collocazione isolata che TSI attribuisce a Raso, OGU che gli indici tradizionali (sempre con l'eccezione di BB) posizionano nel cluster (Raso(Branco(S_Luzia-S_Vicente)));
 - infine, va messo in evidenza come un indice particolarmente complesso dal punto di vista del calcolo, come il TSI, restituisca attraverso la Cluster Analysis (WPGMA) una struttura di dendrogramma in gran parte sovrapponibile con quella ottenuta con l'indice BB, caratterizzato quest'ultimo da una formula estremamente semplice (cfr. Tab. II; Figg. 14-15).

Valutare la "bontà" oggettiva di TSI rispetto ai tradizionali indici di associazione è ancora del tutto prematuro, in quanto, nonostante la sua maggior complessità di calcolo finalizzata a rispettare alcune delle incongruenze che caratterizzano gli indici generalmente utilizzati per il calcolo della somiglianza con dati binari, non risulta ancora sperimentato in modo soddisfacente. Alcuni aspetti critici sono stati tuttavia evidenziati da Deutsch et al. (2006).

Molto recentemente Shils (2006), sulla base del TSI sopra descritto, ha

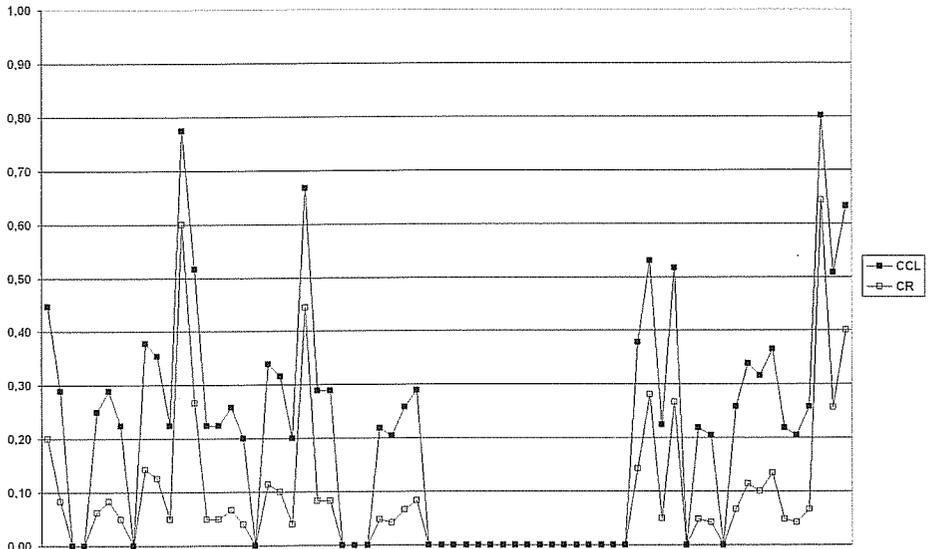


Fig. 9 - Andamento monotono degli indici CCL e CR (cfr. testo).

proposto una versione modificata definita “Tripartite Biogeographical Index” (TBI), nella quale analogamente al TSI vengono prese in considerazione tre “funzioni dei costi”, e più precisamente:

- funzione F (Frequency), intesa come il numero di aree nelle quali un dato taxon viene rilevato, espresso come frazione del numero totale delle aree investigate;
- funzione C (Connectivity), intesa come misura delle connessioni tra aree di distribuzione disgiunte (o popolazioni) di un dato taxon nell’ambito di una data regione geografica; C è funzione dell’isolamento e della dimensione delle aree di distribuzione;
- funzione G (Grouping), che rappresenta il modello di diffusione di un singolo taxon in una data area in relazione alla regione geografica considerata e alla distanza minima necessaria a unire le aree di distribuzione disgiunte.

Va messo in evidenza che per il calcolo del TBI, oltre che alla matrice della presenza/assenza dei singoli taxa nelle unità geografiche considerate, è necessario prendere in esame anche i loro areali completi, in quanto il TBI restituisce un valore di “affinità” tra modelli di distribuzione e non valori di somiglianza tra unità geografiche sulla base della composizione tassonomica dei loro popolamenti. Tuttavia per ulteriori approfondimenti sul TBI si rimanda a Shils (2006).

Il “GECO Coefficient”

Il “GECO Coefficient” (dG), recentemente proposto da Hennig e Hausdorf (2006), prende il nome dal fatto che “the GEographic distance as well as

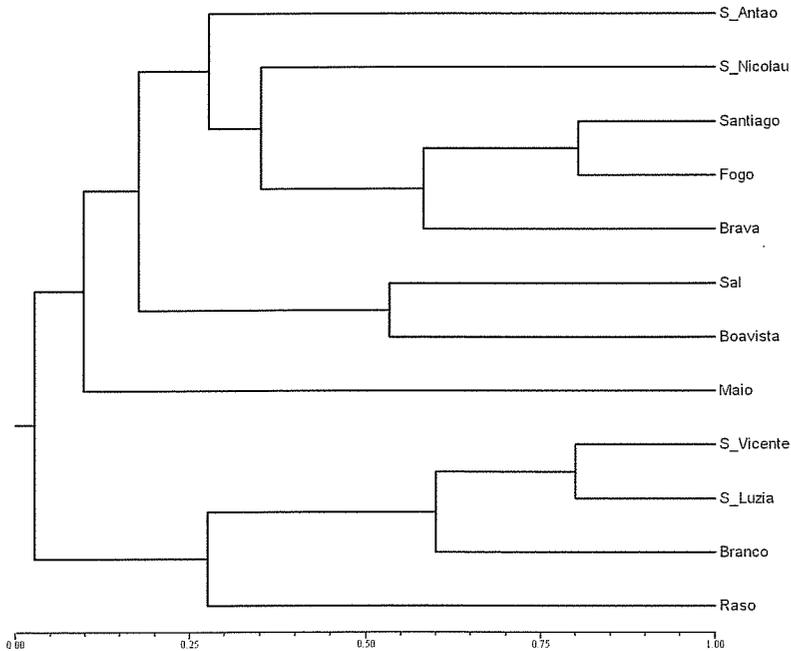


Fig. 10 - Dendrogramma ottenuto con K2 e WPGMA (cfr. testo).

COngruence of ranges is considered". Infatti, la particolarità di questo indice risiede nel fatto che il calcolo della dissimiglianza (distanza) tra distribuzioni di taxa (R-mode analysis) considera non solo la percentuale di unità geografiche condivise, ma anche le relazioni geografiche tra le unità occupate. Per rendere più agevole il confronto con gli altri indici sopra discussi, in questo lavoro il coefficiente GECO viene considerato come un coefficiente di somiglianza (e non di distanza), stabilendo i suoi valori limite da "0" (massima dissimiglianza) a "1" (massima somiglianza).

Rispetto ai tradizionali indici di somiglianza/distanza che non considerano le relazioni spaziali tra le unità geografiche, il coefficiente GECO possiede due vantaggi: 1) riflette meglio di altri indici la somiglianza calcolata tra le unità geografiche di una data regione occupate dai taxa esaminati; 2) risulta meno sensibile agli effetti dovuti a campionamenti insufficienti, quali, ad esempio, presenza di discontinuità nella distribuzione dei taxa considerati.

Questo indice, quindi, si propone di risolvere un aspetto cruciale delle analisi biogeografiche che può essere illustrato prendendo, ad esempio, la situazione schematicamente riportata in Fig. 16. In essa sono considerati quattro taxa che presentano distribuzioni geografiche come riportato nei riquadri A, B, C e D. Applicando i tradizionali indici di associazione, come quelli sopra

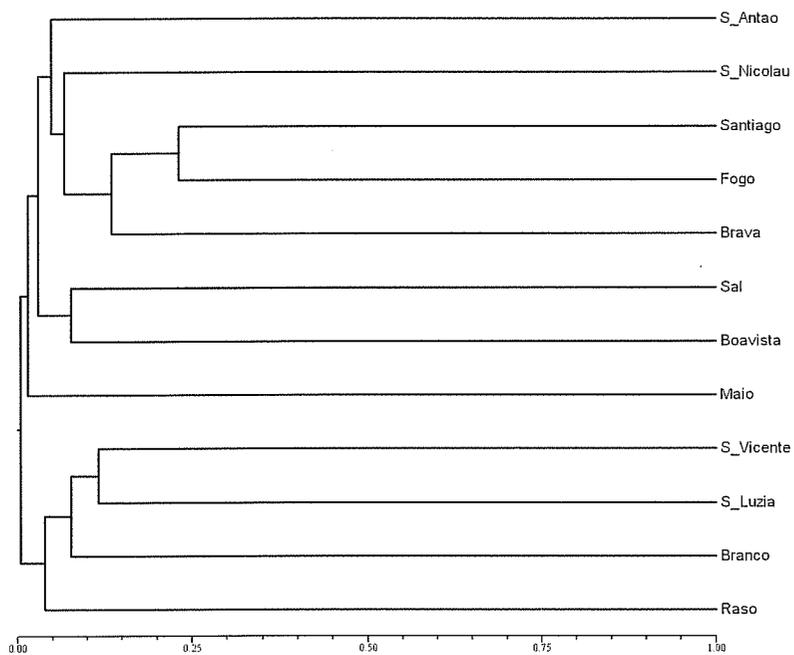


Fig. 11 - Dendrogramma ottenuto con RR e WPGMA (cfr. testo).

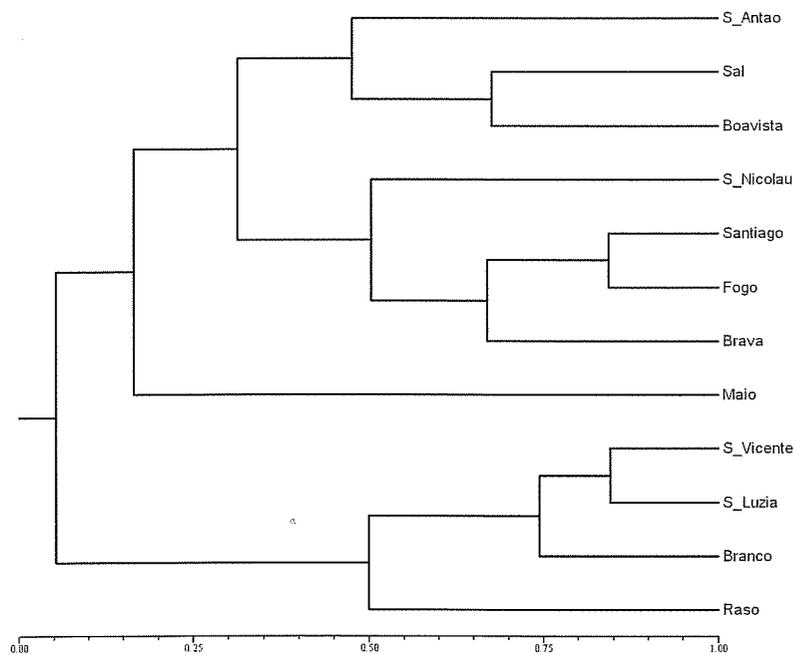


Fig. 12 - Dendrogramma ottenuto con BUB e WPGMA (cfr. testo).

discussi, i confronti a coppie tra le distribuzioni AB, AD e BD restituiscono valori di somiglianza uguali a “0” (oppure valori di distanza uguali a “1”). Tuttavia, sembrerebbe piuttosto ovvio considerare le distribuzioni riportate nei riquadri A e B tra loro come più somiglianti (o meno distanti) rispetto al confronto delle stesse con la distribuzione riportata nel riquadro C. Gli indici solitamente utilizzati per il calcolo della somiglianza/distanza, infatti, non sono in grado di discriminare queste due differenti situazioni. A tal proposito, quindi, Hennig e Hausdorf (2006) hanno proposto di introdurre nel calcolo della somiglianza anche la distanza delle aree considerate, sulla base della seguente formula (modificata per ottenere valori di somiglianza da 0 a 1): $dG(A, B) =$

$$1 - \left[\frac{1}{2} \left(\frac{\sum_{a \in A} \min_{b \in B} u(d_R(a, b))}{|A|} + \frac{\sum_{b \in B} \min_{a \in A} u(d_R(a, b))}{|B|} \right) \right]$$

dove:

- u è una trasformazione monotona crescente con $u(0) = 0$ che nel caso venga considerata come “funzione identità” ovvero $f(u) = u$, allora dG risulta essere la media tra le distanze geografiche medie di tutte le unità di A riguardo alle rispettive unità più prossime di B e le distanze geografiche medie di tutte le unità di B rispetto alle rispettive unità più prossime di A;
- $|A|$ è il numero di unità geografiche di A;
- $|B|$ è il numero di unità geografiche di B;
- dR è la distanza geografica tra le unità geografiche a e b .

Il coefficiente GECO va tuttavia considerato come una generalizzazione dell’indice K2, con il quale coincide quando le distanze tra le unità geografiche vengono considerate uguali a “0”. L’algoritmo per l’applicazione di questo coefficiente è incluso nel package PRABCLUS, implementato nel software statistico “R” e disponibile all’indirizzo <http://cran.r-project.org> (cfr. Hausdorf e Hennig, 2006).

Applicando il coefficiente GECO ai dati riportati in Tab. III, integrati con la matrice delle distanze che intercorrono tra le isole dell’arcipelago di Capo Verde (Tab. VI), è stato ottenuto il dendrogramma di somiglianza delle distribuzioni dei 26 taxa considerati riportato in Fig. 17. Allo scopo di operare un confronto, è stato costruito un secondo dendrogramma (Fig. 18), ottenuto con un’analogha procedura ma utilizzando l’indice K2 (Tab. II) in “R-mode”. Dall’analisi comparativa dei due dendrogrammi, ottenuti con il metodo di clustering WPGMA (Sneath e Sokal, 1973), si può osservare come la struttura degli stessi

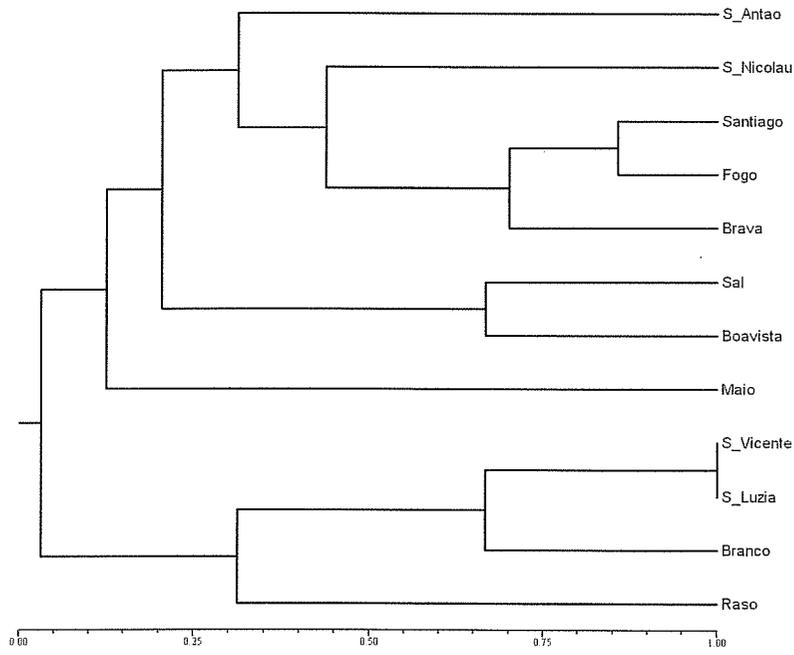


Fig. 13 - Dendrogramma ottenuto con SI e WPGMA (cfr. testo).

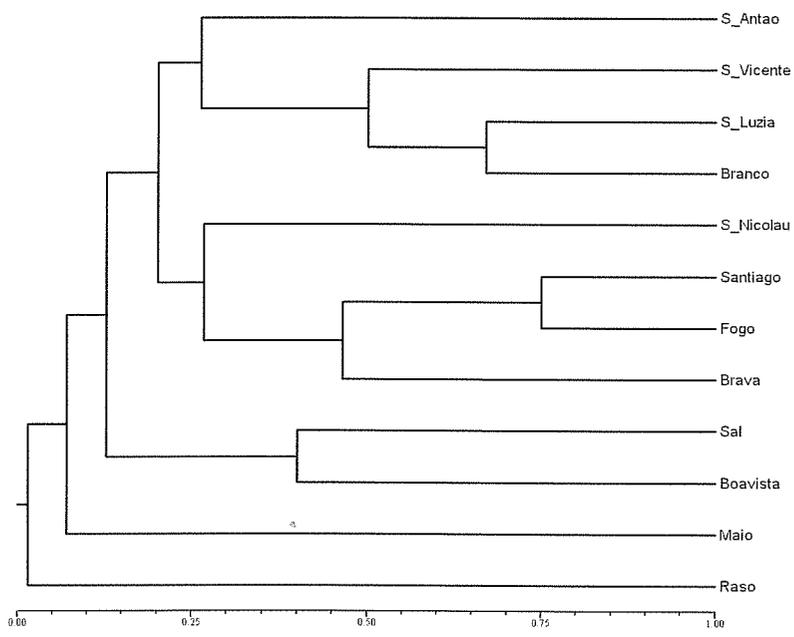


Fig. 14 - Dendrogramma ottenuto con BB e WPGMA (cfr. testo).

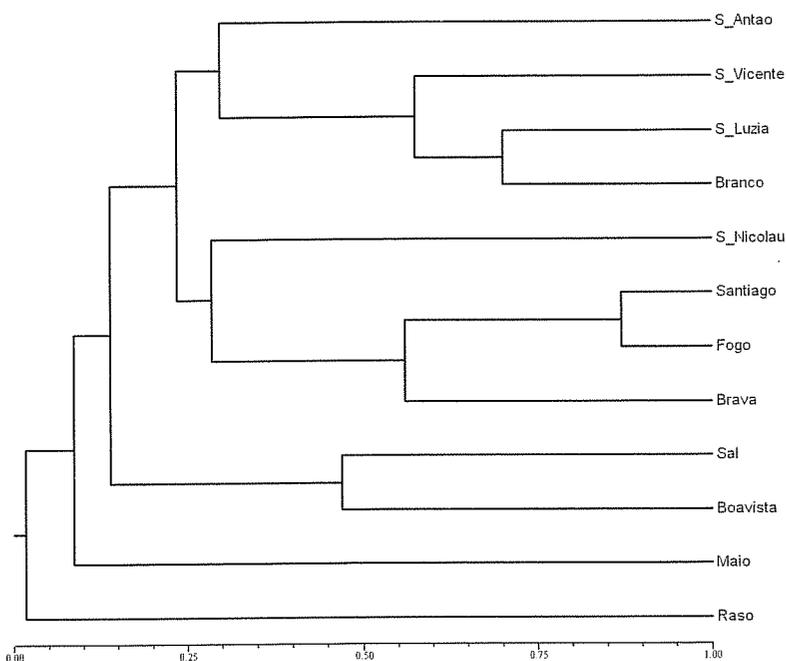


Fig. 15 - Dendrogramma ottenuto con TSI e WPGMA (cfr. testo).

sia in alcuni punti sensibilmente diversa. Ad esempio, la posizione del taxon T5, che GECO colloca nel cluster ((T2-T5)(T18-T14)), mentre K2 lo include nel cluster (((T1-T14)T23)T25). Nel primo caso contribuisce in modo significativo la vicinanza geografica delle isole nelle quali T5 è presente (cfr. Tab. III) rispetto a quelle dove risultano presenti i taxa T2, T18 e T24; mentre nel secondo caso i più elevati valori di somiglianza tra le distribuzioni dei taxa T1, T14, T23 con T5 sono dovuti esclusivamente alla loro copresenza nell'isola di Boavista (cfr. Tab. III).

Un ulteriore esempio della differente "sensibilità" di GECO rispetto a K2, si può evidenziare analizzando la diversa posizione del taxon T12 nei due relativi dendrogrammi (Figg. 17-18). Nel dendrogramma ottenuto con K2 (Fig. 18), si può osservare che T12 viene collocato, in posizione piuttosto distante, nel cluster ((T25;T10)(T6)), mentre in quello ottenuto con GECO viene collocato, a un livello di somiglianza nettamente più elevato, in un cluster che include i taxa T10 e T25. Nel primo caso, la posizione di T12 e T6 in relazione al gruppo (T10-T25) risulta piuttosto spuria in quanto ambedue questi taxa condividono, sia con T10 sia con T25, una sola OGU (Branco per T12 e S_Vicente per T6) e inoltre non risultano presenti in nessuna delle altre isole considerate (cfr. Tab. III). I rispettivi valori di somiglianza, quindi, sono identici (K2 per: T12-T10 = 0,67; T12-T25 =

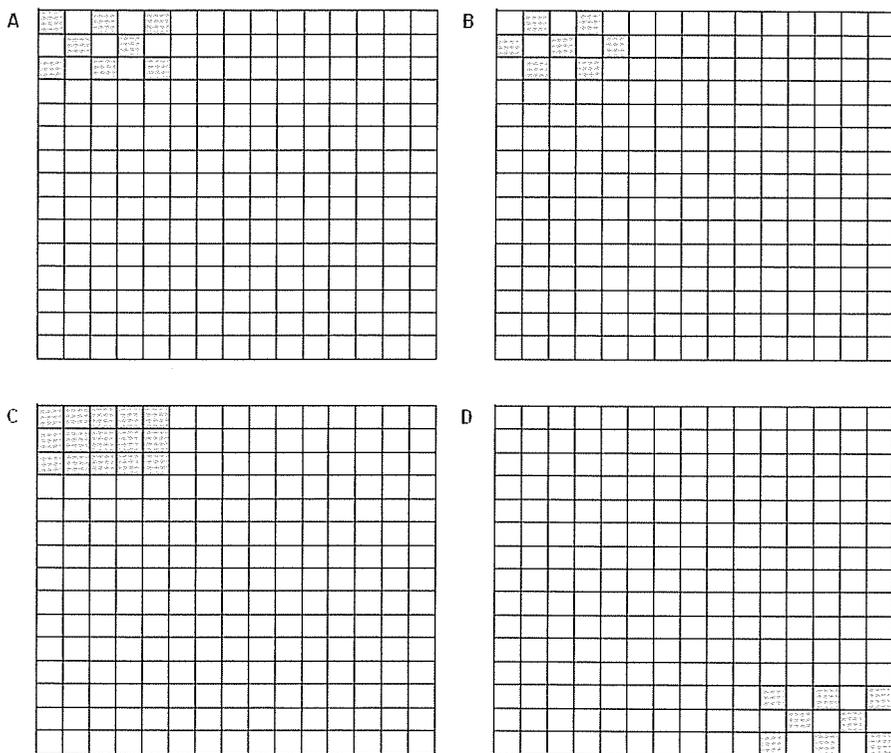


Fig. 16 - Aree di distribuzione A, B, C e D. In grigio sono indicate le unità geografiche occupate dai rispettivi taxa (cfr. testo).

0,62; T6-T10 = 0,67; T6-T25 = 0,62) e di conseguenza le loro posizioni nel cluster dovrebbero risultare equivalenti rispetto al cluster (T10-T25).

Nel caso di GECO (Fig. 17), invece, T12 presenta un valore di somiglianza molto elevato in relazione al gruppo (T10-T25), mentre T6 viene collocato in una posizione molto “esterna” al cluster (((T4-T9-T13)((T10-T25)T12))). Anche in questo caso contribuisce in modo significativo, oltre che la copresenza di T12, T10 e T25 nell’isola di Branco, anche la vicinanza geografica di questa con le isole São Vicente, Santa Luzia e Raso, che vedono la presenza in parte o in tutte dei taxa T10 e T25 (cfr. Tab. III).

Per ulteriori approfondimenti riguardanti il coefficiente GECO si rimanda a quanto riportato in Hennig e Hausdorf (2006) e Hausdorf e Hennig (2006).

CONCLUSIONI

Dalla breve panoramica sopra riportata si può concludere come le differenze realmente importanti nei numerosi indici di somiglianza sinora proposti, possano essere così sintetizzate:

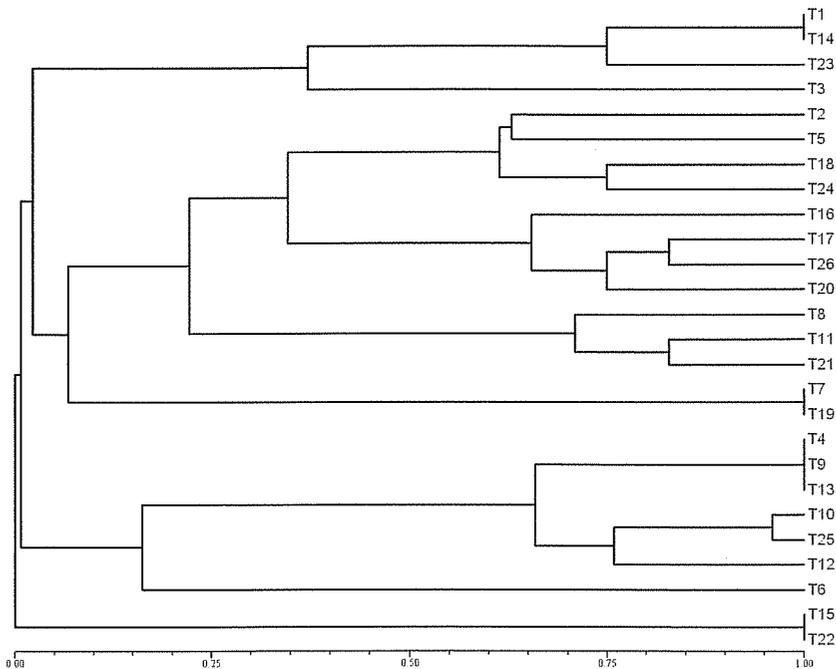


Fig. 17 - Dendrogramma ottenuto con GECO e WPGMA (cfr. testo).

- indici che basano il calcolo della somiglianza in particolar modo sul contributo del numero di attributi condivisi (parametro a), quali CC, CI, LL, SS2 e LW (Tab. II), presentano un comportamento monotono che non modifica la struttura del dendrogramma indipendentemente dal metodo di “clustering” utilizzato, ma soltanto i livelli di somiglianza ai quali si aggregano i vari clusters; anche nei coefficienti CCL e CR viene dato un peso particolare al numero di attributi condivisi dalle due OGU s confrontate. L'utilizzo di questa tipologia di coefficienti è indicato nella maggior parte delle analisi che richiedono un calcolo dei valori di somiglianza tra unità geografiche;
- indici che basano il calcolo della somiglianza sulla percentuale di attributi condivisi relativamente al numero totale di attributi presenti in ciascuna delle OGU s confrontate; in questa categoria rientra il coefficiente K2, che può anch'esso essere considerato di utilizzo generale;
- indici che basano il calcolo della somiglianza anche sul contributo delle coassenze; nelle analisi biogeografiche dovrebbero essere tuttavia considerati soltanto quei coefficienti (BUB e RR) che considerano le coassenze solo nel caso in cui almeno un attributo positivo viene condiviso (copresenze), questo per evitare somiglianze spurie basate esclusivamente sulla condivisione di

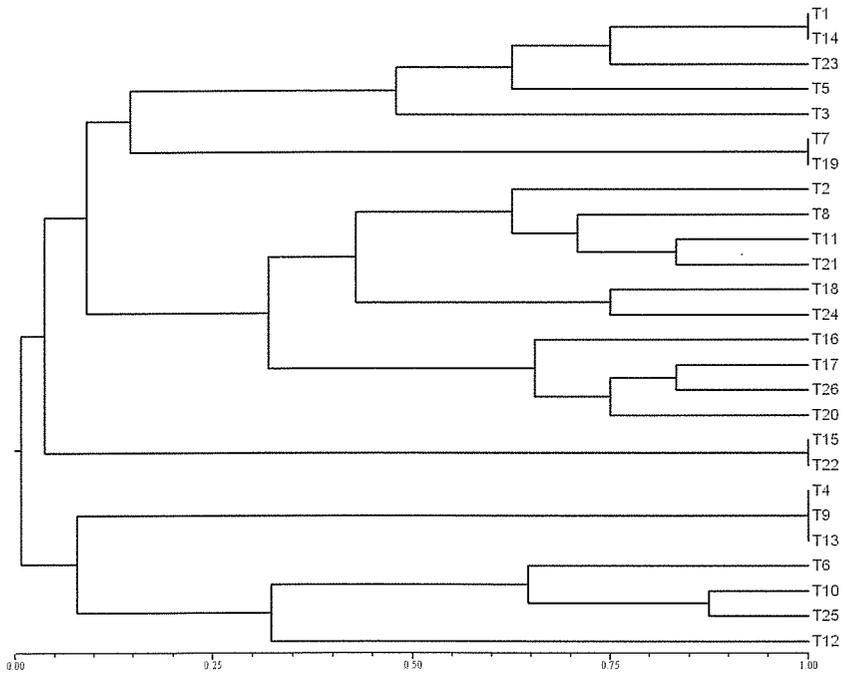


Fig. 18 - Dendrogramma ottenuto con K2 (in R-mode) e WPGMA (cfr. testo).

attributi negativi (coassenze); questa tipologia di indici risulta utile quando le conoscenze derivanti dal campione sono statisticamente significative rispetto a quelle riferite alla situazione reale;

- indici “asimmetrici” che tengono conto del numero di attributi condivisi solo in rapporto al totale degli attributi della OGU che ne presenta il numero minore (SI) oppure in rapporto al totale degli attributi della OGU che ne presenta il numero maggiore (BB); l’utilizzo di questa tipologia di indici può essere indicato quando esistono significative differenze tra le dimensioni delle liste di attributi nelle diverse OGU.

Le tendenze attuali dei nuovi indici finalizzati al calcolo della somiglianza per matrici binarie vanno nella direzione di: a) diminuire o eliminare le incongruenze insite nelle formule dei coefficienti tradizionalmente utilizzati come, ad esempio, si propone il TSI (cfr. Tulloss, 1997); b) di considerare altri fattori, oltre il solo dato di presenza/assenza di attributi nelle OGU, come la reciproca distanza tra le unità geografiche considerate, allo scopo di rendere maggiormente significativo e discriminante il calcolo della somiglianza, come si propone, ad esempio, il coefficiente GECO (cfr. Hennig e Hausdorf, 2006).

BIBLIOGRAFIA

- ARECHAVALETA M., ZURITA N., MARRERO M. C., MARTIN J.L. (eds.) 2005 - Lista preliminar de especies silvestres de Cabo Verde (hongos, plantas y animales terrestres). Consejería de Medio Ambiente y Ordenación Territorial, Gobierno de Canarias.
- BIONDI M. 1988 - Osservazioni comparative sul comportamento di tre indici di similarità per dati binari. *Biogeographia*, 11 (1985): 285-292.
- DEUTSCH R., CHERNER M., GRANT I. 2006 - Significance testing of a cluster of multivariate binary variables: comparison of the tripartite T index to three common similarity measures. *Stat. Methods Med. Res.*, 15: 285-299.
- HAUSDORF B., HENNIG C. 2006 - Biogeographical test of the vicariance model in Mediterranean land snails. *J. Biogeogr.*, 33: 1202-1211.
- HENNIG C., HAUSDORF B. 2006 - A robust distance coefficient between distribution areas incorporating geographic distances. *Syst. Biol.*, 55: 170-175.
- PALUMBO D.J. 1977 - *Statistics in Political and Behavioral Sciences. Revised Edition.* Columbia University Press, New York.
- SCHILS T. 2006 - The tripartite biogeographical index: a new tool for quantifying spatio-temporal differences in distribution patterns. *J. Biogeogr.*, 33: 560-572.
- SHI G.R. 1993 - Multivariate data analysis in palaeoecology and palaeobiogeography - a review. *Palaeogeogr., Palaeoclimatol., Palaeoecol.*, 105: 199-234.
- SNEATH P.H.A., SOKAL R.R. 1973 - *Numerical taxonomy.* Freeman, San Francisco, CA, USA.
- TULLOSS R.E. 1997 - Assessment of similarity indices for undesirable properties and a new tripartite similarity index based on cost functions. In: M.E. Palm, L.H. Chapela (eds.), *Mycology in sustainable development: expanding concepts, vanishing borders*, Parkway Publishers, Boone, NC, USA: 122-143.