



Università degli Studi di Pisa
Dipartimento di Statistica e Matematica
Applicata all'Economia

Report n. 229

**“Un software applicativo per un’analisi di dati
sui marchi genetici (Genetic Markers)”**

Nicola Salvati

Pisa, luglio 2002

- Stampato in Proprio -

INDICE

1. Introduzione.....	pag.	1
2. Identificazione delle popolazioni per la conservazione.....	“	2
3. Il metodo per lo studio della ricchezza allelica.....	“	3
4. Introduzione sul linguaggio R.....	“	7
5. Il programma di calcolo RAREF.....	“	9
6. Un esempio applicativo.....	“	12
7. Considerazioni finali.....	“	17
Riferimenti Bibliografici		

1. Introduzione

Il Consiglio Nazionale delle Ricerche "Istituto Miglioramento Genetico delle Piante Forestali" fa ricorso ai metodi che si basano sui "genetic markers" nello studio delle specie selvatiche, alle quali deve essere data la priorità per la loro conservazione.

Per effettuare queste analisi è utile l'impiego di appositi programmi (software) di calcolo automatico. In questo lavoro ci proponiamo di riportare una descrizione sintetica di una tematica e di un programma software, che sono stati predisposti per aiutare i biologi a trarre conclusioni e quindi ad operare sul territorio.

Il programma è stato scritto in un linguaggio adeguato per le elaborazioni di carattere statistico, ovvero il linguaggio "R", che viene considerato non particolarmente efficiente da un punto di vista di interazione con l'utente, ma che appare adeguato per la semplificazione delle scelte relative alle elaborazioni statistiche. Per l'implementazione dell'algoritmo di calcolo si è fatto riferimento ad un software preesistente ed immesso in rete, che è documentato in un paper realizzato presso Institut National de la Recherche Agronomique (Petit, El Mousadik, Pons, 1996).

In queste pagine, dopo una sintetica introduzione sul problema affrontato, è riportata una descrizione del linguaggio "R" e del programma "RAREF" utili per l'elaborazione dei dati. Infine, a titolo esemplificativo vengono fornite le indicazioni necessarie per svolgere un'applicazione pratica.

2. Identificazione delle popolazioni per la conservazione

Per selezionare le popolazioni di specie selvatiche, alle quali deve essere data la priorità per la conservazione, i criteri genetici prevedono lo studio dei "Genetic Markers" e in particolare della diversità con i quali essi si presentano nella realtà.

Tradizionalmente la misura della diversità si calcola considerando la eterozigotità attesa (heterozygosity) o la percentuale di polimorfo dei loci analizzati.

Petit, El Mousadik e Pons (1996), invece, propongono come misura per la determinazione della priorità la *ricchezza allelica* e per standardizzare i risultati fra diverse popolazioni introducono la tecnica della *rarefazione*. Questa tecnica permette di valutare il numero atteso di differenti alleli fra campioni di numerosità diversa estratti da diverse popolazioni. I tre studiosi rilevano che il contributo di ciascuna popolazione alla diversità totale può essere diviso in due componenti:

- a) la diversità esistente tra le unità che compongono la popolazione;
- b) la diversità esistente fra le popolazioni.

L'analisi della diversità è stata utilizzata su migliaia di specie di piante e di animali, per fornire le informazioni necessarie, al fine di definire le politiche per la loro salvaguardia e, in particolare, per l'individuazione delle aree da tutelare.

In tutti questi casi assume un'importanza fondamentale il criterio genetico che permette di identificare quali popolazioni di una particolare specie hanno priorità di conservazione. Infatti, poiché non tutte le popolazioni hanno la stessa capacità di risposta al cambiamento delle condizioni ambientali, i dati genetici possono assicurare un miglior uso delle risorse disponibili attraverso la massimizzazione delle potenzialità di risposta, in termini di evoluzione, di un insieme di popolazioni.

L'applicazione del metodo basato sulla ricchezza allelica utilizza il numero di alleli per loci e permette di ottenere risultati più significativi rispetto all'applicazione della eterozigotità media perché tale numero dipende strettamente dalla dimensione della popolazione. In altre parole, la ricchezza allelica è considerata un buon indicatore dei mutamenti demografici passati e quindi può essere ritenuto interessante per lo studio della conservazione. La ricchezza allelica deve, comunque, essere standardizzata per tenere conto delle differenze esistenti tra le dimensioni campionarie. Come è stato rilevato da El Mousadik e Petit (1997), la tecnica della rarefazione, introdotta per la prima volta da Hurlbert (1971) negli studi ecologici, permette di confrontare la ricchezza allelica di campioni di numerosità diversa e quindi il suo utilizzo è fondamentale per il raggiungimento degli obiettivi relativi alla conservazione delle specie.

3. Il metodo per lo studio della ricchezza allelica

La *diversità genetica h* di Nei (1973) nella letteratura ecologica viene misurata mediante l'*indice di Simpson* (1949) ed è funzione della probabilità che due alleli campionari siano diversi; Petit, El Mousadik e Pons introducono un nuovo concetto per descrivere l'effetto di ciascuna popolazione sulle misure classiche della diversità, partendo dall'assunzione che i campioni derivano da un vasto insieme di popolazioni indipendenti di numerosità elevata. Essi, prendono in esame le stime corrette (unbiased) della diversità totale (h_T), di quella media (h_S), di quella assoluta ($D_{ST} = h_T - h_S$) e di quella relativa ($G_{ST} = D_{ST} / h_T$).

Per descrivere le variazioni di questi indicatori dovute all'effetto di una specifica popolazione *i-esima*, si distingue tra i valori delle diversità stimate (\hat{h}_T e \hat{h}_S), considerando le *n* popolazioni osservate, e i valori corrispondenti ($\hat{h}_{T/i}$ e $\hat{h}_{S/i}$) calcolati su *n-1* popolazioni, escludendo la popolazione *i-esima*.

A partire da tali considerazioni è opportuno riportare le espressioni necessarie per le soluzioni dei problemi di calcolo e che sono state prese in esame per definire il programma RAREF.

Per ogni campione estratto si calcola il rapporto fra le unità che presentano un certo tipo di allele (fra quelli studiati) e il numero totale di unità che formano il campione; la proporzione viene determinata per valutare se una specie è caratterizzata dalla presenza di alcuni tipi di alleli rispetto ad altri.

Per l'allele *j-esimo* nella popolazione *i-esima* si definisce la frequenza stimata x'_{ij} :

$$x'_{ij} = \frac{1}{n_i - 1} \sum_{l \neq j} x_{il} \quad [1]$$

in un campione di n_i unità estratte casualmente dove n è il numero delle popolazioni studiate.

Negli studi di Petit e Pons (1996) per la *i-esima* popolazione lo stimatore

$$\hat{h}_i = \frac{n_i}{n_i - 1} \left[1 - \sum_j x_{ij}^2 \right] \quad [2]$$

della diversità è ottenuto sottraendo da 1 la sommatoria della proporzione (elevata al quadrato) di unità che si distribuiscono fra i diversi alleli.

La sommatoria che figura nella espressione [2] non assume mai il valore unitario, fatta eccezione per il caso in cui tutte le unità sono concentrate in un allele; perciò quanto più le unità sono equidistribuite fra gli alleli, tanto minore risulta la sommatoria e tanto maggiore è la differenza, ovvero la diversità, all'interno della *i-esima* popolazione.

Dalla formula [2] si ricavano le espressioni relative alla diversità totale e media:

$$\hat{h}_T = 1 - \frac{1}{n(n-1)} \sum_j \sum_{i \neq j} x_{ij} \cdot x_{ij} \quad [3]$$

$$\hat{h}_S = \frac{1}{n} \sum_i \hat{h}_i \quad [4]$$

che possono essere riscritte:

$$\hat{h}_T = \frac{1}{n} \sum_i \hat{h}_T(i) \quad [5]$$

$$\hat{h}_S = \frac{1}{n} \sum_i \hat{h}_S(i) \quad [6]$$

in funzione delle misure $\hat{h}_T(i)$ e $\hat{h}_S(i)$:

$$\hat{h}_T(i) = 1 - \sum_j x_{ij} \cdot x'_{ij} \quad [7]$$

$$\hat{h}_S(i) = \frac{1}{2} (\hat{h}_i + \hat{h}_{S/i}) \quad [8]$$

della diversità nella distribuzione di alleli nella medesima popolazione e della disuguaglianza tra popolazioni, dove $\hat{h}_{S/i}$ rappresenta il valore della diversità media delle n popolazioni esclusa quella i -esima.

Sulla base delle espressioni proposte è possibile definire gli indici:

$$\hat{D}_{ST}(i) = \hat{h}_T(i) - \hat{h}_S(i) \quad [9]$$

$$\hat{G}_{ST}(i) = \frac{\hat{D}_{ST}(i)}{\hat{h}_T(i)} \quad [10]$$

che gli studiosi propongono come stimatori assoluti e relativi della diversità per la *i-esima* popolazione. In particolare l'indice $\hat{D}_{ST}(i)$ dipende dalla differenza esistente tra la diversità totale della *i-esima* popolazione e la media tra la diversità della *i-esima* popolazione e quella delle *n* popolazioni esclusa la *i-esima*, mentre l'indice $\hat{G}_{ST}(i)$ è un indice relativo ottenuto come rapporto fra l'indice precedente e la misura della diversità totale della popolazione *i-esima*.

Da queste formule si può ricavare quella:

$$C_T(i) = \frac{2}{n-2} \cdot \frac{\hat{h}_T - \hat{h}_{T/i}}{\hat{h}_T} \quad [11]$$

relativa al contributo della *i-esima* popolazione alla diversità totale, che può essere scomposta nelle due componenti di diversità: la componente $C_S(i)$, ovvero il contributo alla diversità della *i-esima* popolazione dovuto alla sua diversità rispetto alle altre; la componente $C_D(i)$, ovvero il contributo dovuto alla sua divergenza interna. E' opportuno notare che l'indice $C_T(i)$ è funzione del fattore di correzione $\frac{2}{n-2}$ e il numero delle popolazioni deve essere superiore a 2 perché altrimenti l'indice $C_T(i)$ assume un valore assurdo.

Per concludere, il prospetto degli indici utili per l'analisi, può essere completato considerando il valore della stima dell'errore quadratico medio della diversità stimata della *i-esima* popolazione. A tale proposito notiamo che nel lavoro di Petit (1997) non viene illustrato il metodo di calcolo di tale errore. Perciò, questo indice è stato definito attraverso la seguente espressione,

$$s.e.(\hat{h}_i) = \sqrt{\frac{\sum_j (x_{ij} - \hat{h}_i)^2}{\sum_j x_{ij} - 1}} \quad [12]$$

che è conforme al calcolo tradizionale dell'errore standard nella metodologia statistica.

4. Introduzione sul linguaggio R

Il linguaggio R è uno strumento di calcolo automatico per le elaborazioni di tipo statistico di pubblico dominio (open source) che, nel rispetto dei diritti d'autore, può essere copiato, distribuito e modificato da chiunque. Inizialmente scritto dal Dipartimento di Statistica dell'Università di Auckland, Nuova Zelanda, dal 1997 lo sviluppo di R è affidato ad un gruppo internazionale (R Core Team). Una copia di R può essere scaricata accedendo all'indirizzo web:

<http://cran.r-project.org/bin>

in cui si trovano diverse versioni del linguaggio per diversi sistemi operativi e i relativi manuali d'uso.

Tale linguaggio rappresenta un ambiente statistico per la manipolazione, l'analisi e la rappresentazione grafica di dati ed è un programma interattivo, perché i comandi producono una risposta immediata e prevedono una programmazione object-oriented. Inoltre, l'ambiente R può essere esteso attraverso la costruzione di librerie da parte dell'utente e offre la possibilità di ricorrere alle librerie dinamiche scritte in C, Fortran e linguaggi simili, nonché di creare interfacce grafiche, anche se questa ultima opportunità non è stata sviluppata come è accaduto per altri linguaggi.

A questo punto della trattazione vengono presentati i comandi principali per poter iniziare a lavorare in R, mentre nel paragrafo successivo abbiamo riportato le istruzioni che compongono il programma RAREF e un esempio della sua applicazione.

Dopo aver mandato in esecuzione il linguaggio R dal sistema Windows, appare una nuova finestra con il simbolo ">" che indica che l'ambiente è pronto a ricevere

istruzioni. Per attivare l'aiuto in linea in formato Windows occorre ricorrere alla sequenza tipica di comandi:

```
> option
```

```
> help()
```

nei quali in alcuni casi figurano delle parentesi tonde, che consentono di specificare opzionalmente le scelte dell'utilizzatore.

Per il formato HTML si usa il comando:

```
> help.start()
```

oppure si seleziona la voce *R language* dall'icona *Help* nella linea dei comandi.

Se si desiderano informazioni su una funzione di cui è già noto il nome, ma di cui per esempio non si conosce la sintassi, si può ricorrere al comando `help(funzione)`. Inoltre nell'ambiente R il comando `apropos("stringa di caratteri")`, produce l'elenco di tutte le funzioni e gli oggetti i cui nomi contengono la stringa di caratteri specificata tra virgolette (Es: `> apropos(mean)` permette di avere tutte le informazioni sul comando `mean`).

Per terminare una seduta di lavoro, dal simbolo ">" si digita il comando `q()` oppure selezionando l'icona *File* della linea di comandi si sceglie *Exit*. In R l'uscita dall'ambiente determina la comparsa del messaggio "*Save workspace image?*": se la risposta è *SI*, le modifiche apportate e gli oggetti creati durante la seduta vengono salvati nella *directory* di lavoro e potranno essere recuperati nelle sedute successive. Se la risposta è *NO*, l'ambiente viene chiuso senza alcun salvataggio.

5. Il programma di calcolo RAREF

Di seguito vengono illustrate alcune parti del programma di calcolo RAREF realizzato col linguaggio R e si fornisce una descrizione dell'uso dei sotto-programmi e delle istruzioni necessarie per pervenire ai risultati di un'analisi.

All'inizio di ogni programma è indispensabile inizializzare il sistema R, azzerando la memoria con l'istruzione:

```
rm(list=ls(all=TRUE))
```

per poi eseguire le istruzioni successive, che permettono di richiamare le funzioni realizzate con altri programmi, indicando in modo preciso il percorso (path) e le istruzioni

```
source("c:/Documenti/Lavoro C.N.R/Comandi.r.2.txt")
```

```
matricecalcolo<-matrix(scan(file=nomepathdirec[[1]]))
```

con le quali vengono acquisiti i dati da elaborare.

Nella prima parte del programma si esegue una serie di operazioni funzionali alla determinazione dei dati relativi alle popolazioni e agli aplotipi ai quali vengono applicati le formule che sono state presentate nei paragrafi precedenti. I valori delle frequenze stimate [1] per le $(k-1)$ popolazioni si ottengono eseguendo le istruzioni contenute nel Prospetto 1:

Prospetto 1. Calcolo delle frequenze stimate per le $(k-1)$ popolazioni

```
rig1<-rig  
for (i in 1:rig1) {for (j in 1:col-1) {allele<-0;  
                  for (k in 1:rig) {if (k!=i) allele<-allele+matriceripulital[k,j]};  
mafresti[i,j]<-(allele/(rig-1))}}
```

I cicli di *for* rispetto agli indici *i* e *j* consentono di calcolare la sommatoria delle frequenze e quindi di ottenere la frequenza stimata per l'allele *j-esimo* nella popolazione *i-esima*.

Successivamente si procede nell'elaborazione calcolando la diversità media [3] e quella totale [4] (Prospetto 2).

Prospetto 2. Calcolo della diversità media e totale

```
# calcolo della diversità media #
diversmedia<-0
for (i in 1:rig) {diversmedia<-diverspop[i,1]+diversmedia}
diversmedia<-diversmedia/rig
# calcolo della diversità totale #
diverstot<-0
rig1<-rig
calc<-0
for (i in 1:rig1) {for (j in 1:coll) {for (k in 1:rig) {calc<-
(calc+(matriceripulita1[i,j]*matriceripulita1[k,j]))}}}
diverstot<-(1-((1/(rig*(rig-1)))*calc))
```

A partire da queste informazioni prendiamo in esame la fase principale del programma, che è relativo allo studio dei Genetic Markers, ovvero l'applicazione del metodo della rarefazione, elencando le singole istruzioni (Prospetto 3).

Prospetto 3. Applicazione del metodo della rarefazione

```
# Calcolo rarefazione#
Pbs<-0
rf<-matricecalcolo[3,1]
rfk<-rf-1
Pb<-matrix(0,rig,col1)
Pbk<-matrix(0,rig,1)
for (i in 1:rig) {for (j in 2:col) {if ((npop[i]-matriceripulita[i,j])<rf) Pb[i,j-1]<-0
else
      {Pb[i,j-1]<-(npop[i]-matriceripulita[i,j])/npop[i];
      for (k in 1:rfk)
      {Pb[i,j-1]<-(Pb[i,j-1]*((npop[i]-matriceripulita[i,j])-k))/(npop[i]-k)}}}}
for (i in 1:rig) {for (j in 1:col1) {Pbk[i]<-Pbk[i]+(1-Pb[i,j])}}
for (i in 1:rig) {Pbs<-Pbs+(Pbk[i]/rig)}

# Calcolo RT e RST (procedura RHOST)#
tot<-0
Rho<-0
toti<-matrix(0,col1,1)
for (i in 1:rig) {for (j in 2:col) {tot<-tot+round(matriceripulita[i,j])}}
for (j in 2:col) {for (i in 1:rig) {toti[j-1]<-toti[j-1]+round(matriceripulita[i,j])}}
Pbt<-0
Pbti<-matrix(0,col1,1)
for (j in 1:col1) {if ((tot-toti[j])<rf) Pbti[j]<-0 else
      {Pbti[j]<-(tot-toti[j])/tot;
      for (i in 1:rfk) {Pbti[j]<-Pbti[j]*((tot-toti[j])-i)/(tot-i)}}}
for (j in 1:col1) {Pbt<-Pbt+1-Pbti[j]}
Rho<-(Pbt-Pbs)/(Pbt-1)
```

L'esecuzione delle istruzioni del programma si conclude attraverso il calcolo del contributo alla diversità, che è riconducibile a due fattori: la diversità fra popolazioni e la diversità nelle popolazioni [11]. L'ultima fase della procedura riguarda la scrittura dei risultati su un foglio di lavoro di Excel, che permette qualsiasi altra elaborazione e soprattutto la possibilità di realizzare grafici, per fornire delle ulteriori indicazioni interessanti per lo studio della conservazione delle popolazioni.

6. Un esempio applicativo

Il file di input dei dati è un file di testo che può essere realizzato in Excel, ma che deve essere salvato come file di testo delimitato da tabulazione (opzione di excel: esempio: POP.txt).

La prima riga del file deve indicare il numero di aplotipi, il numero di popolazioni, la misura della rarefazione, che non dovrebbe essere più grande della dimensione del campione di popolazione più piccolo (Petit, El Mousadik, Pons, 1996), e l'indicazione del numero di popolazioni (o sottogruppi) alle quali si riferisce l'insieme dei campioni estratti. Ogni popolazione (o sottogruppo) deve essere contraddistinta con uno stesso numero che deve essere introdotto sempre nella prima colonna del file di input. Le righe successive alla prima del file dei dati sono formate da un numero di colonne pari al numero di aplotipi, alle quali viene aggiunta una colonna con l'identificazione (numero) della popolazione (o sottogruppo) di appartenenza (N.B.: il numero indice deve iniziare da 1 e non può essere nullo o negativo). In totale il numero delle righe del file dei dati sarà pari al numero di campioni estratti con l'aggiunta della prima riga di informazione per il programma RAREF (Prospetto 4).

Nel programma RAREF non ci sono limiti al numero di aplotipi, di campioni, di popolazioni e di sottogruppi se non quelli consentiti dalla dimensione dell'area di lavoro del sistema (WORKSPACE) R, che generalmente offre una opportunità sufficiente per le analisi da effettuare.

Prospetto 4. Esempio di file di input dei dati da elaborare

18	6	10	3			
1	3	0	9	2	8	...
1	0	0	1	3	3	...
2	1	1	0	0	0	...
3	1	3	0	6	10	...

Per utilizzare RAREF è necessario, prima di tutto, entrare all'interno del sistema R (il programma è eseguibile in R) cliccando due volte sull'icona. Una volta aperto il sistema R, nella barra degli strumenti si utilizza l'opzione *File* e all'interno la routine *Source R Code*. Si apre una nuova finestra di dialogo che chiede di indicare il file da considerare come fonte dei dati. Successivamente si ricerca il programma RAREF (registrato su CD nella directory "D:") considerando l'opzione *All File*. Cliccando due volte sul programma o utilizzando l'opzione *Apri*, il sistema R esegue il programma e sul video compare la richiesta del nome del file di input contenente i dati relativi alle popolazioni e viene richiesto anche il percorso da seguire in modo corretto. Per esempio, se il file POP.txt si trova sulla scrivania del computer occorre indicare il percorso (N.B.: entro le virgolette " ", altrimenti il file non viene letto e si ha come risposta un messaggio di errore): "c: / windows / desktop / POP.txt".

Dopo aver indicato il percorso si digita return e sullo schermo compare il numero 2: digitando nuovamente return il programma viene eseguito nell'ambiente R e i risultati appaiono su video e sono salvati su un file Excel chiamato OUTPUT.xls sulla scrivania del computer. Prima di eseguire nuovamente il programma RAREF con dati diversi è necessario cambiare nome al file di output, perché altrimenti i risultati precedenti si perdono, dato che il programma sovrascrive ai vecchi risultati quelli nuovi. Nel Prospetto 5 è riportato un esempio dei risultati ottenuti con il programma in questione.

Prospetto 5. Esempio di file di output del programma di calcolo RAREF

Rarefazione	10									
Nb all	Nb pop	Media Aritmetica	Media armonica							
12	120	4.925	4.887984							
hS	hT	Gst								
0.1275	0.765427	0.831723								
RS	RT	RST								
1.1667778	2.344156	0.875924								
POP	N	Nall	Hi	seHi	Pb-1	Dhs	Dht	Dgst	Dst	
1	5	1	0	0	0	0.064286	0.981513	0.934503	0.917227	
2	5	1	0	0	0	0.064286	0.981513	0.934503	0.917227	
3	4	1	0	0	0	0.064286	0.981513	0.934503	0.917227	
4	5	3	0.7	0.1225	1	0.411345	0.942857	0.563725	0.531513	
	Ct	Cs	Cd	Crt	Crs	Crd				
POP	0.004785	-0.0014	0.006185	0.003273	-0.00104	0.004316				
1	0.004785	-0.0014	0.006185	0.003273	-0.00104	0.004316				
2	0.004785	-0.0014	0.006185	-0.00381	-0.00104	-0.00277				
3	0.003929	0.006285	-0.00236	0.002209	0.005209	-0.003				
4	-0.00098	-0.0014	0.000416	-0.00511	-0.00104	-0.00406				

Essi riguardano una serie di indici, che sono utili per lo studio dei Genetic Markers per una o più popolazioni oggetto di studio. Per ciascuna popolazione l'elenco degli indicatori e delle relative informazioni è riportato nel Prospetto 6.

Prospetto 6. Elenco degli indici per lo studio dei Genetic Markers

RAREFAZIONE	Indica la misura della rarefazione
Nb-all	Numero di aplotipi sottoposti all'analisi
Nb-pop	Numero di popolazioni indagate
Media Aritmetica e armonica	Numero medio di individui per popolazione e media armonica degli aplotipi per individuo
hS	Diversità media
hT	Diversità totale
Gst	Gst medio
RS	Rarefazione calcolata considerando la diversità media
RT	Rarefazione calcolata considerando la diversità totale
RST	$(RT-RS)/(RT-1)$
POP	Numero progressivo della popolazione
N	Numerosità della popolazione
Nall	Numero di aplotipi non nulli per ogni popolazione

Hi	Diversità per la <i>i-esima</i> popolazione
seHi	Errore standard relativo alla diversità della popolazione <i>i-esima</i>
Pb-1	Contributo alla rarefazione della <i>i-esima</i> popolazione
Dhs	Stima della diversità media quando la <i>i-esima</i> popolazione è esclusa dal calcolo ($\hat{h}_s(i)$)
Dht	Stima della diversità totale quando la <i>i-esima</i> popolazione è esclusa dal calcolo ($\hat{h}_t(i)$)
Dgst	Stima della diversità relativa della <i>i-esima</i> popolazione ($\hat{G}_{st}(i)$)
Dst	Stima della diversità assoluta della <i>i-esima</i> popolazione ($\hat{D}_{st}(i)$)
Ct	Contributo della <i>i-esima</i> popolazione alla diversità totale ($\hat{C}_t(i)$)
Cs e Cd	Componenti del contributo alla diversità Totale ($\hat{C}_s(i)$ e $\hat{C}_d(i)$)
Crt	Contributo alla ricchezza all'elica della <i>i-esima</i> popolazione
Crs e Crd	Componenti del contributo alla ricchezza allelica totale della <i>i-esima</i> popolazione

7. Considerazioni finali

A conclusione di questa nota tecnica riportiamo alcune considerazioni di carattere generale su alcune ulteriori possibilità di elaborazione, che possono essere ritenute interessanti per approfondire le analisi utili nel campo di ricerca preso in esame.

Un primo aspetto da considerare riguarda l'importanza di questo tipo di ricerche, che non si fermano ad una analisi metodologica, ma vanno oltre determinando gli interventi su campo. Da qui la necessità di un'analisi statistica per cercare di semplificare l'acquisizione di informazioni utili per prendere le decisioni più corrette.

In tal senso, l'analisi di diversità proposta da Petit, El Mousadik, Pons (1996) può essere considerata una soluzione, senz'altro più sofisticata, della scomposizione della varianza. Infatti, la diversità totale viene ripartita con modalità particolari nelle due componenti: (a) la diversità fra le unità che compongono la popolazione; (b) la diversità fra popolazioni.

L'impiego del programma applicativo realizzato nel linguaggio R, che come si è già detto è molto potente da un punto di vista statistico, anche se è ritenuto piuttosto limitato da un punto di vista interattivo. Non bisogna poi dimenticare la possibilità di utilizzare un help molto semplice ed esemplificativo.

Per quanto riguarda l'aspetto metodologico, si può muovere una critica alle proposte di Petit, El Mousadik, Pons (1996), perché i risultati ottenuti non sono sottoposti a nessuna analisi di significatività, che potrebbe essere molto importante per capire se i risultati sono dovuti al caso, oppure se esiste una certa diversità tra le popolazioni. Questo aspetto è fondamentale al fine di prendere delle decisioni, che comportano interventi sul territorio per determinare le priorità per la conservazione delle specie. Per questo motivo per il futuro, sarebbe certamente importante approfondire lo studio di questo campo per individuare delle nuove metodologie (ad

esempio quelle per lo studio della concentrazione) adeguate all'analisi dei "Genetic Markers".

Riferimenti Bibliografici

- Bortot P., Ventura L., Salvan A. (2000). "Inferenza Statistica: Applicazioni con S-PLUS e R". CEDAM, Padova
- Petit R., El Mousadik A., Pons O. (1996). "Identifying Population for Conservation on the Basis of Genetic Markers". Paper a cura di Institut National de la Recherche Agronomique, Gaziner Cedex, France.