

Problème pratique de statistique n° pps069
Les tortues géantes des Galapagos

A.-B. Dufour & T. Jombart

9 mars 2006

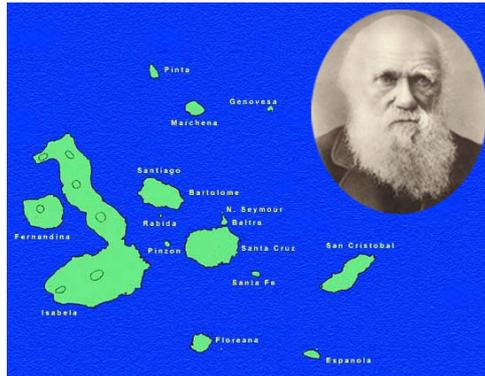
La fiche décrit le contenu de la liste de  qui reproduit les données acquises et publiées par C. Ciofi, M.C. Milinkovitch, J.P. Gibbs, A. Caccone & J.R. Powell (2002, Microsatellite analysis of genetic divergence among populations of giant Galápagos tortoises, *Molecular Ecology*, 11, 2265-2283). 17 populations sont réparties sur 11 localités dans 6 îles. Elles sont positionnées sur une carte schématique. Les estimations des fréquences alléliques sur 10 microsatellites hautement polymorphes donnent un beau problème de génétique spatiale.

Table des matières

1	L'archipel des Galapagos	2
2	Les sites de l'étude	4
3	Les populations étudiées	6
4	Les microsatellites	8
	Références	9

1 L'archipel des Galapagos

La tortue géante est quasi synonyme des Iles Galapagos. En effet, les premiers colons espagnols ont trouvé que la forme de la carapace de ces tortues rappelait la selle d'équitation appelé "galapago". Donc, lorsqu'on parle des îles Galapagos, on parle des îles des tortues géantes.



<http://www.rit.edu/~rhrsbi/GalapagosPages/Tortoise.html>

Il y a 15 races identifiées de ces tortues, toutes généralement considérées comme appartenant à l'espèce *Geochelone elephantopus*. Sur ces 15 races, 4 sont éteintes, et une n'est plus représentée que par un individu (Lonesome George, fig. 1), sur l'île de Pinta. Le genre *Geochelone* lui-même est représenté par un ensemble d'espèces de tortues de tailles petite ou moyenne que l'on trouve en Amérique du Sud, en Afrique, au Madagascar et en Asie. Dans le passé, les espèces géantes de *Geochelone* étaient présentes sur tous les continents excepté l'Australasia. Mais aujourd'hui, il ne reste plus des formes géantes que *elephantopus* de *G.* aux Galapagos et *gigantea* de *G.* sur l'île d'Aldabara.

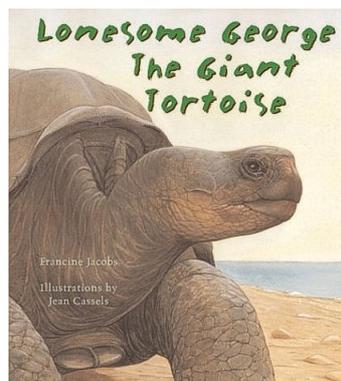


FIG. 1 – <http://www.15degreelab.com/george.jpg>

L'étude de la diversité génétique des tortues des Galapagos s'inscrit dans plusieurs contextes : pour apporter des éléments à la question de leur statut taxonomique, pour inférer l'histoire de la colonisation des îles, et pour leur conservation.

Les données acquises et publiées par C. Ciofi, M.C. Milinkovitch, J.P. Gibbs, A. Caccone & J.R. Powell [1] sont préparées dans l'objet `pps069`.

```
load(url("http://pbil.univ-lyon1.fr/R/pps/pps069.rda"))
library(ade4)
names(pps069)
```

```
[1] "area" "ico" "pop" "misc" "loc" "tab"
```

La liste est constituée de six éléments. L'essentiel est un tableau de fréquences alléliques :

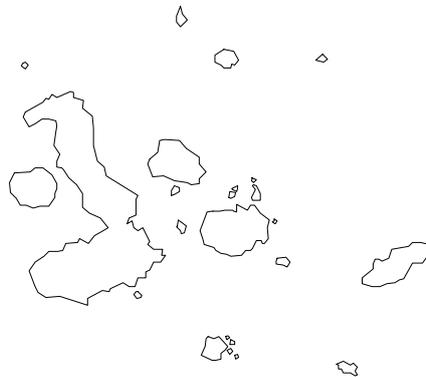
```
dim(pps069$tab)
```

```
[1] 17 211
```

`pps069$tab` contient les fréquences alléliques dans 17 populations pour 10 loci microsatellites. Diverses informations sont attachées aux deux marges de ce tableau (populations et allèles-loci).

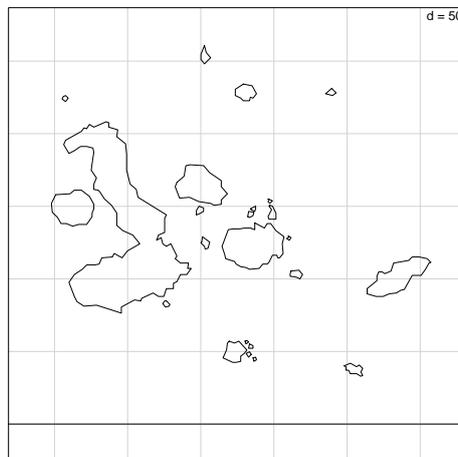
La composante `pps069$area` est un data frame à trois composantes destinées à tracer un fond de carte simplifié.

```
area.plot(pps069$area)
```



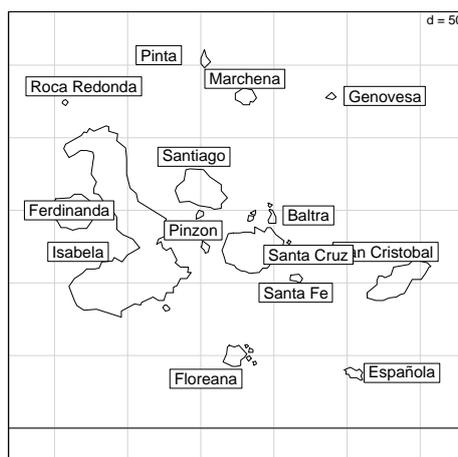
Pour avoir une échelle (carré de 50 km) :

```
s.label(pps069$area[, -1], clab = 0, cpoi = 0, area = pps069$area,
        xlim = c(25, 325))
```



Pour ajouter quelques noms des îles, utiliser `pps069$util` :

```
s.label(pps069$area[, -1], clab = 0, cpoi = 0, area = pps069$area,
        xlim = c(25, 325))
s.label(pps069$misc, clab = 1.25, add.plot = T)
```



2 Les sites de l'étude

Les 17 populations sont issues de 11 localités réparties sur six des îles des Galapagos décrites par les auteurs (*op. cit.* Table 2 page 2270) :

Isabela Les tortues ont été choisies dans 11 localités différentes :

1. PBL Predas Blanca, au nord du volcan Wolf
2. PBR Puerto Bravo, au sud du volcan Wolf
3. VD près du volcan Darwin
4. VA près du volcan Alcedo
5. CAZ la Cazuela, au nord est de la Siera Negra
6. RU Roca Union, au sud de la Siera Negra

7. CR Cabo Rosa, au sud de la Siera Negra
8. LP Las Pampas, à l'ouest de Cerro Azzul
9. LT Las Tablas, à l'ouest de Cerro Azzul
10. LC Los Crateres, à l'est de Cerro Azzul
11. PEG Pegas, au nord est de Cerro Azzul

Santa_Cruz Les tortues ont été choisies dans 2 localités différentes :

1. CRU la Caseta, au sud-ouest de l'île
2. CF Cerro Fatal, au sud-est de l'île

Pinzon PZ

Santiago AGO

San_Cristobal SCR

Espanola ESP

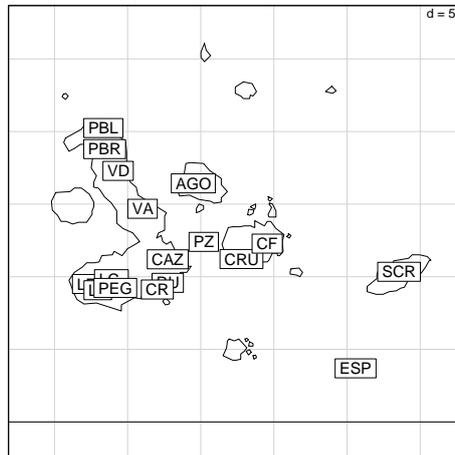
Ce que les auteurs disent des sites d'études a été regroupé dans `pps069$pop`, listé dans le tableau 1.

	x	y	isl	carap	n	subspe
PBL	83.4	202.4	Isabela	SD	20	becki
PBR	84.3	187.9	Isabela	SD	27	becki
VD	93.5	172.9	Isabela	D	24	microphyles
VA	110.2	146.8	Isabela	D	20	vandenburghi
CAZ	127.4	112.0	Isabela	D	20	guntheri
RU	127.4	95.8	Isabela	D	20	guntheri
CR	120.3	91.2	Isabela	D	20	guntheri
LP	72.0	95.0	Isabela	D	20	vicina
LT	79.6	90.7	Isabela	D	20	vicina
LC	88.9	98.3	Isabela	D	20	vicina
PEG	91.6	92.3	Isabela	D	19	vicina
CRU	177.8	112.2	Santa_Cruz	D	20	porteri
CF	195.4	122.8	Santa_Cruz	D	16	porteri
PZ	152.0	124.0	Pinzon	S	20	ephyppium
AGO	144.9	164.3	Santiago	SD	20	darwini
SCR	285.5	103.4	San_Cristobal	S	20	chatamensis
ESP	255.8	36.9	Espanola	S	15	hoodensis

TAB. 1 – Information sur les populations étudiées consignées dans `pps069$pop`. `x` et `y`, coordonnées des sites sur la carte, `isl`, nom de l'île contenant le site, `carap`, type de la carapace des tortues de la population étudiée, `n`, nombre d'individus de l'échantillon, `subspe`, nom de la sous-espèce à laquelle appartient la population étudiée. Chaque population est indiquée par son code.

Les coordonnées sont approximatives et reconstituées à partir du texte. Les points voisins ont été écartés pour plus de lisibilité.

```
s.label(pps069$area[, -1], clab = 0, cpoi = 0, area = pps069$area,
        xlim = c(25, 325))
s.label(pps069$pop, add.p = T, clab = 1.25)
```



3 Les populations étudiées

En chaque site une population est échantillonnée. On distingue parmi ces tortues 15 sous-espèces :

Genus	species	race	Island
<i>Geochelone</i>	<i>elephantopus</i>	<i>elephantopus</i>	Floreana (éteint)
		non décrit	Santa Fe (éteint)
		<i>phantastica</i>	Fernandina (éteint)
		<i>wallacei</i>	Rabida (éteint)
		<i>hoodensis</i>	Espanola
		<i>abingdoni</i>	Pinta
		<i>ephippium</i>	Pinzon
		<i>chatamensis</i>	San Cristobal
		<i>darwinii</i>	Santiago
		<i>vicina</i>	Volcan Cerro Azul Isabela
		<i>guntheri</i>	Volcan Sierra Negra, Isabela
		<i>vandenburghi</i>	Volcan Alcedo, Isabela
		<i>microphyes</i>	Volcan Darwin, Isabela
		<i>becki</i>	Volcan Wolf, Isabela
		<i>porteri</i>	Santa Cruz

Ces races peuvent être divisées en deux grands types morphologiques : la carapace a soit une forme de dôme (fig. 2), soit une forme de selle d'équitation (fig. 3). Ces formes sont aisément différenciées, notamment par l'aspect de la ligne située au dessus du coup, basse chez les premières, voutée chez les secondes.

Enfin, un morphotype intermédiaire est présent sur l'île de Santiago. Cette information est contenue dans le facteur `pps069popcarap` (D : dôme ; S : selle ; DS : intermédiaire). Trois icônes ont été conservées dans `pps069$ico`, ce qui permet de faire des figures avec des symboles.



FIG. 2 – Carapace en forme de dôme.

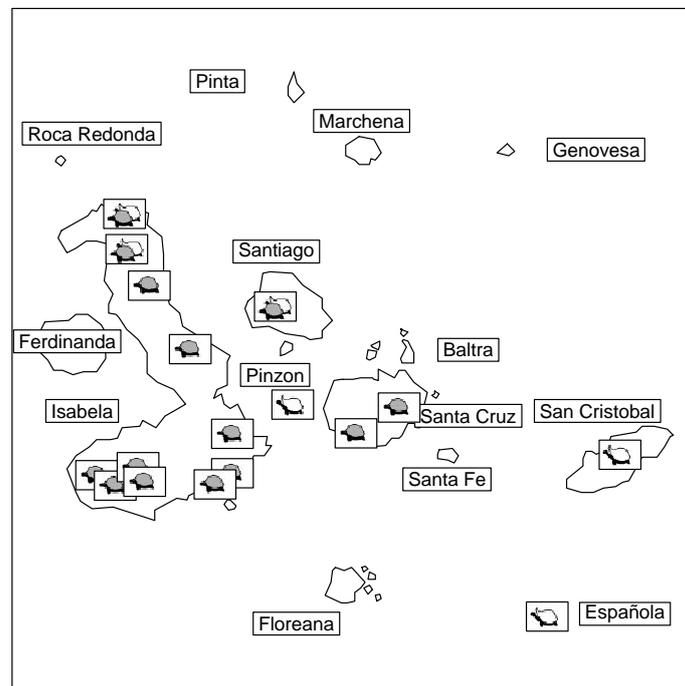


FIG. 3 – Carapace en forme de selle.

```

library(pixmap)
clogo <- 0.5
marold <- par("mar")
par(mar = c(0.1, 0.1, 0.1, 0.1))
area.plot(pps069$area)
s.label(pps069$misc, add.plot = T)
a1 <- 1:length(pps069$ico)
names(a1) <- names(pps069$ico)
a1 <- a1[as.character(pps069$pop$carap)]
for (k in 1:17) {
  a2 <- pps069$ico[[a1[k]]]
  x2 <- pps069$pop[k, 1]
  y2 <- pps069$pop[k, 2]
  lx <- attr(a2, "size")[2] * clogo/2
  ly <- attr(a2, "size")[1] * clogo/2
  addlogo(a2, c(x2 - lx, x2 + lx), c(y2 - ly, y2 + ly))
  rect(x2 - lx, y2 - ly, x2 + lx, y2 + ly)
}
par(mar = marold)

```



4 Les microsattellites

Des différents marqueurs moléculaires existants, les microsattellites sont parmi les plus utilisés à l'heure actuelle [4]. Ce sont des courtes séquences d'ADN (de deux à cinq nucléotides) répétées en tandem sur une longueur souvent inférieure à 100 paires de bases. Ce nombre de répétitions est hautement variable (jusqu'à plus de 30 allèles rapportés dans [3]), ce qui présente le premier avantage des microsattellites. De plus, ces marqueurs sont codominants (on distingue les hétérozygotes des homozygotes). On leur trouve de multiples applications : *"Les microsattellites étant des marqueurs hautement polymorphes présentent de nombreuses applications. Hormis les études de génétique de populations ou de parenté, ils sont utilisés pour la réalisation de cartes génomiques, pour du typage génotypique, et pour établir des empreintes génétiques en biotechnologie ou en criminologie."* [2].

`ggtortoises$loc` contient le nom des 10 microsattellites analysés et le nombre d'allèles pour chaque locus.

```
pps069$loc
```

```
GAL45  GAL50  GAL73  GAL75  GAL94  GAL100  GAL127  GAL136  GAL159  GAL263
   13    37    20    22    13    19    31    12    19    25
```

Le tableau `pps069$tab` contient pour chaque population et pour chaque forme allélique de chaque locus l'effectif observé. Pour retrouver les fréquences alléliques publiées (op.cit. Appendix I), il suffit de diviser par $2n$ où n est le nombre d'individus échantillonnés.

```
t(pps069$tab[, 1:13])
```

	PBL	PBR	VD	VA	CAZ	RU	CR	LP	LT	LC	PEG	CRU	CF	PZ	AGO	SCR	ESP
GAL45.87	11	7	2	8	3	4	6	7	14	11	13	31	27	7	2	7	1
GAL45.95	0	12	0	0	0	0	0	0	0	0	0	0	0	0	0	10	10
GAL45.103	0	0	0	0	0	0	0	0	0	0	0	0	0	0	4	4	0
GAL45.105	0	0	1	0	0	0	0	0	0	0	0	0	0	4	0	0	0
GAL45.107	23	35	27	21	5	11	8	18	12	6	14	3	5	28	20	16	19
GAL45.109	6	0	9	2	0	0	2	0	0	0	1	0	0	1	8	0	0
GAL45.111	0	0	0	9	6	13	7	7	4	16	7	5	0	0	5	3	0
GAL45.113	0	0	0	0	0	2	3	2	0	1	2	1	0	0	0	0	0
GAL45.115	0	0	7	0	14	5	9	2	8	6	0	0	0	0	1	0	0
GAL45.117	0	0	0	0	0	0	2	0	0	0	0	0	0	0	0	0	0
GAL45.119	0	0	2	0	0	0	0	0	0	0	0	0	0	0	0	0	0
GAL45.121	0	0	0	0	10	5	3	4	2	0	1	0	0	0	0	0	0
GAL45.123	0	0	0	0	2	0	0	0	0	0	0	0	0	0	0	0	0

```
round(t(pps069$tab[, 1:13]/2/pps069$pop$n)[, 1:5], dig = 3)
```

	PBL	PBR	VD	VA	CAZ
GAL45.87	0.275	0.130	0.042	0.200	0.075
GAL45.95	0.000	0.222	0.000	0.000	0.000
GAL45.103	0.000	0.000	0.000	0.000	0.000
GAL45.105	0.000	0.000	0.021	0.000	0.000
GAL45.107	0.575	0.648	0.562	0.525	0.125
GAL45.109	0.150	0.000	0.188	0.050	0.000
GAL45.111	0.000	0.000	0.000	0.225	0.150
GAL45.113	0.000	0.000	0.000	0.000	0.000
GAL45.115	0.000	0.000	0.146	0.000	0.350
GAL45.117	0.000	0.000	0.000	0.000	0.000
GAL45.119	0.000	0.000	0.042	0.000	0.000
GAL45.121	0.000	0.000	0.000	0.000	0.250
GAL45.123	0.000	0.000	0.000	0.000	0.050

Le tableau contient donc, au plan statistique 17 lignes et 210 colonnes. Il est massivement multivarié. Cet exemple permet d'apprendre à manipuler ce type d'information.

Références

- [1] M.C. Ciofi, C. Milinkovitch, J.P. Gibbs, A. Caccone, and J.R. Powell. Microsatellite analysis of genetic divergence among populations of giant galapagos tortoises. *Molecular Ecology*, 11 :2265–2283, 2002.
- [2] M. Harry. *Génétique Moléculaire et Évolutive*. 2001.
- [3] P.G. Parker, A.A. Snow, M.D. Schug, G.C. Booton, and P.A. Fuerst. What molecules can tell us about populations : choosing and using a molecular marker. *Ecology*, 79(2) :361–382, 1998.
- [4] C. Schlötterer. The evolution of molecular markers - just a matter of fashion? *Genetics*, 5 :63–69, 2004.