Supplementary materials for *adegenet 1.3-0*: new tools for the analysis of genome-wide SNP data

Thibaut Jombart*, Ismaïl Ahmed

Table 1: Comparison of the performances of `glPca`, `dudi.pca` and `prcomp` in terms of (i) RAM necessary for storing the SNP data (Input size), (ii) RAM used for storing the output of the analysis (Output size) and (iii) time necessary for completing the analysis (Runtime). Results are presented for several datasets, which contain the same number of cells (50,000,000) but which vary according to the number of individuals (n) and SNPs (p).

	Method	Input size (MB)	Output size (MB)	Runtime (min)
$n =$ **50,** $p =$ **1,000,000**				
	glPca	6.03	68.67	2.99
	dudi.pca	381.47	694.29	69.01
	prcomp	381.47	389.13	0.50
$n =$ **100,** $p =$ **500,000**				
	glPca	6.09	34.34	5.66
	dudi.pca	381.47	537.89	17.24
	prcomp	381.47	385.37	0.56
$n =$ **500,** $p =$ **100,000**				
	glPca	6.63	6.91	26.73
	dudi.pca	381.47	412.83	3.71
	prcomp	381.47	384.20	2.21
$n =$ **1000,** $p =$ **50,000**				
	glPca	7.29	3.51	53.14
	dudi.pca	381.47	397.27	2.77
	prcomp	381.47	389.60	3.92

As a second experiment, we investigated how many SNPs could be handled by `glPca`, `dudi.pca` and `prcomp` on a standard R session (Windows 32bits) before the latter crashed. The number of individuals was fixed to 50, while varying the number of SNPs analyzed. The maximum number of SNPs that could be analyzed by `glPca` was comprised between 19,000,000 and 20,000,000, while this range went down to 1,200,000–1,300,000 for `prcomp` and 400,000–500,000 for `dudi.pca`.

*to whom correspondence should be addressed