

Analysing genomic data using adegenet

Thibaut Jombart and Ismail Ahmed

May 30, 2011

Contents

1 Introduction	1
2 Classes of objects	1
2.1 SNPbin : storage of single genomes	1
2.2 genlight : storage of multiple genomes	1
3 In practice	1
3.1 Using accessors	1
3.2 Data conversions	1
3.3 Principal Component Analysis (PCA)	1
3.4 Discriminant Analysis of Principal Components (DAPC)	1

1 Introduction

Modern sequencing technologies now make complete genomes more widely accessible. The subsequent amounts of genetic data pose challenges in terms of storing and handling the data, making former tools developed for classical genetic markers such as microsatellite impracticable on standard computers. Adegenet has developed new object classes dedicated to handling genome-wide polymorphism (SNPs) with minimum rapid access memory (RAM) requirements.

Two new formal classes have been implemented: **SNPbin**, used to store genome-wide SNPs for one individual, and **genlight**, which stored the same information for multiple individuals. In this vignette, we present these classes and show how they can be used for genetic data analysis.

2 Classes of objects

2.1 SNPbin: storage of single genomes

2.2 genlight: storage of multiple genomes

3 In practice

3.1 Using accessors

3.2 Data conversions

3.3 Principal Component Analysis (PCA)

3.4 Discriminant Analysis of Principal Components (DAPC)