snpMatrix 4-digit Series

Hin-Tak Leung

December 14, 2012

1 Introduction

The snpMatrix 4-digit series started with 1.2.2.1 (Mar 15 2008) and reaches 1.19.0.17 (Dec 10 2012).
It consists of changes, bug fixes or enhancements, some of which appeared in later 3-digit series,
and also enhancements which never did, such that the functionality to natively read Illumina's and
DECODE's binary file formats.

Version 1.4.0 (April 2008) to 1.18.0/1.19.0 (October 2011) were briefly bundled with Bioconductor.

2 Details

2.1 Between 1.19.0 and 1.19.0.17

The difference between snpMatrix 1.19.0 and 1.19.0.17 consists of 498 in-tree commits and 385+
out-of-tree commits.

The in-tree commits consists of the following:

- Fixing bugs introduced between 2008 and 2011, by re-analysing the 2007 WTCCC1 data.
  (see also the out-of-tree commits section). Also examples and internal checks to make sure
  that bugs are not re-added.

- As a the side-effect of the bug fixes, a fair number of David Clayton's home-brewed (and some
  buggy) linear algedra routines were replaced with their R LAPACK and BLAS equivalents.
  There is a few percents in speed gain.

- Natively read Illumina's EGT, BPM, and CNV-related binary file formats, as well as some
  of DECODE's.

- Major under-the-hood changes related to link-loaders; A few percents in speed gain, more
  noticeable in PE (i.e. Windows), less so in ELF/MACH-O (Linux/Mac OS X).

The bottom of the in-tree commits is `81b60ada308bf1fe90a6f4e5e308cea2a28b118f`, Feb 14
2011. The in-tree commits have been rebased many times, and not necessarily in chronological
order.

2.1.1   The out-of-tree commits

The out-of-tree commits (385 up to snpMatrix 1.19.0.17, December 2012) started on Aug 26 2011, with `73f134d96760f48aef836068a620bc556234a397`, which resurrect the 2007 infra-structure for analysing WTCCC1 data, to track down bugs introduced between 2008 and 2011.

The first version of this document is the 386th.

Many of the out-of-tree commits are out of tree because of special requirements. For example:

- The "China and Chinese Cities" vignette requires Tibetan and Arabic capabilities on the host platform, as well as a cairo-capable R.

- The "Algorithms and Thailand" vignette and the "China and Chinese Cities" vignette requires LaTeXcapable of Thai and Chinese respectively.

- The "Regression and Migration" vignette requires the WTCCC1 data.

- The "MHC Subset Preparation" vignette requires the UK AMD data.

- The "snpStats Bug" vignettes obviously requires loading other extra R packages.

A   snpStats 4-digit Series

The "less buggy" snpStats 4-digit Series consists of the follow bug fixes, none of which are in the 3-digit series:

4. ld statistics from `ld()` (the older +ld.snps()+ in snpMatrix is not affected).

3. cbind/rbind

2. Input/Output and conversion of uncertain genotypes.

1. Statistical tests plus miscellaneous bugs accumulated between 2008 and 2011.

More details in the "snpStats Bug" vignette.