snpMatrix 4-digit Series

Hin-Tak Leung

January 14, 2015

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[1]). 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 1.19.0.20

snpMatrix 1.19.0.20 ships a snpStats-compatibility mode for the first time. This mode, when set on, is initially synchronized with and emulate less-buggy snpStats 1.17.0.8. The extensive work was done during the period spanning and commemorating 25.89.64 and the 79 days of protests. It is 19 months of work and consists of about 400 commits in total, just under 300 in tree. See the new Compatibility Mode vignette for details.

Small new incompatibility in `show(X.snp.matrix,)` for males. It now does "A", "B", instead of "A/Y", "B/Y".

2.2 1.19.0.18—1.19.0.19

1.19.0.19, June 2013 Various small and misc updates.

1.19.0.18, April 2013 New uncertain option to `read.snps.chiamo()` for generating uncertain genotypes, and `new read.snps.vcf()` routine. New "IO Suite" vignette (requiring 1000 Genome and HapMap data to run).

2.3 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:

---

[1]As of the first draft of this document.

- 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 side-effect of the bug fixing effort, a fair number of David Clayton's home-brewed (and some buggy) linear algebra 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.

2.3.1 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:

8. (Jan 3 2015) GLM-related bug fix. Compared very different routines of same name between the abandoned snpMatrix2 branch and snpStats and found both wrong.

7. (Aug 8 2014) Indeterminisms in data input vignette and imputation vignette, and read.mach(). EM step in imputation can silently gives wrong result (not fixed, just issuing warning for now). Also new ISSUES file listing known but not yet addressed issues.

6. (Aug 24 2013) first non-(internal)-bug-fix release - upgrade bundled zlib from 1.2.5 to 1.2.8 . Note that snpMatrix upgraded to zlib 1.2.8 in May 2013, and to 1.2.7 in Dec 2012, and have additional performance and portability enhancements in the zlib-related code.

5. (Mar 28 2013) second bug in Input/Output and conversion of uncertain genotypes.

4. (Nov 19 2012) ld statistics from `ld()` (the older +ld.snps()+ in snpMatrix is not affected).

3. (Nov 6 2012) cbind/rbind

2. (Sep 1 2012) Input/Output and conversion of uncertain genotypes.

1. (Oct 16 2011) Statistical tests plus miscellaneous bugs accumulated between 2008 and 2011.

More details in the "snpStats Bug" vignette.