RESEARCH

Methods to compute the composite log-likelihood (CLL) of allelic frequencies for the detection of signatures of selection in diploid genomes

Filippo Biscarini[1*], Nelson Nazzicari[1] and Alessandra Stella[1]

[*]Correspondence:
filippo.biscarini@tecnoparco.org
[1]Department of Bioinformatics, PTP, Via Einstein - Loc. Cascina Codazza, Lodi, Italy
Full list of author information is available at the end of the article

Abstract

First part title: Text for this section.

Second part title: Text for this section.

Keywords: composite log-likelihood; signatures of selection; diploid genomes

Background

Selection, both natural and artificial, is one of the major forces that can shape the genome of living organisms and change allele frequencies. A mutation that is beneficial for the adaptation of an organism to its environment, or that is of agricultural or industrial interest, tends to increase in frequency in the population, together with neighbouring genomic regions which tend to be hitch-hiked in the process ([1]). Through the last decade, the on-going genomic revolution has been making available hundreds of thousands of genetic markers for several animal, microbial and plant species. By looking at the allele frequency at marker loci along the genome of populations experiencing different selective pressures, it is possible to identify genomic regions -and ultimately genes- involved in processes such as domestication, adaptation, evolution and artificial selection. There are a number of methods based on allele frequencies that have been developed to detect such signatures of selection. A popular method is Wright's F_{ST} ([2, 3]) that has been applied to studies in humans ([4]), plants ([5]) and animals ([6]). Alternatively, the likelihood of the difference between allele frequencies in different populations can be estimated and used to detect the presence of signatures of selection ([7, 8]). However, there are several ways in which such likelihoods can be computed, and these might differ in computation requirements and statistical properties, such as the sensitivity to detect signals of selection or the behaviour along the margins of the dimensional space. It may therefore be of interest to investigate the statistical properties of different estimators for the likelihood of the difference between allele frequency, and assess how well they are capable of detected actual signals of selection.

In this study we evaluated 5 (check the number!) different ways to estimate the likelihood of the difference in allele frequency between populations: 1) .... The logarithm of the likelihoods thus calculated were then computed and combined across sliding windows along the genome (CLL, composite log-likelihood) in order to detect signatures of selection. The proposed CLL approaches were finally compared

with approaches based on the F_{ST} and on simple Euclidean distances between genotypes. The results of the study are hereby presented and discussed.

Methods

Numerical illustration

Text for this sub-heading ...

Working example

Working example

Text for this sub-sub-heading ...

Sub-sub-sub heading for section Text for this sub-sub-sub-heading ... In this section we examine the growth rate of the mean of Z_0, Z_1 and Z_2. In addition, we examine a common modeling assumption and note the importance of considering the tails of the extinction time T_x in studies of escape dynamics. We will first consider the expected resistant population at vT_x for some $v > 0$, (and temporarily assume $\alpha = 0$)

$$E\big[Z_1(vT_x)\big] = E\bigg[\mu T_x \int_0^{v\wedge 1} Z_0(uT_x)\exp\big(\lambda_1 T_x(v-u)\big)\,du\bigg].$$

If we assume that sensitive cells follow a deterministic decay $Z_0(t) = xe^{\lambda_0 t}$ and approximate their extinction time as $T_x \approx -\frac{1}{\lambda_0}\log x$, then we can heuristically estimate the expected value as

$$\begin{aligned} E\big[Z_1(vT_x)\big] &= \frac{\mu}{r}\log x \int_0^{v\wedge 1} x^{1-u}x^{(\lambda_1/r)(v-u)}\,du \\ &= \frac{\mu}{r}x^{1-\lambda_1/\lambda_0 v}\log x \int_0^{v\wedge 1} x^{-u(1+\lambda_1/r)}\,du \\ &= \frac{\mu}{\lambda_1-\lambda_0}x^{1+\lambda_1/rv}\left(1-\exp\left[-(v\wedge 1)\left(1+\frac{\lambda_1}{r}\right)\log x\right]\right). \quad (1) \end{aligned}$$

Thus we observe that this expected value is finite for all $v > 0$ (also see [9, 10, 11, 12, 13]).

Results and discussion

Text ...

Competing interests

The authors declare that they have no competing interests.

Author's contributions

Text for this section ...

Acknowledgements

Text for this section ...

Author details

[1]Department of Bioinformatics, PTP, Via Einstein - Loc. Cascina Codazza, Lodi, Italy. [2]Marine Ecology Department, Institute of Marine Sciences Kiel, Düsternbrooker Weg 20, 24105 Kiel, Germany.

References

1. Braverman, J.M., Hudson, R.R., Kaplan, N.L., Langley, C.H., Stephan, W.: The hitchhiking effect on the site frequency spectrum of dna polymorphisms. Genetics **140**(2), 783–796 (1995)
2. Wright, S.: The genetical structure of populations. Annals of eugenics **15**(1), 323–354 (1949)
3. Nei, M.: Analysis of gene diversity in subdivided populations. Proceedings of the National Academy of Sciences **70**(12), 3321–3323 (1973)
4. Akey, J.M., Zhang, G., Zhang, K., Jin, L., Shriver, M.D.: Interrogating a high-density snp map for signatures of natural selection. Genome research **12**(12), 1805–1814 (2002)
5. Zhao, K., Wright, M., Kimball, J., Eizenga, G., McClung, A., Kovach, M., Tyagi, W., Ali, M.L., Tung, C.-W., Reynolds, A., *et al.*: Genomic diversity and introgression in o. sativa reveal the impact of domestication and breeding on the rice genome. PLoS One **5**(5), 10780 (2010)
6. Kijas, J.W., Lenstra, J.A., Hayes, B., Boitard, S., Neto, L.R.P., San Cristobal, M., Servin, B., McCulloch, R., Whan, V., Gietzen, K., *et al.*: Genome-wide analysis of the world's sheep breeds reveals high levels of historic mixture and strong recent selection. PLoS biology **10**(2), 1001258 (2012)
7. Nielsen, R., Williamson, S., Kim, Y., Hubisz, M.J., Clark, A.G., Bustamante, C.: Genomic scans for selective sweeps using snp data. Genome research **15**(11), 1566–1575 (2005)
8. Stella, A., Ajmone-Marsan, P., Lazzari, B., Boettcher, P.: Identification of selection signatures in cattle breeds selected for dairy production. Genetics **185**(4), 1451–1461 (2010)
9. Koonin, E.V., Altschul, S.F., Bork, P.: Brca1 protein products: functional motifs. Nat Genet **13**, 266–267 (1996)
10. Kharitonov, S.A., Barnes, P.J.: Clinical Aspects of Exhaled Nitric Oxide. in press
11. Zvaifler, N.J., Burger, J.A., Marinova-Mutafchieva, L., Taylor, P., Maini, R.N.: Mesenchymal cells, stromal derived factor-1 and rheumatoid arthritis [abstract]. Arthritis Rheum **42**, 250 (1999)
12. Jones, X.: Zeolites and synthetic mechanisms. In: Smith, Y. (ed.) Proceedings of the First National Conference on Porous Sieves: 27-30 June 1996; Baltimore, pp. 16–27 (1996). Stoneham: Butterworth-Heinemann
13. Margulis, L.: Origin of Eukaryotic Cells. Yale University Press, New Haven (1970)

Figures

Figure 1 Sample figure title. A short description of the figure content should go here.

Figure 2 Sample figure title. Figure legend text.

Tables

Table 1 Sample table title. This is where the description of the table should go.

	B1	B2	B3
A1	0.1	0.2	0.3
A2	...	..	.
A3	..	.	.

Additional Files

Additional file 1 — Sample additional file title
Additional file descriptions text (including details of how to view the file, if it is in a non-standard format or the file extension). This might refer to a multi-page table or a figure.

Additional file 2 — Sample additional file title
Additional file descriptions text.