

Local Score Based Method On Pool-Sequenced Behavior-Divergent Quail Lines Precisely Detected Selection Signatures Related To Autism

Maria Inès Fariello Rico, Cécile Arnould, Simon Boitard, Patrice Dehais, Thomas Faraut, Elisabeth Duval, Christine Leterrier, Sylvie Mercier, Frederique Pitel, Julien Recoquillay, et al.

▶ **To cite this version:**

Maria Inès Fariello Rico, Cécile Arnould, Simon Boitard, Patrice Dehais, Thomas Faraut, et al.. Local Score Based Method On Pool-Sequenced Behavior-Divergent Quail Lines Precisely Detected Selection Signatures Related To Autism. Journée Régionale de Bioinformatique et Biostatistique, Génopole Toulouse, Jun 2014, Toulouse, France. 2014. hal-01194044

HAL Id: hal-01194044
https://hal.science/hal-01194044

Submitted on 3 Jun 2020

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Local Score Based Method On Pool-Sequenced Behavior-Divergent Quail Lines Precisely Detected Selection Signatures Related To Autism

M.I. Fariello[1,2,3,4,5], C. Arnould[6], S. Boitard[7,8], P. Dehais[10], T. Faraut[1,3,4], E. Lebihan[11], C. Leterrier[6], S. Mercier[11,12], F. Pitel[1,3,4], J. Recoquillay[11], D. Robelin[1,3,4], G. Salin[1,3,4,13], M. SanCristobal[1,3,4,14]

[1] GENPHYSE, UMR1388 INRA-INPT-ENVT-Université de Toulouse, Auzeville, BP32627, 31326, Castanet Tolosan, Cedex, France

Magali.san-cristobal@toulouse.inra.fr

[2] Institut Pasteur, Montevideo, Uruguay, [3]Université de Toulouse INPT ENSAT, France, [4]Université de Toulouse INPT ENVT, France, [5]Universidad de la República, Montevideo, Uruguay, [6]INRA – CNRS – Université de Tours, France, [7]INRA, Jouy-en-Josas, France, [8]MNHN / CNRS / EPHE / UPMC, Paris, France, [9]AgroParisTech, Paris, France, [10]INRA Toulouse, France,[11]INRA Tours, Nouzilly, France, [11]Université Toulouse Le Mirail, France, [12]Institut de Mathématiques de Toulouse, France, [13]INRA Castanet-Tolosan, France, [14]INSA Toulouse, France

[1] Introduction

Detecting genomic regions that have evolved under selection has received much interest these last years. Linkage disequilibrium (LD) leads to the persistence of a footprint of selection around positively selected mutations, so the selection signature is not limited to a single causal mutation, but generally extends to a wider genomic interval including this mutation. Consequently, detection power is expected to increase when searching for such intervals rather than considering markers independently of each other. Haplotypic methods (e.g. Voight et al. (2006), Sabeti et al. (2007), Fariello et al. (2013)) require individual genotype data (for inferring haplotypes) and are rather computationally demanding. On the other hand, single marker statistics have a lot of variability. Sliding window approaches do not require individual genotypes and run faster than haplotype-based methods. However, they are also less powerful and imply to choose a window size, which is usually done arbitrarily. To overcome this problem, alternative approaches to find clusters of high Fst values were proposed by Myles et al. (2008) and Johansson et al. (2010).

The objective of this study is to present a new strategy for detecting footprints of selection, which runs very fast and is suited to the case where only allele frequencies are available. We demonstrate that this novel approach performs well by detecting genes associated to social reinstatement behavior in quail, using unpublished pooled NGS data from two quail lines that have been divergently selected for this trait. This strategy aims at accumulating selection signals via small p-values of single-marker tests, (or equivalently large values of -log10(p-value), further defined as a score) in an automatic manner, using the statistical theory of local scores (e.g. Mercier et al. (2003)). The local score approach plays an important role in bio-informatics, where it is used for computing sequence alignment scores (Karlin and Altschul (1990)).

[2] Materials and Methods

1.1 Quail data

Two divergent lines produced and maintained at the INRA experimental unit 1295 (UE PEAT, F-37380 Nouzilly, France) were used in the experiment. These lines show high (HSR) and low social reinstatement behavior (LSR) (e.g. Mills and Faure (1991), Schweitzer et al. (2011)). Ten individuals from generation 50 of each quail line were pooled and sequenced (paired-ends, 100 bp) on a HiSeq 2000 sequencer (Illumina). In the absence of an available genome sequence for the quail, the reads of the two divergent lines were mapped to the chicken genome assembly (GallusWU2.58). Allele frequencies and then F-LK values (i.e. Fst corrected for population structure, Bonhomme et al. (2010)) were finally computed at all SNPs.

1.2 Model

We propose to start from results of single marker tests along the genome and to highlight segments of adjacent loci with small p-values. More precisely, our strategy is to consider high values of the "score" X = -log10(p-value) (up to an additive constant) and to accumulate them over adjacent loci. The Lindley process h=(h_1,…h_l,…h_L) is defined as h_l=max(0,h_(l-1)+X_l) and h_0=0. The local score is the maximum of the Lindley process...

[3] Results

On the quail data, all methods gave unclear patterns, except the local score that clearly pinpointed 4 short genomic regions (see Figure 2 for GGA1). One to two genes laid in each candidate region. Some of them are known to be related to behavior traits in Human, in particular related to autism, suggesting that the detected regions are not false positives.

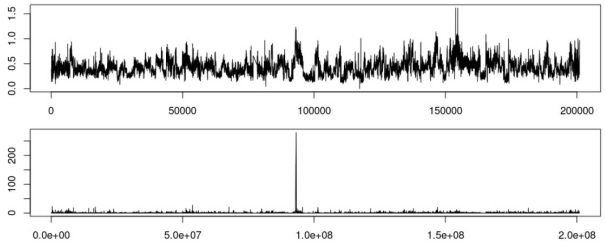


Figure 2: Selection footprints for Quail data on GGA1. Top: windowed Fst (sliding window of 10kb), bottom: local score (Lindley process) of score function -log10(pvalue(Fst))-1.

References

[4] Bonhomme, M., Chevalet, C., Servin, B., et al. (2010). Genetics, 186:241-262.
[5] Fariello, M.I., Boitard, S., Naya, H. et al. (2013). Genetics, 193:929-41.
[6] Guedj, M., Robelin, D., Hoebeke, M., et al. (2006). Stat. Appl Genet. Mol. Biol., 5: Article22.
[7] Johansson, A.M., Pettersson, M.E., Siegel, P.B., Carlborg, O. (2010). PLoS Genet., 6: e1001188.
[8] Karlin, S., Altschul, S.F. (1990). PNAS, 87:2264-8.
[9] Mercier, S., Cellier, D., Charlot, D. (2003). Applied Probability Trust.
[10] Mills, A., Faure, J. (1991). J. Comparative Psychology, 105:25-38.
[11] Myles, S., Tang, K., Somel, M., et al. (2008). Ann. Hum. Genet., 72:99-110.
[12] Sabeti, P.C., Varilly, P., Fry, B. et al. (2007). Nature, 449: 913-918.
[13] Schweitzer, C., Levy, F., Arnould, C. (2011). Animal Behaviour, 81:535-542.
[14] Voight, B.F., Kudaravalli, S., Wen, X., Pritchard, J.K. (2006). PLoS Biol., 4: e72.