University of Massachusetts Medical School
eScholarship@UMMS

GSBS Student Publications | Graduate School of Biomedical Sciences

12-1-2011

On detecting selective sweeps using single genomes

Priyanka Sinha
Ecole Polytechnique Federale de Lausanne

Aslihan Dincer
University of Massachusetts Medical School, aslihan.dincer@umassmed.edu

Daniel Virgil
University of Massachusetts Medical School, Daniel.Virgil2@umassmed.edu

See next page for additional authors

Follow this and additional works at: http://escholarship.umassmed.edu/gsbs_sp

Part of the Bioinformatics Commons, Genetics and Genomics Commons, and the Medicine and Health Sciences Commons

Repository Citation

Sinha, Priyanka; Dincer, Aslihan; Virgil, Daniel; Xu, Guang; Poh, Yu-Ping; and Jensen, Jeffrey D., "On detecting selective sweeps using single genomes" (2011). *GSBS Student Publications.* Paper 1776.
http://escholarship.umassmed.edu/gsbs_sp/1776

This material is brought to you by eScholarship@UMMS. It has been accepted for inclusion in GSBS Student Publications by an authorized administrator of eScholarship@UMMS. For more information, please contact Lisa.Palmer@umassmed.edu.

On detecting selective sweeps using single genomes

Authors
Priyanka Sinha, Aslihan Dincer, Daniel Virgil, Guang Xu, Yu-Ping Poh, and Jeffrey D. Jensen

Comments
Copyright: © 2011 Sinha, Dincer, Virgil, Xu, Poh and Jensen. This is an open-access article distributed under the terms of the Creative Commons Attribution Non Commercial License, which permits non-commercial use, distribution, and reproduction in other forums, provided the original authors and source are credited.

Rights and Permissions
Citation: Sinha P, Dincer A, Virgil D, Xu G, Poh Y-P and Jensen JD (2011) On detecting selective sweeps using single genomes. Front. Gene. 2:85. doi: 10.3389/fgene.2011.00085. Link to article on publisher's website

This article is available at eScholarship@UMMS: http://escholarship.umassmed.edu/gsbs_sp/1776

frontiers in GENETICS

ORIGINAL RESEARCH ARTICLE
published: 01 December 2011
doi: 10.3389/fgene.2011.00085

On detecting selective sweeps using single genomes

Priyanka Sinha[1], Aslihan Dincer[2], Daniel Virgil[2], Guang Xu[2], Yu-Ping Poh[2] and Jeffrey D. Jensen[1,2]*

[1] School of Life Sciences, École Polytechnique Fédérale de Lausanne, Lausanne, Switzerland
[2] Program in Bioinformatics and Integrative Biology, University of Massachusetts Medical School, Worcester, MA, USA

Edited by:
Rinaldo Wellerson Pereira, Universidade Católica de Brasília, Brazil

Reviewed by:
Jacob A. Tennessen, Oregon State University, USA
Andrew Kern, Dartmouth College, USA

***Correspondence:**
Jeffrey D. Jensen, School of Life Sciences, École Polytechnique Fédérale de Lausanne, Lausanne, Switzerland.
e-mail: jeffrey.jensen@epfl.ch

Identifying the genetic basis of human adaptation has remained a central focal point of modern population genetics. One major area of interest has been the use of polymorphism data to detect so-called "footprints" of selective sweeps – patterns produced as a beneficial mutation arises and rapidly fixes in the population. Based on numerous simulation studies and power analyses, the necessary sample size for achieving appreciable power has been shown to vary from a few individuals to a few dozen, depending on the test statistic. And yet, the sequencing of multiple copies of a single region, or of multiple genomes as is now often the case, incurs considerable cost. Enard et al. (2010) have recently proposed a method to identify patterns of selective sweeps using a single genome – and apply this approach to human and non-human primates (chimpanzee, orangutan, and macaque). They employ essentially a modification of the Hudson, Kreitman, and Aguade test – using heterozygous single nucleotide polymorphisms from single individuals, and divergence data from two closely related species (human–chimpanzee, human–orangutan, and human–macaque). Given the potential importance of this finding, we here investigate the properties of this statistic. We demonstrate through simulation that this approach is neither robust to demography nor background selection; nor is it robust to variable recombination rates.

Keywords: selective sweeps, demography, adaptation, statistical inference

INTRODUCTION

When a new beneficial mutation increases in frequency and fixes in the population, linked neutral and nearly neutral mutations will similarly increase in frequency on the selected haplotype (Smith and Haigh, 1974). This results in a number of well-described patterns – including a reduction in local variation, a skewed site frequency spectrum, and spatially elevated levels of linkage disequilibrium (see review of Nielsen, 2005). This phenomenon is known as a selective sweep. Many methods have been proposed utilizing these patterns to detect beneficial fixations, using both polymorphism and divergence data in a variety of statistical frameworks. One of the biggest problems confronting all such methodologies is the challenge of distinguishing these patterns from those produced under neutral non-equilibrium conditions (e.g., population size change and population structure) – a problem that different approaches grapple with to varying degrees of success (see review of Thornton et al., 2007). And yet, despite these differences between approaches – one commonality of polymorphism based test statistics is indeed the need for polymorphism data from multiple sequenced individuals.

Thus, owing to the inherent difficulty and cost of generating data appropriate for existing methods of statistical inference, the recent claims of Enard et al. (2010) have attracted considerable attention. Using a whole-genome shotgun sequence of a single human, they argue that they have sufficient power to identify swept regions of the genome. Extending this analysis to chimpanzee, orangutan, and macaque, they focus upon "sweep hotspots" shared among apes, and argue for the long-term adaptive significance of these genomic regions.

The authors use a method inspired by the Hudson, Kreitman, and Aguade (HKA) test (Hudson et al., 1987). This statistic employs intra-specific polymorphism data and inter-specific divergence data to test for deviations from the Standard Neutral model. Enard et al. (2010) derive their polymorphism data from a single heterozygous individual, and the divergence information from comparison with a single chimp, orangutan, or macaque genome. This approach is termed "K–estimation," where the corrected level of heterozygosity is indicated by the value of a statistic K ($0 \leq K \leq 1$).

They assess two single human individuals, the Venter and Watson genomes. They use a co-occurrence test to find the difference between the expected and observed values for the K-statistic among orthologous genes. Numerous other recent studies have sought to identify and describe genome-wide positive selection in humans using various aspects of polymorphism and divergence data (e.g., Sabeti et al., 2006, 2007; Voight et al., 2006; Williamson et al., 2007) – mostly using single nucleotide polymorphisms (SNP) data from HapMap (Frazer et al., 2007) or Perlegen (Hinds et al., 2005). Enard et al. (2010) find essentially no overlap with the candidate gene sets generated using these other methodologies. Enard et al. (2010) perform forward-time population simulations to validate their method using fregene. However, they only consider two models – an equilibrium model and a single bottleneck model. Here we further evaluate their method and assertions using simulations under several models of positive and negative selection, variable recombination and mutation, and a variety of non-equilibrium models.

A possible complication with the K-statistic is its ability to differentiate between patterns produced by a selective sweep, relative to other neutral processes. Given the essence of the statistical design, significant regions will be those that have reduced heterozygosity relative to flanking genomic sequence. Thus, we investigate whether this approach is robust to: (1) non-equilibrium demographic models, which may greatly increase the variance in heterozygosity across the genome, (2) varying levels of selective constraint across the genome, and (3) simple variation in recombination rate.

MATERIALS AND METHODS

SIMULATIONS

All simulations were performed using the program SFS_CODE (Hernandez, 2008). This is a generalized Wright–Fisher forward population genetic simulation for finite-site mutation models with selection, recombination, and demography. The program and documentation are available for download at: http://sfscode.sourceforge.net/SFS_CODE/SFS_CODE_home/SFS_CODE_home.html

The parameters used are human specific and rescaled for computational efficiency. The mutation rate (μ) is $2.35*10^{-8}$ per site per generation, considering human–chimp divergence to be ~1.13%, divergence time is ~6 mya, and a generation time of 25 years (Gutenkunst et al., 2009). For expediency, the effective population size, N_e, is 500. To account for the estimated $N_e = 10{,}000$ for humans, a rescaling factor of 20 was used. Thus, $\theta = 0.00094$ for all simulations. Similarly, the scaled recombination rate $4N_r = \rho = 0.00074$ (Nielsen et al., 2005). The selection coefficient, s, is evaluated at 0.1, 0.01, and 0.001. All simulations were conducted both in the presence and absence of recombination. Additionally, models of recurrent positive and negative selection were modeled with a fraction 0.01, 0.001, and 0.0001 of sites under selection.

Population bottlenecks are modeled in the following way: a population of constant size N is reduced to size N_b at time t_b (in units of $4N$ generations) in the past and then exponentially increases back to size N. Population bottlenecks are simulated for various times since the reduction ($t_b = 0.1$, 0.54, and 1, in $4N$ generations), and severities (0.02, 0.1, and 0.722).

Enard et al. (2010) chose the size of test region (L) depending on the level of heterozygosity across the genome, and thus it varies by species. We simulate a test region of 100 kb with adjacent 2000 kb genomic flanking regions. For this test region, a ratio r_l is calculated, which is the ratio of polymorphism to divergence for the L region. A similar ratio r_g is calculated for the adjacent genomic region (G) and then a final ratio R_{obs} of $r_{l/rg}$. If R_{obs} is less than 1, then there is said to be a local reduction in heterozygosity. Similarly, a ratio R is computed for 5,000 additional windows of size q, that are randomly sampled within G, but at a distance at least five times q from L. The R_{obs} for test region is ranked among the R values for the adjacent regions. K is the proportion of random windows with R lower than R_{obs}.

K values < 0.05 are statistically significant, and thus reject the model (i.e., are consistent with positive selection). Simulations under models of positive selection were performed in order to characterize the true positive rate; while a variety of simulations under alternative models characterize the false positive rate. These two measures thus describe the performance of the K-statistic.

RESULTS

THE STANDARD NEUTRAL MODEL, WITH VARIABLE RATES OF RECOMBINATION

To account for the effect of recombination rate on patterns of selective sweeps, we simulated with and without recombination. Under the equilibrium neutral model (for $\theta = 0.00094$), the false positive rate is 0.167 ($\rho = 0.00074$) and 0.313 ($\rho = 0$), with and without recombination respectively (**Table 1**). The performance under this Standard Neutral model (where the standard false positive rate may be expected to be 0.05), should be taken as the baseline performance of this statistic. Considering the known genomic variation in recombination rate (Broman et al., 1998; Lenzi et al., 2005), this result suggests a strong bias toward false positive signatures being detected in low recombination rate regions (**Figure 1**).

NON-EQUILIBRIUM NEUTRAL MODELS, THE EFFECT OF POPULATION SIZE CHANGE

A tremendous amount of literature has focused upon the ability of test statistics to distinguish between patterns produced under selective vs. demographic models – and distinguishing these models has often proven difficult (see review of Kelley et al., 2006; Thornton et al., 2007). Here, we examine the performance of the K-statistic under a variety of bottleneck models of varying timing and severity – models of the sort that are commonly estimated for non-African human populations (e.g., Gutenkunst et al., 2009).

For population bottlenecks the false positive rate ranges from 0.12 to 0.153 in presence of recombination and 0.311–0.344 in

Table 1 | *K* value for region containing varying recombination rates ($\theta = 0.00094$).

Recombination rate	%$K < 0.05$
0	0.313
0.000074	0.235
0.00074	0.167
0.0074	0.061

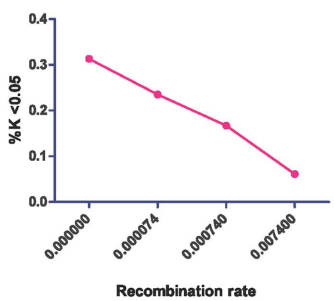

FIGURE 1 | *K* value for region containing varying recombination rates ($\theta = 0.00094$).

absence of recombination (**Table 2**). While there are some common trends – such as the duration of the bottleneck (i.e., the extent of diversity reduction) correlating with an increased false positive rate – it is significant to note that performance is not much worse (relatively) than under the Standard Neutral model itself. Thus, the inherent low power and high false positive rate under simple models of variable recombination rates appears to remain the primary factor.

MODELS OF POSITIVE AND NEGATIVE SELECTION

When the target region is evolving under purifying selection, the false positive rate ranges from 0.117 to 0.143 in presence of recombination, and 0.313–0.344 in the absence of recombination (**Table 3**). As with the demographic models – the primary factor again appears to be the inherent poor performance in the presence of variable recombination rates. However, because this is also a diversity-reducing model, the addition of constrained sites results in an elevated false positive rate – with the effect similarly being more pronounced in regions of low recombination (Begun and Aquadro, 1992).

Finally, under the true model (i.e., the target region is under positive selection), the true positive rate ranges from 0.115 to 0.142 in presence of recombination, and 0.296–0.332 in absence of recombination (**Table 4**). Thus, while there is some power to identify positively selected genomic regions, power does not exceed even the baseline false positive rate under the Standard Neutral model – making all models of true and false positives indistinguishable.

DISCUSSION

From our simulated tests of the *K*-statistic of Enard et al. (2010), a variation of the HKA test designed for detecting positive selection using a single genome, we find that the method is not able to distinguish between models of positive selection, negative selection, or population size change, nor is it robust to local variation in rates of recombination. Indeed, the true and false positive rates under the examined models are roughly similar.

Indeed, the largest determinant in dictating the fraction of rejecting loci is the rate of recombination. While a decrease in the rate of recombination increased the true positive rate, it equally corresponded to an increase in the false positive rate under most models examined. This is owing to the diversity-reducing potential of all of the models examined, with the size of the reduction being dictated by local recombination rates. While the false positive rate was not well explored by Enard et al. (2010) the low power of the statistic is consistent with their results. These results suggest a strong bias toward preferentially identifying low recombination rate regions of the genome – under any true underlying model. And while there is indeed evidence of differences in fine-scale recombination rates between humans and chimps (Ptak et al., 2005; Winckler et al., 2005), results suggest that any overlap between species in selective constraint or rate of crossover will result in false positives in even inter-species comparisons.

The poor performance of this statistic is consistent with the poor overlap with other human genomic scans (Carlson et al., 2005; Voight et al., 2006; Tang et al., 2007; Williamson et al., 2007; Pickrell et al., 2009), with the fraction of overlapping genes varying between 6 and 15%. Thus, despite the novelty of the initial claim, these results strongly suggest that the single genome approach to sweep detection is not robust to any local diversity-reducing model. Thus, the variety of polymorphism based test statistics (see review of Nielsen, 2005; Thornton et al., 2007) for distinguishing the above models using patterns in linkage disequilibrium and the site frequency spectrum appear to be the most promising avenue forward – continuing to necessitate the sampling and sequencing of consistently sampled population data.

Table 2 | *K* value for population bottleneck duration ($\theta = 0.00094$).

Bottleneck duration (4*N* gen)	Bottleneck severity	%*K* < 0.05 ($\rho = 0.00074$)	%*K* < 0.05 ($\rho = 0$)
0.1	0.02	0.119	0.319
	0.1	0.133	0.33
	0.722	0.146	0.344
0.48	0.02	0.144	0.34
	0.1	0.135	0.311
	0.722	0.136	0.339
1	0.02	0.153	0.336
	0.1	0.132	0.313
	0.722	0.12	0.328

Table 3 | *K* values for region under negative selection ($\theta = 0.00094$).

Selection coefficient	Proportion of sites under negative selection	%*K* < 0.05 ($\rho = 0.00074$)	%*K* < 0.05 ($\rho = 0$)
0.1	0.0001	0.129	0.322
	0.001	0.117	0.324
	0.01	0.136	0.334
0.01	0.0001	0.126	0.313
	0.001	0.119	0.313
	0.01	0.143	0.344
0.001	0.0001	0.127	0.334
	0.001	0.143	0.313
	0.01	0.128	0.323

Table 4 | *K* values for region under positive selection ($\theta = 0.00094$).

Selection coefficient	Proportion of sites under positive selection	%*K* < 0.05 ($\rho = 0.00074$)	%*K* < 0.05 ($\rho = 0$)
0.1	0.0001	0.129	0.329
	0.001	0.138	0.313
	0.01	0.131	0.332
0.01	0.0001	0.142	0.33
	0.001	0.132	0.323
	0.01	0.115	0.33
0.001	0.0001	0.127	0.322
	0.001	0.123	0.296
	0.01	0.127	0.33

REFERENCES

Begun, D. J., and Aquadro, C. F. (1992). Levels of naturally occurring DNA polymorphism correlate with recombination rates in *D. melanogaster*. *Nature* 356, 519–520.

Broman, K. W., Murray, J. C., Sheffield, V. C., White, R. L., and Weber, J. L. (1998). Comprehensive human genetic maps: individual and sex-specific variation in recombination. *Am. J. Hum. Genet.* 63, 861–869.

Carlson, C. S., Thomas, D. J., Eberle, M. A., Swanson, J. E., Livingston, R. J., Rieder, M. J., and Nickerson, D. A. (2005). Genomic regions exhibiting positive selection identified from dense genotype data. *Genome Res.* 15, 1553–1565.

Enard, D., Depaulis, F., and Roest Crollius, H. (2010). Humans and non-human primate genomes share hotspots of positive selection. *PLoS Genet.* 6, e1000840. doi:10.1371/journal.pgen.1000840

Frazer, K. A., Ballinger, D. G., Cox, D. R., Hinds, D. A., Stuve, L. L., Gibbs, R. A., Belmont, J. W., Boudreau, A., Hardenbol, P., Leal, S. M., Pasternak, S., Wheeler, D. A., Willis, T. D., Yu, F., Yang, H., Zeng, C., Gao, Y., Hu, H., Hu, W., Li, C., Lin, W., Liu, S., Pan, H., Tang, X., Wang, J., Wang, W., Yu, J., Zhang, B., Zhang, Q., Zhao, H., Zhao, H., Zhou, J., Gabriel, S. B., Barry, R., Blumenstiel, B., Camargo, A., Defelice, M., Faggart, M., Goyette, M., Gupta, S., Moore, J., Nguyen, H., Onofrio, R. C., Parkin, M., Roy, J., Stahl, E., Winchester, E., Ziaugra, L., Altshuler, D., Shen, Y., Yao, Z., Huang, W., Chu, X., He, Y., Jin, L., Liu, Y., Shen, Y., Sun, W., Wang, H., Wang, Y., Wang, Y., Xiong, X., Xu, L., Waye, M. M., Tsui, S. K., Xue, H., Wong, J. T., Galver, L. M., Fan, J. B., Gunderson, K., Murray, S. S., Oliphant, A. R., Chee, M. S., Montpetit, A., Chagnon, F., Ferretti, V., Leboeuf, M., Olivier, J. F., Phillips, M. S., Roumy, S., Sallée, C., Verner, A., Hudson, T. J., Kwok, P. Y., Cai, D., Koboldt, D. C., Miller, R. D., Pawlikowska, L., Taillon-Miller, P., Xiao, M., Tsui, L. C., Mak, W., Song, Y. Q., Tam, P. K., Nakamura, Y., Kawaguchi, T., Kitamoto, T., Morizono, T., Nagashima, A., Ohnishi, Y., Sekine, A., Tanaka, T., Tsunoda, T., Deloukas, P., Bird, C. P., Delgado, M., Dermitzakis, E. T., Gwilliam, R., Hunt, S., Morrison, J., Powell, D., Stranger, B. E., Whittaker, P., Bentley, D. R., Daly, M. J., de Bakker, P. I., Barrett, J., Chretien, Y. R., Maller, J., McCarroll, S., Patterson, N., Pe'er, I., Price, A., Purcell, S., Richter, D. J., Sabeti, P., Saxena, R., Schaffner, S. F., Sham, P. C., Varilly, P., Altshuler, D., Stein, L. D., Krishnan, L., Smith, A. V., Tello-Ruiz, M. K., Thorisson, G. A., Chakravarti, A., Chen, P. E., Cutler, D. J., Kashuk, C. S., Lin, S., Abecasis, G. R., Guan, W., Li, Y., Munro, H. M., Qin, Z. S., Thomas, D. J., McVean, G., Auton, A., Bottolo, L., Cardin, N., Eyheramendy, S., Freeman, C., Marchini, J., Myers, S., Spencer, C., Stephens, M., Donnelly, P., Cardon, L. R., Clarke, G., Evans, D. M., Morris, A. P., Weir, B. S., Tsunoda, T., Mullikin, J. C., Sherry, S. T., Feolo, M., Skol, A., Zhang, H., Zeng, C., Zhao, H., Matsuda, I., Fukushima, Y., Macer, D. R., Suda, E., Rotimi, C. N., Adebamowo, C. A., Ajayi, I., Aniagwu, T., Marshall, P. A., Nkwodimmah, C., Royal, C. D., Leppert, M. F., Dixon, M., Peiffer, A., Qiu, R., Kent, A., Kato, K., Niikawa, N., Adewole, I. F., Knoppers, B. M., Foster, M. W., Clayton, E. W., Watkin, J., Gibbs, R. A., Belmont, J. W., Muzny, D., Nazareth, L., Sodergren, E., Weinstock, G. M., Wheeler, D. A., Yakub, I., Gabriel, S. B., Onofrio, R. C., Richter, D. J., Ziaugra, L., Birren, B. W., Daly, M. J., Altshuler, D., Wilson, R. K., Fulton, L. L., Rogers, J., Burton, J., Carter, N. P., Clee, C. M., Griffiths, M., Jones, M. C., McLay, K., Plumb, R. W., Ross, M. T., Sims, S. K., Willey, D. L., Chen, Z., Han, H., Kang, L., Godbout, M., Wallenburg, J. C., L'Archevêque, P., Bellemare, G., Saeki, K., Wang, H., An, D., Fu, H., Li, Q., Wang, Z., Wang, R., Holden, A. L., Brooks, L. D., McEwen, J. E., Guyer, M. S., Wang, V. O., Peterson, J. L., Shi, M., Spiegel, J., Sung, L. M., Zacharia, L. F., Collins, F. S., Kennedy, K., Jamieson, R., and Stewart, J. (2007). A second generation human haplotype map of over 3.1 million SNPs. *Nature* 449, 851–861.

Gutenkunst, R. N., Hernandez, R. D., Williamson, S. H., and Bustamante, C. D. (2009). Inferring the joint demographic history of multiple populations from multidimensional SNP frequency data. *PLoS Genet.* 5, e1000695. doi:10.1371/journal.pgen.1000695

Hernandez, R. D. (2008). A flexible forward simulator for populations subject to selection and demography. *Bioinformatics* 24, 2786–2787.

Hinds, D. A., Stuve, L. L., Nilsen, G. B., Halperin, E., Eskin, E., Ballinger, D. G., Frazer, K. A., and Cox, D. R. (2005). Whole genome patterns of common DNA variation in three human populations. *Science* 307, 1072–1079.

Hudson, R. R., Kreitman, M., and Aguadé, M. (1987). A test of neutral molecular evolution based on nucleotide data. *Genetics* 116, 153–159.

Kelley, J. L., Madeoy, J., Calhoun, J., Swanson, W., and Akey, J. (2006). Genomic signatures of positive selection in humans and the limits of outlier approaches. *Genome Res.* 16, 980–989.

Lenzi, M. L., Smith, J., Snowden, T., Kim, M., Fishel, R., Poulos, B. K., and Cohen, P. E. (2005). Extreme heterogeneity in the molecular events leading to the establishment of chiasmata during meiosis in human oocytes. *Am. J. Hum. Genet.* 76, 112–127.

Nielsen, R. (2005). Molecular signatures of natural selection. *Annu. Rev. Genet.* 39, 197–218.

Nielsen, R., Williamson, S., Kim, Y., Hubisz, M. J., Clark, A. G., and Bustamante, C. D. (2005). Genomic scans for selective sweeps using SNP data. *Genome Res.* 15, 1566–1575.

Pickrell, J. K., Coop, G., Novembre, J., Kudaravalli, S., Li, J. Z., Absher, D., Srinivasan, B. S., Barsh, G. S., Myers, R. M., Feldman, M. W., and Pritchard, J. K. (2009). Signals of recent positive selection in a worldwide sample of human populations. *Genome Res.* 19, 826–837.

Ptak, S. E., Hinds, D. A., Koehler, K., Nickel, B., Patil, N., Ballinger, D. G., Przeworski, M., Frazer, K. A., and Pääbo, S. (2005). Fine-scale recombination patterns differ between chimpanzee and humans. *Nat. Genet.* 37, 429–434.

Sabeti, P. C., Schaffner, S. F., Fry, B., Lohmueller, J., Varilly, P., Shamovsky, O., Palma, A., Mikkelsen, T. S., Altshuler, D., and Lander, E. S. (2006). Positive natural selection in the human lineage. *Science* 312, 1614–1620.

Sabeti, P. C., Varilly, P., Fry, B., Lohmueller, J., Hostetter, E., Cotsapas, C., Xie, X., Byrne, E. H., McCarroll, S. A., Gaudet, R., Schaffner, S. F., Lander, E. S., International HapMap Consortium, Frazer, K. A., Ballinger, D. G., Cox, D. R., Hinds, D. A., Stuve, L. L., Gibbs, R. A., Belmont, J. W., Boudreau, A., Hardenbol, P., Leal, S. M., Pasternak, S., Wheeler, D. A., Willis, T. D., Yu, F., Yang, H., Zeng, C., Gao, Y., Hu, H., Hu, W., Li, C., Lin, W., Liu, S., Pan, H., Tang, X., Wang, J., Wang, W., Yu, J., Zhang, B., Zhang, Q., Zhao, H., Zhao, H., Zhou, J., Gabriel, S. B., Barry, R., Blumenstiel, B., Camargo, A., Defelice, M., Faggart, M., Goyette, M., Gupta, S., Moore, J., Nguyen, H., Onofrio, R. C., Parkin, M., Roy, J., Stahl, E., Winchester, E., Ziaugra, L., Altshuler, D., Shen, Y., Yao, Z., Huang, W., Chu, X., He, Y., Jin, L., Liu, Y., Shen, Y., Sun, W., Wang, H., Wang, Y., Wang, Y., Xiong, X., Xu, L., Waye, M. M., Tsui, S. K., Xue, H., Wong, J. T., Galver, L. M., Fan, J. B., Gunderson, K., Murray, S. S., Oliphant, A. R., Chee, M. S., Montpetit, A., Chagnon, F., Ferretti, V., Leboeuf, M., Olivier, J. F., Phillips, M. S., Roumy, S., Sallée, C., Verner, A., Hudson, T. J., Kwok, P. Y., Cai, D., Koboldt, D. C., Miller, R. D., Pawlikowska, L., Taillon-Miller, P., Xiao, M., Tsui, L. C., Mak, W., Song, Y. Q., Tam, P. K., Nakamura, Y., Kawaguchi, T., Kitamoto, T., Morizono, T., Nagashima, A., Ohnishi, Y., Sekine, A., Tanaka, T., Tsunoda, T., Deloukas, P., Bird, C. P., Delgado, M., Dermitzakis, E. T., Gwilliam, R., Hunt, S., Morrison, J., Powell, D., Stranger, B. E., Whittaker, P., Bentley, D. R., Daly, M. J., de Bakker, P. I., Barrett, J., Chretien, Y. R., Maller, J., McCarroll, S., Patterson, N., Pe'er, I., Price, A., Purcell, S., Richter, D. J., Sabeti, P., Saxena, R., Schaffner, S. F., Sham, P. C., Varilly, P., Altshuler, D., Stein, L. D., Krishnan, L., Smith, A. V., Tello-Ruiz, M. K., Thorisson, G. A., Chakravarti, A., Chen, P. E., Cutler, D. J., Kashuk, C. S., Lin, S., Abecasis, G. R., Guan, W., Li, Y., Munro, H. M., Qin, Z. S., Thomas, D. J., McVean, G., Auton, A., Bottolo, L., Cardin, N., Eyheramendy, S., Freeman, C., Marchini, J., Myers, S., Spencer, C., Stephens, M., Donnelly, P., Cardon, L. R., Clarke, G., Evans, D. M., Morris, A. P., Weir, B. S., Tsunoda, T., Johnson, T. A., Mullikin, J. C., Sherry, S. T., Feolo, M., Skol, A., Zhang, H., Zeng, C., Zhao, H., Matsuda, I., Fukushima, Y., Macer, D. R., Suda, E., Rotimi, C. N., Adebamowo, C. A., Ajayi, I., Aniagwu, T., Marshall, P. A., Nkwodimmah, C., Royal, C. D., Leppert, M. F., Dixon, M., Peiffer, A., Qiu, R., Kent, A., Kato, K., Niikawa, N., Adewole, I. F., Knoppers, B. M., Foster, M. W., Clayton, E. W., Watkin, J., Gibbs, R. A., Belmont, J. W., Muzny, D., Nazareth, L., Sodergren, E., Weinstock, G. M., Wheeler, D. A., Yakub, I., Gabriel, S. B., Onofrio, R. C., Richter, D. J., Ziaugra, L., Birren, B. W., Daly, M. J., Altshuler, D., Wilson, R. K., Fulton, L. L., Rogers, J., Burton, J., Carter, N. P., Clee, C. M., Griffiths, M., Jones, M. C., McLay, K., Plumb, R. W., Ross, M. T., Sims, S. K., Willey, D. L., Chen, Z., Han, H., Kang, L., Godbout, M., Wallenburg, J. C., L'Archevêque, P.,

Bellemare, G., Saeki, K., Wang, H., An, D., Fu, H., Li, Q., Wang, Z., Wang, R., Holden, A. L., Brooks, L. D., McEwen, J. E., Guyer, M. S., Wang, V. O., Peterson, J. L., Shi, M., Spiegel, J., Sung, L. M., Zacharia, L. F., Collins, F. S., Kennedy, K., Jamieson, R., and Stewart, J. (2007). Genome wide detection and characterization of positive selection in human populations. *Nature* 449, 913–918.

Smith, J. M., and Haigh, J. (1974). The hitch-hiking effect of a favourable gene. *Genet. Res.* 23, 23–35.

Tang, K., Thornton, K. R., and Stoneking, M. (2007). A new approach for using genome scans to detect recent positive selection in the human genome. *PLoS Biol.* 5, e171. doi:10.1371/journal.pbio.0050171

Thornton, K. R., Jensen, J. D., Becquet, C., and Andolfatto, P. (2007). Progress and prospects in mapping recent selection in the genome. *Heredity* 98, 340–348.

Voight, B. F., Kudaravalli, S., Wen, X., and Pritchard, J. K. (2006). A map of recent positive selection in the human genome. *PLoS Biol.* 4, e72. doi:10.1371/journal. pbio.0040072.

Williamson, S. H., Hubisz, M. J., Clark, A. G., Payseur, B. A., Bustamante, C. D., and Nielsen, R. (2007). Localizing recent adaptive evolution in the human genome. *PLoS Genet.* 3, e90. doi:10.1371/journal.pgen.0030090

Winckler, W., Myers, S. R., Richter, D. J., Onofrio, R. C., McDonald, G. J., Bontrop, R. E., McVean, G. A., Gabriel, S. B., Reich, D., Donnelly, P., and Altshuler, D. (2005). Comparison of fine-scale recombination rates in humans and chimpanzees. *Science* 308, 107–111.

Conflict of Interest Statement: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Received: 20 July 2011; paper pending published: 06 September 2011; accepted: 13 November 2011; published online: 01 December 2011.

Citation: Sinha P, Dincer A, Virgil D, Xu G, Poh Y-P and Jensen JD (2011) On detecting selective sweeps using single genomes. Front. Gene. ***2****:85. doi: 10.3389/fgene.2011.00085*

This article was submitted to Frontiers in Evolutionary and Population Genetics, a specialty of Frontiers in Genetics.

Copyright © 2011 Sinha, Dincer, Virgil, Xu, Poh and Jensen. This is an open-access article distributed under the terms of the Creative Commons Attribution Non Commercial License, which permits non-commercial use, distribution, and reproduction in other forums, provided the original authors and source are credited.