Genome Biology

OPINION

The promise and limitations of population exomics for human evolution studies

Jacob A Tennessen[1], Timothy D O'Connor[1], Michael J Bamshad[1,2] and Joshua M Akey[1]*

Abstract

Exome sequencing is poised to yield substantial insights into human genetic variation and evolutionary history, but there are significant challenges to overcome before this becomes a reality.

For the past few decades, advances in molecular biology have continuously refined our understanding of human evolutionary history. A simple model of expansion and global migrations from a single ancestral human population with adaptation at a few protein polymorphisms has transformed into a complex scenario involving introgression among numerous divergent groups, multiple population-specific bottlenecks, and thousands of candidate genomic sites of possible evolutionary importance [1-6]. Although the broad patterns of demographic trends, geographic population structure, and adaptation have now been well established [1-4], emerging genome-scale datasets will enable detailed inferences about particular populations and genes. Major ongoing goals include inferring intracontinental patterns of migration and admixture, reconstructing the history of human population growth and bottlenecks, and categorizing whether polymorphisms are selectively neutral, deleterious, or adaptive (Box 1). Until recently, such questions could be addressed only with the limited statistical power and precision afforded by single nucleotide polymorphism (SNP) arrays or small sets of sequence data. However, exome sequencing has the potential to address many of these questions.

Exome sequencing is a new and powerful technique in which genomic DNA that binds to a predefined target of known exons is sequenced using next-generation technology, in order to capture the protein-coding

*Correspondence: akeyj@uw.edu
[1]Department of Genome Sciences, University of Washington, 3720 15th Ave NE, Box 355065, Seattle, WA 98195-5065, USA
Full list of author information is available at the end of the article

Box 1. Goals and methods of population genetics

Extant patterns of human genetic variation provide information about our demographic and evolutionary history. The goals of population genetics are to infer past events from DNA sequence variation and identify and quantify how evolutionary processes, such as natural selection, population structure, migration, genetic drift, and changes in population size, have shaped human genomic diversity. To this end, numerous population genetics statistics have been developed for analyzing genetic variation. A brief synopsis of population genetic statistics well suited to exome data is as follows.

π: The expected number of differences between two sequences randomly selected from the same locus in a population is represented as π. If π is calculated per base pair, data on both variable and invariant sites, and therefore sequence data rather than SNP array data, are required. Numerous evolutionary inferences rely on π. Its overall magnitude reflects the mutation rate and effective size of a population. Unusually high or low π at a locus can be a signature of natural selection. Most genes in most human populations have per base π values between 10^{-4} and 10^{-3} [13].

Site frequency spectra: A site frequency spectrum represents the relative numbers of variants occurring at all frequencies in a population. The proportion of rare variants as compared with common variants can be used to infer the rate and timing of population growth. Unique spectra for certain genes or certain site classes are thought to reflect variation in the strength and form of natural selection. For example, a selective sweep may eliminate all variation, and all new variants arising after the sweep will be rare initially, resulting in a skewed spectrum with a relative dearth of common variants. Tajima's *D* is a summary statistic of the site frequency spectrum, with negative values indicating a relative excess of rare variants, positive values indicating a relative excess of common variants, and values near zero indicating mutation-drift equilibrium. Site frequency spectra are most accurately inferred with large amounts of unbiased sequence from numerous individuals, as provided by exomics.

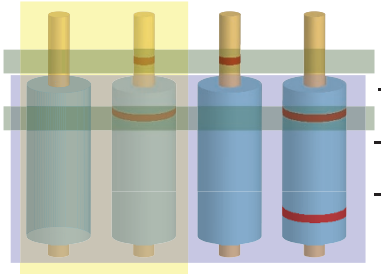
Nonsynonymous/synonymous neutrality tests: Natural selection is expected to act more strongly on nonsynonymous sites than synonymous sites, and there are numerous statistical tests that compare these site classes in order to study selection. Exomes represent the exact portion of the genome where such tests are applicable. For example, the McDonald-Kreitman test [34] compares the ratio of polymorphism at these two site classes with the ratio of interspecies divergence at these two site classes. Under constant purifying selection these two ratios should be similar, so a discrepancy is evidence for adaptive evolution.

portion of the genome [7]. The magnitude and cost-effectiveness of exome datasets vastly overshadow many other methods for studying polymorphism that have recently been popular, such as SNP arrays or single locus resequencing studies**.** Here, we discuss the application of exome data to human population genetics. We argue that exomes will allow many important and detailed analyses that are not possible with SNP arrays because of ascertainment biases. Moreover, although whole-genome sequencing in large population samples is clearly on the horizon, exomes are the most cost effective and practical way of obtaining sufficiently high coverage to rigorously characterize the spectrum of rare variation. However, the absence of noncoding data does limit the application of exomes in nontrivial ways and can lead to misleading inferences if research is not carefully conducted. Thus, we are cautiously optimistic that exomes will address many remaining questions about human evolution, if incompletely.

Exome sequencing – an unbiased measure of polymorphism

Exome sequencing provides an unbiased and complete perspective of coding genetic variation to a degree that has never before been possible. In many respects, 'exomics' combines the most favorable aspects of the other available molecular methods (Figure 1). For example, SNP arrays provide a picture of genome-wide polymorphism in many individuals [3,8,9], but they necessarily suffer from ascertainment biases favoring previously identified SNPs that are common in the populations (primarily of European ancestry) used for variant discovery [10,11]. Gene sequencing eliminates this bias and also provides data on invariant sites for measuring overall polymorphism, π, which underlies numerous population genetic statistics (Box 1). However, studies that examine individual genes illustrate only a small portion of the functional genome. Genome resequencing therefore provides the most complete view of polymorphism [12], but cost, computational complexity, and data storage issues make it difficult at present to sequence thousands of individuals or more at high coverage, which is required for studies of rare variants. Thus, most large-scale genome sequencing so far has used relatively low coverage, biasing the dataset in favor of common variants and confounding demographic and other evolutionary inferences that require an unbiased sample. By contrast, exomics provides a practical way to generate an unbiased picture of variation within a large number of samples at functionally important regions of the genome. By assessing all of the variants within the targeted region, including rare and novel polymorphisms, exome sequencing enables accurate inferences of the site frequency spectrum (Box 1, Figure 2). Similarly, because all protein-coding genes are examined, the relative numbers of very specific types of polymorphism (for example, mutations to each amino acid residue), within narrowly defined site-frequency classes (for example, singletons versus doubletons), can also be estimated accurately.

It is important to note that the ideal filtering strategy used to generate an exome dataset differs slightly between population genetics and phenotype association studies. In association analyses, the goal is usually to maximize the number of putatively real variants, any of which could be causal for the trait in question, and to



Site	Demographic inference	Natural selection inference
Noncoding	True neutral sites	True neutral controls, some causal sites
Common	Ancient events, major trends	Balancing selection, geographic divergence
Invariant	Effective size, mutation rate	Selective sweeps, very conserved regions
Rare	Recent events, subtle trends	Recovery from sweep, deleterious variants

Figure 1. The contributions of different data types to population genetic inferences. Exomes, SNP arrays, and genomes are likely to capture different combinations of these four basic site types: common variants, rare variants, invariant sites, and noncoding sites. Each site type offers unique information relevant to analyses of demography and natural selection. Rare variants and invariant sites captured in exomes are important for numerous evolutionary questions, from estimating the effective population size to detecting positive selective sweeps, and they may be missed by other methods. However, phenotypically causal noncoding variants and truly neutral regions far from genes and free from the effects of selection are absent from exomes.

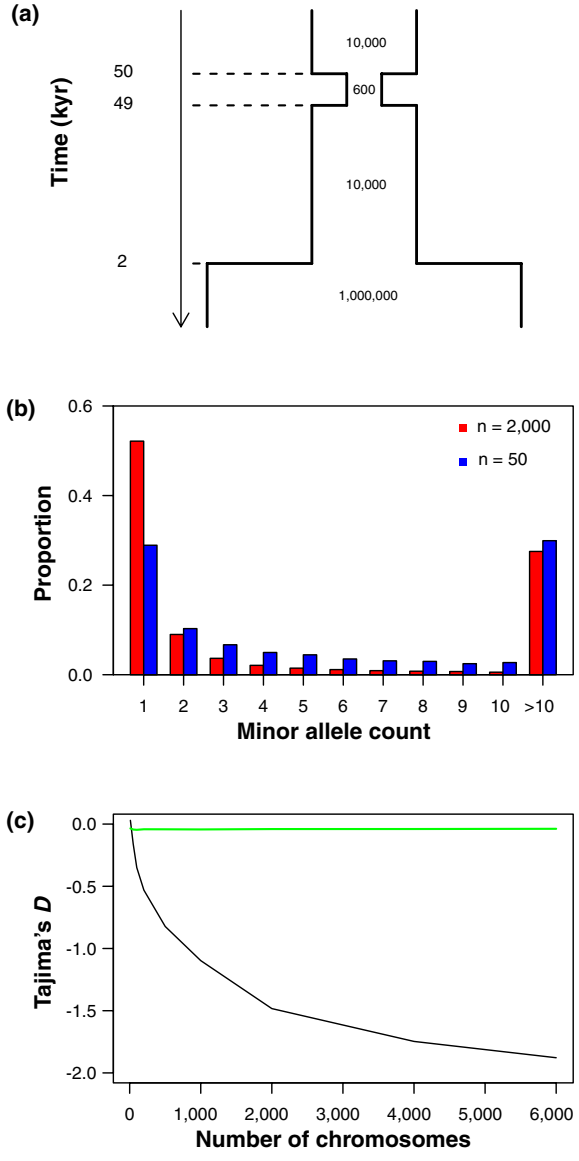


Figure 2. The influence of sample size in detecting recent population expansions. (a) Demographic model used in the simulations. Times of demographic events (in units of thousands of years, kyr) and population sizes are indicated. **(b)** The site frequency spectrum for sample sizes of 50 and 2,000 chromosomes. **(c)** Average value of Tajima's *D* based on 10^4 simulation replicates as a function of sample size (number of chromosomes). The lines denote values of Tajima's *D* from the demographic model in (a) (black line) and a constant sized population (green line). This simulation illustrates the power of large sample sizes for inferring recent demographic events. All information in (b,c) is based on simulating 10 kb of sequence.

ignore invariant sites. However, for endeavors such as resolution of population structure, it is preferable to discard sites with missing data in a substantial proportion of samples in order to minimize clustering of individuals based on 'missingness', defined here as the proportion and identities of genotypes with missing data, and knowledge of invariant sites is essential. As exome sequencing becomes routine and optimized, it will be important to maintain some flexibility in filtering options based on particular research goals.

Detecting natural selection from exome data

One of the most promising applications for exome data is the study of natural selection in humans [13]. Inferring patterns of natural selection on genes is a powerful approach for gauging the functional impact of polymorphisms. Although a nontrivial amount of noncoding DNA is functional, it is clear that exons contain a substantial proportion of the genome's phenotypically relevant sites, subject to strong selective pressures [14]. Natural selection is also easier to study using exons, as many existing statistical tests for estimating selection, such as those based on the ratios of nonsynonymous to synonymous sites, are appropriate only for coding sequence (Box 1). However, most of the signature of a selective event can lie in noncoding regions, even if the target of selection is in an exon. Exome data provide substantial power to detect regions of low polymorphism or high linkage disequilibrium only if exon density in the region is sufficiently high. Even then, estimating the precise length of the region affected by selection is not possible without full sequence data, although sequencing the flanking noncoding areas after identifying an interesting region is always an option.

Positive selection, the fixation of new favorable alleles, is an important evolutionary phenomenon that has proven difficult to thoroughly characterize. Numerous studies have identified genomic regions displaying extreme values in statistical tests of selective neutrality, but the overlap among these lists of candidate regions is often poor, suggesting a high proportion of false positives [4]. In addition, it is often unclear whether outlier SNPs are themselves the targets of selection or merely linked to the true targets [4,15]. Analysis of exome sequences promises to enhance power for resolving these issues. A typical signature of a positive selective sweep includes low π and an excess of rare variants, which can most directly be identified with sequence data. Assuming that the real causal variant is in an exon, it can be pinpointed with high precision. Owing to their rich information content, even a small sample of exomes can show differences between selected and neutral regions and allow adaptive substitutions to be identified [13]. For example, the causal nonsynonymous polymorphism in *SLC24A5*, a gene that influences skin pigmentation, is a clear outlier with respect to both interpopulation divergence and patterns of polymorphism in flanking exons, such that its adaptive significance is apparent in a sample of as few as ten exomes [13,16].

Whether human populations actually harbor a large proportion of adaptive coding variants flanked by regions of low π or skewed site frequency spectra depends on where and how selection usually acts in humans, which is still unresolved. If selection acts primarily on non-coding regions [17] or on standing genetic variation, such that dramatic polymorphism-reducing selective sweeps do not occur [18-20], exomes will have less of an advantage over other methods such as SNP arrays or full genomes for studying positive selection. So far, the clearest example of positive selection inferred from exome data is the hypoxia response gene EPAS1, which has evolved rapidly in high-altitude Tibetan populations [21]. The strongest candidate SNP at EPAS1 is in an intron that happened to be included, and the primary evidence for positive selection is high divergence between populations rather than low polymorphism. The fact that the gene was still identified highlights the versatility of exomes, but SNP-based or noncoding-inclusive approaches might have had similar, if not greater, power to detect selection in this case.

Balancing selection, the maintenance of multiple favorable variants, can also be studied with exome data. Under the classic model of balancing selection, two or more alleles are maintained at intermediate frequency in a population. Most of these cases in humans have probably already been identified because the variants in question would be common, although flanking sequence data can help strengthen or refute the case for balancing selection on a particular SNP, as in the case of the prion protein gene *PRNP*, in which a widely publicized claim of cannibalism-associated balancing selection [22] was shown to be an artifact of ascertainment bias [23]. Under other forms of balancing selection, one allele might be very rare and therefore as yet undiscovered. For example, under fluctuating selection [24], a currently deleterious, and therefore rare, allele may have been advantageous in the past and could be again in the future. Similarly, the equilibrium allele frequencies in the case of over-dominance, or heterozygote advantage, are proportional to the relative selective disadvantages of each homozygote genotype [25]; thus, if one homozygote is quite deleterious (for example, lethal), whereas the other is only slightly less deleterious than the heterozygote, a highly skewed allele frequency will be maintained by balancing selection. It is unknown whether these more complex forms of balancing selection have an important role in the patterns of human genetic diversity, and exomes are ideal for this line of inquiry because their cost-effectiveness allows even rare alleles to be observed.

Purifying selection, the elimination of deleterious mutations, is by far the most common type of selection. Therefore, it is the most relevant to human health because, for the vast majority of functionally relevant polymorphisms in a genome, the derived variant will be deleterious. Distinguishing harmful variants from benign variants is a central goal of disease genetics, and population genetic studies to identify purifying selection are directly relevant to this goal. With a large sample of exomes, it is possible to estimate the probability of deleteriousness for a nonsynonymous variant given its frequency. Assuming that only benign variants ever reach high frequency, the ratio of nonsynonymous to synonymous sites at high frequency can be used to calculate the relative excess of nonsynonymous sites, which are presumed to be deleterious, at lower frequencies [26]. Given the enormous number of variants in an exome dataset, this approach can be tailored to highly specific site classes, based on biochemical properties of the encoded residue or patterns of conservation across species, rather than simply comparing all nonsynonymous and all synonymous polymorphisms. Furthermore, genes with very few nonsynonymous variants overall that do not show evidence of a selective sweep are likely to be under strong purifying selection, so there is an enhanced probability that subsequently discovered rare nonsynonymous variants are deleterious. Such highly conserved genes can be identified only with data on invariant sites from many individuals, which exomes provide.

Population structure and demography

Natural populations are not static and often have complicated demographic histories, including changes in population size and non-random mating leading to geographic structure. Rare variants and unascertained common variants identified from exome sequencing will be a powerful resource for inferences of demographic history. So far, resequencing efforts of smaller subsets of the human genome have already yielded a comprehensive portrait of historical changes in population size, and the relationship between geographically diverse populations, migrations, and admixture [2,27-29]. For example, both African and non-African populations have experienced bottlenecks followed by an exponential increase in population size, although the magnitude of these events has been greater for non-African populations [2,28,29]. Exome sequence data will facilitate more precise estimates of important parameters governing human history, such as the mode and timing of population expansions.

Of particular interest, exome data are well poised to enable new insights into recent demographic events. Because exome sequencing is currently more cost-efficient than whole-genome sequencing, it is possible to study patterns of variation in very large samples. To explore this idea in more detail, we performed a simple coalescent simulation of a population that experienced a bottleneck of moderate intensity 50,000 years ago and a

more recent population expansion 2,000 years ago (Figure 2a). The goal here is not to perfectly recapitulate human demography, but to demonstrate how exome sequence data might facilitate inferences of recent events. From this model, we explored how the site frequency spectrum varies as a function of sample size. As shown in Figure 2b, there is a dramatic shift towards rare alleles, particularly singletons (sites where the minor allele is only observed once in the sample), for larger sample sizes. To quantify this affect more rigorously, we calculated Tajima's *D* statistic (Box 1), which is expected to be negative in cases of an excess of rare variation relative to what is expected in constant sized populations. For small sample sizes (Figure 2c), the recent population expansion is 'invisible' and Tajima's *D* is close to zero, which is the expected value in populations of constant size. However, as sample size increases, Tajima's *D* becomes sharply negative, revealing the recent explosive population growth. Intuitively, these results make sense because the larger sample size provides sufficient numbers of mutations to reveal the recent underlying genealogical structure. Interestingly, in populations of constant size, Tajima's *D* is not influenced by sample size and stochastically varies close to zero (Figure 2c). Thus, because exome sequencing can be performed in large samples, these simple simulations suggest that there is considerable promise in more detailed and quantitative estimates of recent human demographic history.

Moreover, as described above, because exome data do not suffer from the same ascertainment bias inherent in SNP arrays or small-sample datasets, it will possible to explore more nuanced questions related to population structure. For example, an interesting hypothesis to test is whether rare variants have signatures of structure that are different from those derived from common variants. Intuitively, as rare variants are predominantly derived from mutations in the recent past, they may be particularly useful in assessing intracontinental, or perhaps even finer-scale, population structure, even if allele frequency differences at common variants are negligible. Similarly, exome data will also be a powerful resource for understanding how the process and dynamics of admixture manifest themselves in patterns of variation [30] across the genome. At the individual level, exome data may allow reconstruction of the mosaic structure of ancestry blocks (stretches of the genome inherited intact from a parental population [31]), which will provide mechanistic insights into the admixture process and the differences in demographic history of the parental populations [30]. As with SNP array datasets and other genomically incomplete data, haplotypes in unsequenced (noncoding) regions must be inferred from the existing data, with a precision that depends on the density of sequenced (coding) genotypes.

An important general caveat of exome data in understanding human demographic history is that purifying selection acting on deleterious variants will complicate inferences of population parameters, such as effective population sizes, and the site frequency spectrum [32]. A simple strategy to attenuate these concerns is to focus analyses on classes of sites that are expected to be less strongly influenced by purifying selection (such as synonymous sites and targeted introns). However, new methodological approaches that jointly estimate demographic parameters and selection are clearly more desirable and important to develop [33].

Challenges and caveats for population exomics

Although exome datasets remove many of the biases and limitations that have plagued previous population genetic datasets, they can still be misinterpreted if not analyzed appropriately. One potential challenge is presented by cryptic paralogs. Copy number variation is prevalent and remains poorly characterized in humans. Reads from exons that are absent from the capture target, perhaps because they only occur in some individuals, can map to paralogous exons on the capture target, falsely inflating apparent π in these exons. In many cases, these exons can be removed from analysis by filtering on violations of Hardy-Weinberg equilibrium.

Another concern is missing data. It is common to remove invariant sites from exome files in order to reduce them to a manageable size. However, estimates of π require differentiating between truly invariant sites and sites that might be variable but were not sequenced at high coverage in many individuals. For some analyses, it is sufficient to estimate 'missingness' at invariant sites rather than measure it directly, but doing so carries the important caveat that regions of low π could merely be regions of low coverage.

A third challenge is the difficulty of merging datasets. As yet, there is no one accepted definition of the exome. Rather, there are numerous capture targets with different combinations of exons. Even if two targets share the same exon, coverage may be better in one of the targets for a variety of technical reasons. Thus, when sequences from multiple targets are combined into a single dataset, missing data at some sites will be high and highly correlated with the target used. If different populations were sequenced with different targets, analyses of population structure are then confounded. The use of multiple sequencing platforms could potentially cause a similar pattern. Furthermore, multi-sample calling methods for assigning genotypes are more likely to call a variant if it is also seen in other samples, so calling genotypes in batches can cause artifacts if these batches are then merged with each other or with single-sample called genotypes. These effects can be minimized by

excluding sites with a high proportion of missingness, but the best approach is to use the same target and sequencing platform on all samples, and to call genotypes on all samples either all together or else individually.

A fourth caveat is that even with a low overall error rate, the sheer size of the exome means that false positives are inevitable. These can be minimized with strict filters on depth and quality, at the cost of discarding some real variants (for example, increasing the false negative rate). The stringency of filtering depends on the research goal. For most population genetic analyses, a subset of the exome with consistently high-quality data is preferred to a complete exome with a large number of false positives.

A further caveat, perhaps self-evident, is that exomes provide no information about noncoding regions, including many functionally important noncoding sites. Exomics researchers should be careful not to assume that all evolutionarily relevant variation has been captured by exomes. Indeed, some of the most well-documented targets of selection, such as the regulatory region of the lactase gene *LCT*, may leave little detectable signature in exomes [13].

Finally, exomes present the difficulty of a deluge of data. Storing and accessing large exome files is a computational challenge, although exomes are easier to work with than whole genomes. In addition, interpreting the functional consequences of one particular variant among hundreds of thousands is a daunting task. Given that even strict filtering does not eliminate error, it is recommend that sites or regions showing unusual polymorphism patterns be validated with Sanger sequencing before drawing any definitive conclusions about these loci.

Concluding remarks

Exome sequencing represents an important milestone in genomics, and provides a powerful tool for population geneticists that will facilitate estimates of numerous evolutionary parameters with much greater precision than was previously possible. Until large full-genome datasets in all populations of interest are feasible, exomes will represent the best available resource for inferring patterns of human demography and natural selection in an unbiased and comprehensive manner.

Acknowledgements

This work was supported by a research grant (1R01GM076036) from the NIH to JMA and the NHLBI Go Exome sequencing Project (HL-102923) to JMA and MJB.

Author details

[1]Department of Genome Sciences, University of Washington, 3720 15th Ave NE, Box 355065, Seattle, WA 98195-5065, USA. [2]Department of Pediatrics, University of Washington, 1959 NE Pacific St, Box 356320, Seattle, WA 98195-6320, USA

Published: 14 September 2011

References

1. Garrigan D, Hammer MF: **Reconstructing human origins in the genomic era.** *Nat Rev Genet* 2006, **7:**669-680.
2. Schaffner SF, Foo C, Gabriel S, Reich D, Daly MJ, Altshuler D: **Calibrating a coalescent simulation of human genome sequence variation.** *Genome Res* 2005, **15:**1576-1583.
3. Li JZ, Absher DM, Tang H, Southwick AM, Casto AM: **Worldwide human relationships inferred from genome-wide patterns of variation.** *Science* 2008, **319:**1100-1104.
4. Akey JM: **Constructing genomic maps of positive selection in humans: where do we go from here?** *Genome Res* 2009, **19:**711-722.
5. Green RE, Krause J, Briggs AW, Maricic T, Stenzel U, Kircher M, Patterson N, Li H, Zhai W, Fritz MH, Hansen NF, Durand EY, Malaspinas AS, Jensen JD, Marques-Bonet T, Alkan C, Prüfer K, Meyer M, Burbano HA, Good JM, Schultz R, Aximu-Petri A, Butthof A, Höber B, Höffner B, Siegemund M, Weihmann A, Nusbaum C, Lander ES, Russ C, *et al.*: **A draft sequence of the Neandertal genome.** *Science* 2010, **328:**710-722.
6. Reich D, Green RE, Kircher M, Krause J, Patterson N, Durand EY, Viola B, Briggs AW, Stenzel U, Johnson PL, Maricic T, Good JM, Marques-Bonet T, Alkan C, Fu Q, Mallick S, Li H, Meyer M, Eichler EE, Stoneking M, Richards M, Talamo S, Shunkov MV, Derevianko AP, Hublin JJ, Kelso J, Slatkin M, Pääbo S: **Genetic history of an archaic hominin group from Denisova Cave in Siberia.** *Nature* 2010, **468:**1053-1060.
7. Ng SB, Turner EH, Robertson PD, Flygare SD, Bigham AW, Lee C, Shaffer T, Wong M, Bhattacharjee A, Eichler EE, Bamshad M, Nickerson DA, Shendure J: **Targeted capture and massively parallel sequencing of 12 human exomes.** *Nature* 2009, **461:**272-276.
8. The International HapMap Consortium: **A haplotype map of the human genome.** *Nature* 2005, **437:**1299-1320.
9. The International HapMap Consortium: **A second generation human haplotype map of over 3.1 million SNPs.** *Nature* 2007, **449:**851-861.
10. Akey JM, Zhang K, Xiong M, Jin L: **The effect of single nucleotide polymorphism identification strategies on estimates of linkage disequilibrium.** *Mol Biol Evol* 2003, **20:**232-242.
11. Clark AG, Hubisz MJ, Bustamante CD, Williamson SH, Nielsen R: **Ascertainment bias in studies of human genome-wide polymorphism.** *Genome Res* 2005, **15:**1496-1502.
12. The 1000 Genomes Project Consortium: **A map of human genome variation from population-scale sequencing.** *Nature* 2010, **467:**1061-1073.
13. Tennessen JA, Madeoy J, Akey JM: **Signatures of positive selection apparent in a small sample of human exomes.** *Genome Res* 2010, **20:**1327-1334.
14. The Encode Project Consortium: **Identification and analysis of functional elements in 1% of the human genome by the ENCODE pilot project.** *Nature* 2007, **447:**799-816.
15. Grossman SR, Shylakhter I, Karlsson EK, Byrne EH, Morales S, Frieden G, Hostetter E, Angelino E, Garber M, Zuk O, Lander ES, Schaffner SF, Sabeti PC: **A composite of multiple signals distinguishes causal variants in regions of positive selection.** *Science* 2010, **327:**883-886.
16. Lamason RL, Mohideen MA, Mest JR, Wong AC, Norton HL, Aros MC, Jurynec MJ, Mao X, Humphreville VR, Humbert JE, Sinha S, Moore JL, Jagadeeswaran P, Zhao W, Ning G, Makalowska I, McKeigue PM, O'Donnell D, Kittles R, Parra EJ, Mangini NJ, Grunwald DJ, Shriver MD, Canfield VA, Cheng KC: **SLC24A5, a putative cation exchanger, affects pigmentation in zebrafish and humans.** *Science* 2005, **310:**1782-1786.
17. Carroll SB: **Endless forms: the evolution of gene regulation and morphological diversity.** *Cell* 2000, **101:**577-580.
18. Pritchard JK, Pickrell JK, Coop G: **The genetics of human adaptation: hard sweeps, soft sweeps, and polygenic adaptation.** *Curr Biol* 2010, **20:**R208-R215.
19. Hernandez RD, Kelley JL, Elyashiv E, Melton SC, Auton A, McVean G, 1000 Genomes Project, Sella G, Przeworski M: **Classic selective sweeps were rare in recent human evolution.** *Science* 2011, **331:**920-924.
20. Tennessen JA, Akey JM: **Parallel adaptive divergence among geographically diverse human populations.** *PLoS Genet* 2011, **7:**e1002127.
21. Yi X, Liang Y, Huerta-Sanchez E, Jin X, Cuo ZX, Pool JE, Xu X, Jiang H, Vinckenbosch N, Korneliussen TS, Zheng H, Liu T, He W, Li K, Luo R, Nie X, Wu H, Zhao M, Cao H, Zou J, Shan Y, Li S, Yang Q, Asan, Ni P, Tian G, Xu J, Liu X, Jiang T, Wu R, *et al.*: **Sequencing of 50 human exomes reveals adaptation to high altitude.** *Science* 2010, **329:**75-78.

22. Soldevila M, Andrés AM, Ramírez-Soriano A, Marquès-Bonet T, Calafell F, Navarro A, Bertranpetit J: **The prion protein gene in humans revisited: lessons from a worldwide resequencing study.** *Genome Res* 2006, **16:**231-239.
23. Kreitman M, Di Rienzo A: **Balancing claims for balancing selection.** *Trends Genet* 2004, **20:**300-304.
24. Bell G: **Fluctuating selection: the perpetual renewal of adaptation in variable environments.** *Philos Trans R Soc Lond B Biol Sci* 2010, **365:**87-97.
25. Hedrick PW: *Genetics of Populations.* 3rd Edition. Sudbury, MA: Jones and Bartlett, 2005:140.
26. Li Y, Vinckenbosch N, Tian G, Huerta-Sanchez E, Jiang T, Jiang H, Albrechtsen A, Andersen G, Cao H, Korneliussen T, Grarup N, Guo Y, Hellman I, Jin X, Li Q, Liu J, Liu X, Sparsø T, Tang M, Wu H, Wu R, Yu C, Zheng H, Astrup A, Bolund L, Holmkvist J, Jørgensen T, Kristiansen K, Schmitz O, Schwartz TW, *et al.*: **Resequencing of 200 human exomes identifies an excess of low-frequency non-synonymous coding variants.** *Nat Genet* 2010, **42:**969-972.
27. Akey JM, Eberle MA, Rieder MJ, Carlson CS, Shriver MD, Nickerson DA, Kruglyak L: **Population history and natural selection shape patterns of genetic variation in 132 genes.** *PLoS Biol* 2004, **2:**e286.
28. Voight BF, Adams AM, Frisse LA, Qian Y, Hudson RR, Di Rienzo A: **Interrogating multiple aspects of variation in a full resequencing data set to infer human population size changes.** *Proc Natl Acad Sci USA* 2005, **102:**18508-18513.
29. Coventry A, Bull-Otterson LM, Liu X, Clark AG, Maxwell TJ, Crosby J, Hixson JE, Rea TJ, Muzny DM, Lewis LR, Wheeler DA, Sabo A, Lusk C, Weiss KG, Akbar H, Cree A, Hawes AC, Newsham I, Varghese RT, Villasana D, Gross S, Joshi V, Santibanez J, Morgan M, Chang K, Iv WH, Templeton AR, Boerwinkle E, Gibbs R, Sing CF: **Deep resequencing reveals excess rare recent variants consistent with explosive population growth.** *Nat Commun* 2010, **1:**131.
30. Pfaff CL, Parra EJ, Bonilla C, Hiester K, McKeigue PM, Kamboh MI, Hutchinson RG, Ferrell RE, Boerwinkle E, Shriver MD: **Population structure in admixed populations: effect of admixture dynamics on the pattern of linkage disequilibrium.** *Am J Hum Genet* 2001, **68:**198-207.
31. Bryc K, Auton A, Nelson MR, Oksenberg JR, Hauser SL, Williams S, Froment A, Bodo JM, Wambebe C, Tishkoff SA, Bustamante CD: **Genome-wide patterns of population structure and admixture in West Africans and African Americans.** *Proc Natl Acad Sci U S A* 2010, **107:**786-791.
32. McVicker G, Gordon D, Davis C, Green P: **Widespread genomic signatures of natural selection in hominid evolution.** *PLoS Genet* 2009, **5:**e1000471.
33. Williamson SH, Hernandez R, Fledel-Alon A, Zhu L, Nielsen R, Bustamante CD: **Simultaneous inference of selection and population growth from patterns of variation in the human genome.** *Proc Natl Acad Sci U S A* 2005, **102:**7882-7887.
34. McDonald JH, Kreitman M: **Adaptive protein evolution at the Adh locus in *Drosophila*.** *Nature* 1991, **351:**652-654.

doi:10.1186/gb-2011-12-9-127
Cite this article as: Tennessen JA, *et al.*: The promise and limitations of population exomics for human evolution studies. *Genome Biology* 2011, **12:**127.