Int J Primatol (2013) 34:303–314
DOI 10.1007/s10764-013-9663-2

A New Method for Genome-wide Marker Development and Genotyping Holds Great Promise for Molecular Primatology

Christina M. Bergey · Luca Pozzi · Todd R. Disotell · Andrew S. Burrell

Received: 14 December 2012 / Accepted: 29 January 2013 / Published online: 28 February 2013

Abstract Over the last two decades primatologists have benefited from the use of numerous molecular markers to study various aspects of primate behavior and evolutionary history. However, most of the studies to date have been based on a single locus, usually mitochondrial DNA, or a few nuclear markers, e.g., microsatellites. Unfortunately, the use of such markers not only is unable to address successfully important questions in primate population genetics and phylogenetics (mainly because of the discordance between gene tree and species tree), but also their development is often a time-consuming and expensive task. The advent of next-generation sequencing allows researchers to generate large amounts of genomic data for nonmodel organisms. However, whole genome sequencing is still cost prohibitive for most primate species. We here introduce a second-generation sequencing technique for genotyping thousands of genome-wide markers for nonmodel organisms. Restriction site–associated DNA sequencing (RAD-seq) reduces the complexity of the genome and allows inexpensive and fast discovery of thousands of markers in many individuals. Here, we describe the principles of this technique and we demonstrate its application in five primates, *Microcebus* sp., *Cebus* sp., *Theropithecus gelada*, *Pan troglodytes*, and *Homo sapiens*, representing some of the major lineages within the order. Despite technical and bioinformatic challenges, RAD-seq is a promising method for multilocus phylogenetic and population genetic studies in primates, particularly in young clades in which a high number of orthologous regions are likely to be found across populations or species.

C. M. Bergey (✉) · L. Pozzi · T. R. Disotell · A. S. Burrell
Department of Anthropology, New York University, New York, NY 10003, USA
e-mail: christina.bergey@nyu.edu

C. M. Bergey · L. Pozzi · T. R. Disotell
New York Consortium in Evolutionary Primatology, New York, USA

Keywords Genotyping · Nonmodel organisms · Phylogenetics · Population genetics · Restriction site–associated DNA sequencing · Second-generation DNA sequencing · Single-nucleotide polymorphisms

Introduction

In molecular primatology, as in all fields of molecular biology, the development of variable genetic markers is essential for the study of organisms at different levels, from population genetic to phylogeographic to phylogenetic research (Avise 1994). During the first decades of the discipline, an impediment to researchers was the need to develop and type polymorphic markers in a taxon of interest. The markers that resulted from this time-consuming and expensive task were often uninformative when applied outside of the population used in their design, necessitating further rounds of primer design or microsatellite assays (Davey *et al*. 2011). Owing to the bottleneck caused by inefficient marker discovery, many population genetic or phylogenetic studies in molecular primatology have been based on one or few loci, usually mitochondrial DNA or microsatellites (Ting and Sterner 2012). Inferences from such studies can reliably give the evolutionary history of those particular regions of the genome, but they fail to capture the complete complex history of the population adequately given the mosaic nature of genomic evolution (Degnan and Rosenberg 2009; Edwards 2009; Maddison 1997; Maddison and Knowles 2006). Sufficient resolution depends on high marker density, and until recently that goal has been out of reach for many primate researchers (Edwards 2009).

The rapidly decreasing costs of DNA sequencing technology have promised revolutionary gains for primatology (Enard and Pääbo 2004; Goodman *et al*. 2005; Ting and Sterner 2012). A primate researcher benefits from the many nearby sequenced and assembled reference genomes in the order, but genomic studies of nonmodel organisms remain difficult. Though the cost of whole genome sequencing has fallen to a level feasible for many researchers' budgets (Perry *et al*. 2012), sequencing whole genomes for the tens or hundreds of individuals desired in a typical population genetic study is often prohibitively expensive and quite possibly superfluous (McCormack *et al*. 2012). Fortunately, researchers have recently developed techniques that reduce the complexity of the genome and allow for the discovery and genotyping of thousands or tens of thousands of genome-wide makers in many individuals in a single step (Davey *et al*. 2011; McCormack *et al*. 2012). These methods, including restriction site–associated DNA sequencing (RAD-seq), reduced-representation libraries (RRL), complexity reduction of polymorphic sequences (CRoPS), and low coverage genotyping, are reviewed in Davey *et al*. (2011). RAD-seq is one such simple, inexpensive reduced representation technique that allows for the sequencing of small fragments of the genome adjacent to restriction enzyme cut sites (Baird *et al*. 2008). These restriction site–associated DNA tags (RAD tags) were originally developed for use in microarray hybridization genotyping (Miller *et al*. 2007), but an updated protocol substitutes second-generation DNA sequencing to rapidly discover and type single-nucleotide polymorphisms (SNPs) (Baird *et al*. 2008; Etter *et al*. 2011). The lack of reliance on a reference genome and applicability to data sets of many individuals make it a promising technique for

phylogenetic or population genetic studies as well as relatedness analyses and reconstruction of pedigrees in nonmodel organisms, such as many primates.

The RAD-seq Technique

The following is a summary of the RAD tag library preparation protocol of Etter *et al*. (2011) (Fig. 1). The RAD-seq library preparation begins when genomic DNA is digested with a restriction enzyme, such as *Eco*RI or *Psp*XI (Fig. 1a). The P1 adapter is then ligated to the fragments, connected to the sticky end at the restriction enzyme cut site. The P1 adapter contains an amplification site for polymerase chain reaction (PCR), an Illumina sequencing priming site, and an individual-specific barcode of five basepairs (bp) (Fig. 1b). Once the barcode has been added, fragments from multiple individuals can be pooled (Fig. 1c), and the DNA is randomly sheared with a sonicator to have a length distribution predominantly less than 1 kilobase (Fig. 1d). To select for reads that are suitable for sequencing on the Illumina platform, the sheared samples are size selected via agarose gel electrophoresis, extracting fragments between 300 and 500 bp in length. The second adapter, P2, is a Y adapter, meaning its two halves are complementary for only part of their length (Fig. 1e). It is ligated to the fragments and then the fragments are amplified via PCR (Fig. 1f). Because the second adapter has divergent ends, the reverse amplification primer is unable to bind until after the forward amplification primer has filled in its complementary sequence. This ensures that only RAD tags ligated to P1 are able to amplify. After 12–14 cycles of PCR, kept low to minimize the risk of introducing PCR artifacts or biases, the library is ready for final clean-up, quality control, and sequencing.

Previous RAD-seq Studies

RAD-seq is an economical and efficient method for SNP discovery and genotyping. Since its first application by Baird *et al*. (2008) on two model organisms —the fungus *Neurospora crassa* and the three-spined stickleback *Gasterosteus aculeatus*— RAD-seq has been successfully applied to several organisms for which reference genome information was not available.

The ability of RAD-seq technology to identify thousands of orthologous SNPs across multiple individuals at both intra- and interspecific level makes this technique extremely promising for the study of population structure (Emerson *et al*. 2010; Hohenlohe *et al*. 2010; Keller *et al*. 2012), gene flow and hybridization (Hohenlohe *et al*. 2011; Keller *et al*. 2012), phylogeography (Emerson *et al*. 2010), and phylogeny (Rubin *et al*. 2012; Wagner *et al*. 2012). The RAD tag sequencing approach has been particularly used to generate SNP data to address questions in population genomics. For example, a series of studies conducted by Hohenlohe and colleagues investigated parallel adaptation and hybridization in several species of fish (Hohenlohe *et al*. 2010, 2011, 2012), while Emerson *et al*. (2010) identified >3700 SNPs for pitcher plant mosquitoes in eastern North America, providing the first phylogeographic study using RAD sequence data.

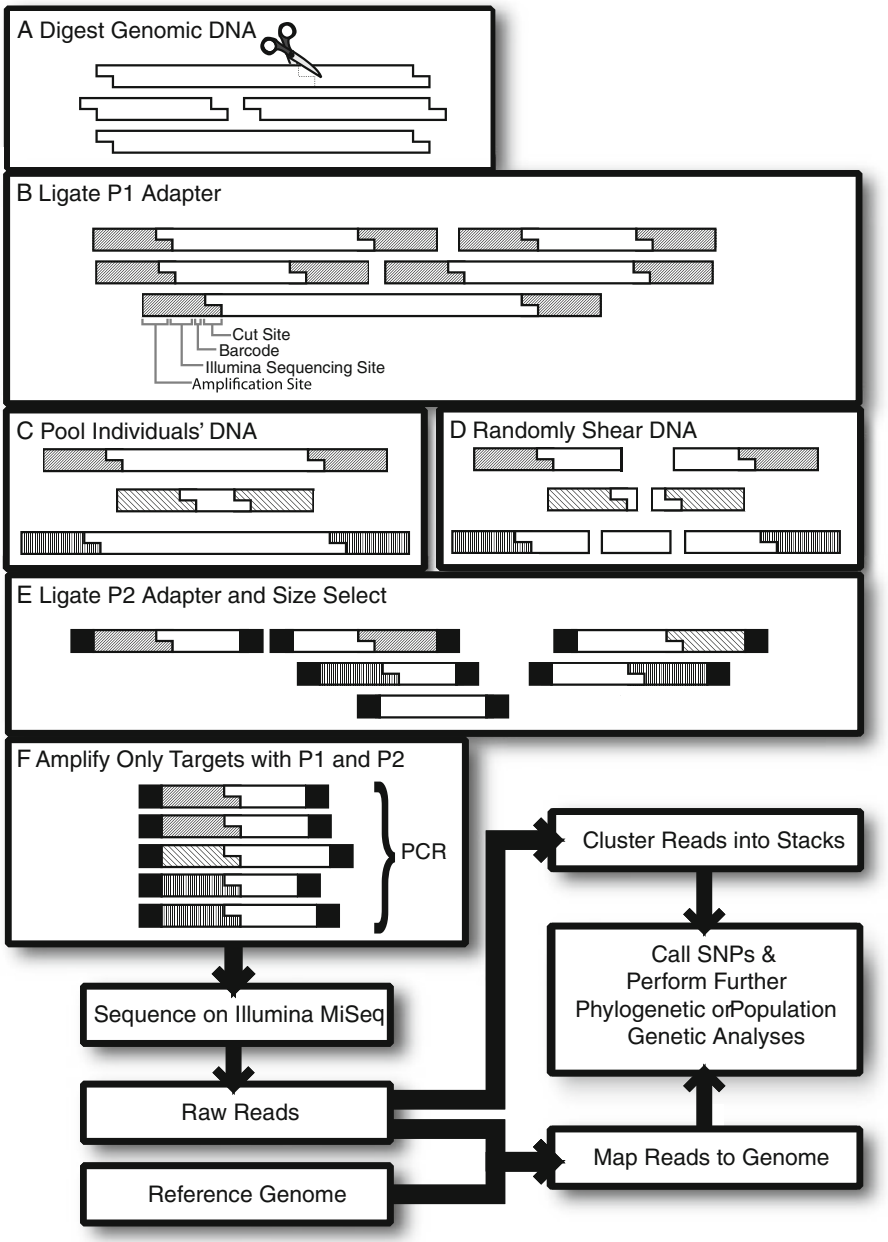

Fig. 1 An overview of the RAD-seq library creation protocol and initial analysis steps.

Although RAD sequencing is more effective in addressing questions at or below the level of a single species, a few recent studies have used this technique in the analysis of phylogenetic questions. Rubin *et al.* (2012) provided a simulation study in which they investigated the accuracy of RAD-seq data to reconstruct phylogenies in organisms with different population sizes and clade ages (*Drosophila*, mammals, and

yeasts). In their study the authors supported the efficiency of RAD-seq data in inferring phylogenies, but they also caution that this approach achieves the best results in younger clades where more orthologous restriction sites are likely to be retained across species. This simulation analysis was confirmed in two recent empirical studies in which a RAD tag sequencing approach was successfully used to reconstruct phylogenetic relationships in two recent but speciose radiations: the African cichlids (Wagner *et al*. 2012) and the *Heliconius* butterflies (Nadeau *et al*. 2012).

Here, we demonstrate the technique in five primates: a lemur, New World monkey, Old World monkey, and two apes, and test the flexibility of the restriction enzyme across the order. We include two hominoids that share a nearby well-annotated, well-assembled reference genome (that of the human), which we use to assess how many possible sequencing sites yielded reads. We also chose subjects with and without a reference genome to demonstrate the two possible analytical techniques: clustering, which does not require a reference genome, and mapping, which does.

Methods

Library Preparation and Sequencing

We digested genomic DNA from five individuals (representing *Microcebus* sp., *Cebus* sp., *Theropithecus gelada*, *Pan troglodytes*, and *Homo sapiens*) with *Psp*XI (New England Biolabs) and used it to create a multiplexed RAD tag library. DNA had been previously isolated from blood or tissue samples in the NYU collection.

Our library preparation method followed that of Etter *et al*. (2011) with the following modifications: the P1 adapter bottom oligonucleotide was modified to have an overhang corresponding to the cut site of *Psp*XI, and a longer P2 adapter suitable for paired end sequencing was used (P2_top: 5′- /5Phos/GAT CGG AAG AGC GGT TCA GCA GGA ATG CCG AGA CCG ATC AGA ACA A-3′; P2_bottom: 5′- CAA GCA GAA GAC GGC ATA CGA GAT CGG TCT CGG CAT TCC TGC TGA ACC GCT CTT CCG ATC*T −3′). Individual-specific barcodes contained in the P1 adapter differed by at least three nucleotides. We chose *Psp*XI based on the results of *in silico* digestion of the human, rhesus macaque, and baboon reference genomes using custom Perl scripts. After 14 cycles of PCR, we sequenced the prepared library as one 150-cycle paired-end run and one 150-cycle single-end run of an Illumina MiSeq at the NYU Langone Medical Center's Genome Technology Center using a spike-in of 30 % PhiX DNA to control for low diversity in the library at the barcode and restriction sites. Other individuals were sequenced alongside those of the present study. Sequences are available to download from the NCBI Short Read Archive (accession number SRP018000).

Sequence Analysis: Clustering and SNP Discovery

As input for the clustering analysis, we combined the first read of the paired-end run and the single-end run reads. We demultiplexed, or separated by barcode, sequence reads and excluded reads without an expected barcode or an intact restriction enzyme

cut site from the analysis. We also removed reads with any quality scores <10, as low-quality scores indicate high probability of an incorrect base call. Using the program Stacks, we clustered all reads into sets that differed by no more than 2 bp and compared closely related sets to detect orthologous loci and SNPs using a maximum likelihood approach (Catchen *et al*. 2011). We tallied orthologous SNPs using VCFtools (Danecek *et al*. 2011).

Sequence Analysis: Assess RAD Tag Coverage

To assess the RAD tag coverage, we mapped human and chimpanzee reads to the highest quality primate reference genome, that of humans. Again, we excluded reads without an expected barcode or an intact restriction enzyme cut site. We aligned reads to the human reference genome (GRCh37/hg19, Human Genome Consortium 2001) using the Burrows–Wheeler aligner (BWA) with default parameters (Li and Durbin 2009). We separately mapped the single-end and paired-end data and then combined the resultant files after alignment. We removed reads that were unmapped or that had low mapping quality using Picard (http://picard.sourceforge.net) and BamTools (Barnett *et al*. 2011).

After performing local realignment around indels with GATK (DePristo *et al*. 2011), we identified SNPs and short indels using SAMtools mpileup and BCFtools (Li *et al*. 2009). We required a minimum coverage of 3 reads and a maximum of 100 to call a SNP or an indel at a given location. We tallied orthologous SNPs using VCFtools (Danecek *et al*. 2011).

To assess how many restriction sites were successfully sequenced and to analyze the degree of overlap between multiplexed individuals' data sets, we first bioinformatically found all possible *Psp*XI cut sites in the human genome using the oligoMatch utility in the USCS Genome Browser program and created a BED file of all regions 1000 bp upstream and downstream (Meyer *et al*. 2012). This allowed us to calculate the coverage of these restriction site–associated regions using BEDtools' multiBamCov (Quinlan and Hall 2010).

Results

We could confidently assign 12.3 million sequencing reads with an intact barcode and restriction enzyme cut site to one of the five primates. Of those reads, 9.1 million passed quality control filtration and were clustered into stacks. By comparing these stacks and including only SNPs that were present in multiple species, we identified 7910 SNPs among all samples. Information for each individual from the clustering analysis is summarized in Table I.

The human and chimpanzee samples could be mapped to a well-annotated, well-assembled reference genome (that of the human), which we used to assess how many possible sequencing sites yielded reads. In the human genome, we found 58,172 possible cut sites for *Psp*XI and 116,344 possible sequencing sites (two per cut site, one upstream and one downstream). Of those possible locations, 111,686 locations (96.00 %) had at least one mapped read present in human, and 91,646 locations (78.77 %) had at least one mapped read present in chimpanzee (Table II). For 90,022

Table I Results of the clustering analysis

Taxon	No. of reads	No. of filtered reads	No. of ind. stacks	Mean coverage (SD)	No. of SNPs in multiple-species stacks[a]
Microcebus sp.	2,830,832	2,025,103	248,324	7.66 (20.71)	13
Cebus sp.	1,946,096	1,427,413	107,829	12.64 (139.08)	56
Theropithecus gelada	1,918,425	1,392,709	136,657	9.70 (17.34)	212
Pan troglodytes	2,616,062	1,910,560	157,775	11.50 (42.27)	5886
Homo sapiens	3,032,823	2,374,733	131,544	17.37 (25.30)	5786

[a] Counts of only SNPs in loci that are present in at least one other taxon

sites (77.38 %), both chimpanzee and human had at least one read. When we restrict the analysis to sites with at least three reads, 109,098 sites (93.77 %) had sequences in human, 89,628 sites (77.04 %) in chimpanzee, and 86,604 sites (74.44 %) in both. From these data, we found 9275 SNPs relative to the human reference genome that were present in both chimpanzee and human data sets.

Discussion

We have demonstrated the RAD-seq technique in five primate taxa using two analytical pipelines: a clustering technique that requires no reference genome and a mapping technique that does. We showed that the method and the restriction enzyme, *Psp*XI, worked in all tested primates. We first applied the clustering method to the reads to infer SNPs, and not surprisingly found greater numbers of shared SNPs in the samples with shorter average pairwise phylogenetic distances (such as the catarrhines) than in those more distantly related to the other samples (such as *Microcebus*). This is illustrative of the shallow time depth for which RAD-seq is best suited, which we discuss in the text that follows, but that should not hinder studies within primate lineages. To demonstrate the mapping approach and test how many potential sequencing sites had associated reads, we mapped the human and chimp reads to the human reference genome, as the apes were sufficiently closely related to the genome to allow mapping. When we compared our data to the predicted enzyme cut sites, a high percentage of sites (96.00 % in human and 78.77 % in chimpanzee) had at least one associated read. We inferred SNPs from these mapped

Table II Results of mapping human and chimpanzee reads to the human reference genome

Taxon	No. of reads	No. of filtered Reads	No. of loci ≥ 1 read	No. of loci ≥ 3 reads	No. of SNPs relative to human genome[a]
Pan troglodytes	3,917,046	2,826,643	91,646	89,628	309,703
Homo sapiens	4,542,978	3,784,192	111,686	109,098	35,651

Read count is higher than in the clustering analysis because of the inclusion of the second reads of the paired-end run

[a] Counts of variants relative to the human reference genome (hg19)

reads, illustrating the approach that many primate researchers can use if nearby well-assembled genomes are available.

There are several advantages in using RAD-seq over other molecular techniques. First, this methodology is quite inexpensive and requires little labwork. The development of a library can be completed in only 2 days of labwork, and all the different steps can be easily performed in a standard molecular laboratory. Also, the possibility to multiplex several individuals in the same Illumina run using either standard barcodes or a custom combinatorial indexing method (Peterson *et al*. 2012) allows researchers to reduce the number of sequencing runs, decreasing the costs even further (Davey *et al*. 2011; McCormack *et al*. 2012; Peterson *et al*. 2012). The cost of library preparation and sequencing for the present study was <$2000, and the cost per individual could be decreased further with more individuals multiplexed, with an enzyme that cuts at fewer sites, or with a higher throughput machine, such as the Illumina HiSeq. The cost of performing Sanger sequencing on PCR products for only 500 loci in 20 individuals has been estimated at $145,800 (Lemmon *et al*. 2012), making RAD-seq orders of magnitude less expensive than traditional PCR-based methods of genotyping.

Second, RAD-seq represents a great improvement in the discovery of molecular markers to be used in population genetics and phylogenetics. Previous studies to date have been based on a single locus (mainly mitochondrial DNA) or a few tens of loci (microsatellites for population genetics or nuclear loci for phylogenetics). RAD-seq techniques can easily produce thousands of independent SNPs in a single run, increasing 100–1000 times the amount of data available to researchers. RAD sequencing can produce a large amount of orthologous SNP data that can be employed in a wide range of studies, including population genomics and demographics, e.g., effective population size estimates, bottlenecks, etc., gene flow and hybridization between closely related species, species boundaries, phylogeography, and phylogeny especially at the intrageneric level (Emerson *et al*. 2010; Hohenlohe *et al*. 2010, 2011; Keller *et al*. 2012; Rubin *et al*. 2012; Wagner *et al*. 2012).

Third, RAD-seq data also have the potential to transform our ability to construct pedigrees and infer relatedness among individuals from wild populations. Genetic data are increasingly common in studies of primate behavior, primarily to determine kin relationships (Di Fiore 2003). The most commonly used genetic markers are microsatellites (also known as simple tandem repeats [STRs]). These are highly variable, often having >10 alleles. However, they are difficult to genotype accurately. SNPs, conversely, are only biallelic but much easier to genotype confidently. The difference in variability means that many more SNPs than microsatellites need to be genotyped in order to infer relationships (Jones *et al*. 2010). Both types of marker are difficult to develop using traditional methods, and typically only about a dozen loci are used in studies of nonmodel species. Because SNPs are not variable enough to be informative at such low numbers, microsatellites have been the marker of choice in behavioral studies of nonmodel organisms. RAD-seq will drastically increase the ease with which SNPs can be discovered and sequenced, making this class of data suddenly viable for behavioral research. Simulations and empirical studies have demonstrated that even moderate numbers of SNPs (<200) allow more accurate estimates of relatedness and pedigrees than small numbers (<20) of microsatellites (Anderson and Garza 2006; Hauser *et al*. 2011). In summary, we believe that the use

of RAD-seq technology will provide extremely valuable information to study recent radiation within primates and to address some major open questions in primatology.

Despite the great potential of the application of RAD sequencing in primatology, there are some possible limitations of this technique, including errors introduced during library preparation, bioinformatic challenges, a requirement of high-quality DNA, and the limited evolutionary distance for which the technique is applicable. Several sources of error are inherent in the RAD-seq technique and must be considered during analysis. These include restriction fragment bias, restriction site heterozygosity, and PCR GC content bias. Much of the bias can be explained by restriction fragment length bias, caused by incomplete shearing, which is less of a problem in rare cutters such as the enzyme chosen for our primate study, *Psp*XI. A review of the potential pitfalls associated with RAD-seq found none of these problems insurmountable and recommended Stacks, the software used in the present study, for analyses (Davey *et al*. 2012).

One of the largest challenges facing a researcher adopting RAD-seq may be in analyzing the relatively large amount of sequence data. Fortunately, many tools for analyzing RAD-seq data are freely available, including all those used in the present study. Most, however, lack graphical user interfaces making familiarity with the command line a prerequisite. With access to a multiprocessor computing cluster, the bioinformatic analysis can be completed with runtimes on the order of hours.

Possibly, the main constraint of employing RAD-seq on a large scale within primates is related to the need for high-quality DNA in order to build the library. In this study we used DNA extracted from tissue or blood. However, most molecular primatologists are limited in their use of invasive samples and more often rely on low-quality samples such as hair or feces. Although not yet available, in theory, RAD-seq protocols using noninvasive samples could be developed. In a recent study, Perry *et al*. (2010) presented a genomic-scale capture protocol to obtain endogenous DNA from primate fecal samples. Capture methods have been also used to obtain low-quantity and poor quality DNA from museum specimens (Mason *et al*. 2011) or even fossils (Burbano *et al*. 2010; Krause *et al*. 2010).

Another possible limitation of the RAD-seq approach is the evolutionary time scale of its application. RAD-seq data in fact might not be suitable for comparing very distantly related taxa (Rubin *et al*. 2012). In their study, Rubin and colleagues showed a negative correlation between phylogenetic accuracy and evolutionary divergence time, suggesting that the age of a clade is a major determinant of the success of the RAD method (Rubin *et al*. 2012). The deep divergences between taxa in fact decrease the number of discoverable RAD loci for two main reasons: first, restriction sites can change over time, reducing the number of orthologous loci retained across distantly related taxa; second, orthology is more difficult to infer based on sequence similarity when evolutionary divergence is high (Rubin *et al*. 2012). This correlation between accuracy and divergence time either reduces the number of orthologus loci available for phylogenetic reconstruction or increases the amount of missing data; both scenarios can affect phylogenetic performance, reducing the support values in many nodes or supporting different topologies. However, despite this drawback, Rubin and colleagues successfully reconstructed the phylogeny of 12 species of *Drosophila*, with a crown age of 40–60 Mya. This result suggests that RAD-seq data might be informative enough to reconstruct the phylogeny of most

lineages within primates (crown age between 65 and 85 Mya; Perelman *et al*. 2011; Steiper and Seiffert 2012; Wilkinson *et al*. 2011).

In conclusion, this study illustrates the value of RAD-seq approach in discovering a large number of independent SNPs that can be used to address many questions in primatology, ranging from population genomics to phylogenetics. Our preliminary study of primates shows the feasibility of this technique across the primate order, even when nearby reference genomes are not available. Future developments in both sequencing technologies and computational tools will address —and most likely overcome— the current limitations of RAD sequencing, making this technique viable for studies of a large number of primate species and populations.

Acknowledgments The present study was supported by a Leakey Foundation General Grant and an NSF Graduate Research Fellowship. We thank the NYU Langone Medical Center's Genome Technology Center for assistance with library preparation and sequencing, as well as two anonymous reviewers and the editors for their helpful comments.

References

Anderson, E. C., & Garza, J. C. (2006). The power of single-nucleotide polymorphisms for large-scale parentage inference. *Genetics, 172*(4), 2567–2582.

Avise, J. C. (1994). *Molecular markers, natural history, and evolution*. New York: Chapman & Hall.

Baird, N. A., Etter, P. D., Atwood, T. S., Currey, M. C., Shiver, A. L., Lewis, Z. A., Selker, E. U., et al. (2008). Rapid SNP discovery and genetic mapping using sequenced RAD markers. *PloS One, 3*(10), e3376.

Barnett, D. W., Garrison, E. K., Quinlan, A. R., Strömberg, M. P., & Marth, G. T. (2011). BamTools: a C++ API and toolkit for analyzing and managing BAM files. *Bioinformatics, 27*(12), 1691–1692.

Burbano, H. A., Hodges, E., Green, R. E., Briggs, A. W., Krause, J., Meyer, M., et al. (2010). Targeted investigation of the Neanderthal genome by array-based sequence capture. *Science, 328*(5979), 723–725.

Catchen, J. M., Amores, A., Hohenlohe, P., Cresko, W., & Postlethwait, J. H. (2011). Stacks: Building and genotyping loci de novo from short-read sequences. *G3, 1*(3), 171–182.

Danecek, P., Auton, A., Abecasis, G., Albers, C. A., Banks, E., DePristo, M. A., Handsaker, R. E., et al. (2011). The variant call format and VCFtools. *Bioinformatics, 27*(15), 2156–2158.

Davey, J. W., Cezard, T., Fuentes-Utrilla, P., Eland, C., Gharbi, K., & Blaxter, M. L. (2012). Special features of RAD sequencing data: Implications for genotyping. *Molecular Ecology*. doi:10.1111/mec.12084.

Davey, J. W., Hohenlohe, P. A., Etter, P. D., Boone, J. Q., Catchen, J. M., & Blaxter, M. L. (2011). Genome-wide genetic marker discovery and genotyping using next-generation sequencing. *Nature Reviews Genetics, 12*(7), 499–510.

Degnan, J. H., & Rosenberg, N. A. (2009). Gene tree discordance, phylogenetic inference and the multispecies coalescent. *Trends in Ecology and Evolution, 24*(6), 332–340.

DePristo, M. A., Banks, E., Poplin, R., Garimella, K. V., Maguire, J. R., Hartl, C., Philippakis, A. A., et al. (2011). A framework for variation discovery and genotyping using next-generation DNA sequencing data. *Nature Genetics, 43*(5), 491–498.

Di Fiore, A. (2003). Molecular genetic approaches to the study of primate behavior, social organization, and reproduction. *Yearbook of Physical Anthropology, 46*, 62–99.

Edwards, S. V. (2009). Is a new and general theory of molecular systematics emerging? *Evolution: International Journal of Organic Evolution, 63*(1), 1–19.

Emerson, K., Merz, C., & Catchen, J. (2010). Resolving postglacial phylogeography using high-throughput sequencing. *Proceedings of the National Academy of Sciences of the USA, 107*(37), 16196–16200.

Enard, W., & Pääbo, S. (2004). Comparative primate genomics. *Annual Review of Genomics and Human Genetics, 5*, 351–378.

Etter, P. D., Bassham, S., Hohenlohe, P. A., Johnson, E., & Cresko, W. A. (2011). SNP discovery and genotyping for evolutionary genetics using RAD sequencing. In V. Orgogozo & M. V. Rockman (Eds.), *Molecular methods for evolutionary genetics* (pp. 157–178). New York: Humana Press.

Goodman, M., Grossman, L. I., & Wildman, D. E. (2005). Moving primate genomics beyond the chimpanzee genome. *Trends in Genetics, 21*(9), 511–517.

Hauser, L., Baird, M., Hilborn, R., Seeb, L. W., & Seeb, J. E. (2011). An empirical comparison of SNPs and microsatellites for parentage and kinship assignment in a wild sockeye salmon (*Oncorhynchus nerka*) population. *Molecular Ecology Resources, 11*(Supplement 1), 150–161.

Hohenlohe, P. A., Amish, S. J., Catchen, J. M., Allendorf, F. W., & Luikart, G. (2011). Next-generation RAD sequencing identifies thousands of SNPs for assessing hybridization between rainbow and westslope cutthroat trout. *Molecular Ecology Resources, 11*(Supplement 1), 117–122.

Hohenlohe, P. A., Bassham, S., Currey, M., & Cresko, W. A. (2012). Extensive linkage disequilibrium and parallel adaptive divergence across threespine stickleback genomes. *Philosophical Transactions of the Royal Society of London B: Biological Sciences, 367*(1587), 395–408.

Hohenlohe, P. A., Bassham, S., Etter, P. D., Stiffler, N., Johnson, E. A., & Cresko, W. A. (2010). Population genomics of parallel adaptation in threespine stickleback using sequenced RAD tags. *PLoS Genetics, 6*(2), e1000862.

Human Genome Sequencing Consortium (2001). Initial sequencing and analysis of the human genome. *Nature*, *409*(6822), 860–921.

Jones, A. G., Small, C. M., Paczolt, K. A., & Ratterman, N. L. (2010). A practical guide to methods of parentage analysis. *Molecular Ecology Resources, 10*(1), 6–30.

Keller, I., Wagner, C. E., Greuter, L., Mwaiko, S., Selz, O. M., Sivasundar, A., et al. (2012). Population genomic signatures of divergent adaptation, gene flow and hybrid speciation in the rapid radiation of Lake Victoria cichlid fishes. *Molecular Ecology*. doi:10.1111/mec.12083.

Krause, J., Fu, Q., Good, J. M., Viola, B., Shunkov, M. V., Derevianko, A. P., et al. (2010). The complete mitochondrial DNA genome of an unknown hominin from southern Siberia. *Nature, 464*(7290), 894–897.

Lemmon, A. R., Emme, S. A., & Lemmon, E. M. (2012). Anchored hybrid enrichment for massively high-throughput phylogenomics. *Systematic Biology, 61*(5), 727–744.

Li, H., & Durbin, R. (2009). Fast and accurate short read alignment with Burrows-Wheeler transform. *Bioinformatics, 25*(14), 1754–1760.

Li, H., Handsaker, B., Wysoker, A., Fennell, T., Ruan, J., Homer, N., Marth, G., et al. (2009). The sequence alignment/map format and SAMtools. *Bioinformatics, 25*(16), 2078–2079.

Maddison, W. P. (1997). Gene trees in species trees. *Systematic Biology, 46*(3), 523.

Maddison, W. P., & Knowles, L. L. (2006). Inferring phylogeny despite incomplete lineage sorting. *Systematic Biology, 55*(1), 21–30.

Mason, V. C., Li, G., Helgen, K. M., & Murphy, W. J. (2011). Efficient cross-species capture hybridization and next-generation sequencing of mitochondrial genomes from noninvasively sampled museum specimens. *Genome Research, 21*(10), 1695–1704.

McCormack, J. E., Hird, S. M., Zellmer, A. J., Carstens, B. C., & Brumfield, R. T. (2012). Applications of next-generation sequencing to phylogeography and phylogenetics. *Molecular Phylogenetics and Evolution*. doi:10.1016/j.ympev.2011.12.007.

Meyer, L. R., Zweig, A. S., Hinrichs, A. S., Karolchik, D., Kuhn, R. M., Wong, M., Sloan, C. A., et al. (2013). The UCSC Genome Browser database: Extensions and updates 2013. *Nucleic Acids Research, 41*(D1), D46–D69.

Miller, M. R., Dunham, J. P., Amores, A., Cresko, W. A., & Johnson, E. A. (2007). Rapid and cost-effective polymorphism identification and genotyping using restriction site associated DNA (RAD) markers. *Genome Research, 17*(2), 240–248.

Nadeau, N. J., Martin, S. H., Kozak, K. M., Salazar, C., Dasmahapatra, K. K., Davey, J. W., et al. (2012). Genome-wide patterns of divergence and gene flow across a butterfly radiation. *Molecular Ecology*. doi:10.1111/j.1365-294X.2012.05730.x.

Perelman, P., Johnson, W. E., Roos, C., Seuánez, H. N., Horvath, J. E., Moreira, M. A. M., et al. (2011). A molecular phylogeny of living primates. *PLoS Genet, 7*(3), e1001342.

Perry, G. H., Marioni, J. C., Melsted, P., & Gilad, Y. (2010). Genomic-scale capture and sequencing of endogenous DNA from feces. *Molecular Ecology, 19*(24), 5332–5344.

Perry, G. H., Reeves, D., Melsted, P., Ratan, A., Miller, W., Michelini, K., Louis, E. E., et al. (2012). A genome sequence resource for the aye-aye (*Daubentonia madagascariensis*), a nocturnal lemur from Madagascar. *Genome Biology and Evolution, 4*(2), 126–135.

Peterson, B. K., Weber, J. N., Kay, E. H., Fisher, H. S., & Hoekstra, H. E. (2012). Double digest RADseq: An inexpensive method for de novo SNP discovery and genotyping in model and non-model species. *PloS One, 7*(5), e37135.

Quinlan, A. R., & Hall, I. M. (2010). BEDTools: A flexible suite of utilities for comparing genomic features. *Bioinformatics, 26*(6), 841–842.

Rubin, B. E. R., Ree, R. H., & Moreau, C. S. (2012). Inferring phylogenies from RAD sequence data. *PloS One, 7*(4), e33394.

Steiper, M. E., & Seiffert, E. R. (2012). Evidence for a convergent slowdown in primate molecular rates and its implications for the timing of early primate evolution. *Proceedings of the National Academy of Sciences of the USA, 109*(16), 6006–6011.

Ting, N., & Sterner, K. N. (2012). Primate molecular phylogenetics in a genomic era. *Molecular Phylogenetics and Evolution*. doi:10.1016/j.ympev.2012.08.021.

Wagner, C. E., Keller, I., Wittwer, S., Selz, O. M., Mwaiko, S., Greuter, L., et al. (2012). Genome-wide RAD sequence data provide unprecedented resolution of species boundaries and relationships in the Lake Victoria cichlid adaptive radiation. *Molecular Ecology.* doi:10.1111/mec.12023.

Wilkinson, R. D., Steiper, M. E., Soligo, C., Martin, R. D., Yang, Z., & Tavaré, S. (2011). Dating primate divergences through an integrated analysis of palaeontological and molecular data. *Systematic Biology, 60*(1), 16–31.